Calculus and Analytical Reasoning

Sherman K. Stein
Douglas B. Meade

MAA
MATHEMATICAL ASSOCIATION OF AMERICA
AMS
American Mathematical Society

Calculus and Analytical Reasoning

Sherman K. Stein

Douglas B. Meade

Please send all requests for permissions to meade@math.sc.edu.

Artwork in figures is the work of Joy Markel, under contract with Sherman K. Stein.

Cover image (sound waves) is from Getty. Courtesy Yaroslav Kushta. Creative #: 1469787278

Dedication

To Hannah and Betsy

To Joshua, Rebecca, and Susanna and to Stuart, Michael, Marybeth, and Bradley

Authors' Rationale While Writing This Book

The book consists of 114 sections and 26 "Calculus Is Everywhere" applications in 18 chapters. There are 480 examples, 937 figures, and more than 5700 exercises.

As we wrote each section of this book, we kept in our minds an image of the student who will be using it. The student will most likely be majoring in a STEM discipline. As such, they will be busy, taking other demanding classes in addition to calculus. The degree programs that require the full three semesters of calculus probably expect their students to have a foundational understanding of vector calculus. Chapter 18, the last in the book, represents the culmination of the theory and applications within the covers of this book.

That image shaped both the exposition and the exercises in each section.

A section begins with a brief introduction. Then it quickly moves to an informal presentation of the central idea on the section, followed by numerous examples. Formal proofs are given only after the student has a feel for the core of the section.

Those proofs are what hold the course together. They also serve as a constant review. For this reason we chose student-friendly proofs, adequately motivated. For instance, instead of the elegant, short proof that absolute convergence of a series implies convergence, we employed a longer, but more revealing proof. We avoid pulling tricks out of thin air; hence our new mativation for the cross product. Where one proof will do, we do not use two. Also, rather than proving the theorem in complete generality, we may treat only a special case, if that case conveys the flow of the general proof. As we assembled the exercises we labeled them R (routine), M (medium), and C (challenging), to make sure each section includes a wide variety of exercises. The R exercises focus on definitions and algorithmic calculations. The M exercises require more thought and the C exercises either demand a deeper understanding or offer an alternative view of the material.

In order to keep each section as short as feasible, we concentrate on the mathematics. We avoid the temptation to bring into the text too many applications. Not only would this make the sections too long to be read by a busy student, but we could not do justice to the applications. Instead, because applications are the reason most students study calculus, each chapter concludes with a thorough treatment of a application in a section called "Calculus is Everywhere" (CIE). Because each CIE stands alone, students and instructors are free to deal with each CIE as they please, depending on time available and on interest: skip it, glance at it, browse through it, or read it carefully. The presence of the Calculus is Everywhere sections allow us to replace exercises that start with a detailed description of an application and end with a trivial bit of calculus.. Our guiding theme is do one thing at a time, whehter it's exposition, an example, or an exercise.

As we worked on each section we asked ourselves several questions: Is it the right length? Does it get to the point quickly? Does it focus on one idea and correspond to one (50-minute) lecture? Are there enough examples? Are there enough exercises, with an appropriate balance of routine, medium, and challenging?

Curvature is treated twice, first (Section 9.6) in the plane, without vectors, and later (Section 15.2), in space, with vectors. We do this for two reasons. First, it provides students with a background for appreciating the vector approach. Second, it reduces the vector-based treatment to a reasonable length.

Many students will use vector analysis in engineering and physics courses. One of us sat in on a sophomore-level electromagnetic course in order to find out how the mathematical concepts were applied and what was expected of the students. That inspired a major revision of significant parts of Chapters 15 to 18.

In addition, the new edition reaches limits and derivatives as early as possible, and as simply as possible. Also, we introduce the permanance property (Theorem 2.5.5 in Section 2.5), which implies that a continuous function that is positive at a number remains positive nearby. This is referred to several times; hence we give it a name.

The controversy about what to do about $\epsilon-\delta$ proofs will never end. Therefore, in our text the instructor is free to choose what to do about such proofs. Limits are introduced in Chapter 2, where they are applied to continuous functions. Derivatives are introduced as an application of limits in the first seven sections of Chapter 3. Then, to keep our treatment student-friendly, we broke our discussion of the precise definition of limits into two sections. Section 3.8 treats limits at infinity because the diagrams are easier and the concept is more accessible. The second section, Section 3.9, deals with limits at a number. A rigorous proof is given there of the permananence property, illustrating the power of the $\epsilon-\delta$ approach to demonstrate something that is not intuitively obvious. A similar approach continues throughout the book by keeping the main exposition student-friendly and the mathematics logically consistent with the more rigorous details appearing in C-level exercises. We believe this gives the instructor and the student opportunities to include as much of these details as is appropriate for them.

We spent a lot of time thinking about and discussing what to do about differential equations. Most modern calculus books have a chapter for differential equations. What they do in that chapter varies widely. We finally agreed on what appears in Chapter 13. This is a short chapter (only Chapter 10) on applications of sequences is shorter). The presentations given in this chapter are only an introduction to some of the main topics of a full introductory course in differential equations: slope fields, separable differential equations, integrating factors for first-order linear differential equations, solutions of homogeneous and nonhomogeneous second-order linear differential equations with constant coefficients, and Euler's method. Having this exposure before taking a separate differential equations course should benefit students.

Throughout the book we include exercises that ask only for computing a derivative or an integral. These exercises are intended to keep your differentiation and integration skills sharp. We do not want to assign exercises that explore a new concept the additional responsiblity of offering extensive practice in calculations. This is another example of our general principle: do only one thing at a time, and do it clearly.

Another of our objectives is to help students to develop their mathematical maturity to a level that allows the student to understand the vector analysis in the final chapter. For instance, we often include an exercise which asks the student to state a theorem in their own words without mathematical symbols. We had found while doing some pro-bono tutoring that students do not read a theorem carefully. No wonder they did not know what to do when a supposedly routine exercise asked them to verify a theorem in a particular case.

We enjoyed putting this book together for you. We hope it is an effective tool in your learning of calculus. Enjoy!

Sample Course Outlines

The chapters and sections have been created to be most compatible with a three semester sequence of courses.

Calculus I would be based on the 37 sections in Chapters 1 to 6. This leaves time to include some of the eight (8) CIE applications at the ends of these chapters, and to include a few sections from Chapters 7 or 8 if you wish to include additional applications or techniques of integration.

Chapters 7 to 12 provide another 37 sections on applications and techniques for evaluating integrals, polar coordinates and plane curves, sequences, and series. To round out the course, there are seven (7) more CIE applications — or use one or more of the CIEs from an earlier chapter to reinforce previous topics. Other options are to include topics from Chapter 13, if you want to provide an introduction to differential equations, or Chapter 14, to get a head start on Calculus III.

The third course, Calculus III or Multivariable Calculus, is supported by Chapters 14 to 18. Because these chapters contain a combined 35 sections, there are opportunities to spend more time on some topics without having to rush to have time for students to absorb the final topics in Chapter 18. There are also nine (9) more CIE applications that can be used to complement the material in these chapters.

The differential equations chapter, Chapter 13, is optional and flexible. It can be omitted without interfering with anything encountered in subsequent chapters. In fact, Chapter 13 could be covered at any time after students have a solid foundation evaluating definite and indefinite integrals (Chapter 8). One idea is to use this material as an opportunity for students to get more practice evaluating integrals. Even if Chapter 13 is omitted, the two CIEs at the end of this chapter can easily be covered anytime after completing Chapter 7.

Acknowledgments

This book took a long time to complete, and benefitted from the contributions of many people.

The first edition of Sherman Stein's *Calculus and Analytic Geometry* was published by McGraw-Hill in 1973. The fifth edition, published in 1992, was produced with the assistance of Anthony Barcellos. In 2005, Robert Ross approached Douglas Meade with the idea of partnering with Stein to create a new calculus book based on the strengths of Stein's original book. In 2009, when economic realities and changes in the publishing world led to McGraw-Hill to lose interest in this project, the authors sought other options.

In 2011, the Mathematical Association of America (MAA) expressed interest in publishing this book. Graphical support for creating the figures was provided by Rebecca Elmo. Bev Ruedi and Carol Baxter worked to support the authors for several years, until the MAA transferred its book publishing to the American Mathematical Society (AMS) and Steve Kennedy became the primary point-of-contact. Upon Steve's retirement in 2022, Sergei Gelfand assumed this role.

All of the figures in the final manuscript were created by Joy Markel.

Many colleagues and other reviewers have read various parts of this book over the years. Daniel Drucker (Wayne State University) was particularly supportive and encouraging during the early years. Philip Yasskin (Texas A&M University) has been a sounding board from the beginning. Anthony Wexler (Mechanical and Aerospace Engineering, UC-Davis) provided valuable insight from the perspective of a non-mathematician.

Early in the project some preliminary solutions were prepared by University of South Carolina undergraduate students Major Brightwell, William (Cole) Franks, Taylor Jones, and Kimberly Selsor, with supervision from Stephen Kaczkowski (SC Governor's School for Science and Mathematics).

The final Solutions Manual is almost entirely attributable to the persistence of Dean Hickerson. Dean was also a very focused and reliable proof reader, both for basic English grammar and for subtle mathematical nuances. His attention to detail has significantly impacted the final product.

We are thankful for each of these people for their contributions over the many years of work that has gone into this project. We apologize to anyone whose name we might have omitted. Thank you to each of you!

The two people who have been most supportive have been our wives, Hannah and Betsy. They may not have contributed much to the mathematics, but they have willingly and lovingly supported and encouraged us throughout this process. They sacrificed so much of their family life, including some canceled and rescheduled vacations. Thank you!

Sherman K. Stein
Douglas B. Meade

Preface for the Instructor

Several principles guided the writing of this text:

First, because students are busy taking other courses we keep the exposition to the point. Successful students do read the explanations, not just the exercises.

Second, do one thing at a time. Most sections focus on one idea and correspond to one one-hour class meeting. In the sections we use just enough applications to motivate the mathematics, and then revisit the same application to continue to develop the mathematical foundation over multiple classes. One example of this is our treatment of Taylor series. First, in Section 5.4, we show how a higher derivative influences the growth of a function. (This is not new information for anyone who has driven a car.) In the next section we build upon this foundation to estimate the error when using a Taylor polynomial to approximate the value of a function. Then, in Sections 6.5 (Exercises 44 and 45) and Section 6.S (Exercise 76), the growth theorem is used to estimate the errors in estimating a definite integral.

Because real-world applications are so important we dedicated twenty-six sections to them under the title "Calculus is Everywhere" (CIE). These topics vary from uniform sprinklers to the path of the rear wheel of a scooter, to planets revolving around stars. Most are placed at the end of the chapter with the most relevant mathematical content.

The purpose of "one topic at a time" also applies to the chapters. For instance, the chapter on multiple integrals covers integrals over plane regions, surfaces, and solid regions. This chapter and the one on vector fields set the stage to treat the final chapter on the three big theorems of vector analysis without interruptions.

Third, write for the students. We pay great attention to motivating theorems and give rigorous proofs for only some of them. And the proofs that are provided are intended to be student friendly. For instance, rather than derive the formulas in polar coordinates for arc length from the one in rectangular coordinates, we use a small triangle that makes the formula easy to grasp and remember. Also, rather than using the equation $a_n = (a_n + |a_n|) - |a_n|$ to show that absolute convergence of a sequence implies convergence, we show that the subsequences of positive and of negative terms converge. In the same spirit, we use geometric series, not the binomial theorem, to differentiate x^n. (Chapter 1 has a section reviewing geometric series.)

Most sections begin with a brief introduction, move to an intuitive discussion, statement of the theorem, examples, and conclude with the proof, followed by a summary of the section. Chapter summaries, which students use, offer overviews that individual sections cannot.

Fourth, help the students develop analytical skills needed to understand the final chapters. To do this we include many conceptual exercises. Some at the very start of an exercise set may ask the student to state a theorem in symbols different from the ones used in the text, or to state it in words and no symbols. This is intended to discourage rote memorization in which symbols are not attached to ideas.

We think that our guiding principles are the basis for a teachable book. That the reaction of the anonymous mathematician who edited a draft version of the manuscript was "I wish I had studied calculus from this book" suggests that we have accomplished what we set out to do.

Preface for the Student

As we wrote this text we kept in mind that students are busy taking several courses, not just calculus. For that reason we kept the sections as short as possible and still explain and illustrate the ideas. Reading the text before beginning the homework not only will save you time but will also help you to master the material.

Calculus consists of concepts and techniques of problem solving and calculation. Software can do the calculations but it cannot grasp an abstract concept or identify steps needed to solve a problem. It is the concepts that hold calculus together in one piece rather than as a wild assortment of disconnected observations. In particular, the proofs of the theorems are not included just to show something is true. After all, some of the theorems have been around for centuries and no one questions their truth. We pay great attention to motivating theorems and give rigorous proofs for only some of them. The proofs that we do provide are intended for student consumption. To reinforce our emphasis on building intuition and understanding, the proofs provide an automatic review and deeper perspective on calculus. The ideas in the proofs are the ideas that serve as a basis for the applications of calculus.

When studying a section, first read it, focusing on the main ideas and examples, and including any proofs. Do not expect to understand everything after just one reading what a theorem says, using the examples to reinforce the main points. Then check your understanding by working some of the exercises. That is far more efficient than the reverse order many students follow.

Almost all chapters end with one or more sections titled “Calculus is Everywhere”, which may not be covered in your class. In these sections you will encounter some of the diverse ways that calculus has been, and is being, applied. If you browse through them, you will see how calculus helps to design a sprinkler that uniformly distributes water to a lawn, to predict the path of the rear wheel of a scooter, to analyze the flow of a liquid in a pipe with diameter less than a millimeter, or to accelerate an interplanetary probe. Applications are the reason most students study calculus, so we gave these topicss a prominent and thorough treatment. We hope you enjoy them.

Though the copy editor removed (most of) our jokes, we think you will find the text has a light touch and is readable. In particular, the Sam-and-Jane exercises may provide an amusing but instructive diversion. These exercises present a mathematical disagreement between two students, Sam and Jane, which the reader is asked to settle. In one case both students are wrong. At this moment, even before they begin to study calculus from our book, they have a dispute.

SAM: When I study math I do the homework first. I read the book only if I have trouble.
JANE: That’s ridiculous. I study the text first. Then the homework is easy.
SAM: Well, I’ll look for an example that’s like the homework problem.
JANE: I study the theorem and proofs first. Then the example make sense. This book has lots of examples, but there are still lots of exercises that are not like any example.
SAM: Oh.
JANE: I bet you’ll be looking for help before long.

Contents

Chapter 1

Pre-Calculus Review

This chapter reviews precalculus concepts needed in all subsequent chapters.

Since calculus is the study of functions, Section 1.1 begins with a review of the terminology and notation used when talking about them. In Section 1.2 fundamental types of functions are reviewed: power functions, exponentials, logarithms, and trigonometric functions. Section 1.3 describes how functions can be combined to create new functions.

The final two sections review two important topics that will be used often, geometric series in Section 1.4 and logarithms in Section 1.5.

1.1 Functions

This section reviews several ideas related to functions: piecewise-defined functions, one-to-one functions, inverse functions, and increasing or decreasing functions.

Definition of a Function

The fundamental notion in mathematics is a **set**. A set is any collection of things, such as numbers, points, or apples. For instance, a set of two apples illustrates the concept "two". The objects in a set are called **members** of that set. Almost as fundamental is the notion of **function**.

Definition: *Function*

Assume that X and Y are sets. A **function** from X to Y assigns one (and only one) member in Y to each member in X.

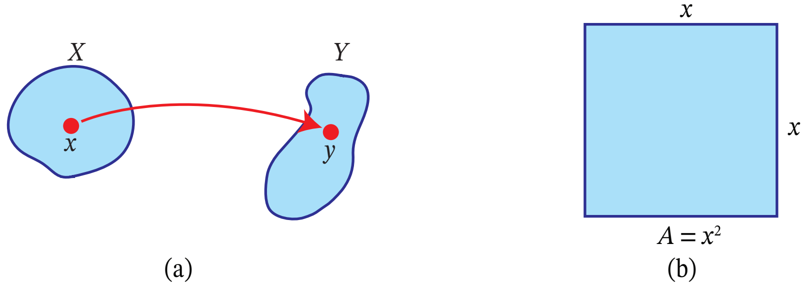


Figure 1.1.1

The notion of a function is illustrated in Figure 1.1.1(b), where the member y in Y is assigned to the member x in X. Usually X and Y will be sets of numbers.

The area A of a square depends on the length of its side x and is given by the formula $A = x^2$. (See Figure 1.1.1(a).)

In this case both X and Y are the set of positive real numbers. The function A assigns to each positive number the positive number x squared. In the notation of functions we may write $A(x) = x^2$. When referring to functions in general we may write $f(x)$ or just f. For the sake of brevity one speaks of the function f or the function $f(x)$, both short for $y = f(x)$ (which is read as "y equals f of x").

If $f(x) = y$, x is called the **input** or **argument** and y is called the **output** or **value** of the function at x. Also, x is called the **independent variable** and y the **dependent variable**.

A function may be given by a formula, as in the function $A(x) = x^2$. Because A depends on x, we say that "A is a function of x." Because A depends on only one number, x, it is called a function of a *single* variable. The area A of a rectangle depends on its length l and width w; it is a function of *two* variables, $A(l, w) = lw$.

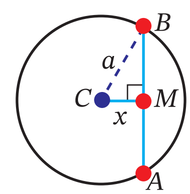
Intervals

Most of the sets we deal with in calculus are intervals. The following are standard notations:

$[a, b]$ the **closed interval** containing all numbers between a and b including both a and b, in short $\{x : a \le x \le b\}$;
(a, b) the **open interval** containing all numbers between a and b excluding both a and b, in short $\{x : a < x < b\}$;
$[a, b)$ the **half-open interval** containing all numbers between a and b including a but not b, in short $\{x : a \le x < b\}$;
$(a, b]$ the **half-open interval** containing all numbers between a and b including b but not a, in short $\{x : a < x \le b\}$;
$[a, \infty)$ the **unbounded interval** containing all numbers larger than or equal to a, in short $\{x : x \ge a\}$;
(a, ∞) the **unbounded interval** containing all numbers larger than a, in short $\{x : x > a\}$;
$(-\infty, a]$ the **unbounded interval** containing all numbers smaller than or equal to a, in short $\{x : x \le a\}$;
$(-\infty, a)$ the **unbounded interval** containing all numbers smaller than a, in short $\{x : x < a\}$;
$(-\infty, \infty)$ the **set of all numbers**, in short $\{x : -\infty < x < \infty\}$.

Ways to Write and Talk about a Function

There are several ways to describe the function that assigns to each argument x the value x^2. One may write $x \mapsto x^2$ (and say "x goes to x^2" or "x is mapped to x^2"). Or one may say simply, "the formula x^2", "the function x^2", or sometimes just "x^2." Using this abbreviation, we might say, "How does x^2 behave when x is large?" Some people object to just "x^2" because they fear that it might be misinterpreted as the number x^2, with no sense of a general assignment. In practice, the context will make it clear whether x^2 refers to a number or to a function.



$f(x)$ = length of AB
Figure 1.1.2

EXAMPLE 1. In the circle of radius a shown in Figure 1.1.2 assume that $f(x)$ is the length of chord AB at a distance x from its center. Find a formula for $f(x)$.

A **circle** is a curve and a **disk** is the region inside a circle.

SOLUTION We are trying to find how the length $|AB|$ varies as x changes. That is, we are looking for a formula for $|AB|$, the length of AB, in terms of x.

Before searching for the formula, it is a good idea to calculate $f(x)$ for some easy inputs. They can serve to check the formula we work out.

In this case $f(0)$ and $f(a)$ can be read at a glance at Figure 1.1.2: $f(0) = 2a$ and $f(a) = 0$. (Why?) Now let us find $f(x)$ for all x in $[0, a]$.

Let M be the midpoint of the chord AB and C the center of the circle. Because $|CM| = x$ and $|CB| = a$, the Pythagorean theorem gives $x^2 + |BM|^2 = a^2$, Hence $|BM| = \sqrt{a^2 - x^2}$ and $|AB| = 2\sqrt{a^2 - x^2}$. Thus

$$f(x) = 2\sqrt{a^2 - x^2}.$$

Does the formula give the correct values at $x = 0$ and $x = a$? ♦

Domain and Range

The set of inputs and the set of outputs of a function are essential parts of the definition of a function. They have special names, which we now introduce.

Definition: ***Domain and range***

Assume that X and Y are sets and f a function from X to Y. The set X is called the **domain** of the function. The set of all outputs of the function is called the **range** of the function. (The range is part or all of Y.)

When the function is given by a formula, the domain is usually understood to consist of all the numbers for which the formula is defined. In Example 1 the domain is $[0, a]$ and the range is $[0, 2a]$.

When using a calculator you must pay attention to the domain corresponding to a function key or command. If you enter a negative number as x and press the $\sqrt{x}$-key to calculate the square root of x you will get no result. It might display an E for "error" or start flashing. Your error was entering a number not in the domain of the square root function.

You can also get into trouble if you enter 0 and press the $1/x$-key. The domain of $1/x$, the **reciprocal function**, consists of all numbers except 0.

Try it. What does your calculator do? Some advanced calculators go into "complex number" mode to handle square roots of negative numbers. *No* calculator, however advanced, can permit division by zero.

Graph of a Function

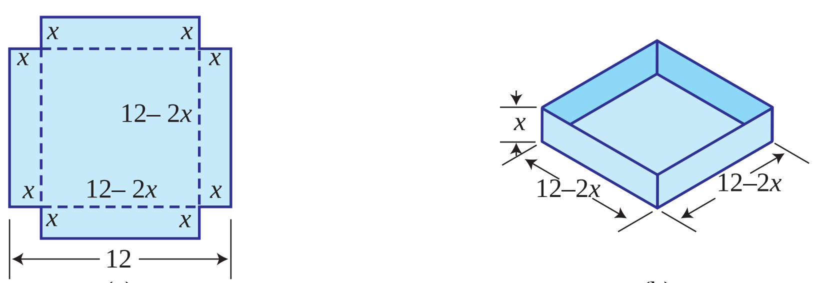
If both the inputs and the outputs of a function are numbers, we can draw a picture of the function, called its **graph**.

Definition: ***Graph of a function***

If f is a function whose inputs and outputs are numbers, the **graph** of f consists of those points (x, y) in the xy-plane such that $y = f(x)$.

The next example illustrates the usefulness of a graph to better understand a problem that sounds complicated when described only with words.

EXAMPLE 2. A tray is to be made from a square piece of cardboard by cutting identical squares from each corner and folding up the flaps. The size of the cardboard is $12'' \times 12''$. Find how the volume of the tray depends on the size of the squares.



(a) (b)

Figure 1.1.3

SOLUTION When the side of each cut-out square is x inches, the resulting tray is shown in Figure 1.1.3(b).

The volume $V(x)$ of the tray is the height, x, times the area of the base $(12-2x)^2$,

$$V(x) = x(12-2x)^2. \tag{1.1.1}$$

The domain of V contains all values of x that lead to an actual tray. This means that x cannot be negative or more than 6, half the length of the sides. Thus, the largest corners that can be cut out have sides of length 6 inches. Therefore the domain of interest is only the interval $[0,6]$. The trays obtained when $x=0$ or $x=6$ are peculiar. What are their volumes?

The short table of inputs and corresponding outputs in Table 1.1.1 will help sketch the graph of (1.1.1).

x (in)	-1	0	1	2	3	4	5	6	7
$V(x)$ (in^3)	-196	0	100	128	108	64	20	0	28

Table 1.1.1

For $x=6$, the quantity $12-2x=0$, so $V(6)=0$. When x is greater than 6, two factors in the formula for $V(x)$ are negative and one is positive. Thus $V(x)$ becomes large for large values of x.

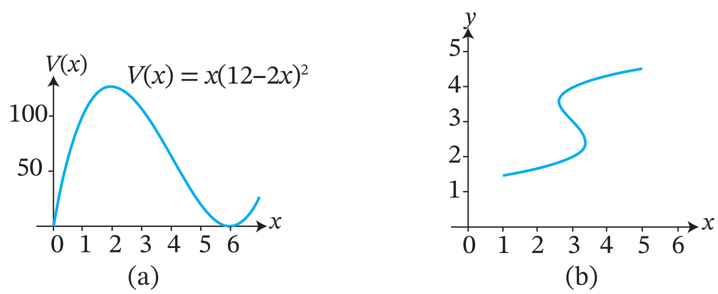


Figure 1.1.4

NOTE: When drawing the graph of a function, as in Figure 1.1.4(a), it is often convenient to have the scale on the vertical axis different from that on the horizontal one. This change preserves the general shape of the graph but stretches or shrinks it vertically.

Whether creating a graph or reading information from a graph, it is important to remember to pay attention to the scale being used on each axis.

For negative x, two factors in (1.1.1) are positive and one is negative. (Which factor is negative?) Thus $V(x)$ is negative and has large absolute value for negative inputs of large absolute value. Figure 1.1.4(a) displays the graph of $V(x)$. Only the part of the graph above the interval $[0,6]$ is meaningful in the tray problem. Other values of x have nothing to do with actual trays. ♦

In order for a curve in the xy-plane to be the graph of a function each vertical line must meet the curve in at most one point. If the vertical line $x=a$ meets the curve twice, say at (a,b) and (a,c), there will be the two outputs, b and c, for the single input a, and it is no longer a function.

Theorem 1.1.1: Vertical Line Test

A set in the xy-plane is the graph of a function f if and only if every vertical line $x=a$ intersects the graph of $y=f(x)$ at most once.

For the graph of a function $y=f(x)$, the number a is in the domain of f if and only if the vertical line $x=a$ intersects the graph.

Figure 1.1.4(b) shows a graph that does not pass the vertical line test. The corresponding input-output table would have three entries for each input x between 2.5 and 3.5, two entries for $x = 2.5$ and $x = 3.5$ and exactly one entry for each input $1 \le x < 2.5$ or $3.5 < x \le 5$. But this graph is the graph of a function when considered only on the domain $[1, 2.5]$ and $[3.5, 5]$.

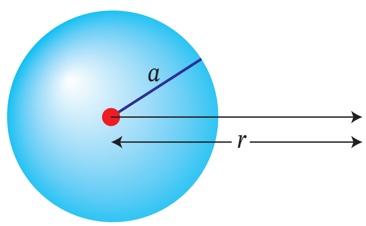
In Example 2 the function is described by a single formula, $V(x) = x(12-2x)(12-2x)$. But a function may be described by different formulas for different intervals or points in its domain, as in the next example.

EXAMPLE 3. A hollow and very thin sphere of radius a has mass M, distributed uniformly throughout its surface. Describe the size of the gravitational force it exerts on a particle of mass m at a distance r from the center of the sphere. (See Figure 1.1.5(a).)

SOLUTION Let $f(r)$ be the force at a distance r from the center of the sphere. In an introductory physics course it is shown that the sphere exerts no force at all on objects in its interior. Thus for $0 \le r < a$, $f(r) = 0$.

The sphere attracts an external particle as though all its mass is at its center. Thus, for $r > a$, $f(r) = GMm/r^2$, where G is a constant whose numerical value depends on the units used for measuring length, time, and mass.

It can be shown by calculus that for a particle on the surface, that is, for $r = a$, the force is $GMm/(2a^2)$. The graph of f is shown in Figure 1.1.5(b). ♦



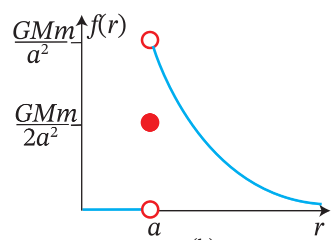


Figure 1.1.5

In a graph that consists of several different pieces, such as Figure 1.1.5(b), the presence of a point on the graph of a function is indicated by a solid dot (•) and the absence of a point by a hollow dot (◦).

The formula for the gravitational force function in Example 3 changes for different parts of its domain, so we write it this way:

$$f(r) = \begin{cases} 0 & \text{if } 0 \le r < a \\ \dfrac{GMm}{2a^2} & \text{if } r = a \\ \dfrac{GMm}{r^2} & \text{if } r > a \end{cases}$$

Such a function is called a **piecewise-defined function**.

Inverse Functions

If we know a particular output of the function $f(x) = x^3$ we can figure out what the input must be. For instance, if $x^3 = 8$, then $x = 2$: we can go backwards from output to input. This is not possible with the function $f(x) = x^2$. If we are told that $x^2 = 25$, we do not know what x is. It can be 5 or -5. However, if we are told that $x^2 = 25$ and that x is positive, then we know that x is 5. This brings us to the notion of a one-to-one function.

Definition: *One-to-One Function*

A function f that does not assign the same output to two different inputs is called a **one-to-one function**. That is, if $f(x_1) = f(x_2)$, then $x_1 = x_2$.

Definition: *Inverse Function*

If f is a one-to-one function, the **inverse function** is the function g that assigns to each output of f the corresponding input. That is, if $f(x) = y$ then $g(y) = x$.

Theorem 1.1.2: Horizontal Line Test

The graph of a one-to-one function never meets a horizontal line more than once. (See Figure 1.1.6.)

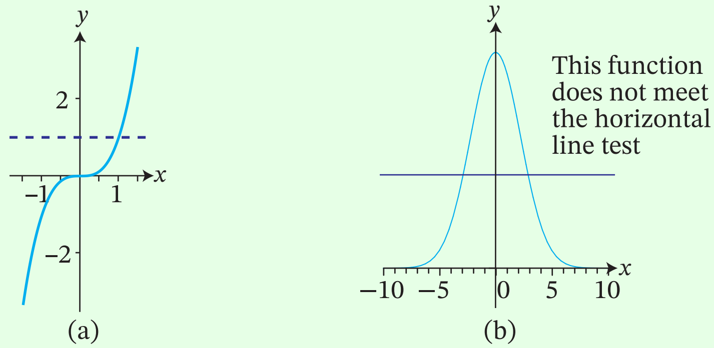


Figure 1.1.6

The function graphed in Figure 1.1.6(a) is one-to-one because it passes the Horizontal Line Test. In Figure 1.1.6(b), since horizontal lines $y = A$ for $0 < A < 1$ intersect the graph of the function twice, the Horizontal Line Test is not passed and so this function is not one-to-one on its domain, $[-10, 10]$.

The function $f(x) = x^3$ is one-to-one on the entire real line. A few entries in the tables for $f(x)$ and its inverse function are shown in Table 1.1.2(a) and (b), respectively.

input	1	2	$\frac{1}{2}$	3	-2
output	1	8	$\frac{1}{8}$	27	-8

(a)

input	1	8	$\frac{1}{8}$	27	-8
output	1	2	$\frac{1}{2}$	3	-2

(b)

Table 1.1.2

In this case the inverse function has a name, the **cube root function**. If $y = x^3$, then x is the cube root of y, so $x = y^{1/3}$.

The inverse function of the one-to-one function f is denoted $\operatorname{inv} f$ or f^{-1}.

Observation 1.1.3: *Notation for Inverse Functions*

The use of $\operatorname{inv} f$ to denote the inverse function of f is based on the fact that many calculators have a button marked inv to indicate the inverse of a function. The mathematical notation for the inverse function of f is f^{-1} or $\operatorname{inv} f$.

NOTE: The -1 *is not an exponent*, and in general the inverse and reciprocal functions are different: f^{-1} is not equal to $\frac{1}{f}$.

EXAMPLE 4. For the constants $m \neq 0$ and b define $f(x) = mx + b$. Show that f is one-to-one and describe its inverse function.

SOLUTION We must show that $f(x_1) = f(x_2)$ forces $x_1 = x_2$. Let us see if it does. If $f(x_1) = f(x_2)$ we have

$$\begin{aligned} mx_1 + b &= mx_2 + b \\ mx_1 &= mx_2 && \text{(subtract } b \text{ from both sides)} \\ x_1 &= x_2 && \text{(divide both sides by } m \neq 0\text{)} \end{aligned}$$

Because $f(x_1) = f(x_2)$ only when $x_1 = x_2$, f is one-to-one.

This problem can also be analyzed graphically. The graph of $y = f(x)$ is the line with slope m and y-intercept b. (See Figure 1.1.7(a).) Since $m \neq 0$, the graph passes the horizontal line test.

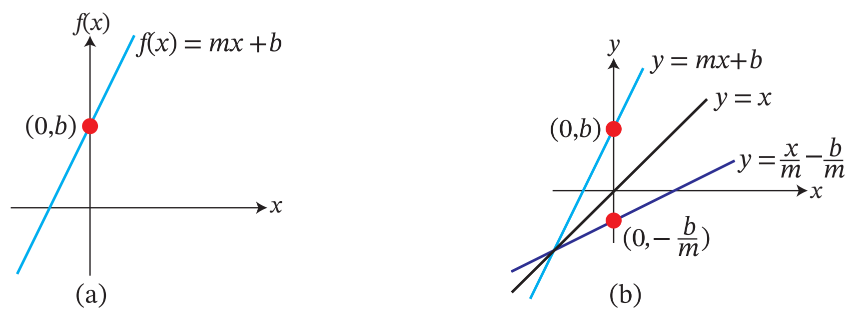


Figure 1.1.7

To find the inverse function, solve the equation $y = f(x)$ to express x in terms of y:

$$\begin{aligned} y &= mx + b \\ y - b &= mx && \text{(subtract } b \text{ from both sides)} \\ \frac{y-b}{M} &= x && \text{(divide by } m \neq 0\text{)} \\ x &= \frac{y}{M} - \frac{b}{M} && \text{(move } x \text{ to left-hand side)} \\ y &= \frac{x}{M} - \frac{b}{M}. && \text{(switch } x \text{ and } y\text{)} \end{aligned}$$

We switched x and y in the last step so we could graph the function with x as the independent variable, which usually appears on the horizontal axis. The final result is the formula for the inverse function:

$$f^{-1}(x) = \frac{x}{M} - \frac{b}{M}.$$

The graph of the inverse function is also a line; its slope is $1/m$, the reciprocal of the slope of the original line, and its y-intercept is $-b/m$. (See Figure 1.1.7(b).) ◆

The Graph of an Inverse Function

When we know the graph of a one-to-one function, it is easy to draw the graph of the inverse function. If (a, b) is a point on the graph of the function f, that is, $b = f(a)$, then (b, a) is a point on the graph of $\operatorname{inv} f$. This is illustrated in Figure 1.1.8(a) in the case with $a = 1$ and $b = 3$.

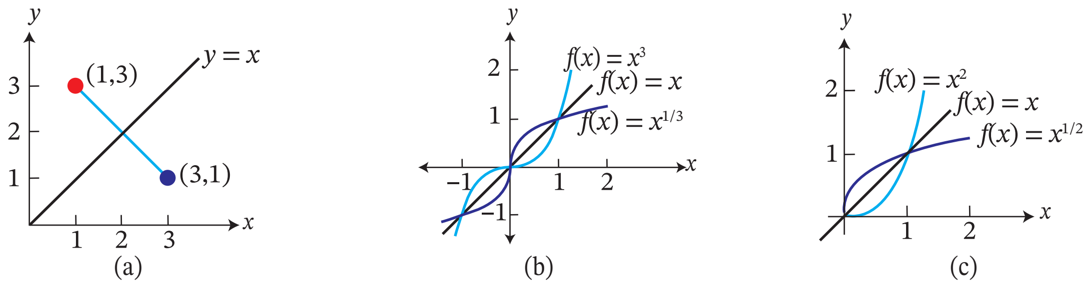


Figure 1.1.8

EXAMPLE 5. Draw the graphs of the inverse of (a) the cubing function given by $f(x) = x^3$, and (b) the squaring function $f(x) = x^2$ restricted to $x \geq 0$.

SOLUTION See Figure 1.1.8(b) and (c). ♦

Decreasing and Increasing Functions

A function is **increasing** on an interval if whenever x_1 and x_2 are in the interval and x_2 is greater than x_1, then $f(x_2)$ is greater than $f(x_1)$. As a pencil moves along the graph of f from left to right, it goes up. This is shown in Figure 1.1.9(a).

In the case of a **decreasing** function, outputs decrease as the input increases: if $x_2 > x_1$ then $f(x_2) < f(x_1)$. (See Figure 1.1.9(b).)

The graph of $f(x) = \sin(x)$ is shown in Figure 1.1.9(c). On its full domain, $[-\pi/2, \pi/2]$ the function is neither a decreasing function nor an increasing function. However, when the domain is restricted to the the interval $[-\pi/2, \pi/2]$ the values of $\sin(x)$ increase. On the interval $[\pi/2, 3\pi/2]$ the values of $\sin(x)$ decrease.

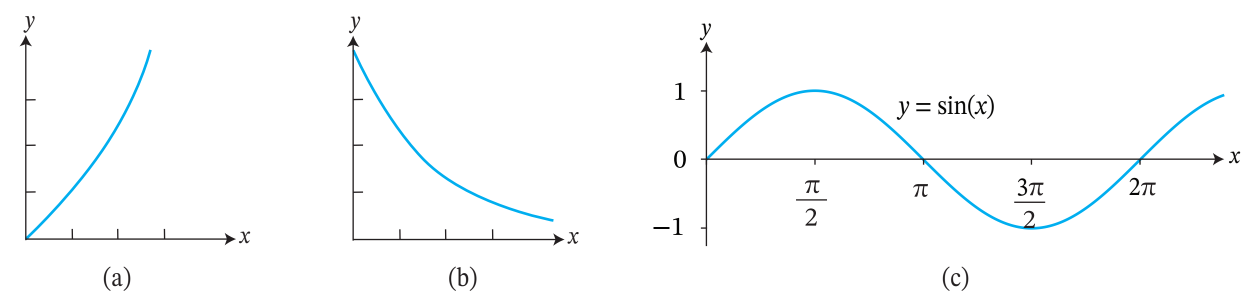


Figure 1.1.9

Monotone is from the Greek, *mono*=single, *tonos*=tone, which also gives us the word 'monotonous').

A function is said to be **monotonic** function on an interval if it is either increasing on the interval or decreasing on the interval. Monotonic functions always pass the horizontal line test.

For $k \neq 0$ and $x > 0$, x^k is a monotonic function. For $k < 0$, x^k is monotone decreasing for $x > 0$; for $k > 0$ it is monotone increasing for $x > 0$. The inverse of x^k is $x^{1/k}$. If $k = 0$, we have a constant function, $x^0 = 1$. It does not pass the horizontal line test, so has no inverse.

Because strict inequalities are used in the definitions of increasing and decreasing, we sometimes say these functions are **strictly increasing** or **strictly decreasing** on an interval. A function f is said to be **nondecreasing** on an interval if whenever x_1 and x_2 are in the interval and x_2 is greater than x_1, then $f(x_2) \geq f(x_1)$. The graph of a nondecreasing function is increasing except on intervals where it is constant. Likewise, f is **nonincreasing** on an interval if whenever x_1 and x_2 are in the interval and $x_2 > x_1$, then $f(x_2) \leq f(x_1)$.

The sign of a function's outputs provides another way to describe some functions. A function that has only positive outputs is called a **positive function**; for instance, 2^x. A **negative function** has only negative outputs; for instance, $-1/(1+x^2)$. A **nonnegative function** has outputs that are either positive or zero; for instance x^2. The outputs of a **nonpositive function** are either negative or zero, for instance, $\sin(x)-1$.

Summary

This section reviewed concepts that will be used throughout the coming chapters: interval, function, domain, range, graph, piecewise-defined function, one-to-one function, inverse function, increasing function, decreasing function, monotonic function, nondecreasing function, nonincreasing function, positive function, negative function, nonnegative function, and nonpositive function.

Every monotonic function has an inverse function and the graph of the inverse function is the reflection across the line $y=x$ of the graph of the original function.

A function can be described in several ways: by a formula, such as $V(x)=x(12-2x)^2$, by a table of values, or by words, such as "the volume of a tray depends on the size of the cut-out squares."

EXERCISES for Section 1.1

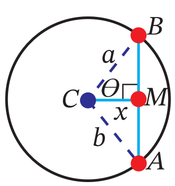


Figure 1.1.10

Exercises 1 to 4 refer to Figure 1.1.10.

1. Express the area of triangle ABC as a function of $x=\overline{CM}$.

2. Express the perimeter of triangle ABC as a function of x.

3. Express the area of triangle ABC as a function of θ.

4. Express the perimeter of triangle ABC as a function of θ.

In Example 2 a tray was formed from an 8" by 11" rectangle by removing squares from the corners. Find and graph the corresponding volume function for trays formed from sheets with the dimensions given in Exercises 5 to 8.

5. 4" by 13" **6.** 5" by 7" **7.** 6" by 6" **8.** 5" by 5"

In Exercises 9 and 10 decide which curves are graphs of (a) functions, (b) increasing functions, and (c) one-to-one functions.

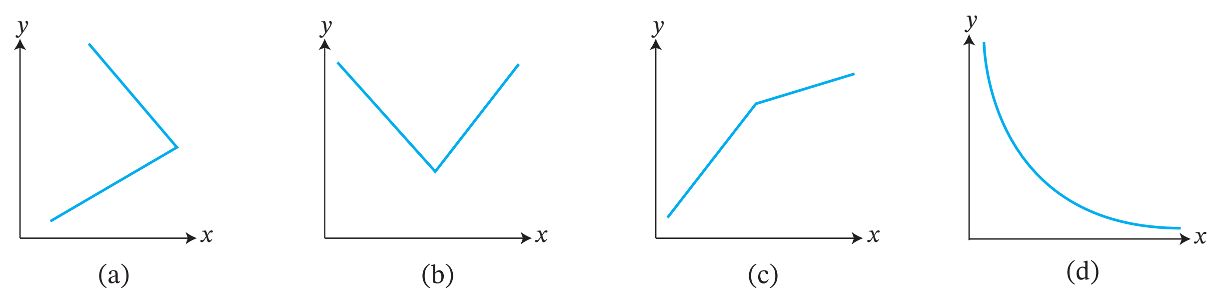


Figure 1.1.11

9. (a) Figure 1.1.11(a) and (b) Figure 1.1.11(b)

10. (a) Figure 1.1.11(c) and (b) Figure 1.1.11(d)

---

11. Let $f(x) = x^3$.

(a) Fill in this table

x	0	1/4	1/2	$-1/4$	$-1/2$	1	2
x^3							

(b) Graph f.
(c) Use the table in (a) to find seven points on the graph of f^{-1}.
(d) Graph f^{-1} (use the same axes as in (b)).

12. Let $f(x) = \cos(x)$, $0 \le x \le \pi$ (angles in radians).

(a) Fill in this table

x	0	$\pi/6$	$\pi/4$	$2\pi/3$	$\pi/2$	$3\pi/4$	π
$\cos(x)$							

(b) Graph f.
(c) Use the table in (a) to find seven points on the graph of inv cos.
(d) Graph inv cos (use the same axes as in (b)).

In Exercises 13 to 18 the functions are one-to-one. Find the formula for each inverse function, expressed in the form $y = g(x)$, so that the independent variable is labeled x.

13. $y = 3x - 2$

14. $y = \frac{x}{2} + 7$

15. $y = x^5$

16. $y = 3\sqrt{x}$

17. $y = \frac{1}{2}(x + \frac{1}{x})$ $(x \ge 1)$

18. $y = \frac{1}{2}(x - \frac{1}{x})$ $(x < 0)$

In Exercises 19 to 23 the slope of line L is given. Let L' be the reflection of L across the line $y = x$. What is the slope of the reflected line, L'? In each case sketch a possible L and its reflection, L'.

19. L has slope 2.

20. L has slope 1.

21. L has slope 0.

22. L has slope -1/3.

23. L has slope -2.

In Exercises 24 to 33 state the formula for the function f and give the domain of the function.

24. $f(x)$ is the perimeter of a circle of radius x.

25. $f(x)$ is the area of a circle of radius x.

26. $f(x)$ is the perimeter of a square of side x.

27. $f(x)$ is the volume of a cube of side x.

28. $f(x)$ is the total surface area of a cube of side x.

29. $f(x)$ is the length of the hypotenuse of the right triangle with legs have lengths 3 and x.

30. $f(x)$ is the length of the side $\overline{AB}$ in the triangle in Figure 1.1.12(a).

31. For $0 \le x \le 4$, $f(x)$ is the length of the path from A to B to C in Figure 1.1.12(b).

32. For $0 \le x \le 10$, $f(x)$ is the perimeter of the rectangle $ABCD$, one side of which has length x, inscribed in the circle of radius 5 shown in Figure 1.1.12(c).

33. A person at point A in a lake is going to swim to the shore ST and then walk to point B. She swims at 1.5 miles per hour and walks at 4 miles per hour. If she reaches the shore at point P, x miles from S, let $f(x)$ denote the time for her combined swim and walk. Obtain an algebraic formula for $f(x)$. (See Figure 1.1.12(d).)

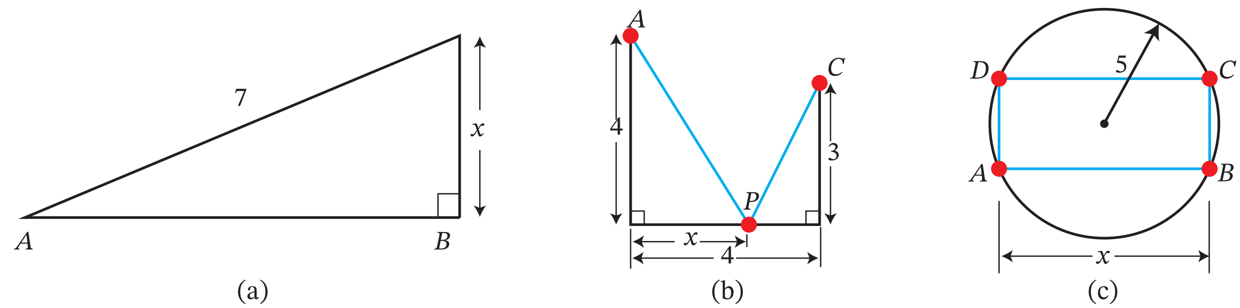


Figure 1.1.12

---

34. A camper at A will walk to the river, put some water in a pail at P, and take it to the campsite at B. Where should P be located to minimize the length of the walk, $AP + PB$? (See Figure 1.1.12(b).)

NOTE: Calculus-based methods for finding minima and maxima will be developed in Chapter 4.

35. The cost of life insurance depends on whether the person is a smoker or a nonsmoker. The following chart lists the annual cost for a male for a million-dollar life insurance policy.

age (yrs)	20	30	40	50	60	70	80
cost for smoker ($)	1150	1164	1944	4344	9864	26500	104600
cost for nonsmoker ($)	396	396	600	1490	3684	10900	41600

(a) Plot the data and sketch the graphs on the same axes for both groups of males.
(b) A smoker at age 20 pays as much as a nonsmoker of about what age?
(c) A smoker pays about how many times as much as a nonsmoker of the same age?

NOTE: A "smoker" is a person who has used tobacco during the previous three years.

36. Let $f(x)$ be the diameter of the largest circle that fits in a $1 \times x$ rectangle.
(a) Graph $y = f(x)$ for $x > 0$. (b) Give a formula for $f(x)$.

37. If f is an increasing function, what, if anything, can be said about f^{-1}?

38. On a typical summer day in the Sacramento Valley the temperature is at a minimum of 60° at 7 AM and a maximum of 95° at 4 PM.

(a) Sketch a graph that shows how the temperature may vary during one day, from midnight to midnight.
(b) A closed shed with little insulation is in the middle of a treeless field. Sketch a graph that shows how the temperature inside the shed may vary during the same period.
(c) Sketch a graph that shows how the temperature in a well-insulated house may vary. Assume that in the evening all the windows and skylights are opened when the outdoor temperature equals the indoor temperature, and closed in the morning when the two temperatures are again equal.

39. The monthly average water and air temperatures in Myrtle Beach, SC, are shown below.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Water Temp (°F)	51	52	57	62	69	77	81	83	80	73	65	55
Air Temp (°F)	56	60	68	76	83	88	91	89	85	77	69	60

REFERENCE: https://weatherspark.com/y/20111/Average-Weather-in-Myrtle-Beach-South-Carolina-United-States-Year-Round

(a) Sketch a graph that shows how the water temperature may vary during one calendar year, that is, from January 1 through December 31.
(b) Sketch a graph that shows how the difference between the air and water temperatures may vary during one calendar year. During what month is the water temperature difference greatest? least?
(c) During February, the water temperature increases 5° in 28 days so the average daily change is $5/28 \approx 0.1786°/\text{day}$. For each month, estimate the average daily change in the water temperature from one day to the next. During which month is this daily change greatest? least?
(d) Repeat (a) and (c) for the air temperature data.

Assume, for convenience, that the temperatures in the table are the temperatures on the first day of each month. In (d), "greatest" means the largest positive change and "least" means the largest negative change.

40. This problem grew out of a question raised by Rebecca Stein-Wexler, daughter of one of the authors, when cutting cloth for a dress. She wanted to cut out two congruent semicircles from a long strip of fabric 44 inches wide and 104 inches long. She guessed that the largest semicircles she could cut would have a radius of about 30 inches.

The radius, r, of the semicircles determines L, the length of fabric used, $L = f(r)$, as shown in Figure 1.1.13.

(a) In the figure it looks as though the radii AB and BC lie on the same straight line. Using geometry, show that they do.
(b) Show that $L = 2r + x$, where $x = |CD|$ is the horizontal distance between A and C.
(c) Draw a picture to show $f(22) = 44$.
(d) For $0 \le r \le 22$, show that $L = 2r$.
(e) For $22 \le r \le 44$, show that $L = 2r + 2\sqrt{r^2 - 484}$.
(f) From (e), deduce that $r = \dfrac{L}{4} + \dfrac{484}{L}$, for $44 \le L \le 88 + 44\sqrt{3}$.
(g) To find the largest semicircles that can be cut from the 104" of fabric available, note that to cut two 44" semicircles from a 44" wide swath of material requires $L = 88 + 44\sqrt{3} \approx 164.2$", which is longer than the length of cloth available. Show that for $L = 104$, r is about 30.654 inches, pretty close to Rebecca's estimate.

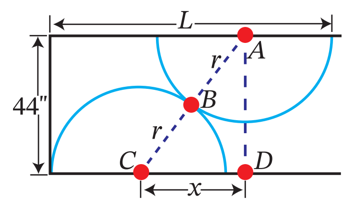


Figure 1.1.13

41. Let $f(x)$ be the length of the segment AB in Figure 1.1.14(a). (a) What are $f(0)$ and $f(a)$? (b) What is $f\left(\frac{a}{2}\right)$? (c) Find the formula for $f(x)$ and explain your solution.

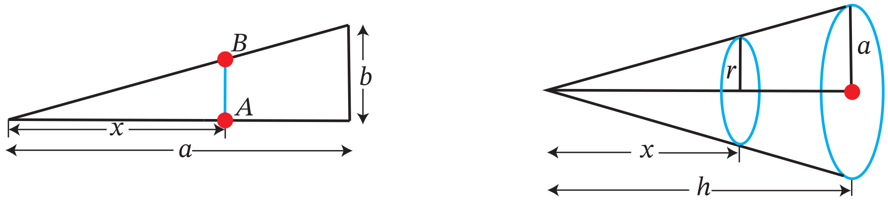


Figure 1.1.14

42. Let $f(x)$ be the area of the right circular cone cross section in Figure 1.1.14(b).
(a) What are $f(0)$ and $f(h)$? (b) Find a formula for $f(x)$ and explain your solution.

43. This is how the cost of a ride in a New York City taxi is calculated. At the start the meter reads \$2.50. For every fifth of a mile, 40 cents is added. Graph the cost as a function of distance travelled. For every two minutes stopped in traffic, 40 cents is added. Graph the cost as a function of distance traveled.

NOTE: The cost also depends on other factors.

44. (a) Find all functions of the form $f(x) = a + bx$, where a and b are constants, such that $f = \text{inv}(f)$.

Be sure to identify any cases where the inverse does not exist.

(b) Sketch the graph of one of the functions found in (a).

1.2 Basic Functions in Calculus

This section describes the basic functions in calculus. In the next section they are used to build more complicated functions.

The Power Functions

The first group of functions consists of the **power functions** x^k where the exponent k is a fixed nonzero number and the base x is the input. When the domain of x^k is restricted to positive numbers, it is one-to-one, and has an inverse $x^{1/k}$ with, again, a domain consisting of all positive numbers.

In Section 1.1 it was shown that the inverse of $f(x) = x^3$ is $f^{-1}(x) = x^{1/3}$ for all x. However, $g(x) = x^4$ does not pass the horizontal line test unless the domain is restricted to, say, nonnegative inputs, $[0,\infty)$. Thus the inverse of $g(x) = x^4$ is $g^{-1}(x) = x^{1/4}$ only for $x \geq 0$.

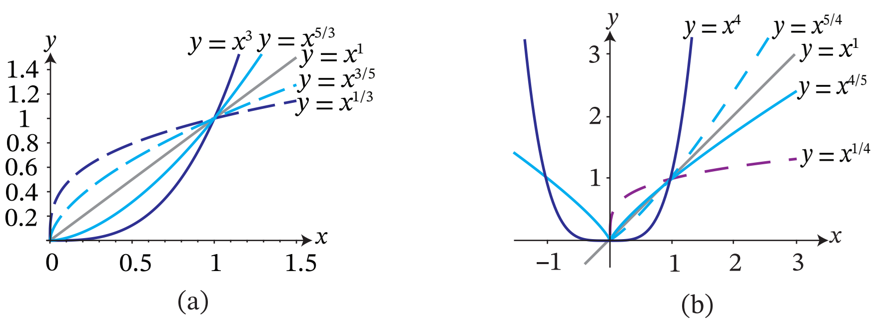


Figure 1.2.1

An important property of power functions is that their inverse functions are also power functions. When the exponent k is an even integer or a rational number (in lowest terms) whose numerator is even (2/3, 4/7, etc.) the graph of $y = x^k$ does not pass the horizontal line test unless the domain is reduced, usually to $[0,\infty)$. And when $k = 0$ we obtain the function x^0, which is constant (with all outputs equal to 1), the very opposite of being one-to-one.

The graphs of several power functions and their inverses are shown in Figure 1.2.1. In Figure 1.2.1(a) the black curve is $y = x$, which is its own inverse, the red curves are $y = x^3$ and $y = x^{1/3}$ (dashed), and the blue curves are $y = x^{5/3}$ and $y = x^{3/5}$ (dashed). In Figure 1.2.1(b) the black curve is $y = x$, the red curves are $y = x^4$ and $y = x^{1/4}$ (dashed), and the blue curves are $y = x^{4/5}$ and $y = x^{5/4}$ (dashed). Note that all of these functions are defined for all x except $y = x^{1/4}$ and $y = x^{5/4}$ which are defined only for $x > 0$.

The Exponential and Logarithm Functions

Next we have the **exponential functions** b^x where the base b is fixed and the exponent x is the input. The inverses of exponential functions are not exponential functions. The inverses are called **logarithms** and are the next class of functions that we will consider.

To be specific, let's take a look at the case $b = 2$ and study $f(x) = 2^x$.

As x increases, so does 2^x. So the function 2^x has an inverse function. (See Figure 1.2.2.) In other words, if $y = 2^x$, then if we know the output y we can determine the input x, the exponent, uniquely. For instance, if $2^x = 8$ then $x = 3$. This is expressed as $3 = \log_2(8)$ and is read as "the logarithm of 8, base 2, is 3."

The table of values of $\log_2(x)$ in Table 1.2.1 helps us graph $y = \log_2(x)$. Putting a smooth curve through the seven points in Table 1.2.1 yields the graph in Figure 1.2.2.

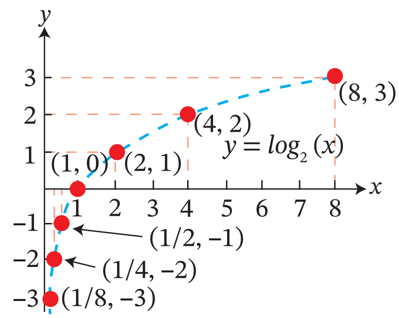


Figure 1.2.2

x	1	2	4	8	1/2	1/4	1/8
$\log_2(x)$	0	1	2	3	-1	-2	-3

Table 1.2.1

As x increases, $\log_2(x)$ grows slowly. For instance $\log_2(1024) = 10$, as every computer scientist knows. For x between 0 and 1, $\log_2(x)$ is negative. As x moves from 1 towards 0, $|\log_2(x)|$ grows very large. For instance, $\log_2(1/1024) = -10$.

Because $\log_2(x)$ is the inverse of the function 2^x, we can sketch the graph of $y = \log_2(x)$ by first sketching the graph of $y = 2^x$ and reflecting it across the line $y = x$.

For any positive base b, except $b = 1$, $\log_b(x)$ is defined similarly. For x and b both positive numbers, the logarithm of x to the base b, denoted $\log_b(x)$, is the power to which we must raise b to obtain x. That is, $y = \log_b(x)$ means $x = b^y$. Then, by the definition of the logarithm, for any $x > 0$,

$$b^{\log_b(x)} = x.$$

NOTE: It is common to use log (without a subscript) to abbreviate $\log_{10}$.

The Trigonometric Functions and Their Inverses

So far we have the power functions, x^k, the exponential functions, b^x, and the logarithm functions, $\log_b(x)$. The last major group of important functions consists of the trigonometric functions, $\sin(x)$, $\cos(x)$, $\tan(x)$, and their inverses (after shrinking their domains to make them one-to-one, as needed).

The trigonometric functions are periodic because their values repeat after a specific period. For example, $\tan(x+\pi) = \tan(x)$ for all x where $\tan(x)$ is defined. (And, when x is an odd multiple of $\pi/2$, $\tan(x)$ and $\tan(x+\pi)$ are both undefined.) A function f is a **periodic function** if there is a nonzero constant k such that $f(x+k) = f(x)$ for all x i n the domain of f. Note that if k is a period, so are $2k, 3k, \ldots$ and $-k, -2k, -3k, \ldots$. The smallest positive period is often singled out as "the period" of f. For example, $\sin(x)$ is periodic with period 2π and $\tan(x)$ is periodic with period π.

NOTE: In calculus angles are generally measured in radians.

$\sin(x)$ **and its inverse**

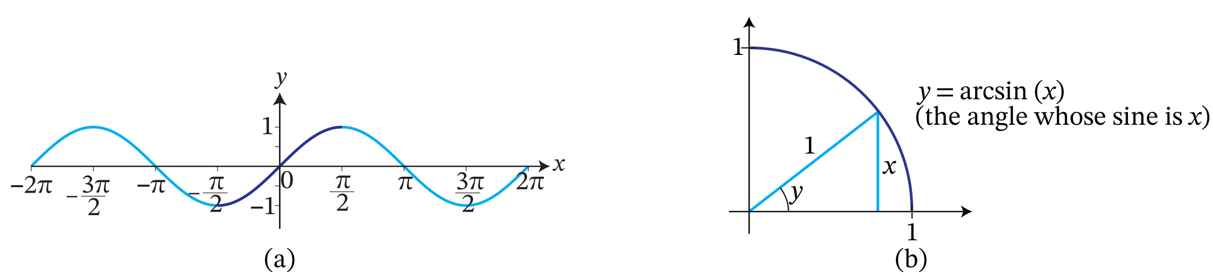


Figure 1.2.3

The graph of the **sine function** $\sin(x)$ has period 2π and is shown in Figure 1.2.3(a). Its range is $[-1, 1]$. On the domain $[-\pi/2, \pi/2]$, $\sin(x)$ is increasing and its values for these inputs already sweep out the full range.

When we restrict the domain of the function $\sin(x)$ to $[-\pi/2, \pi/2]$ it is one-to-one with range $[-1,1]$. This means the sine function has an inverse with domain $[-1,1]$ and range $[-\pi/2, \pi/2]$. The inverse sine function is denoted by $\arcsin(x)$, $\sin^{-1}(x)$, or $\operatorname{inv}\sin(x)$. For x in $[-1,1]$, $\arcsin(x)$ is the angle in $[-\pi/2, \pi/2]$ whose sine is x. In equations:

$$y = \arcsin(x) \iff \sin(y) = x.$$

For instance, $\arcsin(1) = \pi/2$ because the angle in $[-\pi/2, \pi/2]$ whose sine is 1 is $\pi/2$. Similarly, $\sin^{-1}(1/2) = \pi/6$, $\operatorname{inv}\sin(0) = 0$, $\arcsin(-1/2) = -\pi/6$, and $\sin^{-1}(-1) = -\pi/2$. A unit circle helps display these facts, as Figure 1.2.3(b) illustrates.

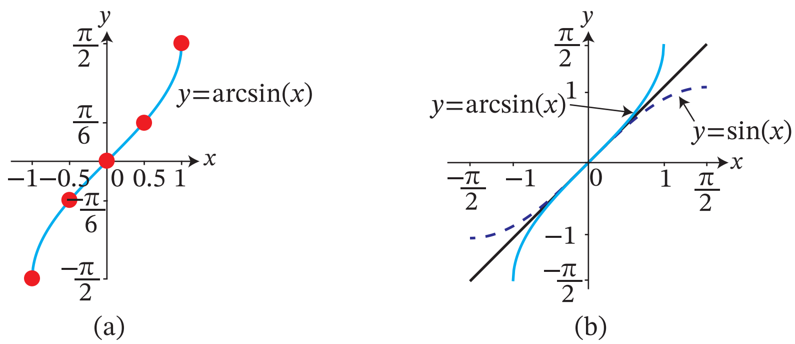


Figure 1.2.4

We could graph $y = \arcsin(x)$ with the aid of those five values. (See Figure 1.2.4(a).) However, it is easier to reflect the graph of $y = \sin(x)$ across the line $y = x$. (See Figure 1.2.4(b).)

We have to remember that while $\sin(\arcsin(x)) = x$ for all $-1 \le x \le 1$, $\arcsin(\sin(x)) = x$ only for $-\pi/2 \le x \le \pi/2$, the largest interval containing zero on which the graph of $y = \sin(x)$ is one-to-one.

$\cos(x)$ and its inverse

The graph of the **cosine function** $\cos(x)$ is shown in Figure 1.2.5(a).

It is not one-to-one, even if we restrict the domain to the domain used for $\sin(x)$, namely $[-\pi/2, \pi/2]$. Because $\cos(x)$ is decreasing on $[0,\pi]$ it is one-to-one on $[0,\pi]$. Moreover, the values of $\cos(x)$ for x in $[0,\pi]$ sweep out all possible values of the cosine function, namely $[-1,1]$.

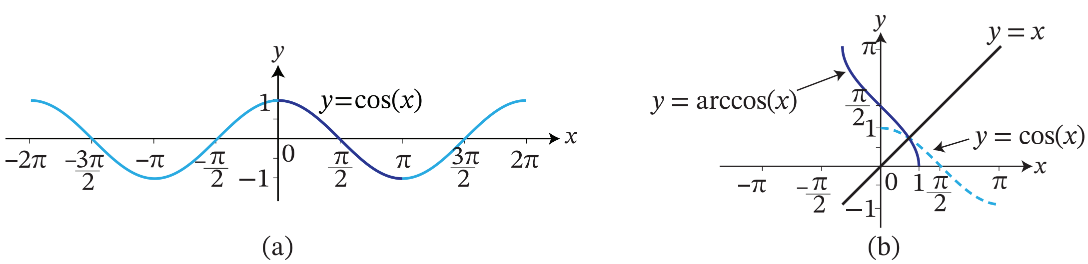


Figure 1.2.5

Because $\cos(x)$ is a one-to-one function on the domain $[0,\pi]$, it has an inverse function, called $\arccos(x)$, $\operatorname{inv}\cos(x)$, or $\cos^{-1}(x)$. Each of these is short for "the angle in $[0,\pi]$ whose cosine is x". For instance, $\arccos(0) = \pi/2$, $\cos^{-1}(1) = 0$, and $\operatorname{inv}\cos(-1) = \pi$. Moreover, because the range of the cosine function is $[-1,1]$, the domain of arccos is $[-1,1]$. Figure 1.2.5(b) shows that the graph of $\arccos(x)$ is obtained by reflecting the graph of $\cos(x)$, with domain $[0,\pi]$, across the line $y = x$.

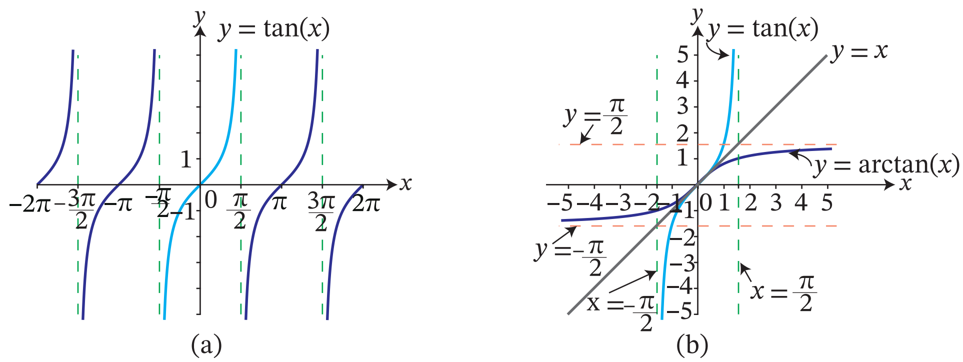


Figure 1.2.6

$\tan(x)$ **and its inverse**

The range of the **tangent function**, $\tan(x) = \sin(x)/\cos(x)$, is $(-\infty, \infty)$. (See Figure 1.2.6(a).)

When the inputs are restricted to $(-\pi/2, \pi/2)$, $\tan(x)$ is one-to-one, and therefore has an inverse function, denoted $\arctan(x)$, $\tan^{-1}(x)$, or $\mathrm{inv}\tan(x)$. The domain of the inverse tangent function is $(-\infty, \infty)$ and its range is $(-\pi/2, \pi/2)$.

For instance, $\tan^{-1}(0) = 0$, $\mathrm{inv}\tan(1) = \pi/4$, and as x increases, $\arctan(x)$ approaches $\pi/2$. Also, $\arctan(-1) = -\pi/4$, and when x is negative and becomes ever more negative (that is, $|x|$ becomes bigger and bigger) $\arctan(x)$ approaches $-\pi/2$. Figure 1.2.6(b) is the graph of $\arctan(x)$. It is the reflection of the blue part of the graph in Figure 1.2.6 across the line $y = x$.

EXAMPLE 1. Evaluate (a) $\sin(\sin^{-1}(0.3))$, (b) $\sin(\tan^{-1}(3))$, and (c) $\tan(\cos^{-1}(0.4))$.

SOLUTION

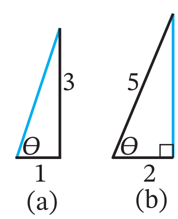


Figure 1.2.7

(a) The expression $\sin^{-1}(0.3)$ is short for "the angle in the interval $[-\pi/2, \pi/2]$ whose sine is 0.3." So the sine of $\sin^{-1}(0.3)$ is 0.3.

(b) To find $\sin(\tan^{-1}(3))$, first draw the angle θ whose tangent is 3 (and which lies in the interval $(-\pi/2, \pi/2)$). Figure 1.2.7(a) shows a simple way to draw this angle. To find the sine of θ, recall that sine equals "opposite/hypotenuse." By the Pythagorean Theorem, the hypotenuse is $\sqrt{3^2 + 1^2} = \sqrt{10}$. From the relation $\tan(\theta) = \text{opposite/adjacent}$, we conclude that

$$\sin(\tan^{-1}(3)) = \frac{3}{\sqrt{10}} \approx 0.9487.$$

The traditional symbol for angles is the Greek letter θ (pronounced "theta").

(c) To evaluate $\tan(\cos^{-1}(0.4))$, first draw an angle whose cosine is $0.4 = \frac{2}{5}$, as in Figure 1.2.7(b), which is based on the fact that cosine equals "adjacent/hypotenuse". To find the tangent of this angle, we need the length of the other leg in Figure 1.2.7(b). By the Pythagorean Theorem it is $\sqrt{5^2 - 2^2} = \sqrt{21}$, from which it follows that

$$\tan(\cos^{-1}(0.4)) = \frac{\sqrt{21}}{2} \approx 2.291.$$

◆

$\csc(x)$**,** $\sec(x)$**, and** $\cot(x)$ **and their inverses**

The **cosecant**, **secant**, and **cotangent** functions are defined in terms of the sine and cosine functions:

$$\csc(x) = \frac{1}{\sin(x)}, \quad \sec(x) = \frac{1}{\cos(x)}, \quad \text{and} \quad \cot(x) = \frac{\cos(x)}{\sin(x)}.$$

Each is defined only when the denominator is not zero. Note that $|\sec(x)| \geq 1$ and $|\csc(x)| \geq 1$. In each case the range consists of two separate intervals: $[1,\infty)$ and $(-\infty,-1]$. These three functions have inverses when restricted to appropriate intervals. Table 1.2.2 contains a summary of the three inverse functions, $\text{arccsc}(x)$, $\text{arcsec}(x)$, and $\arctan(x)$. Figure 1.2.8 shows the graphs of $y = \csc(x)$, $y = \sec(x)$, and $y = \cot(x)$ and their inverses.

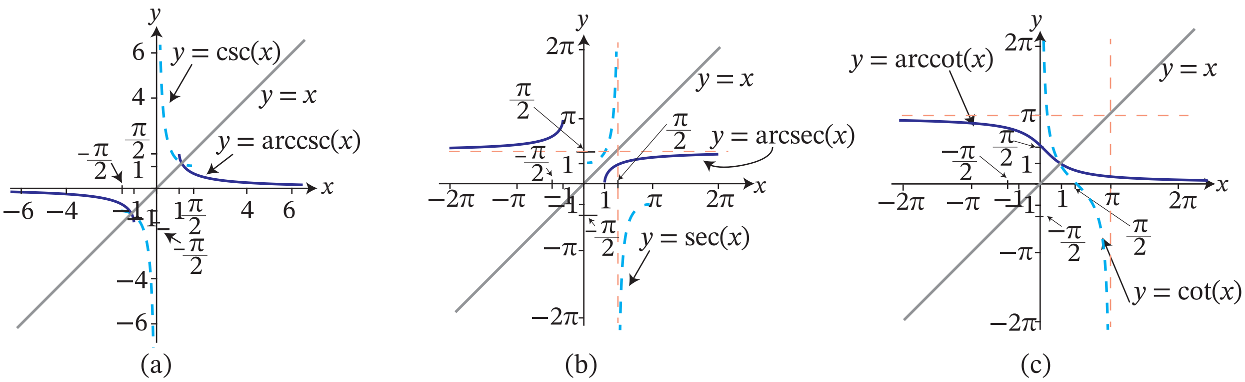


Figure 1.2.8

function	domain (input)	range (output)
$\text{arccsc}(x)$	$(-\infty,-1]$ and $[1,\infty)$	all of $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ except 0
$\text{arcsec}(x)$	$(-\infty,-1]$ and $[1,\infty)$	all of $[0,\pi]$ except $\frac{\pi}{2}$
$\text{arccot}(x)$	$(-\infty,\infty)$	the open interval $(0,\pi)$

Table 1.2.2

Summary

This section reviewed the basic functions in calculus, x^k, b^x, $\sin(x)$, $\cos(x)$, $\tan(x)$, $\csc(x)$, $\sec(x)$, $\cot(x)$, and their inverses, $x^{1/k}$, $\log_b(x)$, $\arcsin(x)$, $\arccos(x)$, arctan, $\text{arccsc}(x)$, $\text{arcsec}(x)$, and $\text{arccot}(x)$. (The inverse of x^k, $k \neq 0$, is just another power function, $x^{1/k}$).

The functions that may be hardest to have a feel for are the logarithms. They really are not as difficult as they sometimes appear. Consider the following two questions: ‘How many 4’s do we add together to get 64?” and “How many 4’s do we multiply together to get 64?” The answer to the first question, 16, is the quotient of 64 and 4. The answer to the second question, 3, is the base 4 logarithm of 64. The graphs of all logarithm functions, $\log_b(x)$ with $b > 1$ all have the same features. Here, we illustrate with $b = 2$ $\log_2(x)$:

- Its graph crosses the x-axis at $(1,0)$ because $\log_2(1) = 0$ $(2^0 = 1)$,
- It is defined only for positive inputs, that is, the domain of $\log_2$ is $(0,\infty)$, because only positive numbers can be expressed in the form 2^x,
- It is an increasing function,
- It grows very slowly as the argument increases: $\log_2(8) = 3$, $\log_2(16) = 4$, $\log_2(32) = 5$, and $\log_2(1024) = 10$,
- For values of x in $(0,1)$, $\log_2(x)$ is negative (if $2^x < 1$, then x is negative),
- For x near 0 (and positive), $|\log_2(x)|$ is large.

The case when the base b is less than 1 is treated in Exercise 50.

EXERCISES for Section 1.2

1. Graph the power function $x^{3/2}$, $x \geq 0$, and its inverse.

2. Graph the power function $\sqrt{x}$ and its inverse.

3. What is the period of $\tan(x)$?

4. What is the period of $\sin(x) + \cos(2x)$?

5. Explain your calculator's response when you try to calculate $\log_{10}(-3)$.

6. Explain your calculator's response when you try to calculate $\arcsin(2)$.

7. (a) Graph 2^x and $\left(\frac{1}{2}\right)^x$ on the same axes. (b) How could the second graph be obtained from the first?

8. (a) Graph 3^x and $\left(\frac{1}{3}\right)^x$ on the same axes. (b) How could the second graph be obtained from the first?

9. For any base b, $b^0 = 1$. What is the corresponding property of logarithms? Explain.

10. For any base b, $b^{x+y} = b^x b^y$. What is the corresponding property of logarithms? Explain.

NOTE: If you have trouble with this exercise, study Section 1.5.

11. Explain why $\log_b\left(\frac{1}{x}\right) = -\log_b(x)$. ("The log of the reciprocal of x is the negative of the log of x.")

12. Explain why $\log_b(c^x) = x\log_b(c)$. ("The log of a number raised to a power x is x times the log of the number.")

13. (a) Evaluate $\log_2(x)$ and $\log_4(x)$ at $x = 1, 2, 4, 8, 16$, and $\frac{1}{16}$.
(b) Graph $\log_2(x)$ and $\log_4(x)$ on the same axes (clearly label each point found in (a)).
(c) Compute $\frac{\log_4(x)}{\log_2(x)}$ for the six values of x in (a).
(d) Explain the phenomenon observed in (c).
(e) How would the graph of $y = \log_4(x)$ be obtained from that for $y = \log_2(x)$?

14. (a) Evaluate $\log_2(x)$ and $\log_8(x)$ at x=1, 2, 4, 8, 16, and 1/8.
(b) Graph $\log_2(x)$ and $\log_8(x)$ on the same axes. Clearly label each point found in (a).
(c) Compute $\frac{\log_8(x)}{\log_2(x)}$ for the six values of x in (a).
(d) Explain the phenomenon observed in (c).
(e) How would you obtain the graph of $\log_8(x)$ from that for $\log_2(x)$?

---

15. Evaluate (a) $\log_{10}(1000)$, (b) $\log_{100}(10)$, (c) $\log_{10}(0.01)$, (d) $\log_{10}\left(\sqrt{10}\right)$, and (e) $\log_{10}(10)$.

16. Evaluate (a) $\log_3(3^{17})$, (b) $\log_3\left(\frac{1}{9}\right)$, (c) $\log_3(1)$, (d) $\log_3\left(\sqrt{3}\right)$, and (e) $\log_3(81)$.

17. Evaluate $5^{\log_5(17)}$.

18. Evaluate $3^{-\log_3(21)}$.

---

19. For positive x near 0, what happens to the functions 2^x, x^2, and $\log_2(x)$?

20. For large values of x what happens to the quotient $\frac{x^2}{2^x}$? Illustrate by using specific values for x.

21. What happens to $\frac{\log_2(x)}{x}$ for large values of x? Illustrate by citing specific x.

In Exercises 22 to 25 the functions are one-to-one. Find the formula for each inverse function, expressed in the form $y = g(x)$, so that the independent variable is labeled x.

22. $y = 3^x$ **23.** $y = 5(2^x)$ **24.** $y = \frac{1}{2}(e^x + e^{-x})\ (x \leq 0)$ **25.** $y = \frac{1}{2}(e^x - e^{-x})$

26. Draw the graphs of $y = \sec(x)$ and $y = \operatorname{arcsec}(x)$ on the same axes.

27. Draw the graphs of $y = \cot(x)$ and $y = \operatorname{arccot}(x)$ on the same axes.

28. Which of the following expressions is equal to $\csc(x)$? (a) $\sin^{-1}(x)$ (b) $(\sin(x))^{-1}$ (c) $\operatorname{inv}\sin(x)$ (d) $\dfrac{1}{\sin(x)}$

In Exercises 29 to 44 evaluate the given expressions.

29. (a) $\sin^{-1}\left(\frac{1}{2}\right)$ (b) $\arcsin(1)$ (c) $\operatorname{inv}\sin\left(-\frac{\sqrt{3}}{2}\right)$ (d) $\arcsin\left(\frac{\sqrt{2}}{2}\right)$

30. (a) $\cos^{-1}(0)$ (b) $\operatorname{inv}\cos(-1)$ (c) $\arccos\left(\frac{1}{2}\right)$ (d) $\arccos\left(-\frac{1}{\sqrt{2}}\right)$

31. (a) $\arctan(1)$ (b) $\operatorname{inv}\tan(-1)$ (c) $\arctan(\sqrt{3})$ (d) $\arctan(1000)$ (approximately)

32. (a) $\operatorname{arcsec}(2)$ (b) $\operatorname{inv}\sec(-2)$ (c) $\operatorname{arcsec}(\sqrt{2})$ (d) $\sec^{-1}(1000)$ (approximately)

33. $\sin\left(\tan^{-1}(2)\right)$ **34.** $\sin\left(\cos^{-1}(0.4)\right)$ **35.** $\tan\left(\tan^{-1}(3)\right)$ **36.** $\tan\left(\sin^{-1}(0.7)\right)$

37. $\tan\left(\sec^{-1}(3)\right)$ **38.** $\sec\left(\tan^{-1}(0.3)\right)$ **39.** $\sin\left(\sec^{-1}(5)\right)$ **40.** $\sec\left(\cos^{-1}(0.2)\right)$

41. $\arctan\left(\tan\left(\frac{\pi}{3}\right)\right)$ **42.** $\arcsin\left(\sin\left(\frac{-3\pi}{4}\right)\right)$ **43.** $\arccos\left(\cos\left(\frac{5\pi}{2}\right)\right)$ **44.** $\operatorname{arcsec}\left(\sec\left(\frac{-\pi}{3}\right)\right)$

45. Approximate each of the following quantities.

(a) $\arcsin(0.3)$ (b) $\arccos(0.3)$ (c) $\arctan(0.3)$ (d) $\dfrac{\arcsin(0.3)}{\arccos(0.3)}$

Observe that (c) and (d) are different.

46. Let k be a period of a function f. Show that $2k$ and $-k$ are also periods of f.

47. *(Richter Scale)* While many reports of earthquake magnitude still refer to the Richter scale, the **moment magnitude scale** has replaced the **Richter scale** because it is more accurate for larger earthquakes. Like the Richter scale, increasing the magnitude by one corresponds to an increase in the amount of energy released by a factor of about 32 (see (e)).

A "major earthquake" typically has a measure of at least 7.5. The earthquake that struck San Francisco and vicinity in 1989 was measured as 6.9. The largest earthquakes in recent years have had a magnitude measured in excess of 9.0: 9.5 in Valdivia, Chile (1960), 9.2 in Southern Alaska (1964), 9.1 off the West Coast of Northern Sumatra, Indonesia (2004), and 9.1 near the East Coast of Honshu, Japan (2011). Each of these earthquakes was more than 1000 times stronger than the 1989 Bay Area earthquake.

In his *Introduction to the Theory of Seismology*, Cambridge, 1965, pp. 271–272, K. E. Bullen explains the Richter scale as follows:

"Gutenburg and Richter sought to connect the magnitude M with the energy E of an earthquake by the formula

$$aM = \log_{10}\left(\frac{E}{E_0}\right)$$

and after several revisions arrived in 1956 at the result $a = 1.5$, $E_0 = 2.5 \times 10^{11}$ ergs."

NOTE: Energy E is measured in ergs. M is the number assigned to the earthquake on the Richter scale. E_0 is the energy of the smallest instrumentally recorded earthquake.

(a) Deduce that $\log_{10}(E) \approx 11.4 + 1.5M$.
(b) Find a formula for E in terms of M.
(c) What is the ratio between the energy of the earthquake that struck Japan in 1933 ($M = 8.4$) and the San Francisco earthquake of 1989 ($M = 6.9$)?
(d) What is the ratio between the energy of the San Francisco earthquake of 1906 ($M = 8.3$) and that of the San Francisco earthquake of 1989 ($M = 6.9$)?
(e) If one earthquake has a Richter measure 1 larger than that of another earthquake, what is the ratio of their energies?
(f) What is the Richter measure of a 10-megaton H-bomb, that is, of an H-bomb whose energy is equivalent to that of 10 million tons of TNT? (One ton of TNT releases an energy of 4.2×10^{16} ergs.)

48. As of 2023, the largest known prime was $2^{82,589,933} - 1$.
(a) When written in decimal notation, how many digits will it have?
(b) How many pages of this book would be needed to print it? (One page can hold 50 70-character lines, a total of 3,500 digits per page.)

NOTE: There is a prize of $250,000 for the discovery of the first prime with at least one billion digits. An online search for "'largest prime number" will find the largest known prime.

49. Say that you have drawn the graph of $y = \log_2(x)$. Jane says that to get the graph of $y = \log_2(4x)$, you just raise that graph 2 units parallel to the y-axis. Sam says, "No, just shrink the x-coordinate of each point on the graph by a factor of 4." Who is right? Why? Explain your answers.

50. Let b be a positive number less than 1.
(a) Sketch the graphs of $y = b^x$ and $y = \log_b(x)$ on the same set of axes.
(b) What is the domain of $\log_b$?
(c) What is the x-intercept? That is, solve $\log_b(x) = 0$.
(d) For what values of x is $\log_b(x)$ positive? negative?
(e) Is $\log_b(x)$ an increasing or decreasing function?
(f) What can be said about the values of $\log_b(x)$ when x is close to zero (and in the domain)?
(g) What can be said about the values of $\log_b(x)$ when x is a large positive number?
(h) What can be said about the values of $\log_b(x)$ when x is a large negative number?

51. Prove that $\log_3(2)$ is irrational, that is, not rational.

1.3 Building More Functions from Basic Functions

This section completes the list of functions needed in calculus. Starting with the basic functions introduced in Section 1.2, we will see how to obtain a function as complicated as

$$f(x) = \frac{\sin(2x) + 3 + 4x + 5x^2}{\log_2(x) + 3^{-5x} + \sqrt{1 + x^3}}.$$

Before we see how to construct new functions from old ones, we introduce one more type of basic function. These functions are so simple that they did not deserve to appear with the functions in the preceding section. They are the constant functions, whose graphs are horizontal lines. (See Figure 1.3.1.)

The Constant Functions

Definition: *Constant Function*

A function $f(x)$ is **constant** if there is a number C such that $f(x) = C$ for all x in its domain. A special constant function is the **zero function**: $f(x) = 0$.

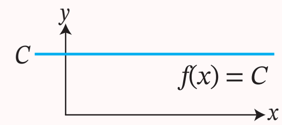


Figure 1.3.1

Using the Four Arithmetic Operations: $+$, $-$, $\times$, /

Let f and g be functions. We can produce other functions from them by using the four operations of arithmetic:

$f+g$: For an input x, the output is $f(x)+g(x)$.
$f-g$: For an input x, the output is $f(x)-g(x)$.
fg: For an input x, the output is $f(x)g(x)$.
f/g: For an input x with $g(x) \neq 0$, the output is $f(x)/g(x)$.

The domains of $f+g$, $f-g$, and fg consist of the numbers that belong to both the domain of f and the domain of g. The function f/g is defined for all numbers x that belong to the domain of f and the domain of g with the extra condition that $g(x) \neq 0$.

With the aid of these constructions any polynomial or rational function can be built up from the simple function $f(x) = x$, called the **identity function**, and the constant functions.

A **polynomial** is a function of the form $p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$, where the coefficients a_0, a_1, a_2, ..., a_n are numbers. If a_n is not zero, the **degree** of the polynomial is n. A **rational function** is the quotient of two polynomials. The domain of a polynomial is the set of all real numbers. The domain of a rational function is all real numbers except those for which the denominator is zero.

EXAMPLE 1. Use addition, subtraction, and multiplication to form the polynomial $F(x) = x^3 + 3x - 7$.

SOLUTION We first build each of the three terms: x^3, $3x$, and 7. The last is just a constant function. Multiplying the identity function x and the constant function 3 gives $3x$. The first term is obtained by first multiplying x and x to obtain x^2 and then multiplying x^2 and x to yield x^3. Adding x^3 and $3x$ gives $x^3 + 3x$. Lastly, subtract the constant function 7 to obtain $x^3 + 3x - 7$.

Each of the three basic functions involved in forming F is defined for all real numbers. The domain of F is also all real numbers: $(-\infty, \infty)$. ♦

Example 1 shows how to build any polynomial using $+$, $-$, and $\times$. Constructing rational functions also requires a use of division.

But how would we build a function like $\sqrt{1+3x}$? This leads us to the most important technique for combining functions to build more complicated ones.

Composite Functions

When f and g are functions, we may use the output of g as an input for f. That is, we can think about $f(g(x))$. For instance, if $g(x) = 1+3x$ and f is the square root function, $f(x) = \sqrt{x}$, then $f(g(x)) = f(1+3x) = \sqrt{1+3x}$. This brings us to the definition of a composite function.

Definition: *Composition of Functions*

Assume X, Y, and Z are sets, g a function from X to Y, and f a function from Y to Z. Then the function that assigns to each element x in X the element $f(g(x))$ in Z is called the **composition of f with g**. It is denoted $f \circ g$, which is read as "f circle g" or as "f composed with g." The function $f \circ g$ is also called a **composite function**.

Thinking of f and g as input-output machines we may consider $f \circ g$ as the machine built by hooking up the machine for f to process the outputs of the machine for g (see Figure 1.3.2).

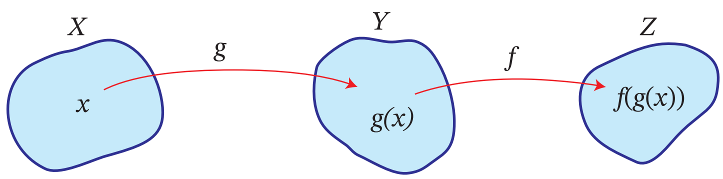


Figure 1.3.2

Most functions we encounter are composite. For instance, $\sin(2x)$ is the composition of $g(x) = 2x$ and $f(x) = \sin(x)$. We can hook up three or more functions to make even more complicated functions. The function $\sin^3(2x) = (\sin(2x))^3$ is built up in three steps:

$$x \longrightarrow 2x \longrightarrow \sin(2x) \longrightarrow (\sin(2x))^3.$$

The first doubles the input, the second takes the sine of its input, and the third cubes its input.

The order matters. If, instead, the cubing is done first, then the sine, and then end by doubling the input, the result is:

$$x \longrightarrow x^3 \longrightarrow \sin(x^3) \longrightarrow 2\sin(x^3),$$

which gives a completely different answer.

Before pressing the sin key, be sure to check that your calculator is in radians mode.

When you enter a function on your calculator or on a computer, you have to be careful of the order in which the functions are applied as you evaluate a composite function. For example, the way to evaluate $\sin(\log(240))$ depends on your calculator. On a traditional scientific calculator you enter 240, press the log key, and then the sin key. On many of the newer graphing calculators you press the sin key, then the log key, then enter 240, followed by two right parentheses,)), and, finally, press the Enter key. The two approaches are different. If you press the sin key before log, you will get $\log(\sin(240))$. For most computer software it is necessary to use parentheses to indicate inputs to functions. In this case you might enter sin(log(240)).

NOTE: Some calculators automatically insert the left parenthesis when the sin and log keys are pressed.

To describe the build-up of a composite function it is convenient to use various letters, not just x, to denote the variables. This is illustrated in Examples 2 to 4.

EXAMPLE 2. Show how the function $\sqrt{4-x^2}$ is built up by the composition of functions. Find its domain.

SOLUTION The function $\sqrt{4-x^2}$ is obtained by applying the square-root function to the function $4-x^2$. To be specific, let $g(x) = 4 - x^2$ and $f(u) = \sqrt{u}$ $(u \geq 0)$. Then

$$f(g(x)) = f(4-x^2) = \sqrt{4-x^2}.$$

The square-root function is defined for all $u \geq 0$ and the polynomial $g(x)$ is defined for all numbers. So $f(g(x))$ is defined only when $g(x) \geq 0$. Now, $g(x) = 4 - x^2 \geq 0$ means $4 \geq x^2$, so $|x| \leq 2$.

Thus, the domain of $\sqrt{4-x^2}$ is the closed interval $[-2, 2]$. ♦

EXAMPLE 3. Express $\dfrac{1}{\sqrt{1+x^2}}$ as a composition of four functions. Find its domain.

SOLUTION Call the input x. First, compute the square of x: x^2. Second, add one: $1+x^2$. Third, take the square root of that output, getting $\sqrt{1+x^2}$. Fourth, take the reciprocal of that result, getting $1/\sqrt{1+x^2}$. In summary, form

$$u = x^2, \qquad \text{then } v = 1+u, \qquad \text{then } w = \sqrt{v}, \qquad \text{then } y = \frac{1}{w}.$$

Given x, we first square the input, then add one, then apply the square-root function, and lastly the reciprocal function.

The domain of the squaring function consists of all real numbers; the domain of the square-root function is $[0,\infty)$; and the domain of the reciprocal function is all numbers except zero. Because $v = 1+x^2 > 0$, $w = \sqrt{v} = \sqrt{1+x^2}$ is defined for all x. Moreover, $w = \sqrt{1+x^2} \geq 1$, so that $y = 1/w = 1/\sqrt{1+x^2}$ is defined for all real numbers x. ♦

The function in Example 3 can also be written as the composition of two functions: $x \longrightarrow 1+x^2 \longrightarrow (1+x^2)^{-1/2}$.

EXAMPLE 4. Let f be the cubing function and g the cube-root function. Compute $(f\circ g)(x)$, $(f\circ f)(x)$, and $(g\circ f)(x)$.

SOLUTION In terms of formulas, $f(x) = x^3$ and $g(x) = \sqrt[3]{x}$.

$$(f\circ g)(x) = f(g(x)) = f(\sqrt[3]{x}) = \left(\sqrt[3]{x}\right)^3 = x.$$
$$(f\circ f)(x) = f(f(x)) = f(x^3) = \left(x^3\right)^3 = x^9.$$
$$(g\circ f)(x) = g(f(x)) = g(x^3) = \sqrt[3]{x^3} = x.$$

The domains of f and g are $(-\infty,\infty)$. Therefore $f\circ g$, $f\circ f$, and $g\circ f$ are defined for all real numbers.

Both $f\circ g$ and $g\circ f$ are the identity function. Whenever g is the inverse of f, $f\circ g$ and $g\circ f$ are the identity function. Each function undoes the action of the other. ♦

EXAMPLE 5. Show that every power function x^k, $x>0$, can be constructed as a composition using exponential and logarithmic functions.

SOLUTION The first step is to write $x = 2^{\log_2(x)}$. Then, $x^k = \left(2^{\log_2(x)}\right)^k$ or, by properties of exponentials, $x^k = 2^{k\log_2(x)}$. So x^k is the composition of three functions: First, find $\log_2(x)$, then multiply by the constant function k, and then raise 2 to the resulting power. ♦

An additional consequence of Example 5 is that it provides a meaning to functions like $x^{\sqrt{2}}$ and $x^{-\pi}$ for $x>0$. We could even go so far as to remove the power functions from our list of basic functions. We chose not to do so because power functions with integer exponents are common and are defined also at negative inputs. Lastly, while it might seem surprising that the power functions can be expressed in terms of exponentials and logarithms, it is more astonishing that trigonometric functions, such as $\sin(x)$, can also be expressed in terms of exponentials, as shown in Chapter 12.

Summary

This section showed how to build more complicated functions from power, exponential, and trigonometric functions and their inverses, and the constant functions. One method is to add, multiply, subtract, or divide outputs. The other is to compose functions so that one function operates on the output of a second function. Composite functions are extremely important in our upcoming study of calculus.

Be careful when composing functions when one of them is a trigonometric function. For instance, what is meant by $\sin x^3$? Is it $\sin(x^3)$ or $(\sin(x))^3$? Do we first cube x, then take the sine, or the other way around? There is a general agreement that $\sin x^3$ stands for $\sin(x^3)$; cube first, then take the sine. In this book we will always use parentheses in "$\sin(x)$" to avoid this ambiguity.

Spoken aloud, $\sin x^3$ is usually "the sine of x cubed," which is ambiguous. We can either insert a brief pause – "sine of (pause) x cubed" – to emphasize that x is cubed rather than $\sin(x)$, or rephrase it as "sine of the quantity x cubed."

On the other hand $(\sin(x))^3$, which is by convention usually written as $\sin^3(x)$, is spoken aloud as "the cube of $\sin(x)$" or "sine cubed of x."

Similar warnings apply to other trigonometric functions and logarithmic functions.

EXERCISES for Section 1.3

The function $y = \sqrt{1+x^2}$ is the composition of $u = 1+x^2$ and $y = \sqrt{u}$. In Exercises 1 to 12 use a similar format to build the given functions as the composition of two or more functions.

1. $\sin(2x)$
2. $\sin^3(x)$
3. $\sin(3x)$
4. $\sin\left(x^3\right)$
5. $\sin^2(x^3)$
6. 2^{x^2}
7. $(x^2+3)^{10}$
8. $\log_{10}(1+x^2)$
9. $\dfrac{1}{x^2+1}$
10. $\cos^3(2x+3)$
11. $\left(\dfrac{2}{3x+5}\right)^3$
12. $\arcsin(\sqrt{x})$

13. These tables show some of the values of functions f and g. Find (a) $f(g(7))$ and (b) $g(f(3))$.

x	1	2	3	4	5
$f(x)$	6	8	9	7	10

x	6	7	8	9	10
$g(x)$	4	3	2	5	1

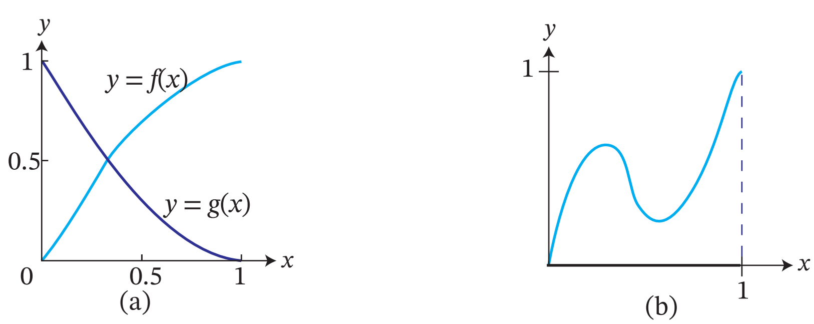


Figure 1.3.3

14. Figure 1.3.3(a) shows the graphs of functions f and g. Estimate (a) $f(g(0.6))$ (b) $f(g(0.3))$ (c) $f(f(0.5))$

15. Figure 1.3.3(b) shows the graph of a function f whose domain is $[0,1]$. Let $g(x) = f(2x)$.
(a) What is the domain of g? (b) Graph $y = g(x)$.

In Exercises 16 to 25 write y as a function of x. Simplify when possible.

16. $u = \sin(x)$, $y = u^2$
17. $u = \sqrt{x}$, $y = u^2$
18. $u = x^3$, $y = \dfrac{1}{u}$
19. $u = 2x^2-3$, $y = \dfrac{1}{u}$

20. $u = \sqrt{x}$, $y = \sin(u)$

21. $u = \log_3(x)$, $y = 3^u$

22. $v = 2x$, $u = v^2 - 1$, $y = u^2$

23. $v = \sqrt{x}$, $u = 1 + v$, $y = u^2$

24. $v = x + x^2$, $u = \sin(v)$, $y = u^3$

25. $v = \tan(x)$, $u = 1 + v^2$, $y = \cos(u)$

---

26. If $f(x) = x^3$, is there a function $g(x)$ such that $(f \circ g)(x) = x$ for all numbers x? If so, is there more than one such function?

27. If $f(x) = x^4$, is there a function $g(x)$ such that $(f \circ g)(x) = x$ for all negative numbers x? If so, is there more than one such function?

28. If $f(x) = x^4$, is there a function $g(x)$ such that $(f \circ g)(x) = x$ for all positive numbers x? If so, is there more than one such function?

29. Let $f(x) = 1/(1-x)$. What is the domain of f? of $f \circ f$? of $f \circ f \circ f$? Show that $(f \circ f \circ f)(x) = x$ for all x in the domain of $f \circ f \circ f$.

30. Find all functions of the form $f(x) = 1/(ax+b)$, where $a \neq 0$, such that $(f \circ f \circ f)(x) = x$ for all x in the domain of $f \circ f \circ f$.

31. Show that there is a function $u(x)$ such that $\cos(x) = \sin(u(x))$.

NOTE: This shows that we did not need to include $\cos(x)$ among our basic functions.

32. Find a function $u(x)$ such that $3^x = 2^{u(x)}$.

33. If f and g are one-to-one, must $f \circ g$ be one-to-one?

34. If $f \circ g$ is one-to-one, must f be one-to-one? Must g be one-to-one?

35. If f has an inverse, $\mathrm{inv} f$, compute $(f \circ \mathrm{inv} f)(x)$ and $((\mathrm{inv} f) \circ f)(x)$.

36. Let $f(x) = 2x^2 - 1$ and $g(x) = 4x^3 - 3x$. (a) Find $(f \circ g)(x)$. (b) Find $(g \circ f)(x)$. (c) Show that $(f \circ g)(x) = (g \circ f)(x)$.

Any two powers, such as x^3 and x^4, commute under composition, their composition in either order being x^{12}. Pairs of polynomials that commute with each other under composition are rare. To convince yourself of this, try to find more examples other than the pair in Exercise 36. Exercises 37 to 40 consider some specific cases. Exercise 41 shows a way to generate many such examples.

37. When $g(x) = x^2$ find all first-degree (linear) polynomials $f(x) = ax + b$, where $a \neq 0$, such that $f \circ g = g \circ f$, that is, $f(g(x)) = g(f(x))$.

38. When $g(x) = x^2$ find all second-degree polynomials $f(x) = ax^2 + bx + c$, where $a \neq 0$, such that $f \circ g = g \circ f$.

39. Let $f(x) = 2x+3$. Find all functions of the form $g(x) = ax+b$, where a and b are constants, such that $f \circ g = g \circ f$.

40. Let $f(x) = 2x + 3$. Find all functions of the form $g(x) = ax^2 + bx + c$, where a, b, and c are constants, such that $f \circ g = g \circ f$.

---

41. This exercise provides lots of opportunities to apply three trigonometric identies: $\cos^2(x) + \sin^2(x) = 1$, $\sin(x+y) = \sin(x)\cos(y) + \cos(x)\sin(y)$, and $\cos(x+y) = \cos(x)\cos(y) - \sin(x)\sin(y)$.

(a) Show that $\sin(2x) = 2\sin(x)\cos(x)$ and $\cos(2x) = 2\cos^2(x) - 1$.

(b) Show that $\sin(3x) = \sin(x)(4\cos^2(x) - 1)$ and $\cos(3x) = 4\cos^3(x) - 3\cos(x)$.

(c) Show that $\sin(4x) = \sin(x)(8\cos^3(x) - 4\cos(x))$ and $\cos(4x) = 8\cos^4(x) - 8\cos^2(x) + 1$.

(d) Show that there are polynomials $Q_n(x)$ and $P_n(x)$ such that $\sin(nx) = \sin(x)Q_n(\cos(x))$ and $\cos(nx) = P_n(\cos(x))$. Show that if this is true for $n = 10$, then it is also true for $n = 11$. for each positive integer n,

(e) Explain why $(P_n \circ P_m)(x) = (P_m \circ P_n)(x)$ for x in $[-1, 1]$.

1.4 Geometric Series

If a and r are real numbers the (infinite) sequence of numbers

$$a, ar, ar^2, ar^3, \ldots$$

is called a **geometric sequence**. Its first term is a. Each term after the first term is obtained by multiplying its predecessor by r, which is called the **ratio**. The n^{th} term is ar^{n-1}.

A finite collection of consecutive terms from a geometric sequence is called a **geometric progression**.

The sum of the first n terms of a geometric sequence is denoted S_n:

S_n is also called a **partial sum** of the sequence $\{a_k\}$.

$$S_n = a + ar + ar^2 + \cdots + ar^{n-1}.$$

There is a short formula for the partial sum of a geometric sequence, which we will use several times.

To find this formula, subtract rS_n from S_n:

$$\begin{array}{rclll} S_n & = & a + & ar + ar^2 + \cdots + ar^{n-1} & \\ rS_n & = & & ar + ar^2 + \cdots + ar^{n-1} + ar^n. & \end{array}$$

The many cancellations give $S_n - rS_n = a - ar^n$. If r is not 1, divide both sides of the equation by $1-r$ to obtain:

Theorem 1.4.1: Short Formula for the Partial Sum of a Geometric Series

$$S_n = \frac{a(1-r^n)}{1-r} \qquad r \neq 1 \tag{1.4.1}$$

EXAMPLE 1. Find (a) $3 + \frac{3}{2} + \frac{3}{4} + \frac{3}{8} + \frac{3}{16} + \frac{3}{32}$ and (b) $1 - \frac{1}{3} + \frac{1}{9} - \frac{1}{27} + \frac{1}{81}$.

SOLUTION

(a) Here $a = 3$, $r = 1/2$, and $n = 6$. The sum is

$$S_6 = \frac{3\left(1-\left(\frac{1}{2}\right)^6\right)}{1-\frac{1}{2}} = 6\left(1-\left(\frac{1}{2}\right)^6\right) = \frac{378}{64} = \frac{189}{32}.$$

(b) In this case $a = 1$, $r = -1/3$, and $n = 5$. So the sum is

$$S_5 = \frac{1\left(1-\left(\frac{-1}{3}\right)^5\right)}{1-\frac{-1}{3}} = \frac{1-\left(\frac{-1}{3}\right)^5}{\frac{4}{3}} = \frac{3}{4}\left(1+\left(\frac{1}{3}\right)^5\right) = \frac{61}{81}.$$

Recognizing a partial sum of a geometric series will prove very useful on our journey through calculus. ♦

We now use the general formula, (1.4.1), to develop the partial sum of another geometric progression that will appear numerous times.

Let x and a be two numbers and consider the sequence

$$x^{n-1}, ax^{n-2}, a^2x^{n-3}, a^3x^{n-4}, \ldots, a^{n-3}x^2, a^{n-2}x, a^{n-1}. \tag{1.4.2}$$

The exponent of x decreases from $n-1$ to 0 while the exponent of a increases from 0 to $n-1$. While it might not look like it at first, (1.4.2) displays the first n terms of a geometric sequence. The first term is x^{n-1} and the ratio is

a/x. Thus, assuming x is not 0 or a,

$$\begin{aligned} x^{n-1} + ax^{n-2} + a^2x^{n-3} + a^3x^{n-4} + \cdots + a^{n-3}x^2 + a^{n-2}x + a^{n-1} &= x^{n-1}\left(\frac{\left(1-\left(\frac{a}{x}\right)^n\right)}{1-\frac{a}{x}}\right) \\ &= \frac{x^{n-1}\left(\frac{x^n-a^n}{x^n}\right)}{\frac{x-a}{x}} \\ &= \frac{x^n-a^n}{x-a}. \end{aligned}$$

This leads us to conclude that

$$x^{n-1} + ax^{n-2} + a^2x^{n-3} + a^3x^{n-4} + \cdots + a^{n-3}x^2 + a^{n-2}x + a^{n-1} = \frac{x^n-a^n}{x-a} \qquad x \neq a \tag{1.4.3}$$

In Chapter 2 we will use (1.4.3) in the reverse order, to express the quotient $(x^n - a^n)/(x-a)$ as a sum of n terms.

Equation (1.4.3) can also be established from the factorizations of $x^n - a^n$:

$$\begin{aligned} x^2 - a^2 &= (x-a)(x+a) \\ x^3 - a^3 &= (x-a)(x^2+ax+a^2) \\ x^4 - a^4 &= (x-a)(x^3+ax^2+a^2x+a^3) \end{aligned}$$

and so on. To establish the last, for instance, multiply out its right-hand side:

$$(x-a)(x^3+ax^2+a^2x+a^3) = (x^4+ax^3+a^2x^2+a^3x) - (ax^3+a^2x^2+a^3x+a^4) = x^4 - a^4.$$

Summary

The key idea of this section is that the sum $a + ar + ar^2 + \cdots + ar^{n-1}$ equals $a(1-r^n)/(1-r)$ as long as r is not 1. If r is 1, then the sum is na, because each summand is a. As a particularly useful case, we have

$$\frac{x^n-a^n}{x-a} = x^{n-1} + ax^{n-2} + a^2x^{n-3} + \cdots + a^{n-2}x + a^{n-1}.$$

EXERCISES for Section 1.4

In Exercises 1 to 6 calculate the sum using the formula for the sum of a geometric progression.

1. $1+3+9+27+81+243$

2. $1-3+9-27+81-243$

3. $2+1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}$

4. $0.5-0.05+0.005-0.0005+0.00005-0.000005+0.0000005$

5. $a^4+a^3b+a^2b^2+ab^3+b^4$

6. $1-\frac{1}{2}+\frac{1}{4}-\frac{1}{8}+\frac{1}{16}-\frac{1}{32}$

In Exercises 7 and 8 write the given polynomial in two different ways as a product of two polynomials. Do not just change the order of the factors.

7. $x^6 - a^6$

8. $x^9 - a^9$

9. Show that $x^4-16=(x^3+2x^2+4x+8)(x-2)$.

10. Show that $x^5-32=(x^4+2x^3+4x^2+8x+16)(x-2)$.

11. This exercise obtains the sum of a geometric progression geometrically. Let r be a positive number less than 1 and n a positive integer.

(a) In the interval $[0,1]$ indicate the numbers $r, r^2, \ldots, r^n$.

(b) The numbers in (a) break the interval $[r^n,1]$ of length $1-r^n$ into n intervals. By adding their lengths show that $1+r+\cdots+r^{n-1}=\dfrac{1-r^n}{1-r}$.

12. (a) Why does $(1-r)+(r-r^2)+(r^2-r^3)+\cdots+(r^{n-1}-r^n)=1-r^n$ for any number r? Look closely at the sum.

(b) From (a) deduce the formula for the sum of the geometric series $1+r+r^2+\ldots+r^{n-1}$ when r is not 1.

13. What happens to $\dfrac{x^3-1}{x^2-1}$ when you choose x nearer and nearer 2? Nearer and nearer 1?
(a) Base your answers on calculations., and (b) Base your answers on geometric series..

14. What happens to $\dfrac{x^5+32}{x+2}$ as x approaches 2? as x approaches -2?
(a) Base your answers on calculations., and (b) Base your answers on geometric series..

In Exercises 15 and 16 Sam and Jane discuss the term r^n that appears in the short formula, (1.4.1), for the sum of a geometric progression.

15. Jane remains a little uncertain, and has been doing some experiments to gather some information that allows her to have the following exchange with Sam.

Sam: When I graph 0.5^n I see a sequence of numbers getting very near 0, as in Figure 1.4.1.
Jane: Maybe you're right. But I computed $(0.999)^{1000}$ and got about 0.37.
Sam: So it looks like those numbers are getting real close to 1/3.
Jane: Why 1/3?
Sam: It's the only number I know near 0.37.
Jane: That's not much of a reason.

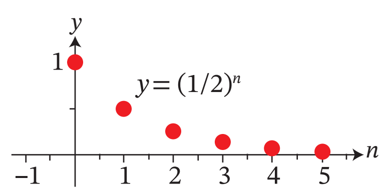


Figure 1.4.1

Based on your calculations, make a conjecture about what happens to 0.999^n as n gets larger and larger.

16. Sam has been thinking, and has the following conversation with Jane.

Sam: I just computed 1.001^n for really large values of n.
Jane: What did you find?
Sam: Well, 1.001^{500} is about 1.65 and 1.001^{1000} is about 2.72.
Jane: So?
Sam: So I think that as n grows, 1.001^n is getting near 3 maybe, or maybe π.
Jane: Well, I just computed 1.001^{2000} and got about 7.38. I think it's getting nearer and nearer 20.

After computing some values of 1.001^n for some larger values of n, offer your own opinion on what happens to 1.001^n as n increases. How do you think 1.000001^n behaves?

---

17. Express 5.14141414 as a fraction. Use the short formula for the sum of a geometric progression.

18. (a) Using your calculator, evaluate the product $2 \cdot \sqrt{2} \cdot \sqrt[4]{2} \cdot \sqrt[8]{2}$.

(b) Each factor in (a) except the first is the square root of its predecessor. Continue the pattern with more factors. Evaluate the product in each case.

(c)

SAM: I think the products will get arbitrarily large.
JANE: Why?
SAM: You're multiplying numbers bigger than 1. So the products keep growing.
JANE: But the factors are getting closer and closer to 1.
SAM: So?
JANE: So maybe the products don't get arbitrarily large.

Decide who is right.

19. (a) What happens to $\frac{1-r^5}{1-r}$ for values of r near 1? (Experiment.)

(b) For a positive integer n, what happens to $\frac{1-r^n}{1-r}$ for values of r near 1? Explain.

1.5 Logarithms

How many 2s must be multiplied to get 32? Whatever the answer is, it is called "the logarithm of 32 to the base 2." Because $2^5 = 32$, the logarithm of 32 to the base 2 is 5. More generally, a logarithm is defined in terms of an exponential function.

Definition: *Definition of Logarithm to the Base b*

If b and c are positive numbers, $b \neq 1$, there is a number d such that

$$b^d = c.$$

The exponent d is called the **logarithm of c to the base b**. It is denoted

$$\log_b(c).$$

It is easy to verify that this definition is consistent with the definition of logarithm given in Section 1.2.

The definition of a logarithm provides the following general formula.

Formula 1.5.1: *Inverse Property of Logarithms*

For all positive numbers b ($b \neq 1$) and c,

$$b^{\log_b(c)} = c.$$

The word "logarithm" comes from the Greek. In a Greek restaurant, to get the bill, one asks the waiter for the "logarismo".

EXAMPLE 1. Find (a) $\log_{10}(1000)$, (b) $\log_2(1024)$, (c) $\log_9(3)$, and (d) $\log_{16}\left(\frac{1}{4}\right)$.

SOLUTION

(a) Because $10^3 = 1000$, $\log_{10}(1000) = 3$.
(b) Because $2^{10} = 1024$, $\log_2(1024) = 10$.
(c) Because $9^{1/2} = 3$, $\log_9(3) = 1/2$.
(d) Because $16^{-1/2} = 1/4$, $\log_{16}(1/4) = -1/2$. ♦

Every property of an exponential function translates into a property of logarithms. For instance, here is how we write the equation $b^{x+y} = b^x b^y$ in the language of logarithms.

Let $c = b^x$ and $d = b^y$. We have

$$x = \log_b(c) \qquad \text{and} \qquad y = \log_b(d). \tag{1.5.1}$$

Because

$$cd = b^x b^y = b^{x+y}$$

we know

$$\log_b(cd) = x + y.$$

Using (1.5.1), we discover the formula for the logarithm of a product:

Formula 1.5.2: *The Logarithm of a Product*

For all positive numbers b ($b \neq 1$), c, and d:

$$\log_b(cd) = \log_b(c) + \log_b(d).$$

This generalizes to the logarithm of the product of several numbers: the logarithm of a product of two or more numbers is the sum of their logarithms.

Logarithms can be used to simplify products, quotients, and powers:

$$\begin{aligned}\log_b\left(\frac{\sqrt{x}(2+x)^3}{\left(1+x^2\right)^5}\right) &= \log_b(\sqrt{x}) + \log_b\left((2+x)^3\right) - \log_b\left(\left(1+x^2\right)^5\right)\\ &= \frac{1}{2}\log_b(x) + 3\log_b(2+x) - 5\log_b\left(1+x^2\right).\end{aligned}$$

In the final expression, most of the exponents and radical sign no longer appear. There is no way to simplify $\log_b(2+x)$ and $\log_b\left(1+x^2\right)$.

Summary

This section reviewed logarithms, which are a different way of talking about exponents. The two key properties of logarithms for a positive base b are $\log_b(xy) = \log_b(x) + \log_b(y)$ and $\log_b(x^y) = y\log_b(x)$.

Figure 1.5.1 is the graph of $y = \log_2(x)$. Notice that as x increases, so does $\log_2(x)$, but very slowly. Also, when x is near 0, $\log_2(x)$ is negative but has large absolute values.

Table 1.5.1 lists the properties of exponential functions and the corresponding properties of logarithms.

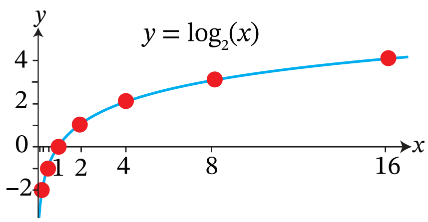


Figure 1.5.1

Exponential Language			Logarithm Language		
b^{x+y}	$=$	$b^x b^y$	$\log_b(cd)$	$=$	$\log_b(c) + \log_b(d)$
b^{x-y}	$=$	$\dfrac{b^x}{b^y}$	$\log_b\left(\dfrac{c}{d}\right)$	$=$	$\log_b(c) - \log_b(d)$
b^0	$=$	1	$\log_b(1)$	$=$	0
b^1	$=$	b	$\log_b(b)$	$=$	1
b^{-x}	$=$	$\dfrac{1}{b^x}$	$\log_b\left(\dfrac{1}{c}\right)$	$=$	$-\log_b(c)$
$(b^x)^y$	$=$	b^{xy}	$\log_b\left(c^d\right)$	$=$	$d\log_b(c)$

Table 1.5.1

EXERCISES for Section 1.5

1. Why is $\log_b(c)$ defined only for positive values of c? $(b > 0)$

2. How is $\log_{b^2}(c)$ related to $\log_b(c)$?

3. Evaluate (a) $\log_b\left(\sqrt{b}\right)$, (b) $\log_b\left(\dfrac{b^3}{\sqrt{b}}\right)$, and (c) $\log_b\left(\sqrt{b}\sqrt[3]{b}b^4\right)$.

4. Simplify $\log_2\left(\dfrac{\left(x^3\right)^5 \sqrt[3]{x+2}\left(1+x^2\right)^{15}}{x^5+7}\right)$.

5. Show that $\dfrac{\log_b(x) - \log_b(y)}{c} = \log_b\left(\left(\dfrac{x}{y}\right)^{1/c}\right)$.

6. What happens to $\dfrac{\log_{10}(x)}{x}$ for large values of x?

In Exercises 7 to 11 establish the given property of logarithms by using a property of exponentials.

Assume $b > 0$.

7. $\log_b(1) = 0$

8. $\log_b(b) = 1$

9. $\log_b\left(\dfrac{1}{c}\right) = -\log_b(c)$ $\quad (c > 0)$

10. $\log_b\left(c^d\right) = d\log_b(c)$ $\quad (c > 0)$

11. $\log_b\left(\dfrac{c}{d}\right) = \log_b(c) - \log_b(d)$ $(c, d > 0)$

In Exercises 12 to 15, use properties of logarithms to express $\log_{10}(f(x))$ as simply as possible.

12. $f(x) = \dfrac{(\cos(x))^7 \sqrt{(x^2+5)^3}}{4 + (\tan(x))^2}$

13. $f(x) = \sqrt{(1+x^2)^5(3+x)^4\sqrt{1+2x}}$

14. $f(x) = \left(x\sqrt{2+\cos(x)}\right)^{x^2}$

15. $f(x) = \sqrt{\dfrac{x(1+x)}{(\sqrt{1+2x})^3}}$

16. (a) Graph $\log_{1/2}(x)$ and $\log_2(x)$. (b) How is $\log_{1/b}(c)$ related to $\log_b(c)$ for any positive number c?

17. Translate the sentence, "She has a five-figure annual income" into logarithms.

18. Show that $\dfrac{\log_b(a+h)-\log_b(a)}{h} = \dfrac{1}{a}\log_b\left(\left(1+\dfrac{h}{a}\right)^{a/h}\right)$.

19. How would you find $\log_5(3^7)$ if your calculator has only a key for logarithms to the base ten?

20. Until the appearance of calculators, slide rules were commonly used for multiplication and division. Now, the International Slide Rule Museum (http://www.sliderulemuseum.com/) is the world's largest repository of slide rule information. To see how a slide rule multiplies two numbers, mark two pieces of paper with the numbers 1, 2, 4, 8, 16, and 32 placed at equal distances apart, as shown in Figure 1.5.2. To multiply, say, 4 times 8, slide the lower paper so its 1 is under the 4. Then the product of 4 and 8 appears above the 8.

(a) Why does the slide rule work?

(b) How would you make a slide rule for multiplying that has all the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 on both scales?

1	2	4	8	16	32

1	2	4	8	16	32

Figure 1.5.2

21. (a) Show that for positive numbers b and c, neither equal to 1, $\dfrac{\log_c(x)}{\log_b(x)} = \log_c(b)$, independent of x $(x>0)$.

(b) What does (a) imply about the graphs of $y=\log_b(x)$ and $y=\log_c(x)$?

22. Rarely is $\log_b(x+y)$ equal to $\log_b(x)+\log_b(y)$.

(a) Show that if $\log_b(x+y)=\log_b(x)+\log_b(y)$, then $y=\dfrac{x}{x-1}$.

(b) Give an example of x and y that satisfy the equation in (a).

NOTE: This Exercise show that while there is an identity for $\log_b(xy)$, there is no identity involving $\log_b(x+y)$.

23. One way to compute b^4 is to start with b and multiply by b three times, obtaining b^2, b^3, and, finally, b^4. But b^4 can be computed with only two multiplications. First compute b^2, then compute $b^2\cdot b^2$. This raises a question encountered when programming a computer. What is the fewest number of multiplications needed to compute b^n? Call that number $m(n)$. For instance, $m(4)=2$.

(a) Show that $m(2)=1$, $m(3)=2$, $m(5)=3$, $m(6)=3$, $m(7)=4$, $m(8)=3$, and $m(9)=4$.

(b) Show that $m(n)\geq\log_2(n)$.

(c) Show that, when n is a power of 2, then $m(n)=\log_2(n)$.

24. Jane says to Sam, "I'm thinking of a whole number in the interval from 1 to 32. You have to find what it is. I'll answer each question 'yes' or 'no'."

(a) What five questions, in order, should Sam ask to be sure he will guess the number?

(b) If, instead, the interval is from 1 to 50, how should Sam modify his questions to be guaranteed to guess the number in the fewest number of questions?

(c) How is this Exercise related to logarithms?

1.S Chapter Summary

This chapter reviewed precalculus material concerning functions. Calculus begins in the next chapter when we answer questions such as "What happens to $(2^x-1)/x$ as x gets very small?". The answers are used in Chapter 3 to settle questions such as "How rapidly does 2^x change for a slight change in x?"

Section 1.1 introduced the terminology of functions: input (argument), output (value), domain, range, independent variable, dependent variable, piecewise-defined function, inverse of a function, graph of a function, decreasing, increasing, nonincreasing, nondecreasing, positive, and monotonic.

Section 1.2 reviewed x^k and its inverse $x^{1/k}$ (constant exponent, variable base), b^x (constant base, variable exponent) and its inverse $\log_b(x)$, and the six trigonometric functions and their inverses. All angles are measured in radians, unless otherwise stated.

Section 1.3 described five ways of getting new functions from functions f and g, namely $f+g$, $f-g$, fg, f/g, and the composition $f\circ g$.

Section 1.4 developed a short formula for a finite geometric sum with first term a and ratio r, $r\neq 1$: $a+ar+ar^2+\cdots+ar^{n-1}=a(1-r^n)/(1-r)$. And, Section 1.5 reviewed the logarithm function with base b, $b>0$ and $b\neq 1$.

EXERCISES for Section 1.S

Exercises 1 to 10 concern logarithms.

1. Evaluate (a) $\log_3\left(\sqrt{3}\right)$, (b) $\log_3\left(3^5\right)$, and (c) $\log_3\left(\frac{1}{27}\right)$.

2. If $\log_4(A)=2.1$, evaluate (a) $\log_4\left(A^2\right)$, (b) $\log_4\left(\frac{1}{A}\right)$, and (c) $\log_4(16A)$.

3. If $\log_3(5)=a$, what is $\log_5(3)$?

4. Find x if $5\cdot 3^x\cdot 7^{2x}=2$.

5. Solve for x: (a) $2\cdot 3^x=7$, (b) $10^{2x}3^{2x}=5$, (c) $3\cdot 5^x=6^x$, and (d) $3^{5x}=2^{7x}$.

6. Why do only positive numbers have logarithms?

NOTE: Chapter 12 shows that negative numbers have logarithms also, but they involve complex numbers.

7. Evaluate (a) $\log_2\left(2^{43}\right)$, (b) $\log_2(32)$, and (c) $\log_2\left(\frac{1}{4}\right)$.

Exercises 8 to 10 concern the relation between logarithms in different bases.

8. Suppose that you want to obtain $\log_2(17)$ in terms of $\log_3(17)$.
(a) Which would be larger, $\log_2(17)$ or $\log_3(17)$? (b) Show that $\log_2(17)=\log_2(3)\log_3(17)$.

9. You can use your calculator with a key for base-ten logarithms to compute logarithms to any base.
(a) Show why $\log_b(x)=\dfrac{\log_{10}(x)}{\log_{10}(b)}$. (b) Compute $\log_2(3)$.

10. Recall that log (with out a subscript) means $\log_{10}$. Using only the log key (and +, -, ×, and /), compute each of the following numbers: (a) $\log_2(6)$ and $\log_6(2)$, (b) the product of $\log_2(6)$ and $\log_6(2)$, and (c) the product of $\log_7(11)$ and $\log_{11}(7)$. (d) Make a conjecture about $\log_a(b)\cdot\log_b(a)$. (e) Show that the conjecture made in (d) is correct.

---

11. When f and g are decreasing functions, which (if any) of the following functions must also be a decreasing function? (a) $f+g$, (b) $f-g$, (c) f/g, (d) f^2, and (e) $-f$.

12. Give an example of (a) an increasing function f with domain $(0,\infty)$ such that $f(f(x))=x^9$ for all x in $(0,\infty)$ and (b) a decreasing function g with domain $(0,\infty)$ such that $g(g(x))=x^9$ for all x in $(0,\infty)$.

13. What type of function is $f\circ g$ if
(a) f and g are increasing, (b) f and g are decreasing, and (c) f is increasing and g is decreasing? Explain.

14. If f is increasing, what (if anything), can be said about $g=\mathrm{inv}(f)$? That is, must the function g be increasing, decreasing, or is there not enough information to answer this?

15. Simplify. (a) $\log_3\left(3^{17.21}\right)$ (b) $\log_5\left(5^{\sqrt{2}}/25^{\sqrt{3}}\right)$ (c) $\log_2\left(4^{123}\right)$ (d) $\log_2\left(\left(4^5\right)^6\right)$ (e) $\tan(\arctan(3))$

16. Graph the given function on the specified interval.
(a) $y = \sin(x)$, x in $[0, 2\pi]$ (b) $y = \sin(3x)$, x in $\left[0, \frac{\pi}{2}\right]$ (c) $y = \sin(x - \pi)$, x in $[0, 2\pi]$ (d) $y = \sin\left(3x - \frac{\pi}{6}\right)$, x in $\left[0, \frac{\pi}{2}\right]$

17. Simplify the given expression. (a) $\tan\left(\arcsin\left(\frac{1}{2}\right)\right)$, (b) $\tan\left(\arctan\left(\frac{-1}{2}\right)\right)$, and (c) $\sin(\arctan(3))$.

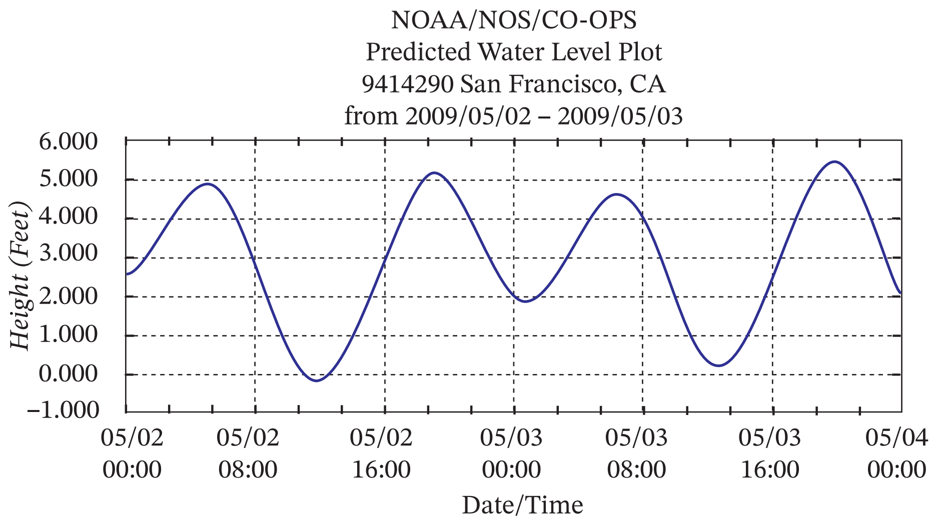


Figure 1.S.1

REFERENCE FOR FIGURE 1.S.1: `http://tidesandcurrents.noaa.gov/gmap3/`

18. The predicted height of the tide at San Francisco for May 2–3, 2009 is shown in Figure 1.S.1.

(a) At what rate was the tide changing at 2 p.m. (14:00) on May 2? Express the answer in feet per hour.

(b) At what date(s) and time(s) was the tide falling the fastest?

(c) At what date(s) and time(s) was it rising the fastest?

(d) At what date(s) and time(s) was it changing most slowly?

(e) How high was the highest tide? The lowest?

19. If your scientific calculator lacks a key to display a decimal approximation to π, how could you use other keys to display it?

20. Imagine that the power key ($\boxed{x^y}$) on your calculator is broken. How would you compute $(2.73)^{3.09}$?

21. Using only the multiplication ($\boxed{\times}$), base-10 exponential ($\boxed{10^x}$), base-10 logarithm ($\boxed{\log_{10}(x)}$), decimal point, and number keys, but no addition or subtraction, how could you compute $0.37 + 1.75$? Carry out the computation.

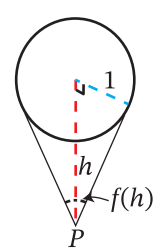


Figure 1.S.2

22. Let a, b, c, d be constants such that $ad - bc \neq 0$.

(a) Show that $y = \dfrac{ax + b}{cx + d}$ is one-to-one.

(b) For which a, b, c, d does the function in (a) equal its inverse function?

23. Figure 1.S.2 shows a circle of radius 1 and a point P at a distance h from it. An arc of the circle is visible from P. That arc subtends an angle.

(a) Express the angle (in radians) as a function of h, $f(h)$.

(b) As P is chosen farther and farther from the circle what happens to $f(h)$?

(c) As P is chosen closer and closer to the circle what happens to $f(h)$?

(Semi-log graphs) In most graphs the scale on the y-axis is the same as the scale on the x-axis, or a constant multiple of it. However, to graph a rapidly increasing function, such as 10^x, it is convenient to distort the y-axis. Instead of plotting the point (x, y) at a height of, say, y inches, plot it at a height of $\log_{10}(y)$ inches. So the datum $(x, 1)$ is drawn at height zero, the datum $(x, 10)$ at height 1 inch, and the datum $(x, 100)$ at 2 inches. Instead of graphing $y = f(x)$, you graph $y = \log_{10}(f(x))$. In particular, if $f(x) = 10^x$, $y = \log_{10}(10^x) = x$: the graph would be a straight line. To avoid having to calculate logarithms, it is convenient to use semi-log graph paper, shown in Figure 1.S.3.

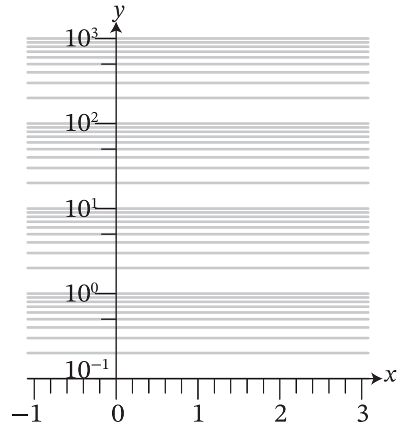


Figure 1.S.3

24. Using semi-log paper, graph $y = 2 \cdot 3^x$.

25. Using semi-log paper, graph $y = \dfrac{2}{3^x}$.

26. Match the curves in Figure 1.S.4 (i-vi) with their equations (a-f).

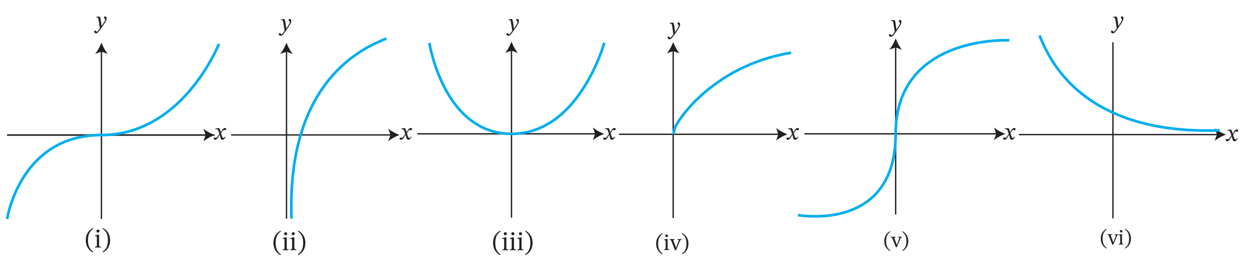


Figure 1.S.4

(a) $y = x^2$ (b) $y = x^3$ (c) $y = \sqrt[3]{x}$ (d) $y = \log_2(x)$ (e) $y = \sqrt{x}$ (f) $y = \left(\dfrac{1}{2}\right)^x$

The equation $y = x - e\sin(x)$, known as **Kepler's equation**, is important in the study of the motion of planets. Here e is the eccentricity of an elliptical orbit, y is related to time, and x is related to an angle.

NOTE: For more information about this problem, search online for "Kepler equation".

27. Consider Kepler's equation with $e = 1$: $f(x) = x - \sin(x)$.
(a) Graph $y = f(x)$. (b) Show that for x in $\left(0, \dfrac{\pi}{2}\right)$, $f(x)$ is an increasing function.

28. (a) For what values of e is the function $f(x) = x - e\sin(x)$ increasing for all numbers x?
(b) Explain why, for the values of e found in (a), even though it cannot be solved explicitly, the equation $y = x - e\sin(x)$ can be solved for x as a function of y, that is, find $x = g(y)$.
(c) How are the graphs of $y = x - e\sin(x)$ and $y = g(x)$ related?

29. The equation $\log_a(b) \cdot \log_b(a) = 1$ makes one wonder, "Is $\log_a(b) \cdot \log_b(c) \cdot \log_c(a) = 1$?" What is the answer?

30. Find all numbers a and b such that $\log_a(b)$ equals $\log_b(a)$.

31. If x is a positive integer and $4 < \log_{10}(x) < 5$, how many digits are in the base-ten representation of x?

32. Denote 723^{723} by c.
(a) What are the first three digits in the decimal representation of c?
(b) How many digits are there in the decimal representation of c?
(c) What is its units digit?

33. Recall that one figure is **similar** to another if one is the other magnified by the same factor in all directions; they are **congruent** if and only if the magnification factor is 1.
(a) Are the graphs of $y = x^2$ and $y = 4x^2$ similar? (b) Are the graphs of $y = x^2$ and $y = 4x^2$ congruent?

(a) three functions that satisfy the equation for all positive values of x and y and (b) one function that does not.

34. $f(x + y) = f(x) + f(y)$ **35.** $f(x + y) = f(x)f(y)$ **36.** $f(xy) = f(x) + f(y)$

37. $f(xy) = f(x)f(y)$ **38.** $f(x) = f(y)$

---

39. Imagine that your calculator fell on the floor and its multiplication and division keys stopped working. However, all the other keys, including the trigonometric, arithmetic, logarithmic, and exponential keys, still functioned. Show how you could use it to calculate the product and quotient of two positive numbers, a and b.

40. A solar cooker can be made in the shape of part of a sphere. The one in Figure 1.S.5 spans only $\pi/3$ (60°) at the center O. For simplicity, take the radius to be 1. Light parallel to OC strikes the cooker at $P = (\cos(\theta), \sin(\theta))$ and is reflected to a point R on the radius OC.
(a) There are two angles of measure θ at P. Why is the top one equal to θ?
(b) Why is the bottom angle at P also θ?
(c) Show that $|OR| = \dfrac{1}{2\cos(\theta)}$.
(d) Show that the heated part of the x-axis has length $\dfrac{1}{\sqrt{3}} - \dfrac{1}{2} \approx 0.077$, about $\dfrac{1}{13}$ of the radius.

CIE 4, ("Reflections on Reflections: Ellipses, Parabolas, and a Solar Cooker") at the end of Chapter 3 describes a parabolic reflector, which reflects all the light to a single point.

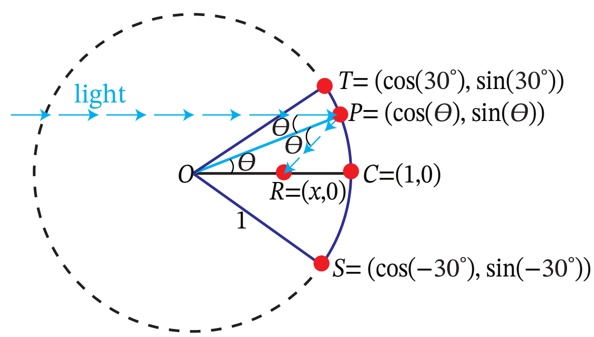


Figure 1.S.5

See also Exercise 4 in CIE 4.

41. A special number, called **Euler's number**, will be introduced in Section 2.2; its definition and approximate value are not needed for this problem. All that you need to know about e is that it is positive, so that logarithms base e are well-defined. Newton computed the logarithms to the base e of 0.8, 0.9, 1.1, and 1.2 to 57 decimal places by hand using a method developed in Section 10.4. Continuing with base e, show how to compute
(a) $\log_e(2)$, using $\log_e(1.2)$, $\log_e(0.8)$, and $\log_e(0.9)$
(b) $\log_e(3)$, using $\log_e(2)$, $\log_e(1.2)$, and $\log_e(0.8)$
(c) $\log_e(4)$, using $\log_e(2)$
(d) $\log_e(5)$, using $\log_e(2)$ and $\log_e(0.8)$
(e) $\log_e(6)$, $\log_e(8)$, $\log_e(9)$, and $\log_e(10)$, using $\log_e(2)$, $\log_e(3)$, and $\log_e(5)$
(f) $\log_e(11)$, using $\log_e(1.1)$, $\log_e(2)$, and $\log_e(5)$
(g) Would the answers to (a) – (f) change if Newton used a base other than e? Explain.

Calculus is Everywhere # 1
Graphs Tell It All

The graph of a function conveys a great deal of information quickly. Here are two examples, both based on numerical data.

The Hybrid Car

A friend of ours bought a hybrid car that runs on electricity at low speeds, on gasoline at higher speeds. The transition between them occurred at about 30 mph. He also purchased the gadget that exhibits "miles-per-gallon" at any instant. With the driver glancing at the speedometer and the passenger watching the gadget, we collected data on fuel consumption (miles-per-gallon) as a function of speed. Figure C.1.1(a) displays what we observed.

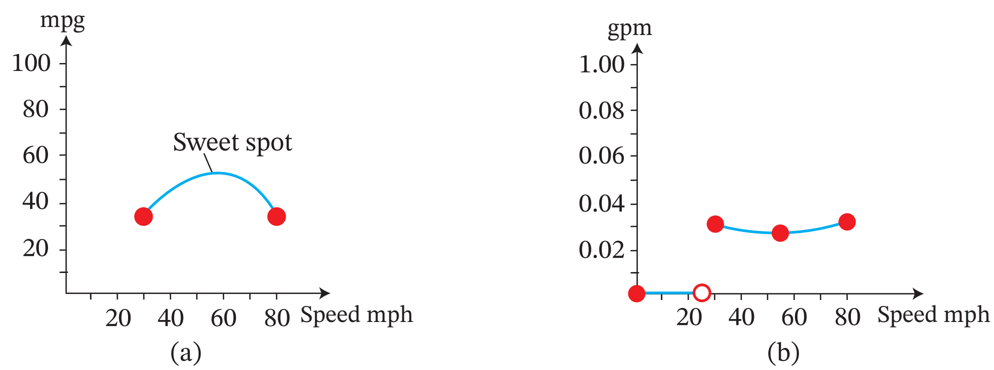


Figure C.1.1

At speeds below 30 mph no gasoline is used. To speak of "miles per gallon" requires division by zero. To avoid this unpleasantness we drew Figure C.1.1(b), which records gallons per mile. The straight-line part of the graph on the speed axis (horizontal) indicates that, at low speeds, no gasoline is used.

The "sweet spot," the speed that maximizes fuel efficiency (as determined by miles-per-gallon), is about 55 mph, while speeds in the range from 40 mph to 70 mph are almost as efficient. However, at 80 mph the car gets only about 30 mpg. In Figure C.1.1(b) a minimum occurs at 55 mph.

Traffic and Accidents

Figure C.1.2 appears in S.K. Stein's, Risk factors of sober and drunk drivers by time of day, *Alcohol, Drugs, and Driving* 5 (1989), pp. 215–227. The vertical scale is described in the paper.

Glancing at the (red, solid) graph labelled "Traffic" in Figure C.1.2 we see that there are peaks at the morning and afternoon rush hours, with minimum traffic around 3 a.m. However, the number of accidents (blue, dashed) is fairly high at that hour. The (black, dotted) "Risk" is measured by the quotient, accidents divided by traffic. It reaches a peak at 1 a.m. This cannot be explained by the darkness at that hour, for the risk rapidly decreases the rest of the night. It turns out that the risk has the same shape as the graph that records the number of drunk drivers.

It is a sobering thought that at any time of day a drunk's risk of being involved in an accident is on the order of one hundred times that of an alcohol-free driver.

Petroleum

The three graphs in Figure C.1.3 show the rate of crude oil production in the United States, the rate at which it was imported, and their sum, the rate of consumption. They are expressed in millions of barrels per day, as a function of time.

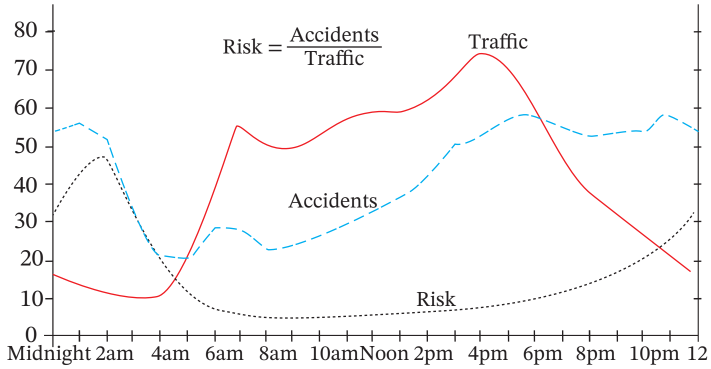


Figure C.1.2

NOTE: A barrel contains 42 gallons, the size of the barrels used to transport oil after it was discovered in Pennsylvania in 1859. This measure for volume of oil is still widely used today.

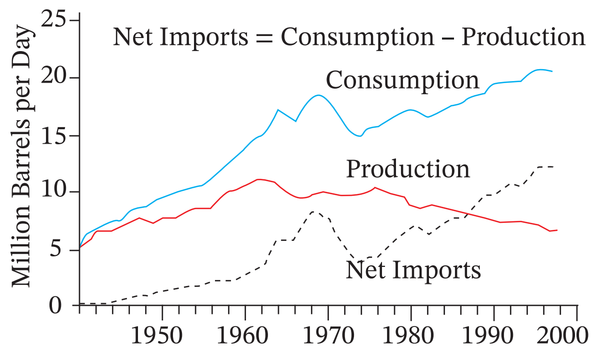


Figure C.1.3

The graphs convey a good deal of history and a warning. In 1950 the United States produced almost enough petroleum to meet its needs, but by 1996 it had to import most of the petroleum it consumed. Moreover, domestic production peaked in 1970.

The total amount of petroleum in the earth is finite; it will run out, and the Age of Oil will end. Geologists, having gone over the globe with a fine-tooth comb, believe they have already found all the major oil deposits. No wonder that the development of alternative sources of energy has become a high priority.

Calculus is Everywhere # 2

How Banks Multiply Money

As of 2015 there were over \$10.6 trillion US dollars in the form of currency, in deposits in accounts, in money market mutual funds, and so on. Given that banks cannot create money, where did all that money come from?

Banks start with our deposits. While they are not allowed to create more money, the regulatory guidelines do provide opportunities for them to leverage these funds for other purposes. Here is one way in which they do this.

When someone makes a deposit at a bank, the bank lends most of it. It cannot lend all of it, for it must keep a reserve to meet the needs of depositors who may withdraw money from their accounts. The government stipulates what this reserve must be, usually between 10 and 20 percent of the deposit.

Let's use 20 percent and see what happens when Sam deposits \$1,000 in a bank.

Our interest in this problem arose from a discussion one of the authors had with his daughter, Susanna Stein. She had taken a graduate-level course in economics and has subsequently verified this explanation with professional economists:

> So, let's say Sam deposits \$1,000 in a bank. Sam's bank must keep \$200 (20% of \$1,000) as reserves, placed in its own vault or at the Federal Reserve Bank. The bank then lends \$800 to Jane, who has an account at the same bank or at another bank. Sam still has the \$1,000 to use as he wants, but a new \$800 are out in the economy for Jane to use as she wants. (Well, she better keep enough of this \$800 to pay the interest on her loan.) I think it's that simple. (See Figure C.2.1.)

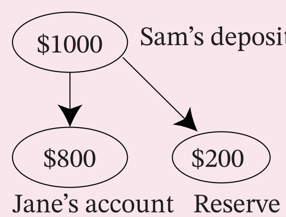


Figure C.2.1

If Sam wants to make a withdrawal from his bank, he still can, either as cash or as a check. Thus Sam's \$1,000 has become \$1,000+\$800=\$1,800.

The process does not have to stop there. The bank is free to lend 80% of Jane's account, which is \$640, and keep the other \$160 (20% of \$800) in reserves. The recipient of the \$640 can then deposit it at a third bank, which must retain 20%, or \$128, but is free to lend 80%, which is \$512. At this point the total of the four loans is

$$\$1,000 + \$800 + \$640 + \$512. \tag{C.2.1}$$

Each summand is 0.8 times the preceding summand. The sum (C.2.1) can be written as

$$1000\left(1 + 0.8 + 0.8^2 + 0.8^3\right).$$

The process goes on indefinitely, through a fifth person, a sixth, and so on. The impact of the initial deposit of \$1000 after n stages is 1000 times the sum

$$1 + 0.8 + 0.8^2 + 0.8^3 + \cdots + 0.8^n. \tag{C.2.2}$$

Being the sum of a geometric progression, with ratio $r = 0.8$, the sum (C.2.2) equals $(1-0.8^{n+1})/(1-0.8)$ and that quotient approaches $1/0.2 = 5$ as n increases. Thus the original \$1000 could create loans whose total value approaches \$5000. Economists say that the **multiplier** is $M = 5$, and the total impact of Sam's initial deposit is five times the initial deposit. While no new money has actually been created, the compounding effect of this infinite sequence of loans from the amounts held in reserve from earlier loans enabled the banks to support an additional \$4000 in loans. The sequence of deposits and loans involves having faith in the future. If this trust is destroyed, then there may be a run on the bank as depositors rush to take their money out. If that disaster can be avoided, then banking is a delightful business.

The concept of the multiplier also appears in measuring economic activity. Assume that the government spends a million dollars on a new road. That amount goes to firms and individuals who build the road. In turn, those firms and individuals spend a portion of that income. This process of earn and spend continues to trickle through the economy. Figure C.2.2 shows additional levels of loans and reserves that are possible with Sam's \$1,000 deposit. The total impact may be much more than the initial amount the government spent. Again, the ratio between the total impact and the initial expenditure is called the **multiplier**.

The mathematics behind the presentation of the multiplier in this CIE is the theory of the geometric series, summing the successive powers of a fixed positive number that is less than 1. Exercise 26 in Section 12.S examines the multiplier without using geometric series. Instead, it uses concepts introduced in later chapters.

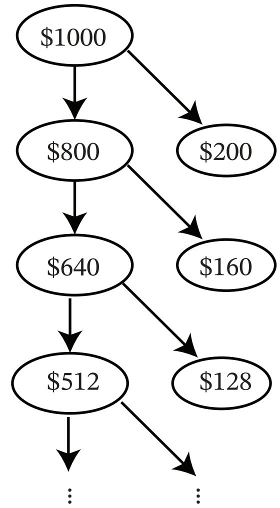


Figure C.2.2

EXERCISES for CIE C.2

1. If the process starting with Sam's \$1,000 goes on and on (forever), what number do the total reserves approach?
  (a) Answer with the aid of geometric series.
  (b) Answer the same question without using geometric series.

2. Consider a deposit of \$$A$ with a reserve rate r $(0 < r < 1)$.
  (a) Use a geometric series to find the total amount loaned and the total reserves.
  (b) Repeat (a) without using geometric series.

3. If the amount a bank must keep on reserve is cut in half, what effect does this have on the multiplier?

Chapter 2

Introduction to Calculus

Two main concepts in calculus are the derivative and the integral. Underlying both is the concept of a **limit**, which this chapter introduces.

The journey starts in Section 2.1 where knowledge about the slope of a line is used to define the slope at a point on a curve. The four limits introduced in Section 2.2 provide the foundation for computing other limits, particularly the ones needed in Chapter 3. The next few sections present a definition of the limit that pertains to cases other than finding the slope of a tangent line (Section 2.3), explores continuous functions (Section 2.4) and presents three properties of continuous functions (Section 2.5). We conclude, in Section 2.6, with a look at graphing functions by hand and with the use of technology.

2.1 Slope at a Point on a Curve

The slope of a (straight) line is the quotient of rise over run, as shown in Figure 2.1.1(a) and (b).

If the line drops as you move from left to right the rise is considered to be negative and the slope is negative (see Figure 2.1.1(b)). It does not matter which point P is chosen on the line. The slope is the same at every point on the line.

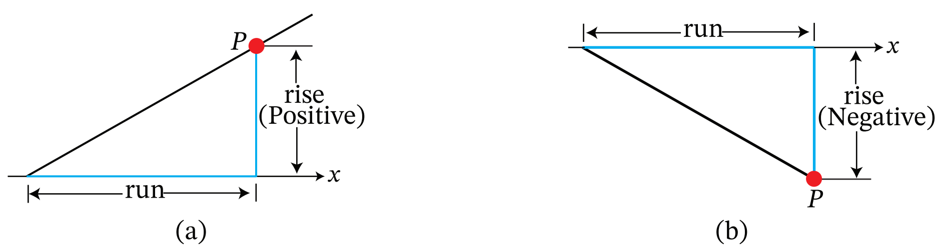


Figure 2.1.1

Hills on US Interstates have slopes that never exceed 6%=0.06. This means the road can rise (or fall) at most 6 feet in 100 (horizontal) feet, see Figure 2.1.2(a). On the other hand the steepest street in San Francisco is Filbert Street, with a slope of 0.315, see Figure 2.1.2(b).

Now consider a line L relative to x- and y-axes, as in Figure 2.1.2(c). Since two points determine the line, they also determine its slope.

To find the slope pick two distinct points on the line, (x_1, y_1) and (x_2, y_2). They determine a rise of $y_2 - y_1$ and a run of $x_2 - x_1$, hence

$$\text{slope} = \frac{\text{rise}}{\text{run}} = \frac{y_2 - y_1}{x_2 - x_1}.$$

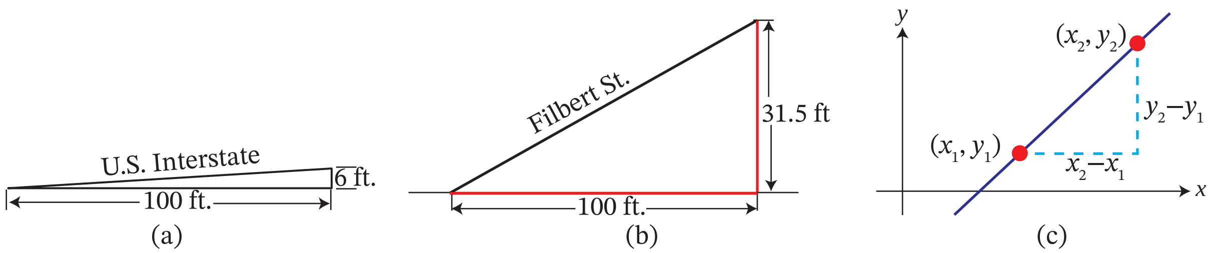


Figure 2.1.2

The rise could be negative, which occurs if $y_2 < y_1$. The slope is positive if both the rise and run are negative.

EXAMPLE 1. Find the slope of the line through $(4, -1)$ and $(1, 3)$.

SOLUTION Figure 2.1.3 shows the two points and the line they determine. Let $(4, -1)$ be (x_1, y_1) and let $(1, 3)$ be (x_2, y_2). So the slope is

$$\frac{3 - (-1)}{1 - 4} = \frac{4}{-3} = -\frac{4}{3}.$$

That the slope is negative is consistent with Figure 2.1.3(a) where the line descends as you go from left to right. ◆

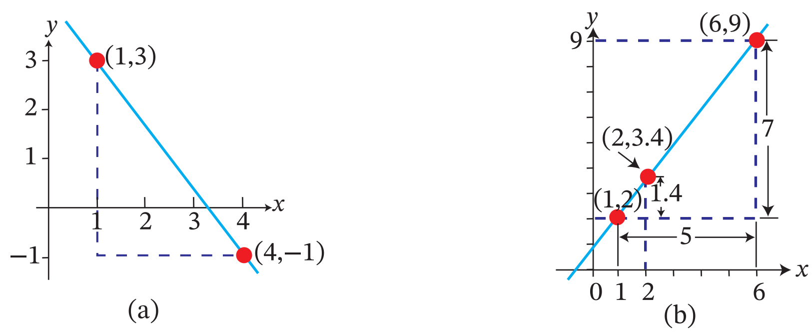


Figure 2.1.3

The slope in Example 1 does not change if $(4, -1)$ is called (x_2, y_2) and $(1, 3)$ is called (x_1, y_1).

If we know a point on a line and its slope we can draw the line. For instance, say we know a line goes through $(1, 2)$ and has slope 1.4, which is 7/5. We draw a triangle with a vertex at $(1, 2)$ and legs parallel to the axes, as in Figure 2.1.3(b). The rise and run of the triangle could be 7 and 5, or 1.4 and 1, or any two numbers in the ratio 1.4 : 1.

If we know a point on a line, say (a, b), and its slope, m, we can draw the line and also write its equation. If our two points are now (x, y) and (a, b), then

$$\text{slope} = \frac{y - b}{x - a} = m.$$

The line's equation is

$$y - b = m(x - a) \qquad \text{or} \qquad y = m(x - a) + b.$$

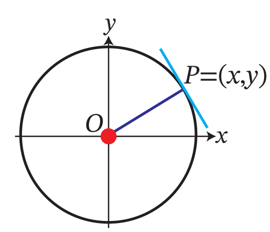


Figure 2.1.4

The Slope at Points on a Circle

A circle with radius 2 and center at the origin $O = (0, 0)$ is shown in Figure 2.1.4. How do we find the tangent line to the circle at $P = (x, y)$? By "tangent line" we mean, informally, the line that most closely resembles the curve near

P. The tangent line is perpendicular to the segment OP, and the slope of OP is $(y-0)/(x-0)$. Exercise 21 shows that the product of the slopes of perpendicular lines is -1 assuming neither line has slope 0. Thus the slope of the tangent line at (x, y) is $-x/y$. For instance, at $(6/5, 8/5)$, a point on the circle, the slope of OP is $(8/5)/(6/5) = 4/3$, which means that the slope of the tangent line at P is $-3/4$.

We say that the slope of the circle at (x, y) is $-x/y$ because that is the slope of the tangent line at this point.

The Slope at a Point on the Curve $y = x^2$

Figure 2.1.5(a) shows the graph of $y = x^2$. How can we find the slope of the tangent line at $(2, 4)$? If we knew the slope, we could draw the tangent.

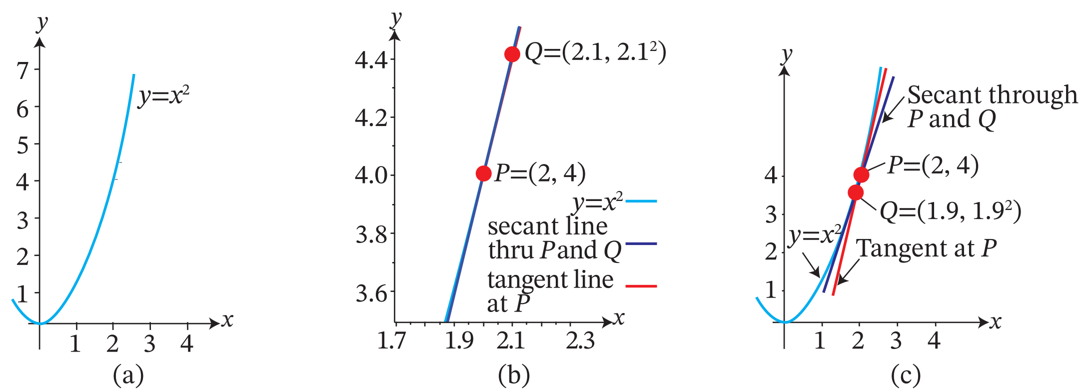


Figure 2.1.5

If we know two points on the tangent, we could calculate its slope. But we know only one point on it, namely $(2, 4)$. To get around this difficulty we will choose a point Q on the parabola $y = x^2$ near P and compute the slope of the line through P and Q. Such a line is called a **secant**. As Figure 2.1.5(b) suggests, when Q is near P the secant line resembles the tangent line at $(2, 4)$. For instance, in Figure 2.1.5(b) we have chosen $Q = (2.1, 2.1^2)$ and computed the slope of the line through P and Q:

$$\text{Slope of secant} = \frac{\text{rise}}{\text{run}} = \frac{\text{change in } y}{\text{change in } x} = \frac{2.1^2 - 2^2}{2.1 - 2} = \frac{4.41 - 4}{0.1} = \frac{0.41}{0.1} = 4.1.$$

Thus an estimate of the slope of the tangent line is 4.1. A careful examination of Figure 2.1.5(b) shows this is an overestimate of the slope of the tangent line. So the slope of the tangent line is less than 4.1.

We can also choose the point Q on the parabola to the left of $P = (2, 4)$. For instance, choosing $Q = (1.9, 1.9^2)$, as in Figure 2.1.5(c), the slope of the line through P and Q is

$$\text{Slope of secant} = \frac{\text{rise}}{\text{run}} = \frac{\text{change in } y}{\text{change in } x} = \frac{1.9^2 - 2^2}{1.9 - 2} = \frac{3.61 - 4}{-0.1} = \frac{-0.39}{-0.1} = 3.9.$$

Inspecting Figure 2.1.5(c) shows that this underestimates the slope of the tangent line. So the slope of the tangent line is greater than 3.9. We have trapped the slope of the tangent line between 3.9 and 4.1. To get closer bounds we could choose Q even nearer to $(2, 4)$.

Figures 2.1.5(b) and (c) are very similar. There are only two differences: the location of the point Q and the graph of the secant line through P and Q. With these two choices of Q the respective secant lines through P and Q are such good approximations of the tangent line at P that they are virtually indistinguishable — that is why they are presented in two separate figures. Fortunately the calculation of the slope of the graph of $y = x^2$ at $(2, 4)$ does not involve the figure.

Using $Q = (2.01, 2.01^2)$ leads to the estimate

$$\frac{2.01^2 - 2^2}{2.01 - 2} = \frac{4.0401 - 4}{0.01} = \frac{0.0401}{0.01} = 4.01$$

and using $Q = (1.99, 1.99^2)$ yields the estimate

$$\frac{1.99^2 - 2^2}{1.99 - 2} = \frac{3.9601 - 4}{-0.01} = \frac{-0.0399}{-0.01} = 3.99.$$

Now we know the slope of the tangent at $(2,4)$ is between 3.99 and 4.01.

To make better estimates we could choose Q even nearer to $(2,4)$, say $(2.0001, 2.0001^2)$. But, still, the slopes we would get would just be estimates.

What we need to know is what happens to the quotient

$$\frac{x^2 - 2^2}{x - 2} \qquad \text{as } x \text{ gets closer and closer to 2.}$$

In the next section we will determine what happens to this quotient as x approaches 2. No experiments will be needed, only a little algebra. This chapter is devoted to answering this and other questions of the same type:

QUESTION: What happens to the values of a function as the function's inputs are chosen nearer and nearer to a fixed number?

The Slope at a Point on the Curve $y = 1/x$

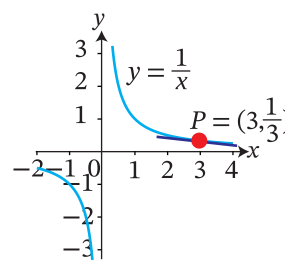


Figure 2.1.6

Figure 2.1.6 shows the graph of $y = 1/x$. Let us estimate the slope of the tangent line to this curve at $(3, 1/3)$.

It is clear that the slope will be negative. We could draw a run-rise triangle on the tangent and get an estimate for the slope. Better estimates are obtained by selecting a nearby point Q on the curve and finding the slope of the secant line through the points P and Q.

We pick $Q = (3.1, 1/(3.1))$. The points $P = (3, 1/3)$ and Q determine a secant whose slope is

$$\frac{\frac{1}{3.1} - \frac{1}{3}}{3.1 - 3} = \frac{-\frac{0.1}{3(3.1)}}{0.1} = -\frac{1}{3(3.1)} = -\frac{1}{9.3} \approx -0.10753.$$

This is an estimate of the slope of the tangent line.

Using $Q = (2.9, 1/2.9)$, we get another estimate:

$$\frac{\frac{1}{2.9} - \frac{1}{3}}{2.9 - 3} = \frac{\frac{0.1}{3(2.9)}}{-0.1} = -\frac{1}{3(2.9)} = -\frac{1}{8.7} \approx -0.11494.$$

By choosing Q nearer $(3, 1/3)$ we could get better estimates.

The Slope at a Point on the Curve $y = \log_2(x)$

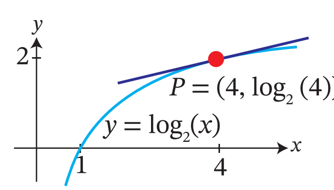


Figure 2.1.7

Figure 2.1.7 shows the graph of $y = \log_2(x)$. Clearly, its slope is positive at all points.

We will make two estimates of the slope at $(4, \log_2(4))$. Before going any further, observe that $(4, \log_2(4)) = (4, 2)$ because $\log_2(4) = \log_2(2^2) = 2$.

For the nearby point Q, let us use $(4.001, \log_2(4.001))$. The slope of the secant through $P = (4,2)$ and Q is

$$\frac{\log_2(4.001) - 2}{4.001 - 4} = \frac{\log_2(4.001) - 2}{0.001}.$$

We use Exercise 9(a) in Section 1.S, and a calculator, to estimate $\log_2(4.001)$ to five decimal places:

$$\log_2(4.001) = \frac{\log_{10}(4.001)}{\log_{10}(2)} \approx \frac{0.60217}{0.30103} \approx 2.00036.$$

So the estimate of the slope of the tangent to $y = \log_2(x)$ at $(4,2)$ is

$$\frac{2.00036 - 2}{0.001} = \frac{0.00036}{0.001} = 0.36.$$

The number 0.36 is an estimate of the slope of the graph of $y = \log_2(x)$ at $P = (4,2)$. It is not the slope there, but even so it could help us draw the tangent at P.

Summary

We introduced a method that uses a nearby point Q on the curve to construct a secant line whose slope estimates the slope of the tangent line to a curve at a point P on the curve. The closer Q is to P, the better the estimate. We applied the technique to the curves $y = x^2$, $y = 1/x$, and $y = \log_2(x)$. In no case did we have to draw the curve. Nor did we find the slope of the tangent except in the special cases of a line and a circle. In most cases we found only estimates. The rest of this chapter develops methods for finding what happens to a function, such as $f(x) = (x^2 - 4)/(x - 2)$, as x gets nearer and nearer a given number.

EXERCISES for Section 2.1

In Exercises 1 and 2 copy the figure and estimate the slope of each line as well as you can. In each case draw a "run–rise" triangle and measure the rise and run with a ruler. (A centimeter ruler is more convenient than one marked in inches.)

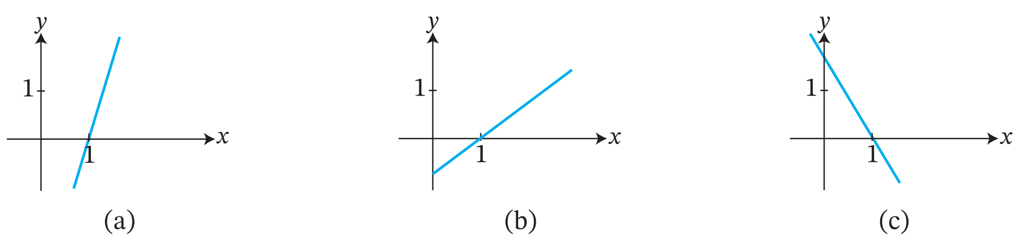


Figure 2.1.8

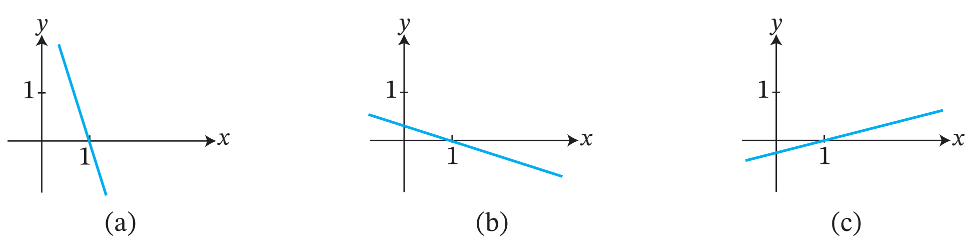


Figure 2.1.9

1. Figure 2.1.8

2. Figure 2.1.9

3. Draw an x-axis and lines of slope $\frac{1}{2}$, 1, 2, 4, 5, -1, and $\frac{-1}{2}$.

4. Draw an x-axis and lines of slope $\frac{1}{3}$, 1, 3, -1, and $\frac{-2}{3}$.

In Exercises 5 to 8 draw the line determined by the given information and give an equation for it.

5. through $(1,2)$ with slope -3

6. through $(1,4)$ and $(4,1)$

7. through $(-2,-4)$ and $(0,4)$

8. through $(2,-1)$ with slope 4

---

9. (a) Graph the line whose equation is $y=2x+3$. (b) Find its slope.

10. (a) Graph the line whose equation is $y=-3x+1$. (b) Find its slope.

11. Estimate the slope of the tangent line to $y=x^2$ at $(1,1)$ using (a) the nearby point $(1.001, 1.001^2)$ and (b) the nearby point $(0.999, 0.999^2)$.

12. Estimate the slope of the tangent line to $y=x^2$ at $(-3,9)$ using (a) the nearby point $(-3.01, (-3.01)^2)$ and (b) the nearby point $(-2.99, (-2.99)^2)$.

13. Estimate the slope of the tangent line to $y=\frac{1}{x}$ at $(1,1)$
- (a) by drawing a tangent line at $(1,1)$ and a rise-run triangle.
- (b) by using the nearby point $\left(1.01, \frac{1}{1.01}\right)$.
- (c) Is the slope of the tangent line smaller or larger than this estimate?

14. Estimate the slope of the tangent line to $y=\frac{1}{x}$ at $(0.5,2)$
- (a) by drawing a tangent line at $(0.5,2)$ and a rise-run triangle.
- (b) by using the nearby point $\left(0.49, \frac{1}{0.49}\right)$.
- (c) Is the slope of the tangent line smaller or larger than this estimate?

15. Estimate the slope of the tangent line to $y=\log_2(x)$ at $(2, \log_2(2))$.
- (a) by drawing a tangent line at $(2, \log_2(2))$ and a rise-run triangle.
- (b) by using the nearby point $ds\left(2.01, \log_2(2.01)\right)$.
- (c) Is the slope of the tangent line smaller or larger than this estimate?

16. Estimate the slope of the tangent line to $y=\log_2(x)$ at $(4,2)$.
- (a) by drawing a tangent line at $(4,2)$ and a rise-run triangle.
- (b) by using the nearby point $ds\left(3.99, \log_2(3.99)\right)$.
- (c) Is the slope of the tangent line smaller or larger than this estimate?

17.
- (a) Graph $y=x^2$ carefully for x in $[-2,3]$.
- (b) Draw the tangent line to $y=x^2$ at $(1,1)$ as well as you can and estimate its slope.
- (c) Using the nearby points $\left(1.1, 1.1^2\right)$ and $\left(0.9, 0.9^2\right)$, determine upper and lower bounds on the slope of the tangent line at $(1,1)$.

18.
- (a) Graph $y=2^x$ carefully for x in $[0,2]$.
- (b) Draw the tangent line to $y=2^x$ at $(1,2)$ as well as you can and estimate its slope.
- (c) Using the nearby points $\left(1.03, 2^{1.03}\right)$ and $(0.97, 2^{0.97})$, determine upper and lower bounds on the slope of the tangent line at $(1,2)$.

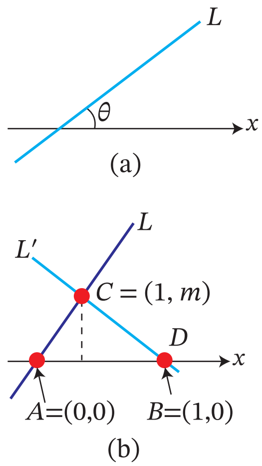


Figure 2.1.10

19. (a) Show that when one computes the slope of the line through $P=(1,2)$ and $Q=(5,3)$, one gets the same answer no matter which is called (x_1,y_1) and which is called (x_2,y_2).
(b) Show that in general both ways of labeling P and Q give the same slope.

20. In Figure 2.1.10(a) the angle between the x-axis and a line that crosses it is called the **angle of inclination** of the line. It is measured counterclockwise from the positive x-axis to the line, as shown in Figure 2.1.10(a). The symbol θ denotes both the angle and its measure, $0<\theta<\pi$. For a line parallel to the x-axis, θ is defined to be 0. Show that $\tan(\theta)$ equals the slope of the line.

21. (This exercise shows that the product of the slopes of perpendicular lines is -1.) One line, L, has positive slope m. For convenience, we assume that line L has positive slope and passes through the origin; L' is the line through $(1,m)$ perpendicular to L of slope m'. The point $(1,m)$ also lies on L. See Figure 2.1.10(b).
(a) Use the fact that triangles $\triangle ABC$ and $\triangle CBD$ are similar to show that L' crosses the x-axis at $(1+m^2,0)$.
(b) Show that the slope of L' is $\dfrac{-1}{m}$. Thus $mm'=-1$.

2.2 Four Special Limits

This section develops the notion of a limit of a function using four examples that play key roles in Chapter 3.

A Limit Involving x^n

Assume that a and n are fixed numbers, with n a positive integer.

> QUESTION: What happens to the quotient $\dfrac{x^n-a^n}{x-a}$ as x is chosen nearer and nearer to a?

In Section 2.1 we met such a question with $n=2$ and $a=2$. A problem on slope led us to examine what happens to $(x^2-4)/(x-2)$ as x approaches 2. Before we examine other cases let us settle that one with a bit of algebra.

We have, for $x\neq 2$,

$$\frac{x^2-2^2}{x-2}=\frac{(x-2)(x+2)}{x-2}=x+2.$$

As x gets nearer and nearer to 2 the expression $x+2$ approaches $2+2=4$. Thus $(x^2-4)/(x-2)$ approaches 4 as x approaches 2, in agreement with our estimates when $x=3.9$ and $x=4.1$. The key to this example is that $x-2$ factors x^2-2^2.

Next, consider the same question with $n=3$ and $a=2$:

> QUESTION: What happens to $\dfrac{x^3-2^3}{x-2}$ as x gets closer and closer to 2?

As x approaches 2, the numerator approaches $2^3-2^3=0$. Because 0 divided by anything (other than 0) is 0 we suspect that the quotient may approach 0. But the denominator approaches $2-2=0$. This is unfortunate because division by zero is not defined.

That x^3-2^3 approaches 0 as x approaches 2 may make the quotient small. That the denominator approaches 0 as x approaches 2 may make the quotient large. How the opposing forces balance determines what happens to the quotient as x approaches 2.

We have already seen that it is pointless to replace x in the quotient by 2 as this leads to $(2^3-2^3)/(2-2)=0/0$, a meaningless expression. Instead, do some experiments and see how the quotient behaves for specific values of x

x	x^3	x^3-2^3	$x-2$	$\frac{x^3-2^3}{x-2}$
1.90	6.859	−1.141	−0.1	11.41
1.99	7.8806	−0.1194	−0.01	11.94
1.999				
2.00	8.0000	0.0000	0.00	undefined
2.001				
2.01	8.1206	0.1206	0.01	12.06
2.10	9.261	1.261	0.1	12.61

Table 2.2.1

near 2, some less than 2, and some more than 2. Table 2.2.1 shows the results as x increases from 1.9 to 2.1. You are invited to fill in the empty places in the table and to add to the list with values of x even closer to 2.

Math is not a spectator sport. Check the calculations recorded in Table 2.2.1. Fill in the empty cells.

The cases with $x = 1.99$ and 2.01, being closest to 2, provide the best estimates of the quotient. They suggest that the quotient the equation approaches a number near 12 as x approaches 2, whether from below or from above.

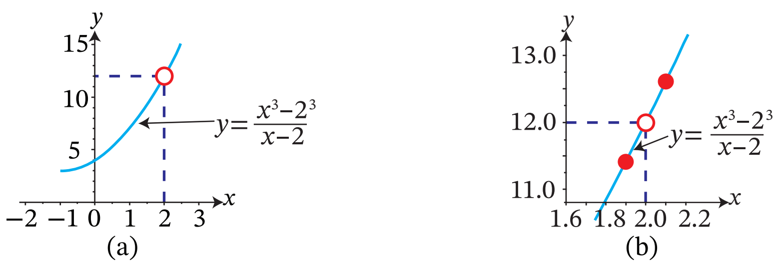


Figure 2.2.1

Figure 2.2.1 provides a graphical view. In (a), the graph of $(x^2-2^2)/(x-2)$ for x between -1 and 3 looks like a parabola with the point corresponding to $x = 2$ deleted because division by zero makes no sense. Confirmation that the values of the quotient approach the same number when x approaches 2 from both the left and from the right is obtained by zooming in to a smaller interval around $x = 2$. In (b), the quotient is plotted for x between 1.6 and 2.4. Assuming the graph accurately represents this function, it seems reasonable that the quotient approaches 12 as x approaches 2.

NOTE: The graphs in Figure 2.2.1(b) and Figure 2.2.2(b) are not straight lines. They look straight only because the viewing windows are so small. That the graphs of many common functions look straight as you zoom in on a point will be important in Section 3.1.

While the numerical and graphical evidence is suggestive, the question can be answered once and for all with a little bit of algebra: x^3-2^3 factors as $(x-2)(x^2+2x+2^2)$. We have

$$\frac{x^3-2^3}{x-2} = \frac{(x-2)(x^2+2x+2^2)}{x-2}. \tag{2.2.1}$$

When x is not 2, (2.2.1) is meaningful, and we can cancel $(x-2)$, showing that

$$\frac{x^3-2^3}{x-2} = x^2+2x+2^2, \qquad \text{for all } x \text{ other than 2.}$$

It is easy to see what happens to x^2+2x+2^2 as x gets nearer and nearer to 2: x^2+2x+2^2 approaches $4+4+4=12$. This agrees with the calculations (see Table 2.2.1).

We say "the limit of $(x^3-2^3)/(x-2)$ as x approaches 2 is 12" and use the shorthand

$$\lim_{x\to 2}\frac{x^3-2^3}{x-2} = \lim_{x\to 2}(x^2+2x+2^2) = 2^2+2^2+2^2 = 12.$$

The key to evaluating the limit in the cases with $n=2$ and $n=3$ is that $x-a$ is a factor of x^2-a^2 and x^3-a^3. Exercises 43 and 44 show that $x-a$ is a factor of x^n-a^n for all positive integers n. Similar algebra, depending on the formula for the sum of a geometric series, yields

Theorem 2.2.1: Special Limit #1: A Limit for x^n

For any positive integer n and fixed number a,

$$\lim_{x\to a}\frac{x^n-a^n}{x-a} = n\cdot a^{n-1}$$

A Limit Involving b^x

Let's try to find the slope of $y=b^x$ at $x=0$ for $b=2$ and for $b=4$. This brings us to needing to be able to answer the following question:

QUESTION: What happens to $\dfrac{2^x-1}{x}$ and to $\dfrac{4^x-1}{x}$ as x approaches 0?

Consider $(2^x-1)/x$ first. As x approaches 0, 2^x-1 approaches $2^0-1=1-1=0$. Since the numerator and denominator in $(2^x-1)/x$ both approach 0 as x approaches 0, we face the same challenge as with $(x^3-2^3)/(x-2)$. There is a battle between two opposing forces.

There are no algebraic tricks to help in this case. Instead, we will rely upon numerical data. While this will be convincing, it is not mathematically rigorous. Later, we will present a way to evaluate this limit that does not depend upon numerical computations.

Table 2.2.2 records some results (rounded to three decimal places) for four choices of x. You are invited to fill in the blanks and to add values of x even closer to 0.

You should also take some time to examine the two graphs of $(2^x-1)/x$ in Figure 2.2.2 to convince yourself that the quotient approaches a single value as x approaches 0 from the left and from the right. The view in Figure 2.2.2(b) provides a better estimate of the y-coordinate of the missing point.

It seems that as x approaches 0, $(2^x-1)/x$ approaches a number close to 0.693. We write

$$\lim_{x\to 0}\frac{2^x-1}{x} \approx 0.693 \qquad \text{rounded to three decimal places.}$$

x	2^x	2^x-1	$\dfrac{2^x-1}{x}$
-0.01	0.993092	-0.006908	0.691
-0.001	0.999307	-0.000693	0.693
-0.0001			
0.0001			
0.001	1.000693	0.000693	0.693
0.01	1.006956	0.006956	0.696

Table 2.2.2

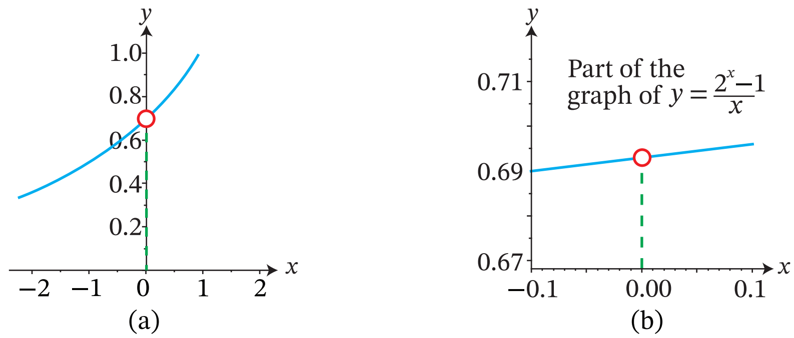


Figure 2.2.2

It is then a simple matter to find

$$\lim_{x\to 0} \frac{4^x - 1}{x}.$$

By factoring of the difference of two squares, $a^2 - b^2 = (a-b)(a+b)$, we have $4^x - 1 = (2^x)^2 - 1^2 = (2^x - 1)(2^x + 1)$. Hence

$$\frac{4^x - 1}{x} = \frac{(2^x - 1)(2^x + 1)}{x} = (2^x + 1)\frac{2^x - 1}{x}.$$

As $x \to 0$, $2^x + 1$ approaches $2^0 + 1 = 1 + 1 = 2$ and $(2^x - 1)/x$ approaches (approximately) 0.693. Thus,

$$\lim_{x\to 0} \frac{4^x - 1}{x} \approx 2 \cdot 0.693 \approx 1.386 \qquad \text{rounded to three decimal places.}$$

We now have strong evidence about the values of $\lim_{x\to 0}(b^x - 1)/(x)$ for $b = 2$ and $b = 4$. They suggest that the larger b is, the larger the limit is. Since $\lim_{x\to 0}(2^x - 1)/x$ is less than 1 and $\lim_{x\to 0}(4^x - 1)/x$ is more than 1, it seems reasonable that there is a value of b such that $\lim_{x\to 0}(b^x - 1)/x = 1$. This special number is called e, **Euler's number**.

Euler named this constant e, but no one knows why he chose this symbol.

We know that e is between 2 and 4 and that $\lim_{x\to 0}(e^x - 1)/x = 1$. It turns out that e is an irrational number with an endless decimal representation that begins $2.718281828\ldots$. In Chapter 3 we will see that e is as important in calculus as π is in geometry and trigonometry.

This finding is important enough that we state it as a theorem.

Theorem 2.2.2: Special Limit #2: Basic Property of e

$$\lim_{x\to 0} \frac{e^x - 1}{x} = 1 \qquad \textit{and} \qquad e \approx 2.71828.$$

In Section 1.2 it was remarked that the logarithm to base b, $\log_b$, can be defined for any base $b > 0$, except $b = 1$. The logarithm to the base $b = e$ deserves special attention: $\log_e(x)$ is called the **natural logarithm**, and is typically written as $\ln(x)$. While most scientists and engineers use $\log(x)$ to refer to $\log_{10}(x)$, mathematicians often use $\log(x)$ to refer to $\log_e(x)$. Thus, in particular,

$$y = \ln(x) \quad \text{is equivalent to} \quad x = e^y.$$

As with any logarithm function, the domain of ln is the set of positive numbers $(0, \infty)$ and the range is the set of all real numbers $(-\infty, \infty)$.

Often the exponential function with base e is written as exp. This notation is convenient when the input is complicated:

$$\exp\left(\frac{\sin^3(\sqrt{x})}{\cos(x)}\right) \qquad \text{is easier to read than} \qquad e^{\sin^3(\sqrt{x})/\cos(x)}.$$

Today, many calculators, spreadsheets, and computer languages use exp to name the exponential function with base e.

A Limit Involving $\sin(x)$

To find the slope of $y = \sin(x)$ at $(0,0)$ we need to know:

QUESTION: What happens to $\dfrac{\sin(x) - \sin(0)}{x - 0} = \dfrac{\sin(x)}{x}$ as x gets nearer and nearer to 0?

Here x represents an angle, measured in radians. In Chapter 3 we will see that in calculus radians are much more convenient than degrees.

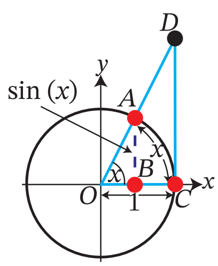


Figure 2.2.3

We first consider positive values of x. Because we are interested in x near 0, we assume that $x < \pi/2$. Figure 2.2.3 identifies both x and $\sin(x)$ on a circle of radius 1, the **unit circle**. Note that x appears twice in Figure 2.2.3, once as a measure of an angle and once as a measure of arc length on a unit circle. This occurs because of the definition of radians.

To get an idea of the value of the limit, try $x = 0.1$. Setting our calculator in the radian mode, we find

$$\frac{\sin(0.1)}{0.1} \approx \frac{0.099833}{0.1} = 0.99833. \tag{2.2.2}$$

Likewise,

$$\frac{\sin(0.01)}{0.01} \approx \frac{0.0099998}{0.01} = 0.99998. \tag{2.2.3}$$

Based on (2.2.2) and (2.2.3), we suspect that maybe the limit is 1.

Geometry and a bit of trigonometry show that $\lim_{x\to 0} \sin(x)/x$ is indeed 1. First, using Figure 2.2.3, we show that $\sin(x)/x$ is less than 1 for x between 0 and $\pi/2$. We know that $\sin(x) = |AB|$. Now, segment AB is shorter than straight-line segment AC since a leg of a right triangle is shorter than its hypotenuse. Also, segment AC is shorter than the circular arc joining A to C, since the shortest distance between two points is a straight line. Thus, $\sin(x) < |AC| < x$. So $\sin(x) < x$. Since x is positive, dividing by x preserves the inequality. We have

$$\frac{\sin(x)}{x} < 1. \tag{2.2.4}$$

Next, we show that $\sin(x)/x$ is greater than something that gets near 1 as x approaches 0. Figure 2.2.3 helps with this step.

The area of triangle $\triangle OCD$ is

$$\frac{1}{2}|OC||CD| = \frac{1}{2}\cdot 1 \cdot |CD| = \frac{1}{2}|CD|.$$

All that remains is to determine the length $|CD|$. The fact that triangles $\triangle OAB$ and $\triangle OCD$ are similar allows us to observe that

$$\frac{\sin(x)}{\cos(x)} = \frac{|CD|}{1}.$$

Thus $|CD| = \sin(x)/\cos(x)$.

Next, the area of $\triangle OCD$ is greater than the area of sector OCA. Since the area of a sector of a disk or radius r subtended by an angle θ is $(\theta/2)r^2$, the area of sector OCA is $x/2 \cdot 1^2 = x/2$, therefore

$$\frac{1}{2}|CD| > \frac{x}{2}$$

which implies that

$$\frac{\sin(x)}{\cos(x)} > x.$$

Now, multiplying by $\cos(x)$, which is positive, and dividing by x (also positive) gives

$$\frac{\sin(x)}{x} > \cos(x). \tag{2.2.5}$$

Putting (2.2.4) and (2.2.5) together, we have

$$\cos(x) < \frac{\sin(x)}{x} < 1. \tag{2.2.6}$$

Since $\cos(x)$ approaches 1 as x approaches 0, $\sin(x)/x$ is squeezed between 1 and something that gets closer and closer to 1. For this reason $\sin(x)/x$ must itself approach 1.

We still must look at $\sin(x)/x$ for $x < 0$ as x gets nearer and nearer to 0. Define u to be $-x$. Then u is positive, and

$$\frac{\sin(x)}{x} = \frac{\sin(-u)}{-u} = \frac{-\sin(u)}{-u} = \frac{\sin(u)}{u}.$$

As x is negative and approaches zero, u is positive and approaches 0. Thus $\sin(x)/x$ approaches 1 as x approaches 0 through positive or negative values.

Theorem 2.2.3: Special Limit #3: The Fundamental Trigonometric Limit

$$\lim_{x\to 0} \frac{\sin(x)}{x} = 1 \qquad \textit{where the angle is measured in radians.}$$

A Limit Involving $\cos(x)$

To find the slope of $y = \cos(x)$ at $x = 0$, recall that $\cos(0) = 1$. This creates interest in the following question:

QUESTION: What happens to $\dfrac{\cos(x) - \cos(0)}{x - 0} = \dfrac{\cos(x) - 1}{x}$ as x gets nearer and nearer to 0?

Knowing that $\lim_{x\to 0} \sin(x)/x = 1$, we can prove the following theorem about $(1 - \cos(x))/x$.

Theorem 2.2.4: Special Limit #4: A Limit Involving Cosine

$$\lim_{x\to 0} \frac{1 - \cos(x)}{x} = 0. \qquad \textit{where the angle is measured in radians.}$$

All we will say about this limit now is that the numerator, $1 - \cos(x)$, is the length of BC in Figure 2.2.3. Exercises 28 and 29 outline how to show the limit is 0.

Summary

This section discussed four important limits:

$$\begin{aligned}
\lim_{x\to a} \frac{x^n - a^n}{x - a} &= n a^{n-1} && (n \text{ a positive integer})\\
\lim_{x\to 0} \frac{e^x - 1}{x} &= 1 && (e \approx 2.71828)\\
\lim_{x\to 0} \frac{\sin(x)}{x} &= 1 && (\text{angle in radians})\\
\lim_{x\to 0} \frac{1 - \cos(x)}{x} &= 0 && (\text{angle in radians}).
\end{aligned}$$

That $\lim_{x\to 0}(e^x - 1)/x = 1$ says, informally, that

$$\frac{\exp(\text{a small number}) - 1}{\text{same small number}} \text{ is near 1.}$$

Each of these limits will be needed in Chapter 3 where we introduce and develop the derivative.

Observation 2.2.5: *The Meaning of* $\lim_{x\to 0} \frac{\sin(x)}{x} = 1$

When x is near 0, $\sin(x)$ and x are both small. That their quotient is near 1 tells us much more, namely, that x is a good approximation of $\sin(x)$.

That means that the difference $\sin(x) - x$ is small, even in comparison to $\sin(x)$. In other words, the relative error

$$\frac{\sin(x) - x}{\sin(x)} \tag{2.2.7}$$

approaches 0 as x approaches 0.

To show this, we evaluate the following limit:

$$\lim_{x\to 0} \frac{\sin(x) - x}{\sin(x)}.$$

We have

$$\lim_{x\to 0} \frac{\sin(x) - x}{\sin(x)} = \lim_{x\to 0} \left(\frac{\sin(x)}{\sin(x)} - \frac{x}{\sin(x)}\right) = \lim_{x\to 0} \left(1 - \frac{x}{\sin(x)}\right) = \lim_{x\to 0} \left(1 - \frac{1}{\left(\frac{\sin(x)}{x}\right)}\right) = 1 - \frac{1}{1} = 0.$$

The relative error in (2.2.7) stays less than 1% for x less than 0.24 radians, just under 14 degrees.

The force acting to return a swinging pendulum to its equilibrium is proportional to $\sin(\theta)$ where θ is the angle that the pendulum makes with the vertical. As one physics book says, "If the angle is small, $\sin(\theta)$ is nearly equal to θ" and it then replaces $\sin(\theta)$ by θ, which is easier to work with.

EXERCISES for Section 2.2

In Exercises 1 to 10 describe the two opposing forces involved in the limit. If you can figure out the limit from results in this section, give it. Otherwise, use a calculator to estimate it.

1. $\lim_{x\to 2} \frac{x^4 - 16}{x - 2}$

2. $\lim_{x\to 0} \frac{\sin(x)}{x\cos(x)}$

3. $\lim_{x\to 0} (1 - x)^{1/x}$

4. $\lim_{x\to 0} (\cos(x))^{1/x}$

5. $\lim_{x\to 0} \frac{\arcsin(x)}{x}$

6. $\lim_{x\to 0} x^x$, $x > 0$

7. $\lim_{x\to 0} \frac{\tan(x)}{x}$

8. $\lim_{x\to 0} \frac{\tan(2x)}{x}$

9. $\lim_{x\to 0} \frac{8^x - 1}{2^x - 1}$

10. $\lim_{x\to 0} \frac{9^x - 1}{3^x - 1}$

Exercises 11 to 15 concern $\lim_{x\to a} \frac{x^n - a^n}{x - a}$.

11. Using the factorization $x^2 - a^2 = (x - a)(x + a)$ find $\lim_{x\to a} \frac{x^2 - a^2}{x - a}$.

12. Using Exercise 11, (a) find $\lim_{x\to 3} \frac{x^2 - 9}{x - 3}$ and (b) find $\lim_{x\to\sqrt{3}} \frac{x^2 - 3}{x - \sqrt{3}}$.

13. (a) By multiplying, show that $(x - a)(x^2 + ax + a^2) = x^3 - a^3$.

(b) Use (a) to show that $\lim_{x\to a} \frac{x^3 - a^3}{x - a} = 3a^2$.

(c) By multiplying, show that $(x - a)(x^3 + ax^2 + a^2x + a^3) = x^4 - a^4$. (This formula can also be found by recognizing the second factor on the left-hand side as a finite geometric sum.)

(d) Use (c) to show that $\lim_{x\to a} \frac{x^4 - a^4}{x - a} = 4a^3$.

14. (a) What is the domain of $\frac{x^2-9}{x-3}$? (b) Graph $\frac{x^2-9}{x-3}$.

RECALL: Use a hollow dot to indicate an absent point in the graph.

15. (a) What is the domain of $\frac{x^3-8}{x-2}$? (b) Graph $\frac{x^3-8}{x-2}$.

Exercises 16 to 19 concern $\lim_{x\to 0} \frac{e^x - 1}{x}$.

16. What is a definition of the number e?

17. Use a calculator to compute $\frac{2.7^x - 1}{x}$ and $\frac{2.8^x-1}{x}$ for $x = 0.001$. How do these results suggest that e is between 2.7 and 2.8.

18. Use a calculator to estimate $\frac{2.718^x - 1}{x}$ for $x = 0.1$, 0.01, and 0.001.

19. Graph $y = \frac{e^x - 1}{x}$ for $x \neq 0$.

Exercises 20 to 30 concern $\lim_{x\to 0} \frac{\sin(x)}{x}$ and $\lim_{x\to 0} \frac{1-\cos(x)}{x}$.

20. Use your calculator to graph $y = \frac{\sin(x)}{x}$.

21. Use your calculator to graph $y = \frac{1-\cos(x)}{x}$.

22. Using the fact that $\lim_{x\to 0} \frac{\sin(x)}{x} = 1$, find the limits of the following as t approaches 0.

(a) $\frac{\sin(3t)}{3t}$, (b) $\frac{\sin(3t)}{t}$, (c) $\frac{\sin(3t)}{\sin(t)}$, and (d) $\frac{\sin^2(x)}{x}$..

23. Why is the arc length from A to C in Figure 2.2.3 equal to x?

24. Why is the length of CD in Figure 2.2.3 equal to $\tan(x)$?

25. Why is the area of triangle $\triangle OCD$ in Figure 2.2.3 equal to $(\tan(x))/2$?

26. An angle of θ radians in a circle of radius r subtends a sector, as shown in Figure 2.2.4. What is the area of the sector with radius r and angle measure θ??

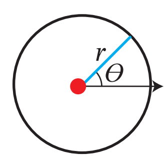


Figure 2.2.4

27. (a) Graph $y = \frac{\sin(x)}{x}$ for x in $[-\pi, 0)$

(b) Graph $y = \frac{\sin(x)}{x}$ for x in $(0, \pi]$.

(c) How are the graphs in (a) and (b) related?

(d) Graph $y = \frac{\sin(x)}{x}$ for $x \neq 0$.

28. When $x = 0$, $\frac{1-\cos(x)}{x}$ is not defined. Estimate $\lim_{x\to 0} \frac{1-\cos(x)}{x}$ by evaluating $\frac{1-\cos(x)}{x}$ at $x = 0.1$ (radians).

29. To find $\lim_{x\to 0} \frac{1-\cos(x)}{x}$ check this algebra and trigonometry:

$$\frac{1-\cos(x)}{x} = \frac{1-\cos(x)}{x}\frac{1+\cos(x)}{1+\cos(x)} = \frac{1-\cos^2(x)}{x(1+\cos(x))} = \frac{\sin^2(x)}{x(1+\cos(x))} = \frac{\sin(x)}{x}\frac{\sin(x)}{1+\cos(x)}.$$

Then show that

$$\lim_{x\to 0}\frac{\sin(x)}{x}\frac{\sin(x)}{1+\cos(x)} = 0.$$

30. Show that $\lim_{x\to 0}\frac{1-\cos(x)}{x^2} = \frac{1}{2}$. This suggests that, for small values of x, $1-\cos(x)$ is close to $\frac{x^2}{2}$, so that $\cos(x)$ is approximately $1-\frac{x^2}{2}$.

(a) Use a calculator to compare $\cos(x)$ with $1-\frac{x^2}{2}$ for $x = 0.2$ and 0.1 radians.

(b) Use a graphing calculator to compare the graphs of $\cos(x)$ and $1-\frac{x^2}{2}$ for x in $[-\pi,\pi]$.

(c) What is the largest interval on which the values of $\cos(x)$ and $1-\frac{x^2}{2}$ differ by no more than 0.1? That is, for what values of x is it true that $\left|\cos(x)-(1-\frac{x^2}{2})\right| < 0.1$?

---

31. Use $\lim_{x\to 0}\frac{e^x-1}{x} = 1$ to find each limit. (a) $\lim_{x\to 0}\frac{e^{10x}-1}{x}$ and (b) $\lim_{x\to 0}\frac{e^{\pi x}-1}{x}$.

32. The limit $\lim_{\theta\to 0}\frac{\sin(4\theta)}{\sin(\theta)}$ appears in The Uniform Sprinkler CIE (CIE 7) at the end of Chapter 5. Find the limit.

33. (a) We examined $\frac{2^x-1}{x}$ only for x near 0. When x is large and positive 2^x-1 is large. So both the numerator and denominator of $\frac{2^x-1}{x}$ are large. The numerator influences the quotient to become large. The large denominator pushes the quotient toward 0. Use a calculator to see how the two forces balance for large values of x.

(b) Sketch the graph of $f(x) = \frac{2^x-1}{x}$ for $x > 0$. (Pay special attention to the behavior of the graph for large values of x.)

34. (a) When x is negative and $|x|$ is large what happens to $(2^x-1)/x$?

(b) Sketch the graph of $f(x) = \frac{2^x-1}{x}$ for $x < 0$. (Pay special attention to the behavior of the graph for large negative values of x.)

35. (a) Using a calculator, explore what happens to $\sqrt{x^2+x}-x$ for large positive values of x.

(b) Show that for $x > 0$, $\sqrt{x^2+x} < x+\frac{1}{2}$.

(c) Using algebra, find what number $\sqrt{x^2+x}-x$ approaches as x increases.

36. Using a calculator, examine the behavior of the quotient $\frac{\theta-\sin(\theta)}{\theta^3}$ for θ near 0.

37. Using a calculator, examine the behavior of the quotient $\frac{\cos(\theta)-1+\frac{\theta^2}{2}}{\theta^4}$ for θ near 0.

Exercises 38 to 41 concern $f(x) = (1+x)^{1/x}$.

38. (a) Why is $(1+x)^{1/x}$ not defined when $x = \frac{-4}{3}$ but is defined when $x = \frac{-5}{3}$? Give an infinite number of $x < -1$ for which it is not defined.

(b) For x near 0, $x > 0$, $1+x$ is near 1. So we might expect $(1+x)^{1/x}$ to be near 1. However, the exponent $1/x$ is large, so perhaps $(1+x)^{1/x}$ is also large. To see what happens, fill in the columns in the table for $x = 0.5$, $x = 0.1$, $x = 0.01$, and $x = 0.001$.

(c) For x near 0 but negative, investigate $(1+x)^{1/x}$ by filling in the table's columns for $x = -0.1$, $x = -0.01$, and $x = -0.001$.

x	1	0.5	0.1	0.01	0.001	−0.001	−0.01	−0.1	−0.5
$1+x$	2								
$1/x$	1								
$(1+x)^{1/x}$	2								

39. Graph $y = (1+x)^{1/x}$ for x in the intervals $(-1,0)$ and $(0,10)$.

Exercises 38 and 39 show that $\lim_{x\to 0}(1+x)^{1/x}$ is about 2.718. This suggests that the number e may equal $\lim_{x\to 0}(1+x)^{1/x}$. In Section 3.5 we show that this is so. Exercise 40 and 41 give persuasive, but incomplete, arguments for this fact. You are asked to find the big hole or unjustified leap in each argument.

40. When $\lim_{x\to a}\frac{f(x)}{g(x)} = 1$ we can write $f(x) \sim g(x)$ near $x = a$. We will read this as "$f(x)$ is close to $g(x)$ when x is near $x = a$." Assume that all we know about the number e is that $\frac{e^x-1}{x} \sim 1$, near 0, that is, $\lim_{x\to 0}\frac{e^x-1}{x} = 1$. Multiplying both sides by x gives $e^x - 1 \sim x$. Adding 1 to both sides of this gives $e^x \sim 1+x$. Finally, raising both sides to the power $1/x$ yields $(e^x)^{1/x} \sim (1+x)^{1/x}$, hence $e \sim (1+x)^{1/x}$. This suggests that $e = \lim_{x\to 0}(1+x)^{1/x}$. The conclusion is correct. Most of the steps are justified. Which steps are justified? Which are not?

41. Assume that $b = \lim_{x\to 0}(1+x)^{1/x}$. We will "show" that $\lim_{x\to 0}\frac{b^x-1}{x} = 1$. First of all, for x near (but not equal to) 0 $b \sim (1+x)^{1/x}$. Then $b^x \sim 1+x$. Hence $b^x - 1 \sim x$. Dividing by x gives $\frac{b^x-1}{x} \sim 1$. Hence $\lim_{x\to 0}\frac{b^x-1}{x} = 1$. Which steps are justified? Which are not?

---

42. An intuitive argument suggested that $\lim_{\theta\to 0}\frac{\sin(\theta)}{\theta} = 1$, which turned out to be correct. Try your intuition on another limit associated with the unit circle shown in Figure 2.2.5.

(a) What do you think happens to the quotient $\frac{\text{Area of triangle } ABC}{\text{Area of shaded region}}$ as θ approaches 0? More precisely, what does your intuition suggest is the limit of the quotient as θ approaches 0?

(b) Estimate the limit in (a) using $\theta = 0.01$.

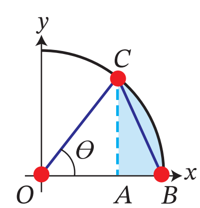


Figure 2.2.5

This limit, which arose during research in geometry, is determined in Exercise 53 in Section 5.6. The authors guessed wrong, as has everyone they asked.

43. For a fixed number a and positive integer n define $P_n(x)$ to be $P_n(x) = x^{n-1} + ax^{n-2} + a^2x^{n-3} + \cdots + a^{n-2}x + a^{n-1}$. As shown in Section 1.4, it equals the quotient $\frac{x^n - a^n}{x-a}$. That is, $(x-a)P_n(x) = x^n - a^n$.

(a) Verify that $(x-a)P_2(x) = x^2 - a^2$. (Compare with Exercise 11.)
(b) Verify that $(x-a)P_3(x) = x^3 - a^3$. (Compare with Exercise 13(a).)
(c) Verify that $(x-a)P_4(x) = x^4 - a^4$. (Compare with Exercise 13(c).)
(d) Explain why $(x-a)P_n(x) = x^n - a^n$ for all positive integers n.

44. Use (1.4.3) in Section 1.4 for the sum of a geometric progression to show that $\lim_{x\to a}\frac{x^n-a^n}{x-a} = na^{n-1}$.

45. Show that $\lim_{x\to 0}\frac{b^x-1}{x} = \ln(b)$ for any positive number b.

2.3 The Limit of a Function

Section 2.2 concerned four important limits:

$$\lim_{x\to a}\frac{x^n-a^n}{x-a}=na^{n-1},\qquad \lim_{x\to 0}\frac{e^x-1}{x}=1,\qquad \lim_{x\to 0}\frac{\sin(x)}{x}=1,\quad \text{and}\quad \lim_{x\to 0}\frac{1-\cos(x)}{x}=0.$$

They are of the form $\lim_{x\to a} f(x)/g(x)$, in which $\lim_{x\to a} f(x)=0$ and $\lim_{x\to a} g(x)=0$. However a limit may have a different form, as illustrated in Exercises 40 and 41 in Section 2.2, which concern $\lim_{x\to 0}(1+x)^{1/x}$.

Limits are fundamental to all of calculus. In this section, we discuss the concept of a limit, beginning with the notion of a one-sided limit.

One-Sided Limits

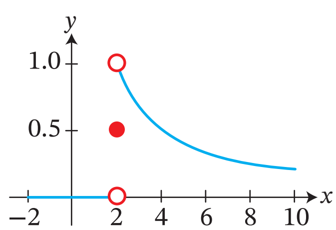


Figure 2.3.1

The domain of the function shown in Figure 2.3.1 is $(-\infty,\infty)$. In particular, it is defined when $x=2$ and $f(2)=1/2$. This is indicated by the solid dot at $(2,1/2)$ in the figure. The hollow dots at $(2,0)$ and $(2,1)$ indicate that these points are not on the graph of this function though some nearby points are on the graph.

Look at the part of the graph for inputs $x>2$, that is, for inputs to the right of 2. As x approaches 2 from the right, $f(x)$ approaches 1. This is expressed as

$$\lim_{x\to 2^+} f(x)=1$$

and is read "the limit of f of x, as x approaches 2 from the right is 1." Similarly, looking at the graph of f in Figure 2.3.1 for x to the left of 2, that is, for $x<2$, we see that the values of $f(x)$ approach a different number, namely, 0. This is expressed as $\lim_{x\to 2^-} f(x)=0$. It might sound strange to say the values of $f(x)$ "approach" 0 since the function values are exactly 0 for all inputs $x<2$. But it is convenient and customary to use the word "approach" even for constant functions.

This illustrates the concept of right-hand and left-hand limits, the two **one-sided limits**.

Definition: *Right-hand limit of* $f(x)$ *at* a

Assume that f is a function defined at least on some open interval (a,c). If, as x approaches a from the right, $f(x)$ approaches a number L, then L is called the **right-hand limit** of $f(x)$ as x approaches a. This is written

$$\lim_{x\to a^+} f(x)=L \qquad \text{or} \qquad f(x)\to L \quad \text{as} \quad x\to a^+.$$

The assertion that $\lim_{x\to a^+} f(x)=L$ is read "the limit of f of x as x approaches a from the right is L" or "as x approaches a from the right, $f(x)$ approaches L."

Definition: *Left-hand limit of* $f(x)$ *at* a

Assume that f is a function defined at least on some open interval (b,a). If, as x approaches a from the left, $f(x)$ approaches a number L, then L is called the **left-hand limit** of $f(x)$ as x approaches a. This is written

$$\lim_{x\to a^-} f(x)=L \qquad \text{or} \qquad f(x)\to L \quad \text{as} \quad x\to a^-.$$

The definitions of one-sided limits do not require that a be in the domain of the function f. If f is defined at a, we do not consider $f(a)$ when examining limits as x approaches a.

The Two-Sided Limit

If the two one-sided limits of $f(x)$ at $x=a$, $\lim_{x\to a^-} f(x)$ and $\lim_{x\to a^+} f(x)$, exist and are equal to L then we say the limit of $f(x)$ as x approaches a is L.

$$\lim_{x\to a} f(x) = L \quad \text{means} \quad \lim_{x\to a^-} f(x) = L \quad \text{and} \quad \lim_{x\to a^+} f(x) = L.$$

For the function graphed in Figure 2.3.1 we found that $\lim_{x\to 2^+} f(x) = 1$ and $\lim_{x\to 2^-} f(x) = 0$. Because they are different, the two-sided limit of $f(x)$ at 2, $\lim_{x\to 2} f(x)$, does not exist.

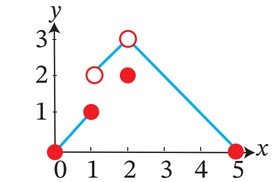


Figure 2.3.2

EXAMPLE 1. Figure 2.3.2 shows the graph of a function f whose domain is the closed interval $[0,5]$. (a) Does $\lim\limits_{x\to 1} f(x)$ exist? (b) Does $\lim\limits_{x\to 2} f(x)$ exist? (c) Does $\lim\limits_{x\to 3} f(x)$ exist?

SOLUTION

(a) Inspection of the graph shows that

$$\lim_{x\to 1^-} f(x) = 1 \qquad \text{and} \qquad \lim_{x\to 1^+} f(x) = 2.$$

Although the two one-sided limits exist, they are not equal. Thus, $\lim_{x\to 1} f(x)$ does not exist. That is, "f does not have a limit as x approaches 1."

(b) Inspection of the graph shows that

$$\lim_{x\to 2^-} f(x) = 3 \qquad \text{and} \qquad \lim_{x\to 2^+} f(x) = 3.$$

Thus $\lim_{x\to 2} f(x)$ exists and is 3. That $f(2) = 2$, as indicated by the solid dot at $(2,2)$, plays no role in our examination of the limit of $f(x)$ as $x \to 2$ (either one-sided or two-sided).

(c) Inspection, once again, shows that

$$\lim_{x\to 3^-} f(x) = 2 \qquad \text{and} \qquad \lim_{x\to 3^+} f(x) = 2.$$

Thus $\lim_{x\to 3} f(x)$ exists and is 2. That $f(3) = 2$ is irrelevant in determining $\lim_{x\to 3} f(x)$. ◆

We now define the two-sided limit without referring to one-sided limits.

Definition: *Limit of* $f(x)$ *at* a.

Assume that f is a function defined at least on the open intervals (b,a) and (a,c), as shown in Figure 2.3.3. If there is a number L such that as x approaches a, from both the right and the left, $f(x)$ approaches L, then L is called the **limit** of $f(x)$ as x approaches a. This is expressed as either

$$\lim_{x\to a} f(x) = L \qquad \text{or} \qquad f(x) \to L \quad \text{as} \quad x \to a.$$

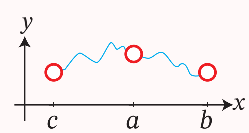


Figure 2.3.3

EXAMPLE 2. For a positive integer n the function $f(x) = \dfrac{x^n - a^n}{x-a}$ is defined for all x other than a. How does it behave for x near a?

SOLUTION In Section 2.2 and its Exercises we found that as x gets closer and closer to a, $f(x)$ gets closer and closer to na^{n-1}. This is summarized by

$$\lim_{x\to a} \frac{x^n - a^n}{x-a} = na^{n-1},$$

which is read as "the limit of $\frac{x^n - a^n}{x-a}$ as x approaches a is na^{n-1}." ♦

EXAMPLE 3. Investigate the one-sided and two-sided limits for the square root function at 0.

SOLUTION The function $\sqrt{x}$ is defined only for x in $[0,\infty)$. We can say that the right-hand limit at 0 exists since $\sqrt{x}$ approaches 0 as $x \to 0$ through positive values of x; that is, $\lim_{x\to 0^+} \sqrt{x} = 0$. Because $\sqrt{x}$ is not defined for negative values of x, the left-hand limit of $\sqrt{x}$ at 0 does not exist. Consequently, the two-sided limit of $\sqrt{x}$ at 0, $\lim_{x\to 0} \sqrt{x}$, does not exist. ♦

EXAMPLE 4. Let $f(x) = 2$ if x is an integer and $f(x) = 1$ otherwise. For which a does $\lim\limits_{x\to a} f(x)$ exist?

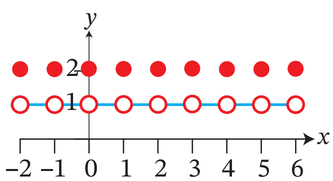


Figure 2.3.4

SOLUTION The graph of f, shown in Figure 2.3.4, will help us decide. If a is not an integer, then for all x sufficiently near a, $f(x) = 1$. So $\lim_{x\to a} f(x) = 1$. Thus the limit exists for all a that are not integers.

Now let a be an integer. In deciding whether $\lim_{x\to a} f(x)$ exists we never consider the value of f at a, namely $f(a) = 2$. For all x sufficiently near an integer a but not equal to a, $f(x) = 1$. Thus, once again, $\lim_{x\to a} f(x) = 1$. The limit exists but is not $f(a)$.

Thus, $\lim_{x\to a} f(x)$ exists and equals 1 for every number a. ♦

EXAMPLE 5. Let $g(x) = \sin\left(\frac{1}{x}\right)$. For which a does $\lim\limits_{x\to a} g(x)$ exist?

SOLUTION To begin, graph the function. The domain of g consists of all x except 0. When x is large, $1/x$ is small, so $\sin(1/x)$ is small. As x approaches 0, $1/x$ becomes large. For instance, when $x = 1/(2n\pi)$, for a nonzero integer n, $1/x = 2n\pi$ and therefore $\sin(1/x) = \sin(2n\pi) = 0$. Thus, the graph for x near 0 crosses the x-axis infinitely often. Similarly, $g(x)$ takes the values 1 and −1 infinitely often for x near 0. The graph is shown in Figure 2.3.5(a).

Does $\lim_{x\to 0} g(x)$ exist? In other words, does $g(x)$ tend toward one specific number as $x \to 0$? No. The function oscillates, taking on all values from −1 to 1 (repeatedly) for x arbitrarily close to 0. Thus $\lim_{x\to 0} \sin(1/x)$ does not exist.

For other values of a, $\lim_{x\to a} g(x)$ does exist and equals $g(a) = \sin(1/a)$. ♦

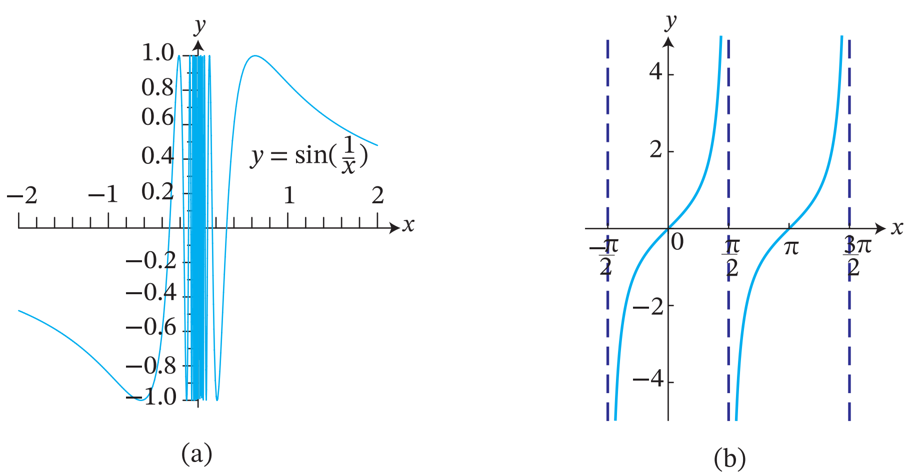


Figure 2.3.5

Infinite Limits at a

A function may assume arbitrarily large values as x approaches a fixed number. One important example is the tangent function. As x approaches $\pi/2$ from the left, $\tan(x)$ takes on arbitrarily large positive values. (See Figure 2.3.5(b).)

We write

$$\lim_{x\to\frac{\pi}{2}^-} \tan(x) = +\infty.$$

However, as $x \to \pi/2$ from the right, $\tan(x)$ takes on negative values of arbitrarily large absolute value. We write

$$\lim_{x\to\frac{\pi}{2}^+} \tan(x) = -\infty.$$

Definition: *Infinite limit of $f(x)$ at a*

Assume that the domain of the function f contains the open interval (a,c). If, as x approaches a from the right, $f(x)$ becomes and remains arbitrarily large and positive, then the limit of $f(x)$ as x approaches a is said to be positive infinity. This is written

$$\lim_{x\to a^+} f(x) = +\infty$$

or sometimes just

$$\lim_{x\to a^+} f(x) = \infty.$$

If, as x approaches a from the left, $f(x)$ becomes and remains arbitrarily large and positive, then we write

$$\lim_{x\to a^-} f(x) = +\infty.$$

Similarly, if $f(x)$ assumes values that are negative and remain arbitrarily large in absolute value, we write either

$$\lim_{x\to a^+} f(x) = -\infty \qquad \text{or} \qquad \lim_{x\to a^-} f(x) = -\infty,$$

depending on whether x approaches a from the right or from the left.

Limits at ∞ and at $-\infty$

Sometimes it is useful to know how $f(x)$ behaves when x is a large positive number or a negative number of large absolute value. The notation for these limits is similar to that for limits at a number, with $x\to\infty$ or $x\to-\infty$ replacing $x\to a$.

EXAMPLE 6. Determine how $f(x) = \dfrac{1}{x}$ behaves for

(a) large positive inputs
(b) negative inputs of large absolute value
(c) small positive inputs
(d) negative inputs of small absolute value

x	$1/x$
10	0.1
100	0.01
1000	0.001

Table 2.3.1

SOLUTION

(a) To get started, make a table of values, such as Table 2.3.1. As x increases through the positive numbers, $1/x$ approaches 0: $\lim_{x\to\infty} f(x) = 0$. This could be read as "as x approaches ∞, $f(x)$ approaches 0."

(b) This is similar to (a), except that the reciprocal of a negative number with large absolute value is a negative number with a small absolute value. Thus, $\lim_{x\to-\infty} f(x) = 0$.
(c) Inputs that are positive and approaching 0 have reciprocals that are positive and large: $\lim_{x\to0^+} f(x) = +\infty$.
(d) Lastly, the reciprocals of inputs that are negative and approaching 0 from the left are negative and arbitrarily large in absolute value: $\lim_{x\to0^-} f(x) = -\infty$. ♦

More generally, for any positive exponent p,

$$\lim_{x\to\infty} \frac{1}{x^p} = 0.$$

Limits of the form $\lim_{x\to\infty} P(x)$ and $\lim_{x\to\infty} P(x)/Q(x)$, where P and Q are polynomials are easy to treat, as the following examples show.

Keep in mind that ∞ is not a number. It is a symbol that tells us that some quantity — either the inputs or the outputs of a function — becomes arbitrarily large, or increases without bound.

EXAMPLE 7. Find $\lim\limits_{x\to\infty} \left(2x^3 - 5x^2 + 6x + 5\right)$.

SOLUTION When x is large, x^3 is much larger than either x^2 or x. With this in mind, we use algebra to determine the limit:

$$2x^3 - 5x^2 + 6x + 5 = \left(2 - \frac{5}{x} + \frac{6}{x^2} + \frac{5}{x^3}\right)x^3.$$

The expression in parentheses approaches 2, while x^3 gets arbitrarily large. Thus

$$\lim_{x\to\infty} (2x^3 - 5x^2 + 6x + 5) = \infty.$$

♦

EXAMPLE 8. Find $\lim\limits_{x\to\infty} \dfrac{2x^3 - 5x^2 + 6x + 5}{7x^4 + 3x + 2}$.

SOLUTION We use the same technique as in Example 7.

$$2x^3 - 5x^2 + 6x + 5 = \left(2 - \frac{5}{x} + \frac{6}{x^2} + \frac{5}{x^3}\right)x^3$$

and

$$7x^4 + 3x + 2 = \left(7 + \frac{3}{x^3} + \frac{2}{x^4}\right)x^4$$

so that

$$\frac{2x^3 - 5x^2 + 6x + 5}{7x^4 + 3x + 2} = \frac{\left(2 - \frac{5}{x} + \frac{6}{x^2} + \frac{5}{x^3}\right)x^3}{\left(7 + \frac{3}{x^3} + \frac{2}{x^4}\right)x^4} = \frac{1}{x}\,\frac{\left(2 - \frac{5}{x} + \frac{6}{x^2} + \frac{5}{x^3}\right)}{\left(7 + \frac{3}{x^3} + \frac{2}{x^4}\right)}.$$

As x gets arbitrarily large, $1/x$ approaches 0, $2 - 5/x + 6/x^2 + 5/x^3$ approaches 2, and $7 + 3/x^3 + 2/x^4$ approaches 7. Thus,

$$\lim_{x\to\infty} \frac{2x^3 - 5x^2 + 6x + 5}{7x^4 + 3x + 2} = 0 \cdot \left(\frac{2}{7}\right) = 0.$$

♦

As Example 8 suggests, the limit of a quotient of two polynomials, $P(x)/Q(x)$, is completely determined by the limit of the quotient of the highest degree terms in $P(x)$ and in $Q(x)$.

If

$$P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \qquad \text{and} \qquad Q(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0,$$

where a_n and b_m are not 0, then

$$\lim_{x\to\infty} \frac{P(x)}{Q(x)} = \lim_{x\to\infty} \frac{a_n x^n}{b_m x^m}.$$

If $m = n$, the limit is a_n/b_m. If $m > n$, the limit is 0. If $n > m$, the limit is infinite, either ∞ or $-\infty$, depending on the signs of a_n and b_m.

Summary

This section introduced the concept of a limit and notations for the various types of limits. One-sided limits are the foundation for the two-sided limit as well as for infinite limits and limits at infinity.

It is important to keep in mind that when deciding whether $\lim_{x\to a} f(x)$ exists, one never involves $f(a)$. Perhaps a is not even in the domain of the function. Even if a is in the domain, the value $f(a)$ plays no role in deciding whether $\lim_{x\to a} f(x)$ exists. Though we write $\lim_{x\to a} f(x) = \infty$ when $f(x)$ gets arbitrarily large when x is near a, the function does not have a limit at a.

EXERCISES for Section 2.3

In Exercises 1 to 4 the graph of a function $y = f(x)$ is given. Decide whether $\lim\limits_{x\to 1^+} f(x)$, $\lim\limits_{x\to 1^-} f(x)$, and $\lim\limits_{x\to 1} f(x)$ exist. Give the value of each limit that exists.

1. Figure 2.3.6(a) **2.** Figure 2.3.6(b) **3.** Figure 2.3.6(c) **4.** Figure 2.3.6(d)

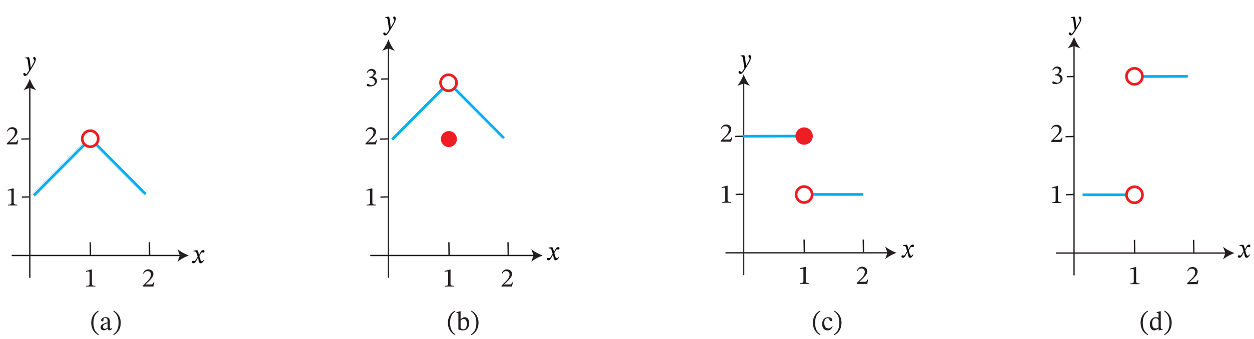


Figure 2.3.6

In Exercises 5 to 12 the limits exist. Find them.

5. $\displaystyle\lim_{x\to 3} \frac{x^2-9}{x-3}$ **6.** $\displaystyle\lim_{x\to 4} \frac{x^2-9}{x-3}$ **7.** $\displaystyle\lim_{x\to 0} \frac{\sin(x)}{x}$ **8.** $\displaystyle\lim_{x\to \frac{\pi}{2}} \frac{\sin(x)}{x}$

9. $\displaystyle\lim_{x\to 0} \frac{e^x-1}{2x}$ **10.** $\displaystyle\lim_{x\to 2} \frac{e^x-1}{2x}$ **11.** $\displaystyle\lim_{x\to 0} \frac{1-\cos(x)}{3x}$ **12.** $\displaystyle\lim_{x\to \pi} \frac{1-\cos(x)}{3x}$

13. (a) Sketch the graph of $y = \log_2(x)$. (b) What are $\lim\limits_{x\to\infty} \log_2(x)$, $\lim\limits_{x\to 4} \log_2(x)$, and $\lim\limits_{x\to 0^+} \log_2(x)$?

14. (a) Sketch the graph of $y = 2^x$. (b) What are $\lim\limits_{x\to\infty} 2^x$, $\lim\limits_{x\to 4} 2^x$, and $\lim\limits_{x\to -\infty} 2^x$?

15. Find $\displaystyle\lim_{x\to a} \frac{x^3-8}{x-2}$ for (a) $a = 1$, (b) $a = 2$, and (c) $a = 3$.

16. Find $\displaystyle\lim_{x\to a} \frac{x^4-16}{x-2}$ for (a) $a = 1$, (b) $a = 2$, and (c) $a = 3$.

17. Find $\displaystyle\lim_{x\to a} \frac{e^x-1}{(x-2)^2}$ for (a) $a = -1$, (b) $a = 0$, (c) $a = 1$, and (d) $a = 2$.

18. Find $\displaystyle\lim_{x\to a} \frac{\sin(x)}{x}$ for (a) $a = \frac{\pi}{6}$, (b) $a = \frac{\pi}{4}$, and (c) $a = 0$.

In Exercises 19 to 24, find the given limit (if it exists).

19. $\lim_{x\to\infty} 2^{-x}\sin(x)$

20. $\lim_{x\to\infty} 3^{-x}\cos(2x)$

21. $\lim_{x\to\infty} \dfrac{3x^5+2x^2-1}{6x^5+x^4+2}$

22. $\lim_{x\to\infty} \dfrac{13x^5+2x^2+1}{2x^6+x+5}$

23. $\lim_{x\to\infty} \dfrac{10x^6+x^5+x+1}{x^6}$

24. $\lim_{x\to\infty} \dfrac{25x^5+x^2+1}{x^3+x+2}$

In Exercises 25 to 27, information is given about functions f and g. In each case decide whether the limit asked for can be determined on the basis of the information. If the limit can be evaluated, give its value. If it cannot be evaluated, show by specific choices of f and g why it cannot.

25. Assume $\lim_{x\to\infty} f(x)=0$ and $\lim_{x\to\infty} g(x)=1$.

Find (a) $\lim_{x\to\infty}(f(x)+g(x))$, (b) $\lim_{x\to\infty}\dfrac{f(x)}{g(x)}$, (c) $\lim_{x\to\infty}\dfrac{f(x)}{(g(x)-1)}$, (d) $\lim_{x\to\infty}\dfrac{g(x)}{f(x)}$, and (e) $\lim_{x\to\infty}\dfrac{g(x)}{|f(x)|}$.

26. Assume $\lim_{x\to\infty} f(x)=\infty$ and $\lim_{x\to\infty} g(x)=\infty$.

Find (a) $\lim_{x\to\infty}(f(x)+g(x))$, (b) $\lim_{x\to\infty}(f(x)-g(x))$, (c) $\lim_{x\to\infty}(f(x)g(x))$, and (d) $\lim_{x\to\infty}\dfrac{g(x)}{f(x)}$.

27. Assume $\lim_{x\to\infty} f(x)=1$ and $\lim_{x\to\infty} g(x)=\infty$.

Find (a) $\lim_{x\to\infty}\dfrac{f(x)}{g(x)}$, (b) $\lim_{x\to\infty}(f(x)g(x))$, and (c) $\lim_{x\to\infty}(f(x)-1)g(x)$.

---

28. Assume $f(x)=\cos\left(\dfrac{1}{x}\right)$. (a) What is the domain of f? (b) Does $\lim_{x\to 0}\cos\left(\dfrac{1}{x}\right)$ exist? (c) Graph $f(x)=\cos\left(\dfrac{1}{x}\right)$.

29. Assume $f(x)=x\sin\left(\dfrac{1}{x}\right)$.

(a) What is the domain of f?
(b) Graph the lines $y=x$ and $y=-x$.
(c) For which x does $f(x)=x$?
When does $f(x)=-x$?
(d) Does $\lim_{x\to 0} f(x)$ exist? If so, what is it?
(e) Does $\lim_{x\to\infty} f(x)$ exist? If so, what is it?
(f) Graph $y=f(x)$.
(g) Does $\lim_{x\to\infty}|f(x)|$ exist? If so, what is it?

30. Assume $f(x)=\dfrac{|x|}{x}$, which is defined except at $x=0$.

(a) What is $f(3)$?
(b) What is $f(-2)$?
(c) Graph $y=f(x)$.
(d) Does $\lim_{x\to 0^+} f(x)$ exist? If so, what is it?
(e) Does $\lim_{x\to 0^-} f(x)$ exist? If so, what is it?
(f) Does $\lim_{x\to 0} f(x)$ exist? If so, what is it?
(g) Graph f.

In Exercises 31 to 33, find $\lim_{h\to 0}\dfrac{f(3+h)-f(3)}{h}$ for each specific function $f(x)$.

31. $f(x)=5x$

32. $f(x)=x^2$

33. $f(x)=e^x$

In Exercises 31 to 33, find $\lim_{h\to 0}\dfrac{f(x+h)-f(x)}{h}$ for each specific function $f(x)$.

34. $f(x)=\sin(x)$

35. $f(x)=\cos(x)$

36. $f(x)=e^{2x}$

---

37. Figure 2.3.7 shows a circle of radius a.

Find (a) $\lim_{\theta\to 0^+} \frac{|BA|}{|\widehat{CA}|}$ (b) $\lim_{\theta\to 0^+} \frac{|BA|}{|CD|}$ (c) $\lim_{\theta\to 0} \frac{\text{Area of } \triangle OAB}{\text{Area of } \triangle ODC}$

$|\widehat{CA}|$ is the length of the arc of the circle from C to A.

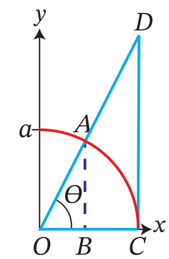


Figure 2.3.7

38. Let $f(x)$ be the diameter of the largest circle that fits in a 1 by x rectangle.
(a) Find a formula for $f(x)$. (b) Graph $y = f(x)$. (c) Does $\lim_{x\to 1} f(x)$ exist?

39. I am thinking of two numbers near 0. What, if anything, can you say about their
(a) product? (b) quotient? (c) difference? (d) sum?

40. I am thinking about two large positive numbers. What, if anything, can you say about their
(a) product? (b) quotient? (c) difference? (d) sum?

41. Sam and Jane are discussing the function $f(x) = \frac{3x^2+2x}{x+5}$.

Sam: For large x, $2x$ is small in comparison to $3x^2$, and 5 is small in comparison to x. So the quotient $\frac{3x^2+2x}{x+5}$ behaves like $\frac{3x^2}{x} = 3x$. Hence, the graph of $y = f(x)$ is very close to the graph of the line $y = 3x$ when x is large.

Jane: Nonsense. After all, $\frac{3x^2+2x}{x+5} = \frac{3x+2}{1+(5/x)}$ which clearly behaves like $3x+2$ for large x. Thus the graph of $y = f(x)$ stays very close to the line $y = 3x+2$ when x is large.

Settle the argument.

42. Sam, Jane, and Wilber are arguing about limits in a case where $\lim_{x\to\infty} f(x) = 0$ and $\lim_{x\to\infty} g(x) = \infty$.

Sam: $\lim_{x\to\infty} f(x)g(x) = 0$, since $f(x)$ is going toward 0.

Jane: Rubbish! Since $g(x)$ gets large, it will turn out that $\lim_{x\to\infty} f(x)g(x) = \infty$.

Wilber: You're both wrong. The two influences will balance out and you will see that $\lim_{x\to\infty} f(x)g(x)$ is near 1.

Settle the argument.

43. Sam and Jane are arguing about limits in a case where $f(x) \geq 1$ for $x > 0$, $\lim_{x\to 0^+} f(x) = 1$ and $\lim_{x\to 0} g(x) = \infty$. What can be said about $\lim_{x\to 0^+} f(x)^{g(x)}$?

Sam: That's easy. Multiply a bunch of numbers near 1 and you get a number near 1. So the limit will be 1.

Jane: Rubbish! Since $f(x)$ may be bigger than 1 and you are multiplying it lots of times, you will get a really large number. There's no doubt in my mind: $\lim_{x\to 0} f(x)^{g(x)} = \infty$.

Settle the argument.

44. An urn contains n marbles. One is green and the remaining $n-1$ are red. When picking one marble at random without looking, the probability is $\frac{1}{n}$ of getting the green marble, and $\frac{n-1}{n}$ of getting a red marble. If you do this experiment n times, each time putting the chosen marble back, the probability of not getting the green marble on any of the n experiments is $\left(\frac{n-1}{n)/}\right)^n$.

(a) For $p(n) = \left(\frac{n-1}{n}\right)^n$, compute $p(2)$, $p(3)$, and $p(4)$ to at least three decimal places.

(b) Show that as $n \to \infty$, $p(n)$ approaches the reciprocal of $\lim_{x\to 0} (1+x)^{1/x}$.

2.4 Continuous Functions

This section introduces the notion of a continuous function. While almost all functions met in practice are continuous, we must always remain alert that a function might not be continuous. We begin with an informal description and then give a more useful working definition.

An Informal Introduction to Continuous Functions

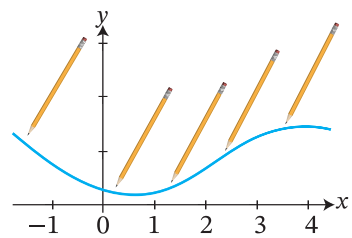


Figure 2.4.1

When drawing the graph of a function defined on some interval, we usually do not have to lift the pencil off of the paper. Figure 2.4.1 shows this typical situation.

A function is said to be **continuous** if, when considered on any interval in its domain, its graph can he traced without lifting the pencil off of the paper. (The domain may consist of several intervals.) According to this definition any polynomial is continuous. So are the basic trigonometric functions, including $y = \tan(x)$, whose graph is shown in Figure 2.3.5(b) of Section 2.3.

You may be tempted to say "But $\tan(x)$ blows up at $x = \pi/2$ and I have to lift my pencil off of the paper to draw the graph." However, $x = \pi/2$ is not in the domain of the tangent function. On every interval in its domain, $\tan(x)$ behaves quite decently; we can sketch its graph without lifting the pencil from the paper. That is why $\tan(x)$ is continuous. The function $1/x$ is also continuous, since it "explodes" only at a number not in its domain, namely at $x = 0$.

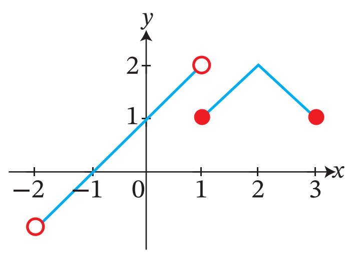


Figure 2.4.2

The function whose graph is shown in Figure 2.4.2 is not continuous. It is defined throughout the interval $(-2,3]$, but to draw its graph you must lift the pencil from the paper near $x = 1$. However, when you consider the function only for x in $[1,3]$, then it is continuous. A formula for it is:

$$f(x) = \begin{cases} x+1 & \text{for } x \text{ in } (-2,1) \\ x & \text{for } x \text{ in } [1,2) \\ -x+4 & \text{for } x \text{ in } [2,3]. \end{cases}$$

It is pieced together from three linear functions.

The Definition of Continuity

x	$f(x)$
0.9	2.93
0.99	2.9954
0.999	2.9999997

Table 2.4.1

Our informal moving pencil notion of a continuous function requires drawing a graph of the function. Our working definition does not require such a graph. Moreover, it generalizes to functions of more than one variable in later chapters.

To get the feeling of this second definition, imagine that you had the information shown in Table 2.4.1 about some function f. What would you expect the output $f(1)$ to be?

It would be quite a shock to be told that $f(1)$ is, say, 625. A reasonable function should present no such surprise. The expectation is that $f(1)$ will be 3. More generally, we expect the output of a function at the input a to be closely connected with the outputs of the function at inputs near a. The functions of interest in calculus usually behave that way. In short, "What you expect is what you get." With this in mind, we define continuity at a number a. We first assume that the domain of f contains an open interval around a.

Definition: *Continuity at a Number a*

Assume that $f(x)$ is defined in some open interval that contains the number a. Then the function f is **continuous at** a if $\lim_{x\to a} f(x) = f(a)$. This means that

1. $f(a)$ is defined (that is, a is in the domain of f).
2. $\lim_{x\to a} f(x)$ exists.
3. $\lim_{x\to a} f(x)$ equals $f(a)$.

As Figure 2.4.3 shows, whether a function is continuous at a depends on its behavior both at a and at inputs near a. Being continuous at a is a local matter, involving perhaps very tiny intervals about a.

Algorithm: Checking for Continuity at a Number

To check whether a function f is continuous at a number a, we ask three questions:

Question 1: Is a in the domain of f?

Question 2: Does $\lim_{x\to a} f(x)$ exist?

Question 3: Does $f(a)$ equal $\lim_{x\to a} f(x)$?

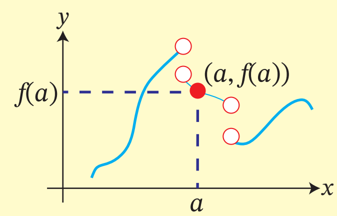


Figure 2.4.3

If all three answers are "yes," we say that f is continuous at a.

If a is in the domain of f and the answer to Question 2 or to Question 3 is "no," then f is said to be **discontinuous at** a.

If a is not in the domain of f, we do not speak of it being continuous or discontinuous there.

We are now ready to define a continuous function.

Definition: *Continuous Function*

Let f be a function whose domain is the x-axis or is made up of open intervals. Then f is a **continuous function** if it is continuous at each number a in its domain. A function that is not continuous is called a **discontinuous function**.

EXAMPLE 1. Use the definition of continuity to decide whether $f(x) = \dfrac{1}{x}$ is continuous.

SOLUTION It is continuous at every point a for which the answers to Questions 1, 2, and 3 are all "yes".

If a is not 0, it is in the domain of f. So, for a not 0, the answer to Question 1 is "yes." Next, for $a \neq 0$,

$$\lim_{x\to a} \frac{1}{x} = \frac{1}{a},$$

so the answer to Question 2 is "yes." Finally, for $x \neq 0$,

$$f(a) = \frac{1}{a},$$

the answer to Question 3 is also "yes." Thus $f(x) = 1/x$ is continuous at every number in its domain. Hence f is a continuous function. This agrees with the moving pencil picture of continuity. ♦

Not every important function is continuous. Let $f(x)$ be the greatest integer that is less than or equal to x. For instance, $f(1.8) = 1$, $f(1.9) = 1$, $f(2) = 2$, and $f(2.3) = 2$. It is often used in number theory and computer science, where it is denoted $[x]$ or $\lfloor x \rfloor$ and called the **floor** of x. People use the floor function every time they answer the question, "How old are you?" The next example examines where the floor function fails to be continuous.

EXAMPLE 2. Graph the floor function, $f(x) = \lfloor x \rfloor$, and find where it is continuous. Is f a continuous function?

SOLUTION We begin with the following table to show the behavior of $f(x) = \lfloor x \rfloor$ for x near 1 or 2.

x	0	0.5	0.8	1	1.1	1.99	2	2.01
$\lfloor x \rfloor$	0	0	0	1	1	1	2	2

For $0 \leq x < 1$, $\lfloor x \rfloor = 0$. At the input $x = 1$ the output jumps to 1 since $\lfloor 1 \rfloor = 1$. For $1 \leq x < 2$, $\lfloor x \rfloor$ remains at 1. Then at 2 it jumps to 2. More generally, $\lfloor x \rfloor$ has a jump at every integer, as shown in Figure 2.4.4.

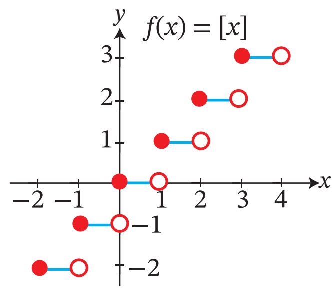


Figure 2.4.4

Let us show that f is not continuous at $a = 2$ by seeing which of the three conditions in the definition are not satisfied. First of all, Question 1 is answered "yes" since 2 lies in the domain of the function; indeed, $f(2) = 2$.

What is the answer to Question 2? Does $\lim_{x\to 2} f(x)$ exist? We see that

$$\lim_{x\to 2^-} f(x) = 1 \qquad \text{and} \qquad \lim_{x\to 2^+} f(x) = 2.$$

Since the left-hand and right-hand limits at $x = 2$ are not equal, $\lim_{x\to 2} f(x)$ does not exist. Question 2 is answered "no."

Already we know that the function is not continuous at $a = 2$. Since the limit does not exist there is no reason to consider Question 3. Because there is a point in the domain where $\lfloor x \rfloor$ is not continuous, it is a discontinuous function. In fact, it is discontinuous at a whenever a is an integer.

Is f continuous at a if a is not an integer? Take the case $a = 1.5$, for instance.

Question 1 is answered "yes," because $f(1.5)$ is defined. (In fact, $f(1.5) = 1$.)

Question 2 is answered "yes," since $\lim_{x\to 1.5} f(x) = 1$.

Question 3 is answered "yes," since $\lim_{x\to 1.5} f(x) = f(1.5) = 1$.

So the floor function is continuous at $a = 1.5$. Similarly, it is continuous at every number that is not an integer.

On any interval that does not include an integer, $\lfloor x \rfloor$ is continuous. For instance, if we consider the function only on the interval $(1.1, 1.9)$, it is continuous there. ♦

Continuity at an Endpoint of an Interval

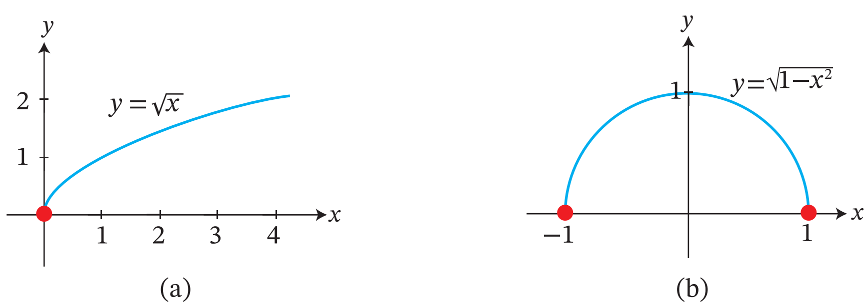


Figure 2.4.5

The functions $f(x) = \sqrt{x}$ and $g(x) = \sqrt{1-x^2}$ are graphed in Figures 2.4.5(a) and (b), respectively. We would like to call both of them continuous. However, there is a technical problem. The number 0 is in the domain of f, but there is no open interval around 0 that lies completely in the domain, as our definition of continuity requires. Since $f(x) = \sqrt{x}$ is not defined for x to the left of 0, we are not interested in numbers x to the left of 0. Similarly, $g(x) = \sqrt{1-x^2}$ is defined only when $1 - x^2 \geq 0$, that is, for $-1 \leq x \leq 1$. To cover this situation we utilize one-sided limits to define the two types of **one-sided continuity**:

Definition: *Continuity from the Right at a Number* a

Assume that $f(x)$ is defined in some closed interval $[a, c]$. Then f **is continuous from the right at** a if

1. $f(a)$ is defined
2. $\lim_{x\to a^+} f(x)$ exists
3. $\lim_{x\to a^+} f(x)$ equals $f(a)$.

For instance, $f(x) = \sqrt{x}$ is continuous on the right at $a = 0$.

Definition: *Continuity from the Left at a Number a*

Assume that $f(x)$ is defined in some closed interval $[b,a]$. Then the function f **is continuous from the left at a** if

1. $f(a)$ is defined
2. $\lim_{x\to a^-} f(x)$ exists
3. $\lim_{x\to a^-} f(x)$ equals $f(a)$

For instance, $f(x)=\sqrt{1-x^2}$ is continuous on the left at $a=1$.

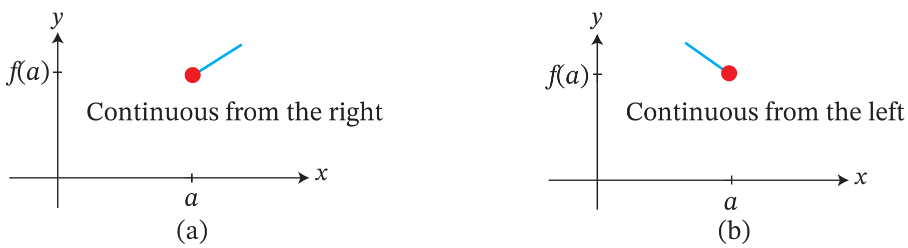


Figure 2.4.6

With these two extra definitions to cover some special cases in the domain, we can extend the definition of continuous function to include those functions whose domains may contain endpoints. We say, for instance, that $\sqrt{1-x^2}$ is continuous because it is continuous at any number in $(-1,1)$, is continuous from the right at -1, and continuous from the left at 1.

These are minor matters that will little concern us in the future. The key point is that $\sqrt{1-x^2}$ and $\sqrt{x}$ are both continuous functions. So are practically all the functions studied in calculus.

The following example reviews the notion of continuity.

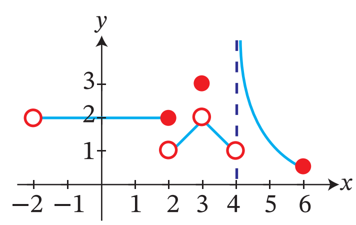


Figure 2.4.7

EXAMPLE 3. Figure 2.4.7 is the graph of a piecewise-defined function $f(x)$ whose domain is the interval $(-2,6]$. Discuss its continuity at (a) 6, (b) 4, (c) 3, (d) 2, (e) 1, and (f) -2.

SOLUTION

(a) Since $\lim_{x\to 6^-} f(x)$ exists and equals $f(6)$, f is continuous from the left at 6.

(b) Since $\lim_{x\to 4} f(x)$ does not exist, f is not continuous at 4.

(c) Inspection of the graph shows that $\lim_{x\to 3} f(x)=2$. However, Question 3 is answered "no" because $f(3)=3$, which is not equal to $\lim_{x\to 3} f(x)$. Thus f is not continuous at 3.

(d) Though $\lim_{x\to 2^-} f(x)$ and $\lim_{x\to 2^+} f(x)$ exist, they are not equal. (The left-hand limit is 2 and the right-hand limit is 1.) Thus $\lim_{x\to 2} f(x)$ does not exist, the answer to Question 2 is "no," and f is discontinuous at $x=2$.

(e) At 1, "yes" is the answer to all three questions: $f(1)$ is defined, $\lim_{x\to 1} f(x)$ exists (it equals 2) and, finally, it equals $f(1)$. f is continuous at $x=1$.

(f) Since -2 is not even in the domain of this function, we do not speak of continuity or discontinuity of f at -2. ◆

As Example 3 shows, a function can fail to be continuous at a given number a in its domain for either of two reasons:

1. $\lim_{x\to a} f(x)$ might not exist
2. when $\lim_{x\to a} f(x)$ does exist, $f(a)$ might not be equal to that limit.

Continuity and Limits

Some limits can be found without any work; for instance, $\lim_{x\to 2} 5^x = 5^2 = 25$. Others offer a challenge, such as, $\lim_{x\to 2}(x^3 - 2^3)/(x-2)$.

If you want to find $\lim_{x\to a} f(x)$, and you know f is a continuous function with a in its domain, then you just calculate $f(a)$. In such a case there is no challenge and the limit is called **determinate**.

The interesting case for finding $\lim_{x\to a} f(x)$ occurs when f is not defined at a. That is when you must consider the influences operating on $f(x)$ when x is near a. You may have to do some algebra or computations. Such limits are called **indeterminate**.

The four limits encountered in Section 2.2, $\lim_{x\to a}(x^n - a^n)/(x-a)$, $\lim_{x\to 0}(b^x - 1)/(x)$, $\lim_{x\to 0}(\sin(x))/x$, and $\lim_{x\to 0}(1-\cos(x))/x$ are indeterminate. Each required some work to find its value.

We list the properties of limits which are helpful in computing limits.

Theorem 2.4.1: Properties of Limits

Let g and h be two functions and assume that $\lim\limits_{x\to a} g(x) = A$ *and* $\lim_{x\to a} h(x) = B$. *Then*

Sum Property

$\lim\limits_{x\to a}\big(g(x)+h(x)\big) = \lim\limits_{x\to a} g(x) + \lim\limits_{x\to a} h(x) = A+B$:
the limit of the sum is the sum of the limits.

Difference Property

$\lim\limits_{x\to a}\big(g(x)-h(x)\big) = \lim\limits_{x\to a} g(x) - \lim\limits_{x\to a} h(x) = A-B$:
the limit of the difference is the difference of the limits.

Product Property

$\lim\limits_{x\to a}\big(g(x)h(x)\big) = \left(\lim\limits_{x\to a} g(x)\right)\left(\lim\limits_{x\to a} h(x)\right) = AB$:
the limit of the product is the product of the limits.

Constant Multiple Property

$\lim\limits_{x\to a}\big(kg(x)\big) = k\left(\lim\limits_{x\to a} g(x)\right) = kA$, *for any constant k.*
This is a special case of the product property when one factor is constant.

Quotient Property

$$\lim_{x\to a}\left(\frac{g(x)}{h(x)}\right) = \frac{\lim\limits_{x\to a} g(x)}{\lim\limits_{x\to a} h(x)} = \frac{A}{B}, \text{ provided } B \text{ is not } 0:$$
the limit of the quotient is the quotient of the limits, provided the denominator is not 0.

Power Property

$$\lim_{x\to a}\left(g(x)^{h(x)}\right) = \left(\lim_{x\to a} g(x)\right)^{\lim_{x\to a} h(x)} = A^B, \text{ provided } A \text{ is positive:}$$
the limit of a varying base to a varying exponent is the limit of the base raised to the limit of the exponent.

NOTE: *Each property remains valid when the two-sided limit is replaced with a one-sided limit.*

EXAMPLE 4. Find $\lim\limits_{x\to 0}\dfrac{(x^4-16)\sin(5x)}{x^2-2x}$.

SOLUTION The denominator can be factored to obtain

$$\frac{(x^4-16)\sin(5x)}{x^2-2x} = \frac{x^4-2^4}{x-2}\cdot\frac{\sin(5x)}{x}.$$

The limit can be rewritten as

$$\lim_{x\to 0} \frac{x^4-2^4}{x-2} \cdot \lim_{x\to 0} \frac{\sin(5x)}{x}.$$

Now, the two separate limits are easily evaluated:

$$\lim_{x\to 0} \frac{x^4-2^4}{x-2} = \frac{-(2^4)}{-2} = -8.$$

and

$$\lim_{x\to 0} \frac{\sin(5x)}{x} = \lim_{x\to 0} 5\frac{\sin(5x)}{5x} = 5\lim_{x\to 0} \frac{\sin(5x)}{5x} = 5\cdot 1 = 5.$$

We conclude that

$$\lim_{x\to 0} \frac{(x^4-16)\sin(5x)}{x^2-2x} = \lim_{x\to 0} \frac{x^4-2^4}{x-2} \cdot \lim_{x\to 0} \frac{\sin(5x)}{5x} = -8\cdot 5 = -40.$$

◆

Summary

This section opened with an informal view of continuous functions, expressed in terms of a moving pencil. It then gave the definition, phrased in terms of limits, which we will use throughout the text. It says that f is continuous at a if $f(a)$ is defined and equals $\lim_{x\to a} f(x)$. The relation between continuity and determinate and indeterminate limits was discussed.

EXERCISES for Section 2.4

In Exercises 1 to 12, which limits can be found at a glance and which require some analysis? That is, decide in each case whether the limit is determinate or indeterminate. Do not evaluate the limit.

1. $\lim_{x\to 0}\left(2^x-1\right)$ **2.** $\lim_{x\to\infty} \frac{2^x-1}{2^x+1}$ **3.** $\lim_{x\to 1} \frac{3^x-1}{2^x-1}$ **4.** $\lim_{x\to 2} \frac{3^x-1}{2^x-1}$

5. $\lim_{x\to\infty} \frac{x}{2^x}$ **6.** $\lim_{x\to 0} \frac{x}{2^x}$ **7.** $\lim_{x\to 0^+} \frac{x^2}{e^x-1}$ **8.** $\lim_{x\to \frac{\pi}{2}^-} (\sin(x))^{\tan(x)}$

9. $\lim_{x\to 0^+} x\log_2(x)$ **10.** $\lim_{x\to 0^+} (2+x)^{3/x}$ **11.** $\lim_{x\to\infty} (2+x)^{3/x}$ **12.** $\lim_{x\to 0^-} \frac{(2+x)^3}{x}$

In Exercises 13 to 16, evaluate the limit.

13. $\lim_{x\to \frac{\pi}{2}} \sin(x)\frac{e^x-1}{x}$ **14.** $\lim_{x\to 0} \cos(x)\frac{e^x-1}{x}$ **15.** $\lim_{x\to 0} \frac{\sin(2x)}{x(\cos(3x))^2}$ **16.** $\lim_{x\to 1} \frac{(x-1)\cos(x)}{x^3-1}$

In Exercises 17 and 18 use the given graph of a function to decide if the given function is continuous on each of the given intervals.

17. The graph of the function f is shown in Figure 2.4.8(a).
(a) $[-2,-1]$, (b) $(-2,-1)$, (c) $(-1,1)$, (d) $[-1,1)$, (e) $(-1,1]$, (f) $[-1,1]$, (g) $(1,2)$, (h) $[1,2)$, (i) $(1,2]$, and (j) $[1,2]$.

18. The graph of the function f is shown in Figure 2.4.8(b).
(a) $[-3,2]$, (b) $(-1,3)$, (c) $(-1,2)$, (d) $[-1,2)$, (e) $(-1,2]$, (f) $[-1,2]$, (g) $(2,3)$, (h) $[2,3)$, (i) $(2,3]$, and (j) $[2,3]$.

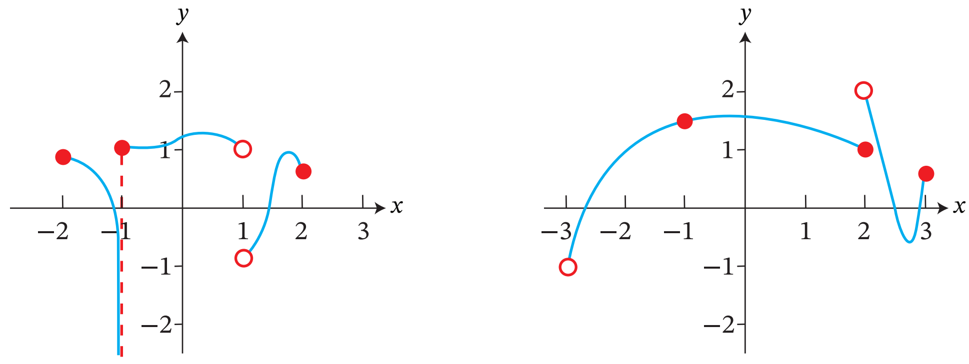


Figure 2.4.8

In Exercises 19 to 22, determine all numbers c for which $\lim_{x\to c} f(x)$ does not exist for the given graph of the function.

19. Figure 2.4.9(a)

20. Figure 2.4.9(b)

21. Figure 2.4.9(c)

22. Figure 2.4.9(d)

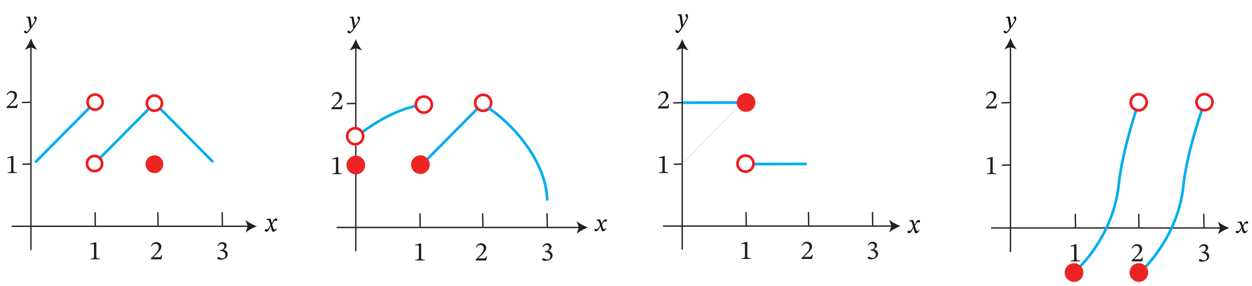


Figure 2.4.9

---

23. Let $f(x) = x + |x|$. (a) Graph f. (b) Is f continuous at -1? (c) Is f continuous at 0?

24. Let $f(x) = 2^{1/x}$ for $x \neq 0$.

(a) Find $\lim_{x\to\infty} f(x)$.

(b) Find $\lim_{x\to-\infty} f(x)$.

(c) Does $\lim_{x\to 0^+} f(x)$ exist?

(d) Does $\lim_{x\to 0^-} f(x)$ exist?

(e) Graph f, using the results from parts (a) to (d).

(f) Can $f(0)$ be defined so that f is continuous for all real numbers?

25. Let $f(x) = x\sin(1/x)$ for $x \neq 0$.

(a) Find $\lim_{x\to\infty} f(x)$.

(b) Find $\lim_{x\to-\infty} f(x)$.

(c) Find $\lim_{x\to 0} f(x)$.

(d) Graph f, using the results from parts (a) to (c).

(e) Can $f(0)$ be defined so that f is continuous for all real numbers?

In Exercises 26 to 28 find equations that the numbers k, p, and m must satisfy to make the function continuous.

26. $f(x) = \begin{cases} \dfrac{\sin(x)}{2x} & x \neq 0 \\ p & x = 0 \end{cases}$

27. $f(x) = \begin{cases} k & x \leq 0 \\ \arcsin(x) & 0 < x \leq 1 \\ p & x > 1 \end{cases}$

28. $f(x) = \begin{cases} \ln(x) & x > 1 \\ k - m\sqrt{x} & 0 < x \leq 1 \\ pe^{-x} & x \leq 0 \end{cases}$

29. (a) Let f and g be two functions defined for all numbers. If $f(x) = g(x)$ when x is not 3, must $f(3) = g(3)$?
(b) Let f and g be two continuous functions defined for all numbers. If $f(x) = g(x)$ when x is not 3, must $f(3) = g(3)$?

Explain your answers.

30. It might be hoped that if the positive number b and the number x are both close to 0, then b^x might be close to some fixed number. If that were so, it would suggest a definition for 0^0. Experiment with various choices of b and x near 0 and on the basis of your data write a paragraph on the theme, "Why 0^0 is difficult to define".

NOTE: This exercise explores the reason 0^0 is difficult to define.

2.5 Three Properties of Continuous Functions

Continuous functions have three properties important in calculus: the extreme value property, the intermediate value property, and the permanence property. All three properties are plausible, and a glance at the graph of a typical continuous function may persuade you that they are obvious. No proofs will be offered: they depend on the precise definitions of limits given in Sections 3.8 and 3.9 and are part of an advanced calculus course.

We will say that a function has a **local or relative maximum** at a point $(c, f(c))$ when $f(c) \geq f(x)$ for x near c. More precisely, there is an open interval I containing c such that if x is in I, and $f(x)$ is defined, then $f(x) \leq f(c)$. Likewise, a function has a **local or relative minimum** at a point $(c, f(c))$ when $f(c) \leq f(x)$ for x near c. Each maximum or minimum is referred to as an **extreme value** or **extremum** of the function.

The plural of extremum is *extrema*.

Extreme Value Property

The first property is that a function continuous throughout the closed interval $[a, b]$ has a value that is the largest somewhere in the interval. (When we refer to an interval $[a, b]$ it is assumed that a and b are numbers with $a < b$.)

Theorem 2.5.1: Maximum Value Property

If f is continuous throughout a closed interval $[a, b]$, then there is at least one number in $[a, b]$ at which f takes on a maximum value. That is, for some number c in $[a, b]$, $f(c) \geq f(x)$ for all x in $[a, b]$.

To persuade yourself that this is plausible, imagine sketching the graph of a continuous function. (See Figure 2.5.1.)

The maximum value property guarantees that a maximum value exists, but it does not tell how to find it. The problem of finding it is addressed in Chapter 4.

There is also a **minimum value property** that states that every continuous function on a closed interval takes on a smallest value. See Figure 2.5.1 for an illustration of this property. The following theorem combines the two properties.

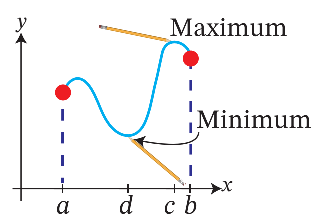


Figure 2.5.1

Theorem 2.5.2: Extreme Value Property

If f is continuous throughout the closed interval $[a, b]$, then there is at least one number in $[a, b]$ at which f takes on a minimum value and there is at least one number in $[a, b]$ at which f takes on a maximum value. That is, for some numbers c and d in $[a, b]$, $f(d) \leq f(x) \leq f(c)$ for all x in $[a, b]$.

EXAMPLE 1. Find all numbers in $[0, 3\pi]$ at which the cosine function, $f(x) = \cos(x)$, takes on a maximum value. Also, find all numbers in $[0, 3\pi]$ at which f takes on a minimum value.

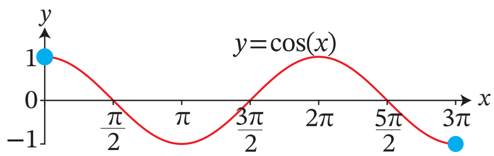


Figure 2.5.2

SOLUTION Figure 2.5.2 is a graph of $f(x) = \cos(x)$ for x in $[0, 3\pi]$. Inspection of the graph shows that the maximum value of $\cos(x)$ for $0 \le x \le 3\pi$ is 1, and it is attained twice: when $x = 0$ and when $x = 2\pi$. The minimum value of $\cos(x)$ is -1, which is also attained twice: when $x = \pi$ and when $x = 3\pi$. ♦

The extreme value property has two assumptions: "f is continuous" and "the domain is a closed interval." If either is removed, the conclusion need not hold.

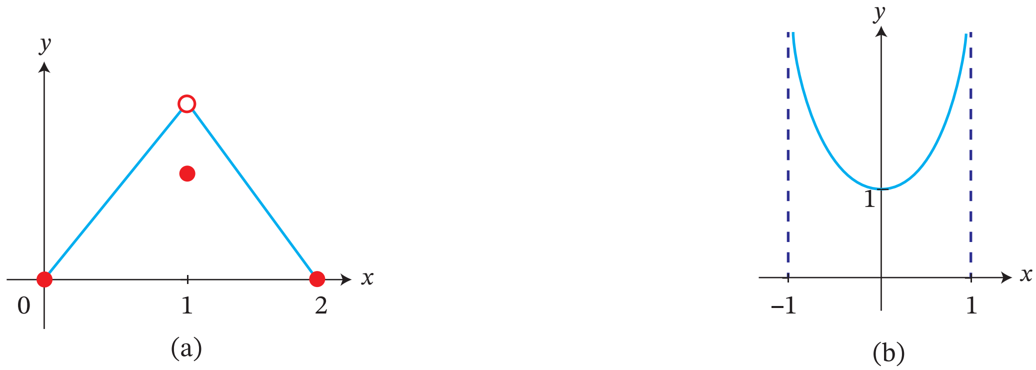


Figure 2.5.3

Figure 2.5.3(a) shows the graph of a function that is not continuous, is defined on a closed interval, but has no maximum value. And although $f(x) = 1/(1-x^2)$ is continuous on $(-1, 1)$, it has no maximum value, as a glance at Figure 2.5.3(b) shows. This does not violate the extreme value property because the domain $(-1, 1)$ is not a closed interval.

Intermediate Value Property

Imagine graphing a continuous function f defined on the closed interval $[a, b]$. As your pencil moves from the point $(a, f(a))$ to the point $(b, f(b))$ the y-coordinate of the pencil point goes through all values between $f(a)$ and $f(b)$. Similarly, if you hike all day, starting at an altitude of 5,000 feet and ending at 11,000 feet, you must have been at 7,000 feet at least once during the day. In mathematical terms, "a function that is continuous throughout an interval takes on all values between any two of its values".

Theorem 2.5.3: Intermediate Value Property

Assume f is continuous throughout the closed interval $[a, b]$. Let m be any number such that $f(a) \le m \le f(b)$ or $f(a) \ge m \ge f(b)$. Then there is at least one number c in $[a, b]$ such that $f(c) = m$.

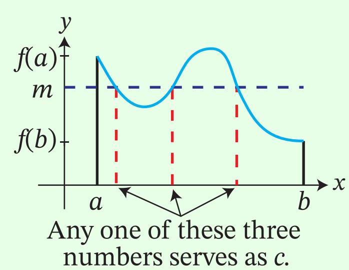


Figure 2.5.4

Pictorially, the intermediate value property asserts that if m is between $f(a)$ and $f(b)$, a horizontal line of height m must meet the graph of f at least once, as shown in Figure 2.5.4.

Even though the intermediate value property guarantees the existence of c, it does not tell how to find it. To find c we must be able to solve an equation, namely, $f(x) = m$.

EXAMPLE 2. Use the intermediate value property to show that the equation $2x^3 + x^2 - x + 1 = 5$ has a solution in the interval $[1,2]$.

SOLUTION If $P(x) = 2x^3 + x^2 - x + 1$, then $P(1) = 2\cdot 1^3 + 1^2 - 1 + 1 = 3$ and $P(2) = 2\cdot 2^3 + 2^2 - 2 + 1 = 19$. Since P is continuous on $[1,2]$ and $m = 5$ is between $P(1) = 3$ and $P(2) = 19$, the intermediate value property says there is at least one number c between 1 and 2 such that $P(c) = 5$.

To get a more accurate estimate for c such that $P(c) = 5$, find a shorter interval for which the intermediate value property can be applied. For instance, $P(1.2) = 4.696$ and $P(1.3) = 5.784$. By the intermediate value property, there is a number c in $[1.2, 1.3]$ such that $P(c) = 5$. ♦

EXAMPLE 3. Show that the equation $-x^5 - 3x^2 + 2x + 11 = 0$ has at least one real root. In other words, show that the graph of $y = -x^5 - 3x^2 + 2x + 11$ crosses the x-axis.

SOLUTION
Let $f(x) = -x^5 - 3x^2 + 2x + 11$. We wish to show that there is a number c such that $f(c) = 0$. To use the intermediate value property, we need to find an interval $[a,b]$ for which 0 is between $f(a)$ and $f(b)$, that is, where one of $f(a)$ and $f(b)$ is positive and the other is negative. Then we could apply the property, with $m = 0$.

We show that there are numbers a and b with $a < b$, such that $f(a) > 0$ and $f(b) < 0$. Because $\lim_{x\to\infty} f(x) = -\infty$, $f(x)$ is negative for x large and positive. Thus, there is a positive number b such that $f(b) < 0$. Similarly, $\lim_{x\to-\infty} f(x) = \infty$, so when x is negative and of large absolute value, $f(x)$ is positive. Hence there is a negative number a such that $f(a) > 0$. Thus there are numbers a and b, with $a < b$, such that $f(a) > 0$ and $f(b) < 0$. For instance, $f(-1) = 7$ and $f(2) = -29$.

The number 0 is between $f(-1)$ and $f(2)$. Since f is continuous on the interval $[-1,2]$, there is a number c in $[-1,2]$ such that $f(c) = 0$. This number c is a solution to the equation $-x^5 - 3x^2 + 2x + 11 = 0$.

The graph of $y = f(x)$ in Figure 2.5.5 confirms these findings. ♦

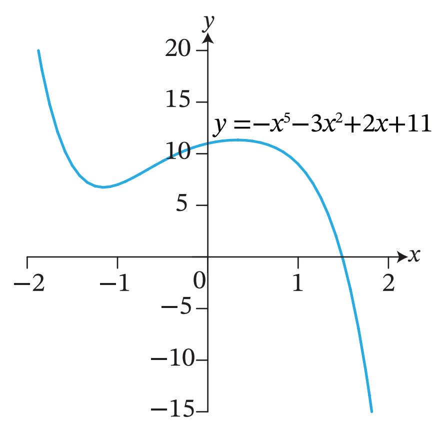


Figure 2.5.5

Observation 2.5.4: *Roots of Polynomials*

The reasoning used in Example 3 shows any polynomial of odd degree has a real root. The argument does not hold for polynomials of even degree; the equation $x^2 + 1 = 0$, for instance, has no real solutions.

EXAMPLE 4. Use the intermediate value property to show there is a negative number such that $\ln(x+4) = x^2 - 3$.

SOLUTION To show that there is a negative number $x = c$ where $\ln(x+4)$ and $x^2 - 3$ have the same value, first look for intersections of the graphs of $y = \ln(x+4)$ and $y = x^2 - 3$ in Figure 2.5.6. The graphs appear to intersect at two different values of x, one of which is negative.

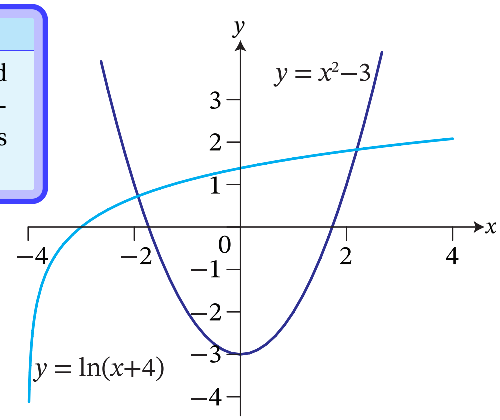


Figure 2.5.6

Showing there is a negative number x for which $\ln(x+4) = x^2 - 3$ is equivalent to showing the difference $\ln(x+4) - (x^2-3)$ is zero for at least one negative number x. This reduces the problem to showing that the function $f(x) = \ln(x+4) - x^2 + 3 = 0$ for some input x, with $x < 0$.

We will proceed as in the previous example. We want to find negative numbers a and b such that $f(a)$ and $f(b)$ have opposite signs.

Because $\ln(x+4)$ is defined only for $x+4 > 0$, that is, for $x > -4$, we search for a and b by making a table of values of $f(x)$ for some arguments in $(-4, 0]$.

x	-3	-2	-1	0
$f(x)$	-6	-0.307	3.099	4.386

Table 2.5.1 t1-5-a

We see that $f(-2)$ is negative and $f(-1)$ is positive. Since 0 lies between $f(-2)$ and $f(-1)$ and f is continuous on $[-2,-1]$, the intermediate value property asserts that there is a number, c, in $[-2,-1]$ such that $f(c) = 0$. It follows that $\ln(c+4) = c^2 - 3$. ♦

In Example 4 the intermediate value property does not provide an exact value for c. As previously noted, the graphs of $y = \ln(x+1)$ and $y = x^2 - 3$ in Figure 2.5.6 suggest that there are two points of intersection, but only one with a negative input. The graph, and the table of values, suggest that the intersection point occurs when the input is close to -2. Calculations on a calculator or computer show that $c \approx -1.931$.

Permanence Property

The extreme value property as well as the intermediate value property involve the behavior of a continuous function throughout an interval. The next property concerns the local behavior of a continuous function.

Assume that a continuous function f is defined on an open interval that contains a and that $f(a) = p$ is positive. Then it seems plausible that f remains positive in some open interval that contains a. We can say something stronger:

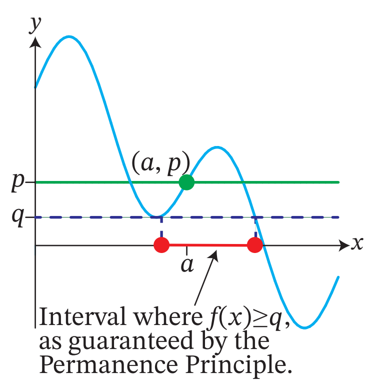


Figure 2.5.7

Theorem 2.5.5: Permanence Property

Assume that the domain of function f contains an open interval that includes the number a. Assume that f is continuous at a and that $f(a) = p$. Let q be any number less than p. Then there is an open interval including a such that $f(x) \geq q$ for all x in that interval.

Observation 2.5.6: *An Application of the Permanence Property*

In particular, if $f(x)$ is continuous on an open interval containing $x = a$, and $f(a)$ is positive, then there is an open interval containing a where $f(x)$ remains positive.

To persuade yourself that the permanence property is plausible, imagine what the graph of $y = f(x)$ looks like near $(a, f(a))$, as in Figure 2.5.7. In Section 3.9 it will be revealed that the permanence property is a consequence of the definition of continuity.

Summary

This section stated the extreme value property, the intermediate value property, and the permanence property. While formal proofs were not provided, graphical motivations for each of the hypotheses were provided.

These properties of continuous functions have produced two important observations: that every odd polynomial must have at least one real root and that a continuous function that is positive at one point must remain positive on an open interval containing that point.

Additional consequences of these properties will be encountered in upcoming chapters.

EXERCISES for Section 2.5

1. Find the maximum value of $\cos(x)$ and the value(s) of x at which the maximum occurs in (a) $\left[0, \frac{\pi}{2}\right]$ and (b) $[0, 2\pi]$.

2. Does the function $f(x) = \frac{x^3 + x^4}{1 + 5x^2 + x^6}$ have maximum and minimum values for x in $[1,4]$? If so, use a graphing device to estimate the extreme values.

3. Does the function $g(x) = 2^x - x^3 + x^5$ have maximum and minimum values for x in $[-3,10]$? If so, use a graphing device to estimate the extreme values.

4. Does the function $h(u) = u^3$ have a maximum value for u in (a) $[2,4]$? (b) $[-3,5]$? (c) $(1,6)$?
In each case, if so, where does the maximum occur and what is the maximum value?

5. Does the function $u(t) = t^4$ have a minimum value for t in (a) $[-5,6]$? (b) $(-2,4)$? (c) $(3,7)$? (d) $(-4,4)$?
In each case, if so, where does the minimum occur and what is the minimum value?

6. Does the function $r(t) = 2 - t^2$ have maximum and minimum values for t in $(-1,1)$? If so, where?

7. Does the function $h(x) = 2 + x^2$ have maximum and minimum values for x in $(-1,1)$? If so, where?

8. Show that $x^5 + 3x^4 + x - 2 = 0$ has at least one solution in $[0,1]$.

9. Show that $x^5 - 2x^3 + x^2 - 3x = -1$ has at least one solution in $[1,2]$.

In Exercises 10 to 14 verify the intermediate value property for the function f, the interval $[a,b]$, and the value m. Find all values of c in each case.

10. $f(x) = 3x + 5$, $[a,b] = [1,2]$, $m = 10$.

11. $f(x) = x^2 - 2x$, $[a,b] = [-1,4]$, $m = 5$.

12. $f(x) = \sin(x)$, $[a,b] = \left[\frac{\pi}{2}, \frac{11\pi}{2}\right]$, $m = -1$.

13. $f(x) = \cos(x)$, $[a,b] = [0,5\pi]$, $m = \frac{\sqrt{3}}{2}$.

14. $f(x) = x^3 - x$, $[a,b] = [-2,2]$, $m = 0$.

---

15. Use the intermediate value property to show $3x^3 + 11x^2 - 5x = 2$ has a solution.

16. Show that $2^x = 3x$ has a solution in $[0,1]$.

17. Does $x + \sin(x) = 1$ have a solution?

18. Does $x^3 = 2^x$ have a solution?

19. Define $f(x) = \frac{1}{x}$, $a = -1$, $b = 1$, and $m = 0$. We have $f(a) \le 0 \le f(b)$. Is there at least one c in $[a,b]$ such that $f(c) = 0$? If so, find c; if not, does this imply the intermediate value property sometimes does not hold?

20. Use the intermediate value property to show there is a positive number such that $\ln(x+4) = x^2 - 3$.

21. Let $f(x) = x^2$. Then $f(2) = 4$. Find an interval (a,b) containing 2 such that $f(x) \ge 3.8$ for all x in (a,b).

Exercise 22 and Exercise 23 are related.

22. Let $P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ be a polynomial of odd degree n and with positive leading coefficient a_n. Show that there is at least one real number r such that $P(r) = 0$. The number r is called a **root** of $P(x)$.

23. The **factor theorem** from algebra asserts that the number r is a root of a polynomial $P(x)$ if and only if $x - r$ is a factor of $P(x)$. For instance, 2 is a root of the polynomial $x^2 - 3x + 2$ and $x - 2$ is a factor of it: $x^2 - 3x + 2 = (x-2)(x-1)$.

(a) Use the factor theorem and Exercise 22 to show that every polynomial of odd degree has a degree 1 factor.

(b) Show that none of x^2+1, x^4+1, and $x^{100}+1$ has a first-degree factor.
(c) Verify that $x^4+1=\left(x^2+\sqrt{2}x+1\right)\left(x^2-\sqrt{2}x+1\right)$. (It can be shown using complex numbers that every polynomial with real coefficients is the product of polynomials of degrees at most 2 with real coefficients.)

---

24. Let $f(x)=a_nx^n+\cdots+a_1x+a_0$ where a_n and a_0 have opposite signs. Show that $f(x)$ has a positive root, that is, that $f(x)=0$ has a positive solution.

25. The continuous function f defined on the x-axis satisfies the equation

$$f(x+y)=f(x)+f(y) \text{ for all } x \text{ and } y.$$

For any constant c, $f(x)=cx$ satisfies the equation since $c(x+y)=cx+cy$. This exercise shows that f must be of that form. To begin, define c to be $f(1)$. Show that

(a) $f(2)=2c$.
(b) $f(0)=0$.
(c) $f(-1)=-c$.
(d) for any positive integer n, $f(n)=cn$.
(e) for any negative integer n, $f(n)=cn$.
(f) $f\left(\frac{1}{2}\right)=\frac{c}{2}$.
(g) for any nonzero integer n, $f\left(\frac{1}{n}\right)=\frac{c}{n}$.
(h) for any integer m and any positive integer n, $f\left(\frac{m}{n}\right)=\frac{m}{n}c$.
(i) for any irrational number x, $f(x)=cx$.

This is where the continuity of f enters. Parts (h) and (i) together complete the solution.

26. (a) If f is a continuous function defined for all real numbers, is there necessarily a number x where $f(x)=x$?
(b) If f is a continuous function with domain $[0,1]$ such that $f(0)=1$ and $f(1)=0$, is there necessarily a number x such that $f(x)=x$?

27. Assume that f is a continuous function defined on $(-\infty,\infty)$ with $f(0)=1$ and $f(2x)=f(x)$ for all numbers x. (a) Give an example of such a function f. (b) Find all such functions and explain.

Exercises 28 to 35 concern convex sets and show how the intermediate value property gives geometric information. In them you will need to define various functions geometrically. You may assume these functions are continuous.

Definition: ***Convex Sets and Curves***

A set in the plane bounded by a curve is **convex** if, for any two points P and Q, the line segment joining them also lies in the set. (See Figure 2.5.8(a).) The boundary of a convex set we will call a **convex curve**.

Disks, triangles, and parallelograms are convex sets. The quadrilateral shown in Figure 2.5.8(b) is not convex.

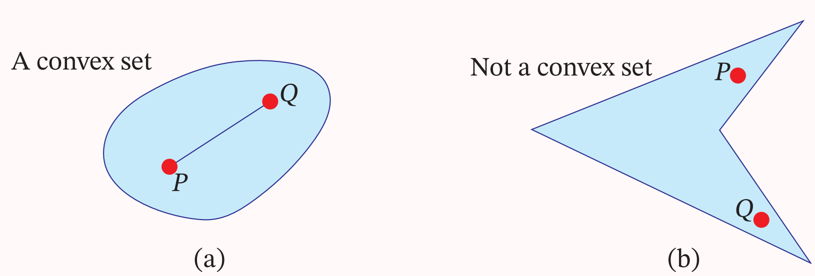


Figure 2.5.8

28. Let L be a line in the plane and let K be a bounded convex set. Show that there is a line parallel to L that cuts K into two pieces with equal areas.

29. (a) Introduce an x-axis perpendicular to L with its origin on L. Each line parallel to L and meeting K crosses the x-axis at a number x. Label the line L_x. Denote by a the smallest and by b the largest value of x. (See Figure 2.5.9.) Assume the area of K is A.

(b) Let $A(x)$ be the area of K to the left of the line L_x. What is $A(a)$? $A(b)$?

(c) Use the intermediate value property to show that there is an x in $[a, b]$ such that $A(x) = \frac{A}{2}$.

(d) Why does (c) show that there is a line parallel to L that cuts K into two pieces of equal areas?

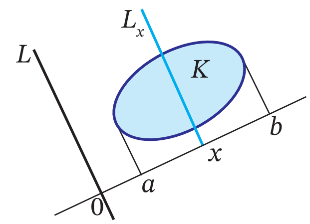


Figure 2.5.9

30. Solve the preceding exercise by applying the intermediate value property to the function $f(x) = A(x) - B(x)$, where $B(x)$ is the area to the right of L_x.

31. Assume P is a point and K is a convex set. Is there a line through P that cuts K into two pieces of equal areas?

32. Assume K is a convex set in the plane. Show that there is a line that simultaneously cuts K into two pieces of equal areas and cuts the boundary of K into two pieces of equal lengths.

33. Given a convex set, K, in the plane. Show that there are two perpendicular lines that cut K into four pieces of equal areas.

34. Given two convex sets, K_1 and K_2, in the plane. Is there a line that simultaneously cuts K_1 into two pieces of equal areas and cuts K_2 into two pieces of equal areas?

This is sometimes called the “two pancakes” problem.

35. Assume K is a convex set in the plane with a boundary that does not contain any line segments. A polygon is said to **circumscribe** K if each edge of the polygon is tangent to the boundary of K.

(a) Is there necessarily a circumscribing equilateral triangle? If so, how many?

(b) Is there necessarily a circumscribing rectangle? If so, how many?

(c) Is there necessarily a circumscribing square?

2.6 Techniques for Graphing

One way to graph a function $f(x)$ is to compute $f(x)$ at several inputs x, plot the points $(x, f(x))$, and draw a curve through them. This may be tedious and, if you choose inputs that give misleading information, may result in an inaccurate graph.

Another way is to use a graphing device such as a graphing calculator or an app on a smartphone. These display only a portion of the graph and might display a part of the graph that is misleading or of little interest. At points with large function values, the graph may be distorted by the calculator's choice of scale.

It pays to be able to get an idea of the shape of a graph quickly, without having to compute too many values. This section describes some shortcuts.

Intercepts

The x-coordinates of the points where the graph of a function meets the x-axis are the x**-intercepts** of the function. The y-coordinate of the point where a graph meets the y-axis is the y**-intercept** of the function.

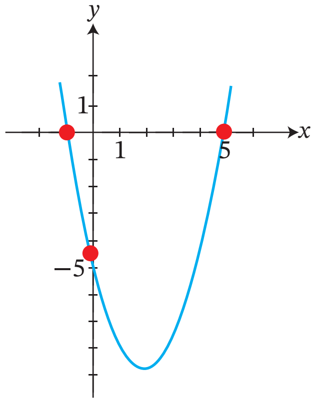


Figure 2.6.1

EXAMPLE 1. Find the intercepts of the graph of $y = x^2 - 4x - 5$.

SOLUTION To find the x-intercepts, set $y = 0$, obtaining $0 = x^2 - 4x - 5$. This quadratic factors into two linear factors:

$$0 = x^2 - 4x - 5 = (x - 5)(x + 1),$$

which is satisfied when $x = 5$ or $x = -1$. There are two x-intercepts, 5 and -1. (If the equation did not factor easily, the quadratic formula could be used.)
To find the y-intercept, set $x = 0$, obtaining

$$y = 0^2 - 4 \cdot 0 - 5 = -5.$$

There is only one y-intercept, namely -5.

The intercepts give us three points on the graph. Tabulating a few more points produces the parabola in Figure 2.6.1, where the intercepts are shown as well. ♦

If $f(x)$ is not defined when $x = 0$, there is no y-intercept. If $f(x)$ is defined when $x = 0$, then it's easy to get the y-intercept: evaluate $f(0)$. While there is at most one y-intercept, there may be many x-intercepts. To find them, solve the equation $f(x) = 0$. That is,

Observation 2.6.1: *Finding Intercepts of* $y = f(x)$

To find the y-intercept (if there is one), compute $f(0)$.
To find the x-intercepts (if there are any), solve the equation $f(x) = 0$.

Symmetry of Odd and Even Functions

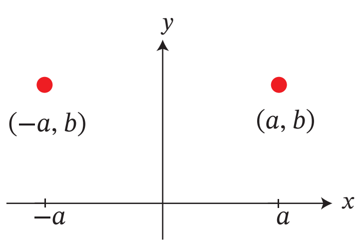
Some functions have the property that when you replace x by $-x$ you get the same value. For instance, the function $f(x) = x^2$ has this property since

$$f(-x) = (-x)^2 = x^2 = f(x).$$

So does $f(x) = x^n$ for any even integer n. There are other functions, such as $3x^4 - 5x^2 + 6$, $\cos(x)$, and $e^x + e^{-x}$, that also have this property.

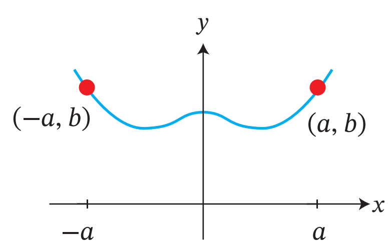
Definition: *Even Function*

A function f wih the property that $f(-x) = f(x)$ is called an **even function**.



Two points on the graph of an even function

(a)



Graph of an even function

(b)

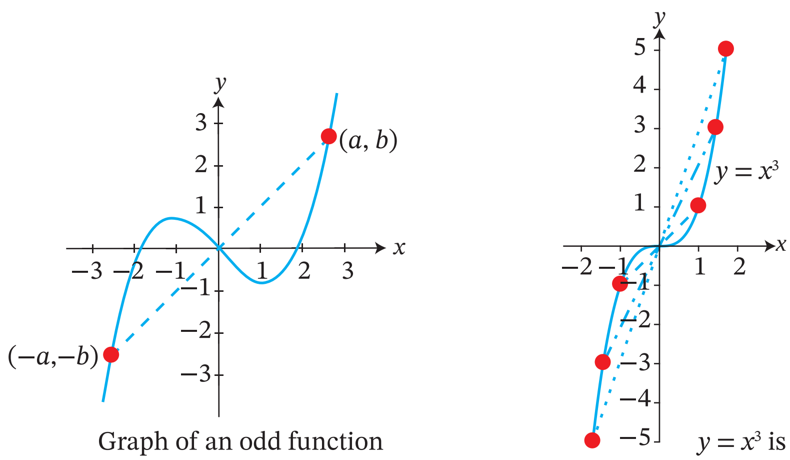
Figure 2.6.2

For an even function f, if $f(a) = b$, then $f(-a) = b$ also. In other words, if the point (a, b) is on the graph of f, so is the point $(-a, b)$, as indicated in Figure 2.6.2(a).

The graph of an even function f is **symmetric with respect to the y-axis**, as shown in Figure 2.6.2(b). If you notice that a function is even, you can save half the work in finding its graph. Graph it for positive x and then get the part for negative x by reflecting across the y-axis. If you wanted to graph $y = x^4/(1-x^2)$, for example, stick to $x > 0$, and then reflect the result across the y-axis.

Definition: *Odd Function*

A function f with the property that $f(-x) = -f(x)$ is called an **odd function**.



Graph of an odd function
(a)

$y = x^3$ is an odd function
(b)

Figure 2.6.3

The function $f(x) = x^3$ is odd since

$$f(-x) = (-x)^3 = -(x^3) = -f(x).$$

For any odd integer n, $f(x) = x^n$ is an odd function. The sine function is also odd, since $\sin(-x) = -\sin(x)$.

If the point (a, b) is on the graph of an odd function, so is $(-a, -b)$, since

$$f(-a) = -f(a) = -b.$$

(See Figure 2.6.3(a).) Because the origin $(0, 0)$ is the midpoint of the segment whose ends are (a, b) and $(-a, -b)$, the graph is said to be **symmetric with respect to the origin**.

If you graph an odd function for positive x, you can obtain the graph for negative x by reflecting it point by point through the origin. For example, if you graph $y = x^3$ for $x \geq 0$, as in Figure 2.6.3(b), you can complete the graph by reflection with respect to the origin, as indicated by the dashed lines.

Most functions are neither even nor odd. For instance, $x^3 + x^4$ is neither even nor odd since $(-x)^3 + (-x)^4 = -x^3 + x^4$, which is neither $x^3 + x^4$ nor $-(x^3 + x^4)$.

Asymptotes

If $\lim_{x\to\infty} f(x) = L$ where L is a real number, the graph of $y = f(x)$ gets arbitrarily close to the horizontal line $y = L$ as x increases. The line $y = L$ is called a **horizontal asymptote** of the graph of f. (See Figure 2.6.4.)

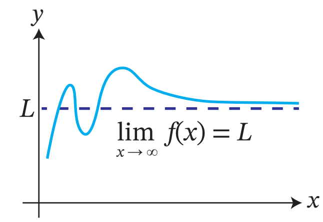


Figure 2.6.4

If a graph has an asymptote, we can draw the asymptote and use it as a guide in drawing the graph.

If $\lim_{x\to a} f(x) = \infty$, then the graph resembles the vertical line $x = a$ for x near a. The line $x = a$ is called a **vertical asymptote** of the graph of $y = f(x)$. The same term is used if

$$\lim_{x\to a} f(x) = -\infty, \qquad \lim_{x\to a^+} f(x) = \infty \text{ or } -\infty, \quad \text{or} \quad \lim_{x\to a^-} f(x) = \infty \text{ or } -\infty.$$

Figure 2.6.5 presents some examples with vertical asymptotes.

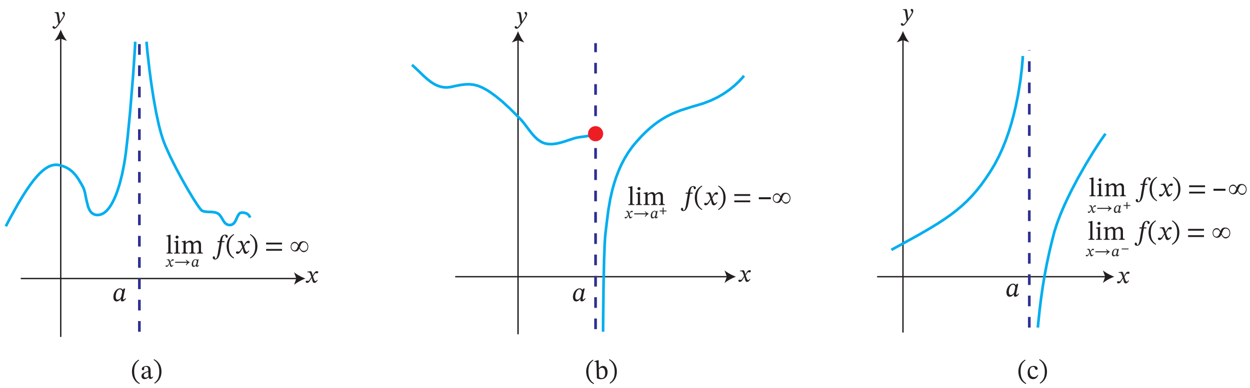


Figure 2.6.5

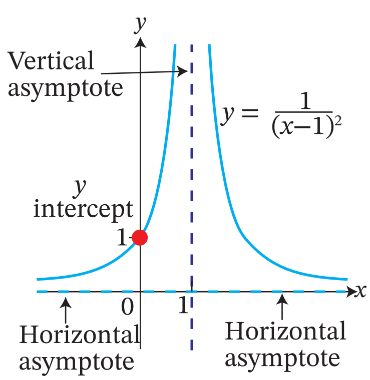


Figure 2.6.6

EXAMPLE 2. Graph $f(x) = \dfrac{1}{(x-1)^2}$.

SOLUTION To see if there is any symmetry, check whether $f(-x)$ is $f(x)$ or $-f(x)$. We have

$$f(-x) = \frac{1}{(-x-1)^2} = \frac{1}{(x+1)^2}.$$

Since $1/(x+1)^2$ is neither $1/(x-1)^2$ nor $-1/(x-1)^2$, the function $f(x)$ is neither even nor odd. Therefore the graph is not symmetric with respect to the y-axis or with respect to the origin.

To determine the y-intercept compute $f(0) = 1/(0-1)^2 = 1$. The y-intercept is 1. To find any x-intercepts, solve the equation $f(x) = 0$, that is,

$$\frac{1}{(x-1)^2} = 0.$$

Since no number has a reciprocal equal to zero, there are no x-intercepts.

To search for a horizontal asymptote examine

$$\lim_{x\to\infty} \frac{1}{(x-1)^2} \qquad \text{and} \qquad \lim_{x\to-\infty} \frac{1}{(x-1)^2}.$$

Both limits are 0. The line $y = 0$, that is, the x-axis, is an asymptote both to the right and to the left. Since $1/(x-1)^2$ is positive, the graph lies above the asymptote.

To discover any vertical asymptotes, find where the function $1/(x-1)^2$ blows up — that is, becomes arbitrarily large (in absolute value). This happens when the denominator $(x-1)^2$ becomes zero. Solving $(x-1)^2 = 0$ we find $x = 1$. The function is not defined for $x = 1$. The line $x = 1$ is a vertical asymptote.

To determine the shape of the graph near the line $x = 1$, examine the one-sided limits: $\lim_{x\to1^+} 1/(x-1)^2$ and $\lim_{x\to1^-} 1/(x-1)^2$. Since the square of a nonzero number is always positive, we see that $\lim_{x\to1^+} 1/(x-1)^2 = \infty$ and $\lim_{x\to1^-} 1/(x-1)^2 = \infty$. All this information is displayed in Figure 2.6.6. ♦

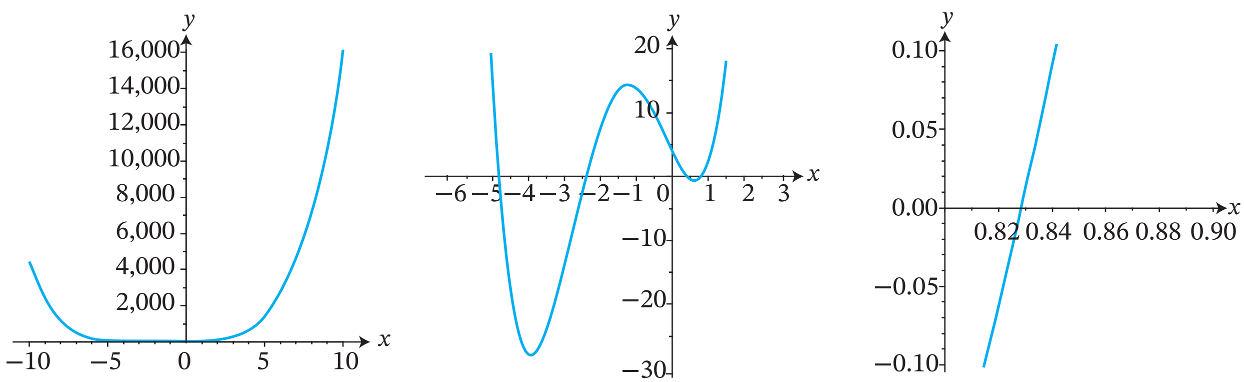
Technology-Assisted Graphing

A graphing utility needs to know the function and the viewing window. We will illustrate the effect of using the wrong viewing window, and provide some general rules for selecting a good viewing window.

The **viewing window** is the portion of the xy-plane described as $[a, b] \times [c, d]$ when the window extends horizontally from $x = a$ to $x = b$ and vertically from $y = c$ to $y = d$. The graph of a function $y = f(x)$ is created by evaluating $f(x)$ for a sample of numbers x between a and b. The point $(x, f(x))$ is added to the plot. It is customary to connect the points to form the graph of $y = f(x)$.

EXAMPLE 3. Find a viewing window that shows the general shape of the graph of $y = x^4 + 6x^3 + 3x^2 - 12x + 4$. Use graphs to estimate the rightmost x-intercept.

SOLUTION Figure 2.6.7(a) is typical of the first plot of a function. Choose a wide x interval, here $[-10, 10]$, and let the graphing software choose an appropriate vertical range. While this view is useless for estimating any specific x-intercept, it is tempting to say that any x-intercepts will be between $x = -6$ and $x = 3$. Figure 2.6.7(b) is the graph of this function on the viewing window $[-6, 3] \times [-30, 20]$. Now four x-intercepts are visible. The rightmost one occurs around $x = 0.8$. Figure 2.6.7(c) is the result of zooming in on this part of the graph where we see that the rightmost x-intercept is about 0.83.



viewing window: $-10 \le x \le 10$,
y range chosen by graphing software

(a)

viewing window: $-3 \le x \le 6$,
$-30 \le y \le 20$

(b)

viewing window: $0.8 \le x \le 0.9$,
$-0.1 \le y \le 0.1$

(c)

Figure 2.6.7

A computer algebra system shows that the four x-intercepts are 0.8284, 0.4142, −2.4142, and −4.8284 (to four decimal places). ♦

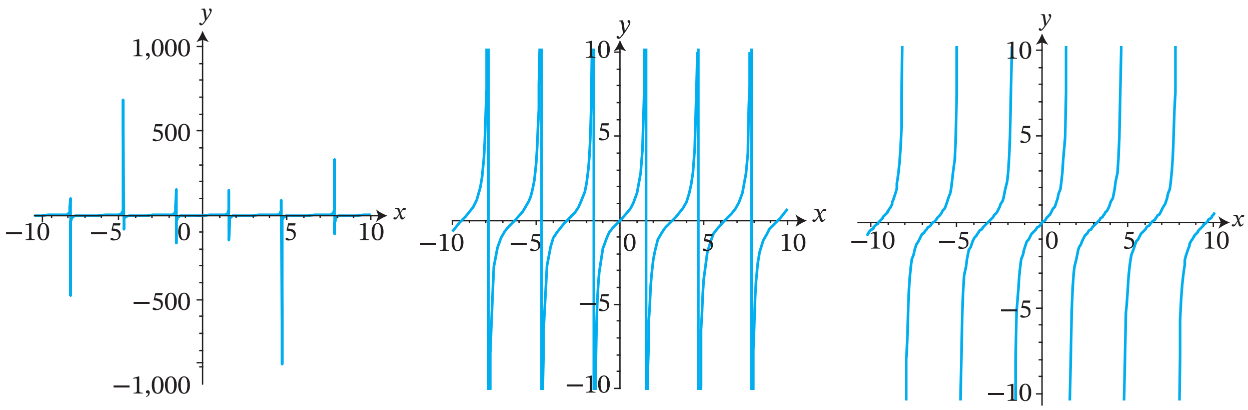
Generating a collection of points and connecting the dots can sometimes lead to ridiculous results, as in Example 4.

EXAMPLE 4. Find a viewing window that shows the general shape and periodicity of the graph of $y = \tan(x)$.

SOLUTION A computer-generated plot of $y = \tan(x)$ for x between −10 and 10 with no restriction on the vertical height of the viewing window is shown in Figure 2.6.8(a). The graph is not periodic and it does not look like the graph of a trigonometric function.

The default vertical height is long: $[-1000, 1000]$. Reducing this by a factor of 100, that is, to $[-10, 10]$, yields Figure 2.6.8(b). The graph is periodic and exhibits the expected periodic behavior.

To understand this plot, realize that the software selects a sample of input values from the domain, computes the value of tangent of each input, then connects the points in order of the input values. The tangent of the last input smaller than $\pi/2$ is large and positive and the tangent of the first input larger than $\pi/2$ is large and negative.



viewing window:
$-10 \leq x \leq 10$, $-10 \leq y \leq 10$,
y range chosen by graphing software

(a)

viewing window:
$-10 \leq x \leq 10$, $-10 \leq y \leq 10$

(b)

viewing window:
$-10 \leq x \leq 10$, $-10 \leq y \leq 10$,
and vertical asymptotes detected

(c)

Figure 2.6.8

Neither point is in the viewing window, but the line segment connecting them passes through the viewing window and appears as the vertical line at $x = \pi/2$ in Figure 2.6.8(b). Because the tangent is not defined for every odd multiple of $\pi/2$, similar reasoning explains the other vertical lines at every odd multiple of $\pi/2$

These segments are not part of the graph. Figure 2.6.8(c) shows the graph of $y = \tan(x)$ with them removed. ♦

Example 4 illustrates why we must remain alert when using technology. We have to check that the results are consistent with what we already know.

The next example shows that sometimes it is not possible to show all of the important features of a function in a single graph.

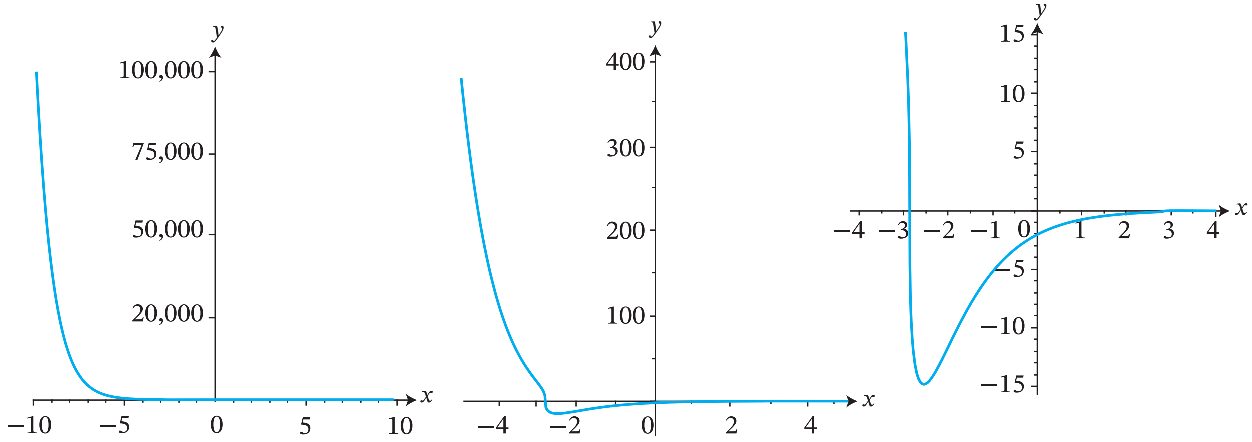
EXAMPLE 5. Use one or more graphs to show all major features of the graph of $y = \sqrt[3]{x^2-8}\, e^{-x}$.

SOLUTION The graph of this function on $[-10, 10]$ with the vertical window chosen by the software is shown in Figure 2.6.9(a). In this window, the exponential function dominates the graph.

At $x = 0$ the value of the function is $(0-8)^{1/3}e^0 = -2$. To get enough detail to see both the positive and negative values of the function, zoom in by reducing the x interval to $[-5, 5]$. The result is Figure 2.6.9(b). Reducing the x interval to $[-4, 4]$ and specifying the y interval as $[-15, 15]$ gives Figure 2.6.9(c).

We could continue to adjust the viewing window until we find suitable views. A more systematic approach is to look at the graphs of $y = \sqrt[3]{x^2-8}$ and $y = e^{-x}$ separately, but on the same pair of axes, as in Figure 2.6.10(a), where the solid red curve is $y = \sqrt[3]{x^2-8}$ and the dashed blue curve is $y = e^{-x}$. The exponential growth of e^{-x} for negative values of x stretches (vertically) the graph of $y = \sqrt[3]{x^2-8}$ to the left of the y-axis while the exponential decay for $x > 0$ (vertically) compresses the graph of $y = \sqrt[3]{x^2-8}$ to the right of the y-axis.

It is prudent to produce two separate plots. To the left of the y-axis, with a viewing window of $[-4, 0] \times [-15, 100]$, the graph of the function is shown in Figure 2.6.10(b). This graph shows the root at $x = -2\sqrt{2}$. To the right of the y-axis, with a shorter viewing window of $[0, 4] \times [-2.2, 0.2]$, the graph is as shown in Figure 2.6.10(c). This view shows the root at $x = 2\sqrt{2}$. ♦



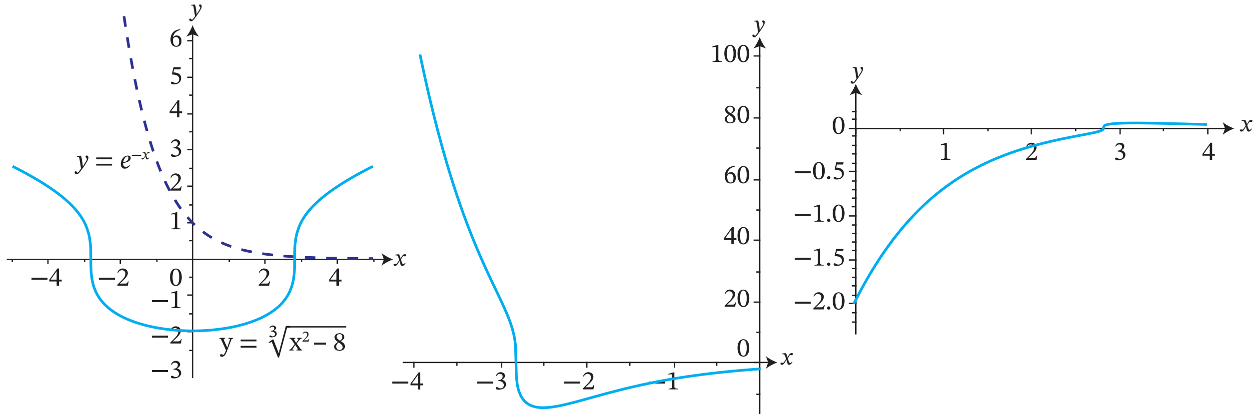
viewing window: $-10 \le x \le 10$,
y range chosen by graphing software

viewing window: $-4 \le x \le 4$,
$-20 \le y \le 400$

viewing window: $-4 \le x \le 4$,
$-15 \le y \le 15$

(a) (b) (c)

Figure 2.6.9



(a) (b) (c)

Figure 2.6.10

Summary

The first half of this section presents three tools for making a quick sketch of the graph of $y = f(x)$ by hand.

(a) *Check for intercepts.* Find $f(0)$ to get the y-intercept. Solve $f(x) = 0$ to get the x-intercepts.

(b) *Check for symmetry.* Is $f(-x)$ equal to $f(x)$ or $-f(x)$?

(c) *Check for asymptotes.* If $\lim_{x\to\infty} f(x) = L$ or $\lim_{x\to-\infty} f(x) = L$, then $y = L$ is a horizontal asymptote. If $\lim_{x\to a} f(x) = +\infty$ or $-\infty$, then $x = a$ is a vertical asymptote. This is also the case whenever $\lim_{x\to a^+} f(x)$ or $\lim_{x\to a^-} f(x)$ is $+\infty$ or $-\infty$.

The second half of the section provided some pointers for using an automatic graphing utility. The examples illustrate the importance of selecting an appropriate viewing window.

EXERCISES for Section 2.6

1. Show that these are even functions.
(a) $f(x) = x^2 + 2$, (b) $f(x) = \sqrt{x^4 + 1}$, (c) $f(x) = \frac{1}{x^2}$, and (d) $f(x) = x^3 \sin(5x)$.

2. Show that these are even functions.
(a) $f(x) = 5x^4 - x^2$, (b) $f(x) = \cos(2x)$, (c) $f(x) = e^x + e^{-x}$, and (d) $f(x) = \sqrt{1 - x^2}\tan^2(x)$.

3. Show that these are odd functions.
(a) $f(x) = x^3 - x$, (b) $f(x) = x + \frac{1}{x}$, (c) $f(x) = e^x - e^{-x}$, and (d) $f(x) = \cos(5x)\sin(3x)$.

4. Show that these are odd functions.
(a) $f(x) = 2x + \frac{1}{2x}$, (b) $f(x) = \tan(x)\cos(2x)$, (c) $f(x) = x^{5/3}$, and (d) $f(x) = \frac{x^3}{\cos(2x)}$.

5. Show that these functions are neither odd nor even.
(a) $f(x) = 3 + x$, (b) $f(x) = (x+2)^2$, and (c) $f(x) = \frac{x}{x+1}$.

6. Show that these functions are neither odd nor even.
(a) $f(x) = 2x - 1$, (b) $f(x) = e^x$, and (c) $f(x) = x^2 + \frac{1}{x}$.

7. Label each function as even, odd, or neither.
(a) $f(x) = x + x^3 + 5x^4$, (b) $f(x) = 7x^4 - 5x^2$, and (c) $f(x) = e^x - e^{-x}$.

8. Label each function as even, odd, or neither.
(a) $f(x) = \frac{1+x}{1-x}$, (b) $f(x) = \ln(x^2 + 1)$, and (c) $f(x) = \sqrt[3]{x^2 + 1}$.

In Exercise 9 to 18 find the x- and y-intercepts, if there are any.

9. $y = 2x + 3$ **10.** $y = 3x - 7$ **11.** $y = x^2 + 3x + 2$ **12.** $y = 2x^2 + 5x + 3$ **13.** $y = 2x^2 + 1$

14. $y = x^2 + x + 1$ **15.** $y = \sin(x+1)$ **16.** $y = \ln(x^2 + 1)$ **17.** $y = \frac{x^2 - 1}{x^2 + 1}$ **18.** $y = e^{\cos(x)}$

In Exercises 19 to 24 find the horizontal and vertical asymptotes.

19. $y = \frac{x+2}{x-2}$ **20.** $y = \frac{x-2}{x^2 - 9}$ **21.** $y = \frac{x}{x^2 + 1}$

22. $y = \frac{3}{1 + e^{-x}}$ **23.** $y = \frac{\sin(2x)}{x}$ **24.** $y = \frac{x}{x^2 + 2x + 1}$

In Exercises 25 to 38, graph the given function.

25. $y = \frac{1}{x-2}$

26. $y = \frac{1}{x+3}$

27. $y = \frac{1}{x^2-1}$

28. $y = \frac{x}{x^2-2}$

29. $y = \frac{x^2}{1+x^2}$

30. $y = \frac{1}{x^3+x^{-1}}$

31. $y = \frac{3}{1+e^{-x}}$

32. $y = \frac{e^{-x}}{3+e^{-x}}$

33. $y = \frac{e^{-x/2}}{1+e^{-x}}$

34. $y = \frac{2+e^{-x}}{3+e^{-2x}}$

35. $y = \frac{1-e^x}{1+e^x}$

36. $y = \frac{2-e^{-2x}}{3+e^{-3x}}$

37. $y = \frac{1}{x(x-1)(x+2)}$

38. $y = \frac{x+2}{x^3+x^2}$

Use a graphing utility to sketch graphs of the functions in Exercises 39 to 57. Be sure to indicate the viewing window used.

39. $(x^2+x-6)\ln(x+2)$

40. $(x^2-x+6)\ln(x+2)$

41. $(x^2+4)\ln(x+1)$

42. $(x^2-4)\ln(x+1)$

43. $\frac{x^3}{x^2-4}\arctan\left(\frac{x}{5}\right)$

44. $\frac{x^2-4}{x^3}\arctan\left(\frac{x}{5}\right)$

45. $\frac{x^3-3x}{x^2-4}$

46. $\frac{x^3-2x}{x^2-4}$

47. $\frac{\sin(x)}{x}$

48. $\frac{\sin(2x)}{x}$

49. $\frac{\sin(2x)}{3x}$

50. $\frac{\sin(x)}{3x}$

51. $\frac{x-\arctan(x)}{x^3}$

52. $\frac{x-\arctan(x)}{x^3+x}$

53. $\frac{x-\arctan(x)}{x^3-1}$

54. $\frac{x-\arctan(x)}{x^3+1}$

55. $\frac{5x^3+x^2+1}{7x^3+x+4}$

56. $\frac{x^3-3x}{x^2-4}\arctan\left(\frac{x}{4}\right)$

57. $\frac{x^3-2x}{x^2-4}\arctan\left(\frac{x}{4}\right)$

Exercises 58 to 64 concern even and odd functions.

58. If two functions are odd, what can be said about (a) their sum? (b) their product? (c) their quotient?

59. If two functions are even, what can be said about (a) their sum? (b) their product? (c) their quotient?

60. If f is odd and g is even, what can be said about (a) $f+g$? (b) fg? (c) f/g?

61. What, if anything, can you say about $f(0)$ if f is defined for all real numbers and f is (a) an even function? (b) an odd function?

62. Which polynomials are even? Explain.

63. Which polynomials are odd? Explain.

64. Is there a function that is both odd and even? Explain.

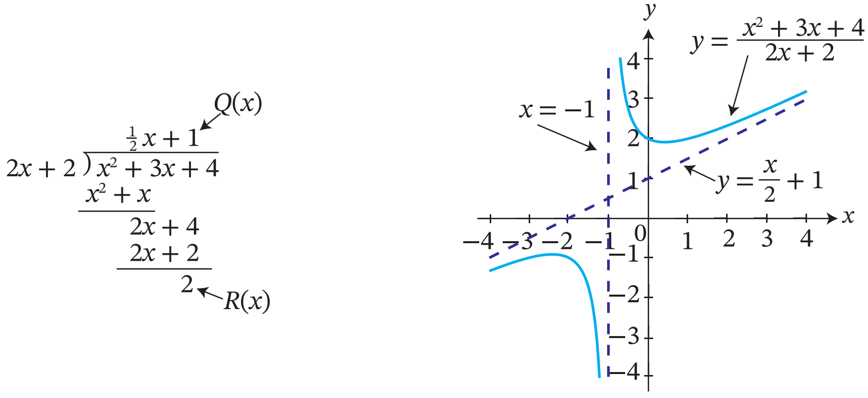
Exercises 65 to 68 concern **tilted asymptotes**. Let $A(x)$ and $B(x)$ be polynomials such that the degree of $A(x)$ is 1 more than the degree of $B(x)$. When you divide $B(x)$ into $A(x)$, you get a quotient $Q(x)$, which is a polynomial of degree 1, and a remainder $R(x)$, which is a polynomial of degree less than the degree of $B(x)$.

For example, Figure 2.6.11(a) shows the long division when $A(x) = x^2+3x+4$ and $B(x) = 2x+2$. Thus

$$x^2+3x+4 = \left(\frac{1}{2}x+1\right)(2x+2)+2$$

This tells us that

$$\frac{x^2+3x+4}{2x+2} = \frac{1}{2}x+1+\frac{2}{2x+2}.$$



(a) (b)

Figure 2.6.11

When x is large, $2/(2x+2) \to 0$. Thus the graph of $y = \dfrac{x^2+3x+4}{2x+2}$ is asymptotic to the line $y = \dfrac{1}{2}x+1$. (See Figure 2.6.11(b).)

Whenever the degree of $A(x)$ exceeds the degree of $B(x)$ by exactly 1, the graph of $y = A(x)/B(x)$ has a tilted asymptote. It can be found by dividing $B(x)$ into $A(x)$, obtaining a quotient $Q(x)$ and a remainder $R(x)$. Then

$$\frac{A(x)}{B(x)} = Q(x) + \frac{R(x)}{B(x)}.$$

The asymptote is $y = Q(x)$. In Exercises 65 through 68 graph the function, showing all asymptotes.

65. $y = \dfrac{x^2}{x-1}$ **66.** $y = \dfrac{x^3}{x^2-1}$ **67.** $y = \dfrac{x^2-4}{x+4}$ **68.** $y = \dfrac{x^2+x+1}{x-2}$

Read the directions for your graphing software to learn how to graph a piecewise-defined function. Then use your graphing utility to sketch graphs of the functions in Exercises 69 and 70.

69. $y = \begin{cases} x^2 - x & x < 1 \\ \sqrt{x-1} & x \geq 1 \end{cases}$

70. $y = \begin{cases} \frac{\sin(x)}{x} & x < 0 \\ \sin x & 0 \leq x \geq \pi \\ x-2 & x > \pi \end{cases}$

Some graphing utilities have trouble plotting functions with fractional exponents. General rules when graphing $y = x^{p/q}$ where p/q is a positive fraction in lowest terms are:

- If p is even and q is odd, then graph $y = |x|^{p/q}$.
- If p and q are both odd, then graph $y = \dfrac{|x|}{x}|x|^{p/q}$.

Use this advice and a calculator to sketch the graph of each function in Exercises 71 to 74.

71. $y = x^{1/3}$ **72.** $y = x^{2/3}$ **73.** $y = x^{4/7}$ **74.** $y = x^{3/7}$

---

75. Assume you already have drawn the graph of $y = f(x)$. Explain, in words, how to obtain the graph of $y = g(x)$ from the graph of $y = f(x)$ if

(a) $g(x) = f(x) + 2$?
(b) $g(x) = f(x) - 2$?
(c) $g(x) = f(x-2)$?
(d) $g(x) = f(x+2)$?
(e) $g(x) = 2f(x)$?
(f) $g(x) = 3f(x-2)$?
(g) $g(x) = 3f(2x)$?
(h) $g(x) = f(2x) + e^{-x^2}$?

76. Let $P(x)$ be a polynomial of degree m and $Q(x)$ a polynomial of degree n. For which m and n does the graph of $y = P(x)/Q(x)$ have a horizontal asymptote?

77. Is there a constant k such that $f(x) = \dfrac{1}{3^x - 1} + k$ (a) is odd? (b) is even?

78. Are there functions f defined for all x such that $f(-x) = \dfrac{1}{f(x)}$? If so, how many? If not, explain why not.

79. Are there functions f defined for all x such that $f(-x) = 2f(x)$? If so, how many? If not, explain why not.

2.S Chapter Summary

One concept underlies calculus: the limit of a function. For a function defined near a (but not necessarily at a) we ask, "What happens to $f(x)$ as x gets nearer and nearer to a?" If the values get nearer and nearer to a number, we call it the limit of the function as x approaches a. Limits, which do not appear in arithmetic, or algebra, or trigonometry, distinguish calculus.

When $f(x) = (2^x - 1)/x$, which is not defined at $x = 0$, we conjectured on the basis of numerical evidence that $f(x)$ approaches 0.693 (to three decimals). With that information we found that $(4^x - 1)/x$ approaches 2(0.693), which is larger than 1. We then defined e as that number (between 2 and 4) such that $(e^x - 1)/x$ approaches 1 as x approaches 0. The number e is as important in calculus as π is in geometry or trigonometry. Its value to three decimal places is about 2.718 and it is called **Euler's number**. Scientific calculators have a key for e^x because it is the most convenient exponential for calculus, as will become clear in the next chapter.

Historical Note: Three Important Bases for Logarithms

While logarithms can be defined for any positive base (other than 1), three numbers have been used most often: 2, 10, and e. Logarithms to the base 2 are used in information theory, for they record the number of "yes – no" questions needed to pinpoint a piece of information. Base 10 has been used for centuries to assist in computations. Since the decimal system is based on powers of 10, certain convenient numbers had obvious logarithms; for instance, $\log_{10}(1000) = \log_{10}(10^3) = 3$. Tables of logarithms to several decimal places facilitated the calculations of products, quotients, and roots. To multiply two numbers, you looked up their logarithms, and then searched the table for the number whose logarithm is the sum of the two logarithms.

The calculator made the tables obsolete, just as it sent the slide rule into museums. However, an online search for "slide rule". returns a list of more than 15 million websites full of history, instruction, and sentiment. The number e is the most convenient base for logarithms in calculus. As early as 1728, Euler used it for that purpose.

When angles are measured in radians,

$$\lim_{x\to 0} \frac{\sin(x)}{x} = 1 \qquad \text{and} \qquad \lim_{x\to 0} \frac{1-\cos(x)}{x} = 0.$$

These limits serve as the basis of the calculus of trigonometric functions developed in the next chapter. The simplicity of the first limit is one reason that in calculus and its applications angles are measured in radians. If angles were measured in degrees, the first limit would be $\pi/180$, which would complicate computations.

Most of the functions of interest in later chapters are continuous. The value of such a function at a number a in its domain is the same as the limit of the function as x approaches a.

A continuous function has three properties, which will be referred to often:

- On a closed interval a continuous function attains a maximum value and a minimum value.
- On a closed interval a continuous function takes on all values between its values at the end points of the interval.
- If a continuous function is positive at some number and defined at least on an open interval containing that number, then it remains positive at least on some open interval. More generally, if $f(a) = p > 0$, and q is less than p, then $f(x) \geq q$ at least on some open interval containing a.
  Similar statements holds when $f(a)$ is negative.

A quick sketch of the graph of a continuous function makes the three properties plausible.

EXERCISES for Section 2.S

1. Define Euler's constant, e, and give its decimal value to five places.

In Exercises 2 to 5 state the given property in your own words, using as few mathematical symbols as possible.

2. The maximum-value property.

3. The intermediate value property.

4. The permanence property.

5. The converse of the maximum-value property.

---

6. (a) Verify that $x^5 - y^5 = (x-y)(x^4 + x^3y + x^2y^2 + xy^3 + y^4)$. (b) Use (a) to find $\lim_{x\to a} \frac{x^5 - a^5}{x-a}$.

7. If $f(x) = \frac{1}{x+2}$ for x not equal to -2, is there a continuous function $g(x)$, defined for all x, that equals $f(x)$ when x is not -2? Explain your answer.

8. If $f(x) = \frac{2^x - 1}{x}$ for x not equal to 0, is there a continuous function $g(x)$, defined for all x, that equals $f(x)$ when x is not 0? Explain your answer.

9. If $f(x) = \sin(1/(x-1))$ for x not equal to 1, is there a continuous function $g(x)$, defined for all x, that equals $f(x)$ when x is not 1? Explain your answer.

10. If $f(x) = x\sin(1/x)$ for x not equal to 0, is there a continuous function $g(x)$, defined for all x, that equals $f(x)$ when x is not 0? Explain your answer.

11. Show $\lim_{x\to 1} \frac{x^{1/3} - 1}{x-1} = \frac{1}{3}$ by writing the denominator as the difference of two cubes, $x - 1 = \left(x^{1/3}\right)^3 - 1$, and using the factorization $u^3 - 1 = (u-1)(u^2 + u + 1)$.

12. Use the factorization in Exercise 6 to find $\lim_{x\to a} \frac{x^{-5} - a^{-5}}{x-a}$.

13. Assume $b > 1$. If $\lim_{x\to 0} \frac{b^x - 1}{x} = L$, find $\lim_{x\to 0} \frac{(1/b)^x - 1}{x}$.

14. By sketching a graph, show that if a function is not continuous it may not

(a) have a maximum even if its domain is a closed interval,

(b) satisfy the intermediate value theorem even if its domain is a closed interval,

(c) have the permanence property even if its domain is an open interval.

15. Let g be an increasing function such that $\lim_{x\to a} g(x) = L$.

(a) Sketch the graph of a function f whose domain includes an open interval around L such that $f\left(\lim\limits_{x\to a} g(x)\right)$ and $\lim\limits_{x\to a} f(g(x))$ both exist but are not equal.

(b) What property of f would assure us that the two limits in (a) would be equal?

The limit $\lim\limits_{x\to a} \dfrac{x^n - a^n}{x-a}$ was evaluated by using the factorization of $x^n - a^n$. If $h = x - a$, the limit can be written as $\lim\limits_{h\to 0} \dfrac{(a+h)^n - a^n}{h}$. Exercises 16 and 17 show the algebra needed to evaluate this new limit.

16. (a) Show that $(a+h)^2 = a^2 + 2ah + h^2$. (b) Use (a) to evaluate $\lim\limits_{h\to 0} \dfrac{(a+h)^2 - a^2}{h}$.

17. (a) Show that $(a+h)^3 = a^3 + 3a^2h + 3ah^2 + h^3$. (b) Use (a) to evaluate $\lim\limits_{h\to 0} \dfrac{(a+h)^3 - a^3}{h}$.

---

18. For the positive integer k the symbol $k!$, called "k factorial," is the product of all the integers from 1 through k. The binomial theorem states that for any positive integer n

$$(a+h)^n = \binom{n}{0} a^0 h^n + \binom{n}{1} a^1 h^{n-1} + \binom{n}{2} a^2 h^{n-2} + \cdots + \binom{n}{n-2} a^{n-2} h^2 + \binom{n}{n-1} a^{n-1} h^1 + \binom{n}{n} a^n h^0$$

where $\dbinom{n}{m} = \dfrac{n!}{m!(n-m)!}$ for $m = 0, 1, 2, \ldots, n$. Show that for any positive integer n, $\lim\limits_{h\to 0} \dfrac{(a+h)^n - a^n}{h} = na^{n-1}$.

In Exercises 19 to 22 find each limit.

19. $\lim\limits_{x\to\infty} \dfrac{\ln(5x)}{\ln(4x^2)}$

20. $\lim\limits_{x\to\infty} \dfrac{\ln(5x)}{\ln(4x)}$

21. $\lim\limits_{x\to\infty} \dfrac{\log_2(x^2)}{\log_4(x)}$

22. $\lim\limits_{x\to\infty} \dfrac{\log_3(x^5)}{\log_9(x)}$

---

23. Find $\lim\limits_{h\to 0} \dfrac{(e^2)^h - 1}{h}$ by factoring the numerator.

24. Assuming that $\lim\limits_{x\to 0^+} x^x = 1$ and that $\lim\limits_{x\to\infty} \ln(x) = \infty$, find:

(a) $\lim\limits_{x\to 0^+} x\ln(x)$, (b) $\lim\limits_{x\to\infty} \dfrac{\ln(x)}{x}$, (c) $\lim\limits_{x\to\infty} x^{1/x}$, (d) $\lim\limits_{x\to\infty} \dfrac{\ln(x)}{x^k}$, k a positive constant,

(e) $\lim\limits_{x\to\infty} \dfrac{x}{e^x}$, (f) $\lim\limits_{x\to\infty} \dfrac{x^n}{e^x}$, n a positive integer, and (g) $\lim\limits_{x\to\infty} \dfrac{(\ln(x))^n}{x}$, n a positive integer.

---

25. Let $2000^x = 2001^{2000}$.

(a) Without doing any calculations, estimate x.

(b) Use logarithms to solve for x.

26. Now, suppose $x^{2001} = 2001^{2000}$.

(a) Without doing any calculations, estimate x.

(b) Use logarithms to solve for x.

---

27. Let $f = g + h$, where g is an even function and f is an odd function. Express g and h in terms of f.

28. If f is an odd function and g is an even function, what, if anything, can be said about

(a) fg, (b) f^2, (c) $f+g$, (d) $f+f$, and (e) $\dfrac{f}{g}$. Explain.

Exercises 29 to 31 provide an early glimpse of an area problem that will be addressed with calculus. They require only a basic understanding of geometric series and limits.

29. Consider the region R bounded by the graphs of $y = 2^{-x}$, $y = 0$, $x = 0$, and $x = 1$. As shown in Figure 2.S.1(a),

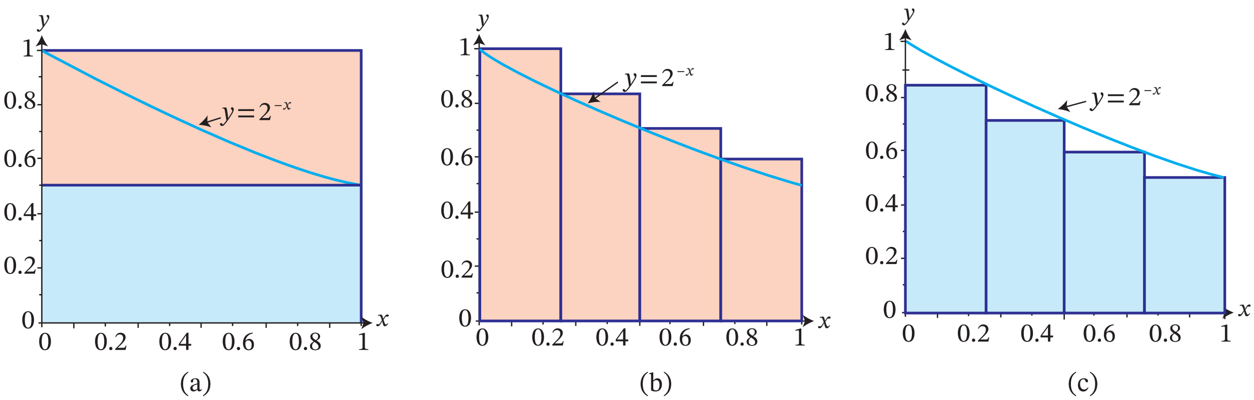


Figure 2.S.1

the region R lies inside a square of area 1 and contains a rectangle of area $1/2$. So its area is between $1/2$ and 1.

(a) Use the rectangles in Figure 2.S.1(b) to estimate the area of R.
(b) Use the rectangles in Figure 2.S.1(c) to estimate the area of R.
(c) Use the answers in (a) and (b) to find an interval that contains the area of R.

30. Instead of the four rectangles used in the preceding exercise use 100 rectangles, all of the same width, to get even closer estimates of the area of R.

31. Let R be the region under $y = 2^{-x}$, to the right of the y-axis, and above the x-axis. Decide if the area of R is finite or infinite.

---

32. **SAM:** Did you know that the AB limit comes right out of the $A+B$ limit, at least for positive functions?
JANE: I don't believe it.
SAM: All I do is write $g(x)$ as $e^{\ln(g(x))}$ and $h(x)$ as $e^{\ln(h(x))}$.
JANE: I see. But you also use the continuity of the exponential and natural logarithm functions.
SAM: So I do. Even so it's neat.

Check that Sam is right.

33. Using an approach like the one in Exercise 32, obtain the A^B limit from the AB limit for two positive functions.

34. The graph of a function f whose domain is $[2,4]$ and range is $[1,3]$ is shown in Figure 2.S.2(a). Sketch the graphs of the following functions and state their domains and ranges.

(a) $g(x) = -3f(x)$
(b) $g(x) = f(x+1)$
(c) $g(x) = f(x-1)$
(d) $g(x) = 3 + f(x)$
(e) $g(x) = f(2x)$
(f) $g(x) = f(x/2)$
(g) $g(x) = f(2x-1)$

35. For a constant k, find $\displaystyle\lim_{h\to 0} \frac{(e^k)^h - 1}{h}$.

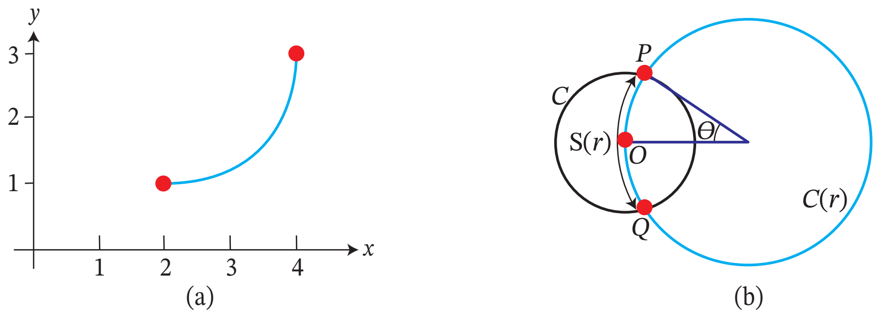


Figure 2.S.2

36. This exercise obtains $\lim_{\theta\to 0}\frac{\sin(\theta)}{\theta}$ without using areas. Figure 2.S.2(b) shows a circle C of radius 1 centered at the origin and a circle $C(r)$ of radius $r > 1$ that passes through the center of C. Let $S(r)$ be the part of $C(r)$ that lies within C. The ends of this curve are P and Q. Let θ be the angle subtended by the top half of $S(r)$ at the center of $C(r)$. As $r \to \infty$, $\theta \to 0$. Define $A(\theta)$ to be the length of the arc $S(r)$ as a function of θ. CONTRIBUTED BY: G. D. Chakerian

(a) Looking at Figure 2.S.2(b), determine $\lim_{\theta\to 0} A(\theta)$.

(b) Show that $A(\theta)$ is $\frac{\theta}{\sin(\theta/2)}$.

(c) Combining (a) and (b), show that $\lim_{\theta\to 0}\frac{\sin(\theta)}{\theta} = 1$.

37. Assume you have a table of logarithms to the base 10. Explain how to use the table to compute

(a) the square root of a number , (b) the cube of a number, and (c) the quotient of two numbers.

38. It is exactly 4:00pm. The hour hand points to the 20 minutes and the minute hand points points to 0 minutes. An hour later the hour hand points to 25 minutes and the minute hand again points to 0 minutes (or 60 minutes, if you wish).

(a) Explain why both hands point in the same direction exactly once between 4:00pm and 5:00pm.

(b) Show that at any time t between 4:00pm and 5:00pm, the minute hand points to t minutes and the hour hand points to $20 + \frac{t}{12}$ minutes.

(c) Graph the positions of the hour and minute hands introduced in (b) on the same axes.

(d) Deduce that the two hands point in the same direction when $t = \frac{240}{11}$ minutes. At what time does this occur?

Calculus is Everywhere # 3

Bank Interest and the Annual Percentage Yield

The Truth in Savings Act, passed in 1991, requires a bank to post the Annual Percentage Yield (APY) on deposits. It depends on the interest rate and on how often the bank computes the interest earned. Imagine that you open an account on January 1 by depositing \$1000. The bank pays interest monthly at the rate of 5 percent a year. How much will there be in your account at the end of the year? For simplicity, assume all the months have the same length.

To begin, we find out how much there is in the account at the end of the first month. The account then has the initial amount, \$1000, plus the interest earned during January. Because there are 12 months, the interest rate in each month is 5 percent divided by 12, which is 0.05/12 per month. So the interest earned in January is \$1000 times 0.05/12. At the end of January the account has

$$\$1000 + \$1000\left(\frac{0.05}{12}\right) = \$1000\left(1 + \frac{0.05}{12}\right).$$

The initial deposit is magnified by the factor $1 + 0.05/12 \approx 1.00417$.

The amount in the account at the end of February is found the same way, but the initial amount is $\$1000(1 + 0.05/12)$ instead of \$1000. Again the amount is magnified by the factor $1 + 0.05/12$ to become

$$\$1000\left(1 + \frac{0.05}{12}\right)^2.$$

The amount at the end of March is

$$\$1000\left(1 + \frac{0.05}{12}\right)^3$$

and at the end of the year the account has grown to

$$\$1000\left(1 + \frac{0.05}{12}\right)^{12},$$

which is about \$1051.16.

The deposit earned \$51.16. If instead the bank computed the interest only once, at the end of the year, the deposit would earn only 5 percent of \$1000, which is \$50. The depositor benefits when the interest is computed more than once a year, so-called compound interest.

A competing bank may offer to compute the interest every day. In that case, the account would grow to

$$\$1000\left(1 + \frac{0.05}{365}\right)^{365},$$

which is about \$1051.27, eleven cents more than the first bank offers. More generally, if the initial deposit is A, the annual interest rate is r, and interest is computed n times a year, the amount at the end of the year is

$$A\left(1 + \frac{r}{n}\right)^n. \tag{C.3.1}$$

In the examples, A is \$1000, r is 0.05, and n is 12 and then 365.

Suppose that A is 1 and r is a generous 100 percent, that is, $r = 1$, and (C.3.1) becomes

$$\left(1 + \frac{1}{n}\right)^n. \tag{C.3.2}$$

How does (C.3.2) behave as n increases?

n	$\left(1+\frac{1}{n}\right)^n$	$\left(1+\frac{1}{n}\right)^n$
1	$\left(1+\frac{1}{1}\right)^1$	2.00000
2	$\left(1+\frac{1}{2}\right)^2$	2.25000
3	$\left(1+\frac{1}{3}\right)^3$	2.37037
10	$\left(1+\frac{1}{10}\right)^{10}$	2.59374
100	$\left(1+\frac{1}{100}\right)^{100}$	2.70481
1000	$\left(1+\frac{1}{1000}\right)^{1000}$	2.71692

Table C.3.1

Table C.3.1 shows a few values of (C.3.2) to five decimal places.

The base, $1+1/n$, approaches 1 as n increases, suggesting that (C.3.2) may approach a number near 1. However, the exponent gets large, so we are multiplying many numbers, all larger than 1. It turns out that as n increases $(1+1/n)^n$ approaches the number e defined in Section 2.2. We can write

$$\lim_{x\to 0^+}(1+x)^{1/x}=e. \tag{C.3.3}$$

Knowing this we can figure out what happens when an account opens with \$1000, the annual interest rate is 5 percent, and the interest is compounded so often that we say it is compounded continuously. In that case we would be interested in

$$1000\lim_{n\to\infty}(1+0.05/n)^n.$$

We can use (C.3.3) to evaluate this expression by noting that $n=(n/0.05)(0.05)$, which gives

$$\left(1+\frac{0.05}{n}\right)^n=\left(\left(1+\frac{0.05}{n}\right)^{\frac{n}{0.05}}\right)^{0.05}. \tag{C.3.4}$$

The expression in the largest parentheses has the form $(1+x)^{1/x}$. Therefore, as n increases, $(1+0.05/n)^{n/0.05}$ approaches e and (C.3.3) approaches $e^{0.05}\approx 1.05127$. No matter how often interest is compounded, the \$1000 would never grow beyond \$1051.27.

The definition of e given in Section 2.2 has no obvious connection to the fact that $\lim_{x\to 0^+}(1+x)^{1/x}$ equals the number e. It seems obvious, by thinking in terms of bank accounts, that as n increases, so does $(1+1/n)^n$. However, as n increases, the base decreases and the exponent increases, producing two competing influences. Without thinking about bank accounts, try showing that $(1+1/n)^n$ does increase.

This is confirmed in Exercise 86 in Section 5.6.

EXERCISES for CIE C.3

1. A dollar is deposited at the beginning of the year in an account that pays an interest rate of 100% a year. Let $f(t)$, for $0 \le t \le 1$, be the amount in the account at time t. Graph the function if the bank pays

(a) only simple interest, computed only at $t = 1$.

(b) compound interest, twice a year computed at $t = \frac{1}{2}$ and 1.

(c) compound interest, three times a year computed at $t = \frac{1}{3}, \frac{2}{3}$, and 1.

(d) compound interest, four times a year computed at $t = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$, and 1.

(e) Are the functions in (a), (b), (c), and (d) continuous?

(f) One could expect the account that is compounded more often than another would always have at least as much throughout the year as the one computed less often. Is that the case in the examples in this problem? (For instance, compare the graphs in (b) and (c) for $0 \le t \le 1$.)

Chapter 3

The Derivative

The two main applied concepts of calculus are defined with the aid of limits. In this chapter we meet the first of these: the derivative of a function. The derivative tells how rapidly or slowly a function changes. For instance, if the function describes the position of a moving object, the derivative tells us its velocity.

The definition of a derivative rests on the notion of a limit. The limits examined in Chapter 3 are the basis for finding the derivatives of all functions of interest.

The goals of this chapter are twofold: to develop those techniques and to explore the meaning of a derivative.

3.1 Velocity and Slope: Two Problems with One Theme

This section discusses two problems that at first glance may seem unrelated. The first concerns the slope of a tangent line to a curve. The second involves velocity. A little arithmetic will show that they are both different versions of one mathematical idea: the derivative.

Slope

Our first problem is important because it is related to finding the straight line that most closely resembles a given graph near a point on the graph.

EXAMPLE 1. What is the slope of the tangent line to the graph of $y = x^2$ at the point $(2, 4)$? (See Figure 3.1.1(a).)

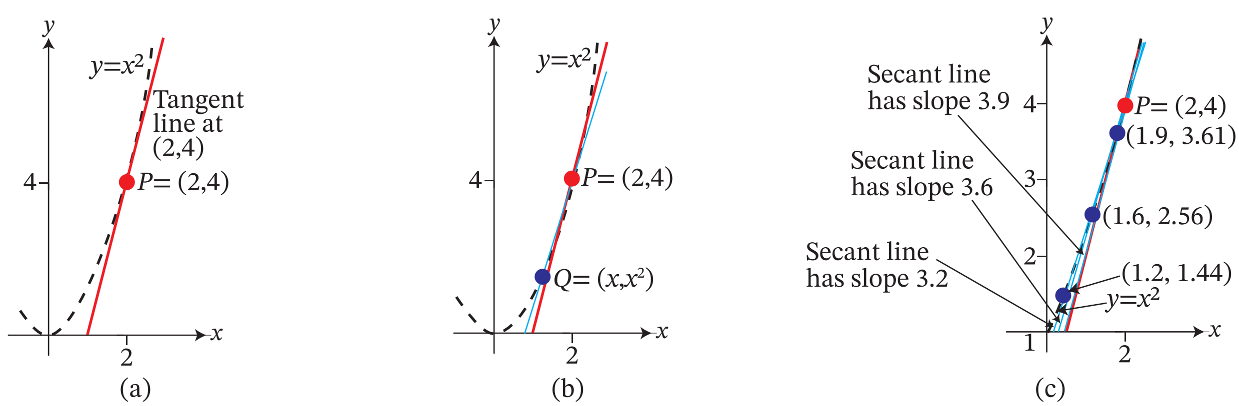


Figure 3.1.1

SOLUTION In Section 2.1 we used a point Q on the curve near $P = (2,4)$ to determine a line that closely resembles the tangent line at $(2,4)$. (See Figure 3.1.1(b).) Using $Q = (2.01, 2.01^2)$ and also $Q = (1.99, 1.99^2)$, we found that the slope of the tangent line is between 4.01 and 3.99. But, we did not find the slope of the tangent at $(2,4)$. Rather than making more estimates by choosing points nearer $(2,4)$, such as $Q = (2.00001, 2.00001^2)$, it is simpler to consider a general point Q — and then look at the limit of the slopes of secant lines through points $P = (2,4)$ and $Q = (x, x^2)$ as Q approaches P, that is, as x approaches 2. Figure 3.1.1(c) shows visually how secant lines become better approximate the tangent line as the (blue) points $Q = (x, x^2)$ move closer to the (red) point $P = (2,4)$.

To find the slope of the secant line through $P = (2,4)$ and $Q = (x, x^2)$ when x is close to 2, but not equal to 2. It has slope

$$\frac{x^2 - 2^2}{x - 2} = \frac{(x+2)(x-2)}{x-2} = x + 2.$$

To find out what happens to the quotient as Q moves closer to P (and x moves closer to 2) apply the techniques of limits developed in Chapter 2. We have

$$\lim_{x\to 2} \frac{x^2 - 2^2}{x - 2} = \lim_{x\to 2} \frac{(x+2)(x-2)}{x-2} = \lim_{x\to 2}(x+2) = 4.$$

Thus, we expect the tangent line to $y = x^2$ at $(2,4)$ to have slope 4. ◆

We never had to make any estimates with specific choices of the nearby point Q. We did not even have to draw the curve.

Velocity

If an airplane or automobile is moving at a constant velocity, we know that distance traveled equals velocity times time. Thus

$$\text{velocity} = \frac{\text{distance traveled}}{\text{elapsed time}}.$$

If the velocity is not constant, we still may speak of its average velocity, which is defined as

$$\text{average velocity} = \frac{\text{distance traveled}}{\text{elapsed time}}.$$

If a trip from San Francisco to Los Angeles, a distance of 400 miles, takes 8 hours, the average velocity is 400/8 or 50 miles per hour.

Suppose that up to time t_1 you have traveled a distance D_1, while up to time t_2 you have traveled a distance D_2, where $t_2 > t_1$. Then during the time interval $[t_1, t_2]$ the distance traveled is $D_2 - D_1$. Thus the average velocity during the time interval $[t_1, t_2]$, which has duration $t_2 - t_1$, is

$$\text{average velocity} = \frac{D_2 - D_1}{t_2 - t_1}.$$

The arithmetic of average velocity is the same as that for the slope of a line.

The next problem shows how to find the velocity at any instant for an object whose velocity is not constant.

EXAMPLE 2. A rock initially at rest falls $16t^2$ feet in t seconds. What is its velocity after 2 seconds? Whatever it turns out to be, it will be called the **instantaneous velocity**.

SOLUTION To start, make an estimate by finding the average velocity of the rock during a short time interval, say from 2 to 2.01 seconds. At the start of this interval the rock has fallen $16(2^2) = 64$ feet. By its end it has fallen $16(2.01^2) = 16(4.0401) = 64.6416$ feet. So, during 0.01 seconds the rock fell 0.6416 feet. Its average velocity is

$$\text{average velocity} = \frac{64.6416 - 64}{2.01 - 2} = \frac{0.6416}{0.01} = 64.16 \text{ feet per second.}$$

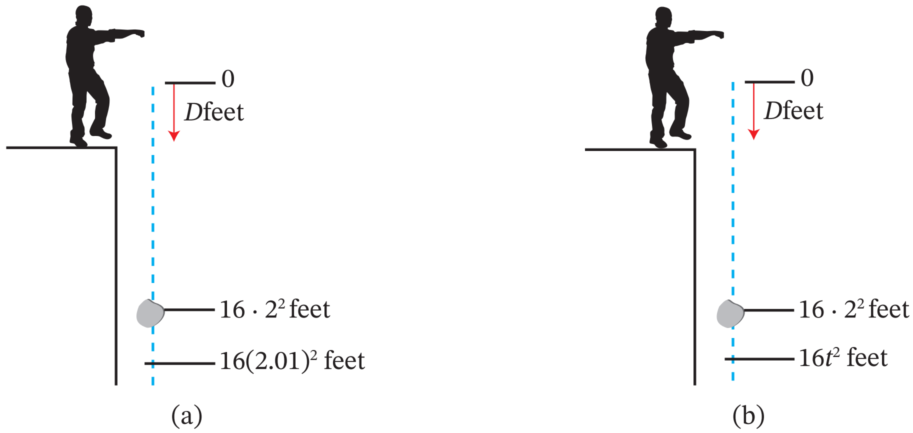


Figure 3.1.2

This is an estimate of the velocity at time $t = 2$ seconds. (See Figure 3.1.2(a).)

Rather than make another estimate with the aid of a shorter interval of time, let us consider the typical time interval from 2 to t seconds, $t > 2$. (Although we will keep $t > 2$, estimates could just as well be made with $t < 2$.) During $t - 2$ seconds the rock travels $16(t^2) - 16(2^2) = 16(t^2 - 2^2)$ feet, as shown in Figure 3.1.2(b). The average velocity of the rock is

$$\text{average velocity} = \frac{16t^2 - 16(2^2)}{t-2} = \frac{16(t^2 - 2^2)}{t-2} \text{ feet per second.}$$

When t is close to 2, what happens to the average velocity? It approaches

$$\lim_{t\to 2} \frac{16(t^2 - 2^2)}{t-2} = 16 \lim_{t\to 2} \frac{t^2 - 2^2}{t-2} = 16 \lim_{t\to 2}(t+2) = 16 \cdot 4 = 64 \text{ feet per second.}$$

We say that the (instantaneous) velocity at time $t = 2$ is 64 feet per second. ♦

Even though Examples 1 and 2 seem unrelated, their solutions turn out to be practically identical: the slope in Example 1 is approximated by the quotient

$$\frac{x^2 - 2^2}{x - 2}$$

and the velocity in Example 2 is approximated by the quotient

$$\frac{16t^2 - 16(2^2)}{t-2} = 16 \cdot \frac{t^2 - 2^2}{t-2}.$$

The only difference between the solutions is that the second quotient has a factor of 16 and x is replaced with t. This may not be too surprising, since the functions involved, x^2 and $16t^2$ differ by a factor of 16. (That the independent variable is named t in one case and x in the other does not affect the computations.)

The Derivative of a Function

In both the slope and velocity problems we were lead to studying similar limits. For the function x^2 it was

$$\frac{x^2 - 2^2}{x - 2} \quad \text{as } x \text{ approaches 2.}$$

For the function $16t^2$ it was

$$\frac{16t^2 - 16(2^2)}{t-2} \quad \text{as } t \text{ approaches 2.}$$

In both cases we formed the change in outputs divided by change in inputs and then found the limit of this quotient as the change in inputs became smaller and smaller. This can be done for other functions, and brings us to one of the two key ideas in calculus, the derivative of a function.

Definition: *Derivative of a Function at a Number a*

Let f be a function that is defined in an open interval that contains the number a. If

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

exists, it is called the **derivative of f at a**, and is denoted $f'(a)$. In this case the function f is said to be **differentiable at a**.

EXAMPLE 3. Find the derivative of $f(x) = 16x^2$ at $x = 2$.

SOLUTION In this case, $f(x) = 16x^2$ for any input x. By definition, its derivative at $x = 2$ is

$$\lim_{x \to 2} \frac{f(x) - f(2)}{x - 2} = \lim_{x \to 2} \frac{16x^2 - 16(2^2)}{x - 2} = 16 \lim_{x \to 2} \frac{x^2 - 2^2}{x - 2} = 16 \lim_{x \to 2}(x + 2) = 64.$$

We say that "the derivative of the function $f(x)$ at 2 is 64" and write $f'(2) = 64$. ♦

Now that we have the derivative of f, we can define the slope of its graph at a point $(a, f(a))$ as the value of the derivative, $f'(a)$. Then we define the **tangent line** at $(a, f(a))$ as the line through $(a, f(a))$ whose slope is $f'(a)$.

Read $f'(a)$ as "f prime at a" or "the derivative of f at a."

EXAMPLE 4. Find the derivative of e^x at a.

SOLUTION We must find

$$\lim_{x \to a} \frac{e^x - e^a}{x - a}. \tag{3.1.1}$$

The limit is not obvious. Let us write x as $a + h$ and see what happens as h approaches 0. The denominator $x - a$ is just h. Then (3.1.1) now reads

$$\lim_{h \to 0} \frac{e^{a+h} - e^a}{h}.$$

This form of the limit is more convenient:

$$\begin{aligned}
\lim_{h \to 0} \frac{e^{a+h} - e^a}{h} &= \lim_{h \to 0} \frac{e^a e^h - e^a}{h} && (\text{ law of exponents }) \\
&= e^a \lim_{h \to 0} \frac{e^h - 1}{h} && (\text{ factor out a constant }) \\
&= e^a \cdot 1 && (\text{ Section 2.2 }) \\
&= e^a.
\end{aligned}$$

So the limit is e^a. That is, the derivative of e^x is e^x itself. ♦

Differentiability and Continuity

Definition: *Differentiable Function*

If a function is differentiable at each point in its domain the function is said to be **differentiable**.

A small piece of the graph of a differentiable function at a looks like part of a straight line. You can check this by zooming in on the graph of a function of your choice. Differential calculus can be described as the study of functions whose graphs locally look almost like a line.

It comes as no surprise that a differentiable function is always continuous.

To show that a function is continuous at an argument a in its domain we must show that $\lim_{x\to a} f(x)$ equals $f(a)$, which amounts to showing $\lim_{x\to a}(f(x)-f(a))$ equals 0. To relate this limit to $f'(a)$ we rewrite it:

$$\begin{aligned}
\lim_{x\to a}(f(x)-f(a)) &= \lim_{x\to a}\frac{f(x)-f(a)}{x-a}(x-a) && (\text{ multiply and divide by } x-a \text{ (nonzero) })\\
&= \lim_{x\to a}\left(\frac{f(x)-f(a)}{x-a}\right)\lim_{x\to a}(x-a) && (\text{ product property of limits })\\
&= f'(a)\cdot 0 && (\text{ definition of the derivative })\\
&= 0.
\end{aligned}$$

So, $\lim_{x\to a} f(x) = f(a)$, which means f is continuous at a.

A function can be continuous yet not differentiable. For instance, $f(x) = |x|$ is continuous but not differentiable at 0, as Figure 3.1.3 suggests.

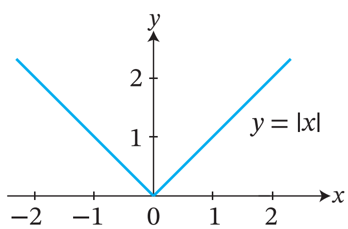


Figure 3.1.3

Summary

From a mathematical point of view, the problems of finding the slope of the tangent line and the velocity of the rock are the same. In each case estimates lead to the same type of quotient, $\left(f(x)-f(a)\right)/(x-a)$. The behavior of this difference quotient is studied as x approaches a. In each case the answer is a limit, called the derivative of the function at the given number, a. Finding the derivative of a function is called "differentiating" the function.

EXERCISES for Section 3.1

1. Let g be a function and b a number. Define the "derivative of g at b".

2. How is the tangent line to the graph of f at $(a, f(a))$ defined?

3. (a) Find the slope of the tangent line to $y = x^2$ at $(4,16)$. (b) Draw the tangent line to the curve at $(4,16)$.

4. (a) Find the slope of the tangent line to $y = x^2$ at $(-1,1)$. (b) Draw the tangent line to the curve at $(-1,1)$.

Exercise 5 to 14 concern slope. Use the technique of Example 1 to find the slope of the tangent line to the graph of the given curve at the specified point.

5. $y = x^2$ at the point $(3,3^2) = (3,9)$

6. $y = x^2$ at $\left(\frac{1}{2},\left(\frac{1}{2}\right)^2\right) = \left(\frac{1}{2},\frac{1}{4}\right)$

7. $y = x^3$ at $(2,2^3) = (2,8)$

8. $y = x^3$ at $(-2,(-2)^3) = (-2,-8)$

9. $y = \sin(x)$ at $(0,\sin(0)) = (0,0)$

10. $y = \cos(x)$ at $(0,\cos(0)) = (0,1)$

11. $y = \cos(x)$ at $\left(\frac{\pi}{4},\cos\left(\frac{\pi}{4}\right)\right) = \left(\frac{\pi}{4},\frac{\sqrt{2}}{2}\right)$

12. $y = \sin(x)$ at $\left(\frac{\pi}{6},\sin\left(\frac{\pi}{6}\right)\right) = \left(\frac{\pi}{6},\frac{1}{2}\right)$

13. $y = 2^x$ at $(1,2^1) = (1,2)$

14. $y = 4^x$ at $\left(\frac{1}{2},4^{1/2}\right) = \left(\frac{1}{2},2\right)$

---

15. (a) Graph $y = \frac{1}{x}$ and, by eye, draw the tangent at $\left(2, \frac{1}{2}\right)$.
(b) Using a ruler, measure a rise-run triangle to estimate the slope of the tangent line drawn in (a).
(c) Using no pictures at all, find the slope of the tangent line to the curve $y = \frac{1}{x}$ at $\left(2, \frac{1}{2}\right)$.

16. (a) Sketch the graph of $y = x^3$ and the tangent line at $(0,0)$.
(b) Find the slope of the tangent line to the curve $y = x^3$ at the point $(0,0)$.

17. (a) Sketch the graph of $y = x^2$ and the tangent line at $(1,1)$.
(b) Find the slope of the tangent line to the curve $y = x^2$ at the point $(1,1)$.

In Exercises 18 to 21 use the method of Example 2 to find the velocity of the rock after

18. 3 seconds **19.** 1 second **20.** $\frac{1}{2}$ second **21.** $\frac{1}{4}$ second

---

22. An object travels t^3 feet in the first t seconds.
(a) How far does it travel during the time interval from 2 to 2.1 seconds?
(b) What is the average velocity during that time interval?
(c) Let h be any positive number. Find the average velocity of the object from time 2 to $2+h$ seconds.
(d) Find the velocity of the object at 2 seconds by letting h approach 0 in the result found in (c).

23. An object travels t^3 feet in the first t seconds.
(a) Find the average velocity during the time interval from 3 to 3.01 seconds.
(b) Find its average velocity during the time interval from 3 to t seconds, $t > 3$.
(c) By letting t approach 3 in the result found in (b), find the velocity of the object at 3 seconds.

Exercises 24 and 25 illustrate a different notation to find the slope of the tangent line: $f'(a) = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h}$.

24. For the parabola $y = x^2$:
(a) Find the slope of the line through $P = (2,4)$ and $Q = (2+h, (2+h)^2)$, where $h \neq 0$.
(b) Show that as h approaches 0, the slope in (a) approaches 4.

25. For the curve $y = x^3$.
(a) Find the slope of the line through $P = (2,8)$ and $Q = \left(1.9, 1.9^3\right)$.
(b) Find the slope of the line through $P = (2,8)$ and $Q = \left(2.01, 2.01^3\right)$.
(c) Find the slope of the line through $P = (2,8)$ and $Q = \left(2+h, (2+h)^3\right)$, where $h \neq 0$.
(d) Show that as h approaches 0, the slope in (a) approaches 12.

---

26. For the curve $y = \sin(x)$.
(a) Find the slope of the line through $P = (0,0)$ and $Q = (-0.1, \sin(-0.1))$.
(b) Find the slope of the line through $P = (0,0)$ and $Q = (0.01, \sin(0.01))$.
(c) Find the slope of the line through $P = (0,0)$ and $Q = (h, \sin(h))$, where $h \neq 0$.
(d) Show that as h approaches 0, the slope in (c) approaches 1.
(e) Use (d) to draw the tangent line to $y = \sin(x)$ at $(0,0)$.

27. For the curve $y = \cos(x)$.
(a) Find the slope of the line through $P = (0,1)$ and $Q = (-0.1, \cos(-0.1))$.
(b) Find the slope of the line through $P = (0,1)$ and $Q = (0.01, \cos(0.01))$.

(c) Find the slope of the line through $P = (0,1)$ and $Q = (h, \cos(h))$, where $h \neq 0$.
(d) Show that as h approaches 0, the slope in (c) approaches 0.
(e) Use (d) to draw the tangent line to $y = \cos(x)$ at $(0,1)$.

28. For the curve $y = 2^x$.
(a) Find the slope of the line through $P = (2, 2^2)$ and $Q = \left(1.9, 2^{1.9}\right)$.
(b) Find the slope of the line through $P = (2, 2^2)$ and $Q = \left(2.1, 2^{2.1}\right)$.
(c) Find the slope of the line through $P = (2, 2^2)$ and $Q = \left(2+h, 2^{2+h}\right)$, where $h \neq 0$.
(d) Show that the slope of the curve $y = 2^x$ at $\left(2, 2^2\right)$ is approximately $4(0.69315) = 2.7726$.
(e) Use (d) to draw the tangent line to the curve at $(2,4)$.

29. For the curve $y = e^x$.
(a) Find the slope of the line through $P = \left(-0.5, e^{-0.5}\right)$ and $Q = \left(-0.6, e^{-0.6}\right)$.
(b) Find the slope of the line through $P = \left(-0.5, e^{-0.5}\right)$ and $Q = \left(-0.49, e^{-0.49}\right)$.
(c) Find the slope of the line through $P = \left(-0.5, e^{-0.5}\right)$ and $Q = \left(-0.5+h, e^{-0.5+h}\right)$, where $h \neq 0$.
(d) Show that as h approaches 0, the slope in (c) approaches $e^{-0.5}$.

30. Show that the slope of the curve $y = 2^x$ at $(3,8)$ is approximately $8(0.69315) = 5.5452$.

31. (a) Use the method of this section to find the slope of the curve $y = x^3$ at $(1,1)$.
(b) What does the graph of $y = x^3$ look like near $(1,1)$?

32. (a) Use the method of this section to find the slope of the curve $y = x^3$ at $(-1,-1)$.
(b) What does the graph of $y = x^3$ look like near $(-1,-1)$?

33. (a) Draw the curve $y = e^x$ for x in the interval $[-2,1]$.
(b) Using a straightedge, draw the tangent line at $(1,e)$ as well as you can.
(c) Estimate the slope of the tangent line by measuring its rise and run.
(d) Using the derivative of e^x, find the slope of the curve at $(1,e)$.

34. (a) Sketch the curve $y = e^x$ for x in $[-1,1]$.
(b) Where does the curve cross the y-axis?
(c) What is the (smaller) angle between the graph of $y = e^x$ and the y-axis at the point found in (b)?

The phrase "slope of the graph of $y = f(x)$" is often shortened to "slope of $y = f(x)$," as in Exercises 35 and 36.

35. With the aid of a calculator, estimate the slope of $y = 2^x$ at $x = 1$, using the intervals
(a) $[1,1.1]$, (b) $[1,1.01]$, (c) $[0.9,1]$, and (d) $[0.99,1]$.

36. With the aid of a calculator, estimate the slope of $y = \dfrac{x+1}{x+2}$ at $x = 2$, using the intervals
(a) $[2,2.1]$, (b) $[2,2.01]$, (c) $[2,2.001]$, and (d) $[1.999,2]$.

---

37. Estimate the derivative of $\sin(x)$ at $x = \dfrac{\pi}{3}$ (a) to two decimal places and (b) to three decimal places.

38. Estimate the derivative of $\ln(x)$ at $x = 2$ (a) to two decimal places and (b) to three decimal places.

The ideas common to both slope and velocity also appear in other applications. Exercises 39 to 43 present the same ideas in biology, economics, and physics.

39. A bacterial culture has a mass of t^2 grams after t minutes of growth.
(a) How much does it grow during the time interval $[2,2.01]$?
(b) What is the average rate of growth during the time interval $[2,2.01]$?
(c) What is the instantaneous rate of growth when $t = 2$?

40. A thriving business has a profit of t^2 million dollars in its first t years. Thus from time $t = 3$ to time $t = 3.5$ (the first half of its fourth year) it has a profit of $(3.5)^2 - 3^2$ million dollars, giving an annual rate of

$$\frac{(3.5)^2 - 3^2}{0.5} = 6.5 \text{ million dollars per year.}$$

(a) What is its annual rate of profit during the time interval $[3, 3.1]$?
(b) What is its annual rate of profit during the time interval $[3, 3.01]$?
(c) What is its instantaneous rate of profit after 3 years?

The mass of the leftmost x centimeters of the nonhomogeneous string shown in Figure 3.1.4 is x^2 grams. For instance, the string in the interval $[0, 5]$ has a mass of $5^2 = 25$ grams and the string in the interval $[5, 6]$ has mass $6^2 - 5^2 = 11$ grams. The **average density** of any part of the string is its mass divided by its length, $\frac{\text{total mass}}{\text{length}}$ grams per centimeter.

Exercises 41 to 43 concern the density of this string.

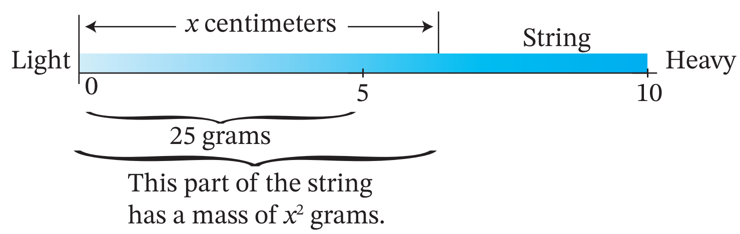


Figure 3.1.4

41. Consider the leftmost 5 centimeters of the string, the middle 2 centimeters of the string, and the rightmost 2 centimeters of the string.
(a) Which piece has the largest mass?
(b) Which piece is densest?

42. (a) What is the mass of the string in the interval $[3, 3.01]$?
(b) Using the interval $[3, 3.01]$, estimate the density at 3.
(c) Using the interval $[2.99, 3]$, estimate the density at 3.
(d) By considering intervals of the form $[3, 3+h]$, h positive, find the density 3 centimeters from the left end.
(e) By considering intervals of the form $[3+h, 3]$, h negative, find the density 3 centimeters from the left end.

43. (a) What is the mass of the string in the interval $[2, 2.01]$?
(b) Using the interval $[2, 2.01]$, estimate the density at 2.
(c) Using the interval $[1.99, 2]$, estimate the density at 2.
(d) By considering intervals of the form $[2, 2+h]$, h positive, find the density 2 centimeters from the left end.
(e) By considering intervals of the form $[2+h, 2]$, h negative, find the density 2 centimeters from the left end.

---

44. (a) Graph the curve $y = 2x^2 + x$.
(b) By eye, draw the tangent line to the curve at the point $(1, 3)$. Using a ruler, estimate its slope.
(c) Sketch the line that passes through the point $(1, 3)$ and the point $\left(x, 2x^2 + x\right)$.
(d) Find the slope of the line in (c).
(e) Letting x get closer and closer to 1, find the slope of the tangent line at $(1, 3)$.
(f) How close was your estimate in (b)?

45. An object travels $2t^2 + t$ feet in t seconds.
(a) Find its average velocity during the interval of time $[1, x]$, for $x > 1$.
(b) Letting x get closer and closer to 1, find the velocity at time 1.

46. Find the slope of the tangent line to the curve $y = x^2$ of Example 1 at $P = (x, x^2)$. To do this, consider the slope of the line through P and the nearby point $Q = (x+h, (x+h)^2)$ and let h approach 0.

47. Find the velocity of the falling rock of Example 2 at any time t. To do this, consider the average velocity during the time interval $[t, t+h]$ and then let h approach 0.

48. Does the tangent line to the curve $y = x^2$ at the point $(1,1)$ pass through the point $(6,12)$?

49. (a) Graph the curve $y = 2^x$ as well as you can for $-2 \le x \le 3$.
(b) Using a straightedge, draw as well as you can a tangent to the curve at $(2,4)$. Estimate its slope by using a ruler to draw and measure a rise-and-run triangle.
(c) Using a secant line through $(2,4)$ and $(x, 2^x)$ for x near 2, estimate the slope of the tangent line to the curve at $(2,4)$.

50. Estimate the slope of the tangent line to the graph of $f(x) = \sin(x)$ at $\left(\frac{\pi}{4}, \frac{\sqrt{2}}{2}\right)$ to two decimal places.

51. (a) Sketch the curve $y = x^3 - x^2$.
(b) Using the method of the nearby point, find the slope of the tangent line to the curve at the point $(a, a^3 - a^2)$.
(c) Find all points on the curve where the tangent line is horizontal.
(d) Find all points on the curve where the tangent line has slope 1.

52. Repeat Exercise 51 for the curve $y = x^3 - x$.

53. An astronaut is traveling from left to right along the curve $y = x^2$. When she shuts off the engine, she will fly off along the line tangent to the curve at the point where she is at the moment the engine turns off. At what point should she shut off the engine in order to reach the point (a) to reach the point $(4,9)$? (b) to reach the point $(4,-9)$?

54. See Exercise 53. Where can an astronaut who is traveling from left to right along $y = x^3 - x$ shut off the engine and pass through the point $(2,2)$?

55. **SAM:** I don't like the book's definition of the derivative.
JANE: Why not?
SAM: I can do it without limits, and more easily.
JANE: How?
SAM: Just define the derivative of f at a as the slope of the tangent line at $(a, f(a))$ on the graph of f.
JANE: Something must be wrong with that.

Does Sam have a valid point? Explain.

3.2 Derivatives of the Basic Functions

In this section we use the definition of the derivative to find the derivatives of the important functions x^a (a rational), e^x, $\sin(x)$, and $\cos(x)$. We also introduce some of the standard notations for the derivative. For convenience, we begin by repeating the definition of the derivative.

Definition: *Derivative of a Function at a Number*

Assume that the function f is defined in an open interval containing a. If

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} \tag{3.2.1}$$

exists, it is called the **derivative of f at a**. A common notation for the derivative of f at a is $f'(a)$.

The symbol "$f'(a)$" is read aloud as "f prime at a" or "the derivative of f at a." The symbols $f'(x)$ and $f'(2)$ are read similarly.

There are several notations for the quotient that appears in (3.2.1) and also for the derivative. Sometimes it is convenient to use $a+h$ instead of x and let h approach 0. Then, (3.2.1) reads

$$\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}. \tag{3.2.2}$$

Expression (3.2.2) says the same thing as (3.2.1): determine how the quotient, change in output divided by change in input, behaves as the change in input gets smaller and smaller.

Sometimes it is useful to call the change in output "Δf" and the change in input "Δx." That is, $\Delta f = f(x) - f(a)$ and $\Delta x = x - a$. Then

$$f'(a) = \lim_{\Delta x\to 0}\frac{\Delta f}{\Delta x}.$$

Δ, pronounced del-tə, is the upper-case Greek letter corresponding to the Latin "D". In mathematics, "Δf," read 'delta eff," generally indicates a difference or change in f.

There is nothing sacred about the letters a, x, and h. One could say

$$f'(x) = \lim_{t\to x}\frac{f(t)-f(x)}{t-x} \qquad \text{or} \qquad f'(x) = \lim_{u\to x}\frac{f(u)-f(x)}{u-x}$$

or, even using altogether different variable names,

$$f'(\theta) = \lim_{\phi\to\theta}\frac{f(\phi)-f(\theta)}{\phi-\theta}.$$

The notation $f'(x)$ reminds us that f', like f, is a function. For each input x the derivative, $f'(x)$, is the output. The derivative of the function f can also be written as $D(f)$.

For instance, the derivative of the squaring function, x^2, is

$$D\left(x^2\right) = \lim_{u\to x}\frac{u^2-x^2}{u-x} = \lim_{u\to x}\frac{(u-x)(u+x)}{u-x} = \lim_{u\to x}(u+x) = 2x.$$

The derivative of a specific function, in this case x^2, is denoted $(x^2)'$ or $D(x^2)$. Then, $D(x^2) = 2x$ is read aloud as "the derivative of x^2 is $2x$." This is shorthand for "the derivative of the function that assigns x^2 to x is the function that assigns $2x$ to x." Since the value of the derivative depends on x, the derivative is a function.

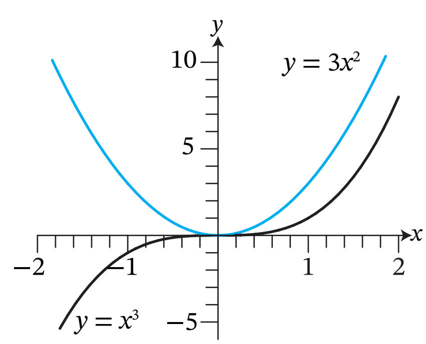


Figure 3.2.1

EXAMPLE 1. Find the derivative of x^3 at a.

SOLUTION

$$(x^3)' = \lim_{x\to a}\frac{x^3-a^3}{x-a} = 3a^2.$$

This limit was evaluated by noticing that it is one of the four limits in Section 2.2. We can write $(x^3)' = 3x^2$, $D(x^3) = 3x^2$ or $d(x^3)/dx = 3x^2$. Figure 3.2.1 shows the graphs of $y = x^3$ and its derivative, $y = 3x^2$. ◆

In the same manner, $\lim_{x\to a}(x^n - a^n)/(x-a) = n\cdot a^{n-1}$ implies that for any positive integer n, the derivative of x^n is nx^{n-1}. The exponent n becomes the coefficient and the exponent of x shrinks from n to $n-1$:

Formula 3.2.1: *Derivative of* x^n

For any positive integer n, $\qquad (x^n)' = nx^{n-1}$

The next example treats an exponential function with a fixed base.

EXAMPLE 2. Find the derivative of 2^x.

SOLUTION

$$D(2^x) = \lim_{h\to 0} \frac{2^{(x+h)} - 2^x}{h} = \lim_{h\to 0} \frac{2^x 2^h - 2^x}{h} = \lim_{h\to 0} 2^x \frac{2^h - 1}{h} = 2^x \lim_{h\to 0} \frac{2^h - 1}{h}.$$

In Section 2.2 we found that $\lim_{h\to 0}(2^h - 1)/h \approx 0.693$. Thus,

$$D(2^x) \approx (0.693)2^x.$$

♦

No one wants to remember the (approximate) constant 0.693, which appears when we use base 2. Recall that in Section 3.1 we found that the derivative of e^x is e^x, with no multiplying constant. The next formula emphasizes this fact.

Formula 3.2.2: *Derivative of* e^x

$$D(e^x) = e^x$$

Next, we turn to trigonometric functions.

EXAMPLE 3. Find the derivative of $\sin(x)$.

SOLUTION A key step in this evaluation of many trigonometric limits using the definition is to recall the summation formulas for sin and cos:

$$\sin(a+b) = \sin(a)\cos(b) + \cos(a)\sin(b) \quad \text{and} \quad \cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b).$$

Then, applying the definition of the derivative at a number to $f(x) = \sin(x)$:

$$\begin{aligned} D(\sin(x)) &= \lim_{h\to 0} \frac{\sin(x+h) - \sin(x)}{h} \\ &= \lim_{h\to 0} \frac{\sin(x)\cos(h) + \cos(x)\sin(h) - \sin(x)}{h} \\ &= \lim_{h\to 0} \frac{\sin(x)(\cos(h) - 1) + \cos(x)\sin(h)}{h} \\ &= \lim_{h\to 0} \sin x \frac{\cos(h) - 1}{h} + \cos(x)\frac{\sin(h)}{h}. \end{aligned}$$

In Section 2.2 we found that $\lim_{h\to 0}\sin(h)/h = 1$ and $\lim_{h\to 0}(1 - \cos(h))/h = 0$. Thus $\lim_{h\to 0}(\cos(h) - 1)/h = 0$ and

$$D(\sin x) = (\sin x)(0) + (\cos x)(1) = \cos(x).$$

♦

This result from Example 3 is important enough to be added to our collection of known derivatives.

Formula 3.2.3: *Derivative of* $\sin(x)$

$$D(\sin(x)) = \cos(x)$$

If we graph $y = \sin(x)$ (see Figure 3.2.2), and consider its shape, the formula $D(\sin(x)) = \cos(x)$ is not a surprise. For instance, for x in $(-\pi/2, \pi/2)$ the slope of the graph of $y = \sin(x)$ is positive. So is $\cos(x)$. For x in $(\pi/2, 3\pi/2)$ the slope of the sine curve is negative. Again, the same as the graph of $y = \cos(x)$. Since $\sin(x)$ has period 2π, we would expect its derivative also to have period 2π. Indeed, $\cos(x)$ has period 2π.

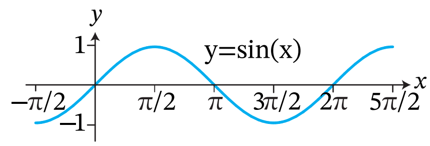


Figure 3.2.2

Similarly, applying the definition of the derivative to $y = \cos(x)$ and the summation identity for cosine, we determine that a simple formula for the derivative of the cosine function.

Formula 3.2.4: *Derivative of* $\cos(x)$

$$D(\cos(x)) = -\sin(x)$$

Derivatives of Other Power Functions

We showed that if n is a positive integer, $D(x^n) = nx^{n-1}$. Now let us find the derivative of power functions x^n where n is not a positive integer.

EXAMPLE 4. Find the derivative of $x^{-1} = \frac{1}{x}$.

SOLUTION Before we calculate the necessary limit, let's pause to see how the slope of $y = 1/x$ behaves. Figure 3.2.3 shows that the slope is always negative. For x near 0, the absolute value of the slope is large, but when $|x|$ is large, the slope is near 0.

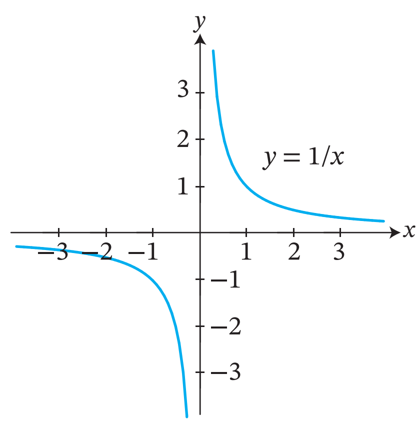


Figure 3.2.3

Now, let us find the derivative of $1/x$:

$$\begin{aligned} D\left(\frac{1}{x}\right) &= \lim_{t\to x} \frac{1/t - 1/x}{t-x} && \text{(definition of derivative of } 1/x \text{)} \\ &= \lim_{t\to x} \frac{1}{t-x}\left(\frac{x-t}{xt}\right) && \left(\text{ since } \frac{1}{t} - \frac{1}{x} = \frac{x-t}{xt} \right) \\ &= \lim_{t\to x} \frac{-1}{xt} \\ &= -\frac{1}{x^2}. \end{aligned}$$

As a check, we see that $-1/x^2$ is always negative, has large absolute value when x is near 0, and is near 0 when $|x|$ is large. ♦

It is worth memorizing that

Formula 3.2.5: *Derivative of* x^{-1}

$$D\left(\frac{1}{x}\right) = -\frac{1}{x^2} \qquad (\text{ for } x \neq 0 \text{ })$$

or, in power notation,

$$D\left(x^{-1}\right) = -x^{-2} \qquad (\text{ for } x \neq 0 \text{ })$$

The second form fits the pattern established for positive integers n, $D(x^n) = nx^{n-1}$.

EXAMPLE 5. Find the derivative of $x^{2/3}$.

SOLUTION Once again we use the definition of the derivative:

$$D(x^{2/3}) = \lim_{t\to x} \frac{t^{2/3} - x^{2/3}}{t-x}.$$

To find the limit, use a bit of algebra. Write the four terms $t^{2/3}$, $x^{2/3}$, t, and x as powers of $t^{1/3}$ and $x^{1/3}$. Thus

$$D\left(x^{2/3}\right) = \lim_{t\to x} \frac{\left(t^{1/3}\right)^2 - \left(x^{1/3}\right)^2}{\left(t^{1/3}\right)^3 - \left(x^{1/3}\right)^3}.$$

Because $a^2 - b^2 = (a-b)(a+b)$ and $a^3 - b^3 = (a-b)(a^2+ab+b^2)$, we find

$$\begin{aligned}
D\left(x^{2/3}\right) &= \lim_{t\to x} \frac{\left(\left(t^{1/3}\right) - \left(x^{1/3}\right)\right)\left(\left(t^{1/3}\right) + \left(x^{1/3}\right)\right)}{\left(\left(t^{1/3}\right) - \left(x^{1/3}\right)\right)\left(\left(t^{1/3}\right)^2 + \left(t^{1/3}\right)\left(x^{1/3}\right) + \left(x^{1/3}\right)^2\right)} && \text{(factor numerator and denominator)}\\
&= \lim_{t\to x} \frac{\left(t^{1/3}\right) + \left(x^{1/3}\right)}{\left(t^{1/3}\right)^2 + \left(t^{1/3}\right)\left(x^{1/3}\right) + \left(x^{1/3}\right)^2} && \text{(cancel common factors)}\\
&= \frac{\left(x^{1/3}\right) + \left(x^{1/3}\right)}{\left(x^{1/3}\right)^2 + \left(x^{1/3}\right)\left(x^{1/3}\right) + \left(x^{1/3}\right)^2} && \text{(evaluate limit as } t\to x \text{)}\\
&= \frac{2x^{1/3}}{3x^{2/3}} = \frac{2}{3}x^{-1/3} && \text{(simplify).}
\end{aligned}$$

In summary, $D\left(x^{2/3}\right) = (2/3)x^{-1/3}$. This formula also follows the pattern we found for $D(x^n)$ for $n = 1, 2, 3, \ldots$ and $n = -1$. The exponent of x becomes the coefficient and the exponent of x is lowered by 1. ♦

The method used in Example 5 applies to any positive rational exponent. In the next two sections we will show how the result extends first to negative rational exponents and then to irrational exponents. In all cases the formula will be the same. We state the general result here, but remember that — so far — we have justified it only for positive rational exponents and for -1.

Formula 3.2.6: *Derivative of Power Functions* x^a

For any fixed number a, $D(x^a) = ax^{a-1}$

The formula holds for values of x where both x^a and x^{a-1} are defined. For instance, $x^{1/2} = \sqrt{x}$ is defined for $x \geq 0$, but its derivative $x^{-1/2}/2$ is defined only for $x > 0$.

Because the derivative of the square root function occurs so often we emphasize its formula

Formula 3.2.7: *Derivative of the Square Root Function*

$$D\left(x^{1/2}\right) = \frac{1}{2}x^{-1/2} \qquad (\text{ for } x > 0)$$

or, in terms of the usual square root sign,

$$D\left(\sqrt{x}\right) = \frac{1}{2\sqrt{x}} \qquad (\text{ for } x > 0)$$

Another Notation for the Derivative

We have used the notations f' and $D(f)$ for the derivative of a function f. There is another notation that is also convenient.

If $y = f(x)$, the derivative is denoted by the symbols dy/dx or df/dx. In this notation the derivative of x^3 is written $d\left(x^3\right)/dx$. If the function is expressed in terms of another letter, such as t, we would write $d\left(t^3\right)/dt$.

The symbol dy/dx is read as "the derivative of y with respect to x" or "dee y, dee x."

Keep in mind that in the notations df/dx and dy/dx, the symbols df, dy, and dx have no meaning by themselves. The symbol dy/dx should be thought of as a single entity, like the numeral 8, which we do not think of as formed of two 0's.

In the study of motion, Newton's **dot notation** is sometimes encountered. If x is a function of time t, then $\dot{x}$ denotes the derivative dx/dt.

Summary

In this section we saw why limits are important in calculus. We need them to define the derivative of a function. The definition can be stated in several ways, but each one says, informally, "look at how a small change in input changes the output." Here is the formal definition, in various notations:

$$f'(a) = \lim_{h\to 0}\frac{f(a+h)-f(a)}{h}, \qquad f'(x) = \lim_{x\to 0}\frac{f(x+h)-f(x)}{h}, \qquad f'(x) = \lim_{t\to x}\frac{f(t)-f(x)}{t-x}, \quad \text{and} \quad f'(x) = \lim_{\Delta x\to 0}\frac{\Delta f}{\Delta x}.$$

The following derivatives should be memorized. However, if one is forgotten it can be recovered by using the definition as a limit.

Function	Derivative
$f(x)$	$f'(x)$
x^a	ax^{a-1}
e^x	e^x
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$

EXERCISES for Section 3.2

1. Show that $D(\cos(x)) = -\sin(x)$.

Using the definition of the derivative, compute the appropriate limit to find the derivatives of the functions in Exercises 2 to 12.

2. $\dfrac{1}{x+2}$ **3.** $2x - x^2$ **4.** 3^x **5.** $6x^3$ **6.** $x^{4/3}$ **7.** $5x^2$

8. $4\sin(x)$ **9.** $2e^x + \sin(x)$ **10.** $x^2 + x^3$ **11.** $\dfrac{1}{2x+1}$ **12.** $\dfrac{1}{x^2}$

13. Use the formulas obtained for the derivatives of e^x, x^a, $\sin(x)$, and $\cos(x)$ to evaluate the derivative of the given function at the given input. (a) e^x at $x = -1$ (b) $x^{1/3}$ at $x = -8$ (c) $\sqrt[3]{x}$ at $x = 27$ (d) $\cos(x)$ at $x = \dfrac{\pi}{4}$ (e) $\sin(x)$ at $x = \dfrac{2\pi}{3}$

14. Use the formulas obtained for the derivatives of e^x, x^a, $\sin(x)$, and $\cos(x)$ to evaluate the derivative of the given function at the given input. (a) e^x at $x = 0$ (b) $x^{2/3}$ at $x = -1$ (c) $\sqrt{x}$ at $x = 25$ (d) $\cos(x)$ at $x = -\pi$ (e) $\sin(x)$ at $x = \dfrac{\pi}{3}$

15. State the definition of the derivative of a function in words, using no mathematical symbols.

16. State the definition of the derivative of $g(t)$ at b as a mathematical formula, with no words.

In Exercises 17 to 22 use the definition of the derivative to show that the given equation is correct.

17. $D\left(e^{-x}\right) = -e^{-x}$

18. $D\left(e^{3x}\right) = 3e^{3x}$

19. $D\left(\dfrac{1}{\cos(x)}\right) = \dfrac{\sin(x)}{\cos^2(x)}$

20. $D(\tan(x)) = 1+\tan^2(x) = \sec^2(x)$

21. $D(\sin(2x)) = 2\cos(2x)$

22. $D\left(\cos\left(\dfrac{x}{2}\right)\right) = \dfrac{-1}{2}\sin\left(\dfrac{x}{2}\right)$

23. Use a limit to show that $D\left((x^{-5}\right) = -5x^{-6}$.

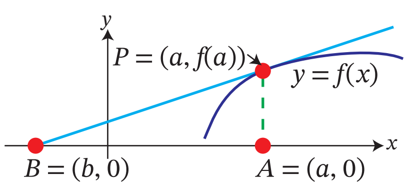


Figure 3.2.4

Let f be a differentiable function and a a number such that $f'(a)$ is not zero. The tangent to the graph of f at $P = (a, f(a))$ meets the x-axis at a point $B = (b, 0)$, see Figure 3.2.4. The point $A = (a, 0)$ is the projection of P onto the x-axis. The **subtangent** of f is the line segment AB. Its length is $|a-b|$. Exercises 24 and 25 involve the subtangent of a function.

24. Show that for e^x the length of the subtangent is the same for all values of a.

25. Find the length of the subtangent at $(a, f(a))$ for any differentiable function f. Assume $f'(a) \neq 0$.

26. This Exercise shows why in calculus angles are measured in radians. Let $\text{Sin}(x)$ denote the sine of an angle of x degrees and let $\text{Cos}(x)$ denote the cosine of an angle of x degrees.

(a) Graph $y = \text{Sin}(x)$ on the interval $[-180, 360]$, using the same scale on the x- and y-axes.

(b) Find $\lim_{x\to 0} \dfrac{\text{Sin}(x)}{x}$.

(c) Find $\lim_{x\to 0} \dfrac{1-\text{Cos}(x)}{x}$.

(d) Using the definition of the derivative, differentiate $\text{Sin}(x)$.

3.3 Shortcuts for Computing Derivatives

This section develops methods for finding the derivative of a function, or what is called **differentiating** a function. Using them will make it easy to find, for instance, the derivative of

$$\frac{(3+4x+5x^2)e^x}{\sin(x)}$$

without going back to the definition of the derivative and finding the limit of a complicated quotient.

We first find the derivative of a constant function.

The Derivative of a Constant Function

Theorem 3.3.1: Constant Rule

The derivative of a constant function $f(x) = C$ is 0.

$$\frac{dC}{dx} = (C)' = 0$$

Proof of the Constant Rule (Theorem 3.3.1)
Let C be a number and let f be the constant function, $f(x) = C$ for all inputs x. By the definition of a derivative, with h replaced by the more descriptive Δx,

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

Since f has the same output C for all inputs, $f(x + \Delta x) = C$ and $f(x) = C$. Thus,

$$\begin{aligned} f'(x) &= \lim_{\Delta x \to 0} \frac{C - C}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{0}{\Delta x} \\ &= \lim_{\Delta x \to 0} 0 && (\text{ since } \Delta x \neq 0 \text{ }) \\ &= 0. \end{aligned}$$

This shows the derivative of a constant function is 0 for all x.

From two points of view, the constant rule is no surprise. Since the graph of $f(x) = C$ is a horizontal line, it coincides with each of its tangent lines, which have slope 0, as can be seen in Figure 3.3.1. Also, if we think of x as time and $f(x)$ as the position of a particle at time x, the constant rule implies that a stationary particle has zero velocity.

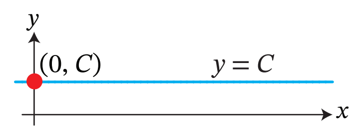


Figure 3.3.1

Derivatives of $f + g$ and $f - g$

The next theorem asserts that if the functions, f and g have derivatives, so does their sum $f + g$ and

$$\frac{d}{dx}(f + g) = \frac{df}{dx} + \frac{dg}{dx}.$$

In other words, "the derivative of the sum is the sum of the derivatives." Equivalently, $(f + g)' = f' + g'$ and $D(f + g) = D(f) + D(g)$. Similar formulas hold for the derivative of $f - g$.

Theorem 3.3.2: Sum Rule and Difference Rule

*If f and g are differentiable functions, then so are $f + g$ and $f - g$. The **sum rule** and **difference rule** for computing their derivatives are*

$$\begin{aligned} (f + g)' &= f' + g' && (\textit{ sum rule }) \\ (f - g)' &= f' - g' && (\textit{ difference rule }) \end{aligned}$$

Proof of the Sum Rule (Theorem 3.3.2)
To justify this we go back to the definition of the derivative. To begin, we give the function $f + g$ the name u, that is, $u(x) = f(x) + g(x)$. We will examine

$$\lim_{\Delta x \to 0} \frac{u(x + \Delta x) - u(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x}. \tag{3.3.1}$$

To evaluate the right-hand side of (3.3.1), express Δu in terms of Δf and Δg. Here are the details:

$$\begin{aligned} \Delta u &= u(x + \Delta x) - u(x) \\ &= \big(f(x + \Delta x) + g(x + \Delta x)\big) - \big(f(x) + g(x)\big) && (\text{ definition of } u \text{ }) \\ &= \big(f(x) + \Delta f\big) + \big(g(x) + \Delta g\big) - \big(f(x) + g(x)\big) && (\text{ definition of } \Delta f \text{ and } \Delta g \text{ }) \\ &= \Delta f + \Delta g \end{aligned}$$

All told, $\Delta u = \Delta f + \Delta g$. The change in u is the change in f plus the change in g.

We can now complete the evaluation of the limit in (3.3.1):

$$\begin{aligned}\lim_{\Delta x\to 0}\frac{\Delta u}{\Delta x} &= \lim_{\Delta x\to 0}\frac{\Delta f+\Delta g}{\Delta x} && (\text{ definition of } u\text{ })\\ &= \lim_{\Delta x\to 0}\frac{\Delta f}{\Delta x}+\lim_{\Delta x\to 0}\frac{\Delta g}{\Delta x} && (\text{ sum property of limits })\\ &= f'(x)+g'(x) && (\text{ definition of derivative }).\end{aligned}$$

Thus, $u = f + g$ is differentiable and $u'(x) = f'(x) + g'(x)$. A similar argument applies to $f - g$. ●

It should not be a surprise to learn that the sum and difference rules in Theorem 3.3.2 extend to any finite number of differentiable functions. For example, $(f + g + h)' = f' + g' + h'$ and $(f - g - h)' = f' - g' - h'$.

EXAMPLE 1. Using the sum rule, differentiate $x^2 + x^3 + \cos(x) + 3$.

SOLUTION To find this derivative of a sum of three terms, differentiate each term and add the results:

$$\begin{aligned}D\left(x^2+x^3+\cos(x)+3\right) &= D(x^2)+D(x^3)+D(\cos(x))+D(3)\\ &= 2x^{2-1}+3x^{3-1}+(-\sin(x))+0\\ &= 2x+3x^2-\sin(x).\end{aligned}$$

♦

EXAMPLE 2. Differentiate $x^4 - \sqrt{x} - e^x$.

SOLUTION This derivative is found by differentiating the first term and then subtracting the derivatives of the second and third terms.

$$\begin{aligned}\frac{d}{dx}\left(x^4-\sqrt{x}-e^x\right) &= \frac{d}{dx}\left(x^4\right)-\frac{d}{dx}\left(\sqrt{x}\right)-\frac{d}{dx}\left(e^x\right)\\ &= 4x^3-\frac{1}{2\sqrt{x}}-e^x.\end{aligned}$$

♦

The Derivative of a Product fg

The following theorem, concerning the derivative of the product of two functions, may be surprising, for it turns out that the derivative of the product is not the product of the derivatives. The formula is more complicated than the one for the derivative of the sum. It asserts that "the derivative of the product is the derivative of the first function times the second plus the first function times the derivative of the second."

The formula for $(fg)'$ was discovered by Leibniz in 1676, but his first ideas for this result were wrong.

Theorem 3.3.3: Product Rule

If f and g are differentiable functions, then so is their product fg. Its derivative is given by

$$(fg)' = f'g + fg'$$

Proof of the Product Rule (Theorem 3.3.3)

The proof is similar to that for the sum and difference rules. This time we give the product fg the name u. Then we express Δu in terms of Δf and Δg. Finally, we determine $u'(x)$ by examining $\lim_{\Delta x\to 0}\frac{\Delta u}{\Delta x}$. These steps are practically forced upon us.

We have $u(x) = f(x)g(x)$ and $u(x+\Delta x) = f(x+\Delta x)g(x+\Delta x)$. Rather than subtract $u(x)$ from $u(x+\Delta x)$ directly, we write $f(x+\Delta x) = f(x)+\Delta f$ and $g(x+\Delta x) = g(x)+\Delta g$. Then

$$\begin{aligned} u(x+\Delta x) &= \big(f(x+\Delta x)\big)\big(g(x+\Delta x)\big) \\ &= \big(f(x)+\Delta f\big)\big(g(x)+\Delta g\big) \\ &= f(x)g(x) + (\Delta f)g(x) + f(x)\Delta g + (\Delta f)(\Delta g).i \end{aligned}$$

Hence

$$\begin{aligned} \Delta u &= u(x+\Delta x) - u(x) \\ &= f(x)g(x) + (\Delta f)g(x) + f(x)(\Delta g) + (\Delta f)(\Delta g) - f(x)g(x) \\ &= (\Delta f)g(x) + f(x)(\Delta g) + (\Delta f)(\Delta g) \end{aligned}$$

and

$$\begin{aligned} \frac{\Delta u}{\Delta x} &= \frac{(\Delta f)g(x) + f(x)(\Delta g) + (\Delta f)(\Delta g)}{\Delta x} \\ &= \frac{\Delta f}{\Delta x}g(x) + f(x)\frac{\Delta g}{\Delta x} + \Delta f\frac{\Delta g}{\Delta x}. \end{aligned}$$

As $\Delta x \to 0$, $\Delta g/\Delta x \to g'(x)$ and $\Delta f/\Delta x \to f'(x)$. Furthermore, because f is differentiable and hence continuous, $\Delta f \to 0$ as $x \to 0$. It follows that

$$\lim_{\Delta x\to 0}\frac{\Delta u}{\Delta x} = f'(x)g(x) + f(x)g'(x) + 0\cdot g'(x).$$

Therefore, u is differentiable and

$$u' = f'g + fg'.$$

Observation 3.3.4: *Graphical Proof of the Product Rule*

Figure 3.3.2 illustrates the product rule and its proof. With f, Δf, g, and Δg taken to be positive, the inner rectangle has area $u = fg$ and the whole rectangle has area $u+\Delta u = (f+\Delta f)(g+\Delta g)$. The shaded region whose area is Δu is made up of rectangles of areas $f\cdot(\Delta g)$, $(\Delta f)\cdot g$, and $(\Delta f)\cdot(\Delta g)$.
The little corner rectangle, of area $(\Delta f)\cdot(\Delta g)$, is negligible in comparison with the other two rectangles. Thus, $\Delta u \approx (\Delta f)g + f(\Delta g)$, which suggests the formula for the derivative of a product.

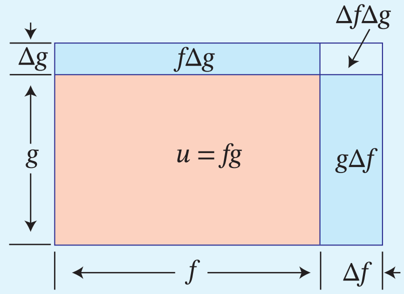


Figure 3.3.2

EXAMPLE 3. Find $D\left(\left(x^2+x^3+\cos(x)+3\right)\left(x^4-\sqrt{x}-e^x\right)\right)$.

SOLUTION The function to be differentiated is the product of the functions differentiated in Examples 1 and 2.

$$\begin{aligned} &D\left(\left(x^2+x^3+\cos(x)+3\right)\left(x^4-\sqrt{x}-e^x\right)\right) \\ &\quad= \left(D\left(x^2+x^3+\cos(x)+3\right)\right)\left(x^4-\sqrt{x}-e^x\right) + \left(x^2+x^3+\cos(x)+3\right)\left(D\left(x^4-\sqrt{x}-e^x\right)\right) \\ &\quad= \left(2x+3x^2-\sin(x)\right)\left(x^4-\sqrt{x}-e^x\right) + \left(x^2+x^3+\cos(x)+3\right)\left(4x^3-\frac{1}{2\sqrt{x}}-e^x\right). \end{aligned}$$

Derivative of a Constant Times f

A special case of the formula for the product rule occurs so frequently that it is singled out in the constant multiple rule.

Theorem 3.3.5: Constant Multiple Rule

If C is a constant function and f is a differentiable function, then Cf is differentiable and its derivative is given by

$$(Cf)' = C(f')$$

In other notations, $\frac{d(Cf)}{dx} = C\frac{df}{dx}$ *and* $D(Cf) = C\,D(f)$.

In words, the derivative of a constant times a function is the constant times the derivative of the function.

Proof of the Constant Multiple Rule (Theorem 3.3.5)

Because we have a product of two differentiable functions, C and f, we may use the product rule. We have

$$\begin{aligned}(Cf)' &= (C')f + C(f') && \text{(derivative of a product)}\\ &= 0\cdot f + Cf' && \text{(derivative of constant is 0)}\\ &= C(f').\end{aligned}$$

The constant multiple rule asserts that "it is legal to move a constant factor outside the derivative symbol."

EXAMPLE 4. Find $D(6x^3)$.

SOLUTION Using the constant multiple rule for differentiation, Theorem 3.3.5,

$$\begin{aligned}D(6x^3) &= 6D(x^3) && \text{(6 is a constant)}\\ &= 6\cdot 3x^2 && (\ D(x^n) = nx^{n-1}\ )\\ &= 18x^2.\end{aligned}$$

With a little practice, one could immediately write $D(6x^3) = 18x^2$. ♦

EXAMPLE 5. Find $D\left(\frac{\sqrt{x}}{11}\right)$.

SOLUTION While this may look like a quotient, it is also a product of a constant and a power, so is easily differentiated with the constant multiple and power rule (for a rational exponent):

$$D\left(\frac{\sqrt{x}}{11}\right) = D\left(\frac{1}{11}\sqrt{x}\right) = \frac{1}{11}D(\sqrt{x}) = \frac{1}{11}\frac{1}{2\sqrt{x}} = \frac{1}{22}x^{-1/2}.$$

♦

Example 5 generalizes to the fact that for a nonzero C,

Theorem 3.3.6: Constant Division Rule

For any differentiable function f and any nonzero constant C,

$$\left(\frac{f}{C}\right)' = \frac{f'}{C}.$$

The formula for the derivative of the product extends to the product of several differentiable functions.

Theorem 3.3.7: Generalized Product Rule

Let f, g, and h be differentiable functions. Then

$$(fgh)' = (f')gh + f(g')h + fg(h').$$

NOTE: *Each summand contains the derivative of one of the functions in the product.*

The next example illustrates the use of this extension of the product rule.

EXAMPLE 6. Differentiate $\sqrt{x}e^x\sin(x)$.

SOLUTION By a direct application of Theorem 3.3.7 with $f(x) = \sqrt{x}$, $g(x) = e^x$, and $h(x) = \sin(x)$:

$$\begin{aligned}\left(\sqrt{x}e^x\sin(x)\right)' &= \left(\sqrt{x}\right)' e^x\sin(x) + \sqrt{x}\left(e^x\right)'\sin(x) + \sqrt{x}e^x\left(\sin(x)\right)' \\ &= \left(\frac{1}{2\sqrt{x}}\right)e^x\sin(x) + \sqrt{x}e^x\sin(x) + \sqrt{x}e^x\cos(x)\end{aligned}$$

◆

EXAMPLE 7. Differentiate $6t^8 - t^3 + 5t^2 + \pi^3$.

SOLUTION The independent variable in this polynomial is t, and the polynomial is to be differentiated with respect to t. By the sum rule, differentiation of any polynomial is done "term-by-term". (Note that π^3 is a constant.)

$$\begin{aligned}\frac{d}{dt}\left(6t^8 - t^3 + 5t^2 + \pi^3\right) &= \frac{d}{dt}\left(6t^8\right) - \frac{d}{dt}\left(t^3\right) + \frac{d}{dt}\left(5t^2\right) + \frac{d}{dt}\left(\pi^3\right) \\ &= 48t^7 - 3t^2 + 10t + 0 \\ &= 48t^7 - 3t^2 + 10t.\end{aligned}$$

◆

Derivative of a Reciprocal $1/g$

Often one needs the derivative of the reciprocal of a function g, that is, $(1/g)'$.

Theorem 3.3.8: Reciprocal Rule

If g is a differentiable function, then

$$\left(\frac{1}{g}\right)' = -\frac{g'}{g^2}, \qquad \textit{where } g(x) \neq 0.$$

Proof of the Reciprocal Rule (Theorem 3.3.8)

Again we go back to the definition of the derivative.

Assume $g(x) \neq 0$ and let $u(x) = 1/g(x)$. Then $u(x+\Delta x) = 1/g(x+\Delta x) = 1/(g(x)+\Delta g)$. Thus

$$\begin{aligned}\Delta u &= u(x+\Delta x) - u(x) \\ &= \frac{1}{g(x)+\Delta g} - \frac{1}{g(x)} \\ &= \frac{g(x) - (g(x)+\Delta g)}{g(x)(g(x)+\Delta g)} && (\text{ common denominator }) \\ &= \frac{-\Delta g}{g(x)(g(x)+\Delta g)} && (\text{ cancellation in numerator }).\end{aligned}$$

Then

$$\begin{aligned}
u'(x) &= \lim_{\Delta x\to 0}\frac{\Delta u}{\Delta x} \\
&= \lim_{\Delta x\to 0}\frac{-\Delta g/\big(g(x)(g(x)+\Delta g)\big)}{\Delta x} \\
&= \lim_{\Delta x\to 0}\frac{-\Delta g/\Delta x}{g(x)(g(x)+\Delta g)} && \left(\text{ algebra: } \frac{(a/b)}{c}=\frac{(a/c)}{b}\ \right) \\
&= \frac{\lim\limits_{\Delta x\to 0}\left(\dfrac{-\Delta g}{\Delta x}\right)}{\lim\limits_{\Delta x\to 0}(g(x)(g(x)+\Delta g))} && (\text{ quotient rule for limits }) \\
&= \frac{-g'(x)}{g(x)^2}. && \left(\ g(x) \text{ is continuous, } \lim_{\Delta x\to 0}\Delta g=0\ \right)
\end{aligned}$$

●

EXAMPLE 8. Find $D\left(\dfrac{1}{\cos(x)}\right)$.

SOLUTION In this case, $g(x)=\cos(x)$ and $g'(x)=-\sin(x)$. Therefore, by the reciprocal rule:

$$\begin{aligned}
D\left(\frac{1}{\cos(x)}\right) &= \frac{-(-\sin(x))}{(\cos(x))^2} \\
&= \frac{\sin(x)}{\cos^2(x)} && (\text{ for all } x \text{ with } \cos(x)\neq 0\ ).
\end{aligned}$$

♦

Example 8 gives a formula for the derivative of $\sec(x)$, which is defined as $1/\cos(x)$.

$$D(\sec(x)) = D\left(\frac{1}{\cos(x)}\right) = \frac{\sin(x)}{\cos^2(x)} = \frac{\sin(x)}{\cos(x)}\frac{1}{\cos(x)} = \tan(x)\sec(x).$$

Therefore,

Formula 3.3.1: *Derivative of* $\sec(x)$

$$D(\sec(x)) = \sec(x)\tan(x) \qquad (\text{ for all } x \text{ in the domain of } \sec(x)\ )$$

EXAMPLE 9. Show that the power rule, in Section 3.2, is valid when a is a negative rational number. That is, show that $D\left(x^{-p/q}\right) = -\dfrac{p}{q}x^{-p/q-1}$ for any positive integers p and q, with $q\neq 0$.

SOLUTION The reciprocal rule to find the derivative of $x^{-p/q}$ Write $x^{-p/q}$ as $1/x^{p/q}$, then

$$D\left(x^{-p/q}\right) = D\left(\frac{1}{x^{p/q}}\right) = \frac{-D(x^{p/q})}{(x^{p/q})^2} = \frac{-\frac{p}{q}x^{\frac{p}{q}-1}}{x^{\frac{2p}{q}}} = -\frac{p}{q}x^{\left(\frac{p}{q}\right)-1-2\left(\frac{p}{q}\right)} = -\frac{p}{q}x^{-p/q-1}.$$

♦

The Derivative of a Quotient f/g

EXAMPLE 10. Derive a formula for the derivative of the quotient $\frac{f}{g}$.

SOLUTION Write the quotient f/g as a product, $f \cdot \frac{1}{g}$. Assuming f and g are differentiable functions, we may use the product and reciprocal rules to find

$$\begin{aligned}
\left(\frac{f(x)}{g(x)}\right)' &= \left(f(x)\frac{1}{g(x)}\right)' && \text{(rewrite quotient as product)}\\
&= f'(x)\left(\frac{1}{g(x)}\right) + f(x)\left(\frac{1}{g(x)}\right)' && \text{(product rule)}\\
&= f'(x)\left(\frac{1}{g(x)}\right) + f(x)\left(\frac{-g'(x)}{g(x)^2}\right) && \text{(reciprocal rule, assuming } g(x)\neq 0 \text{)}\\
&= \frac{f'(x)}{g(x)} - \frac{f(x)g'(x)}{g(x)^2} && \text{(algebra)}\\
&= \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}. && \text{(algebra: common denominator)}
\end{aligned}$$

♦

Example 10 is the proof of the **quotient rule**. A simple case of the quotient rule has already been used to find the derivative of $\sec(x) = 1/\cos(x)$ (Example 8). The quotient rule will be used to find the derivative of $\tan(x) = \sin(x)/\cos(x)$ (Example 11). Because the quotient rule is used so often, it should be committed to memory

Theorem 3.3.9: Quotient Rule

Let f and g be differentiable functions at x, and assume $g(x) \neq 0$. Then the quotient f/g is differentiable at x, and

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2} \qquad \textit{(whenever } g(x)\neq 0 \textit{)}$$

EXAMPLE 11. Find the derivative of the tangent function.

SOLUTION Recalling that $\tan(x) = \sin(x)/\cos(x)$ for all x for which $\cos(x)$ is not zero, the derivative of the tangent function is found by using the quotient rule to differentiate $\tan(x) = \sin(x)/\cos(x)$ for all x for which $\cos(x)$ is not zero, that is, for all x except odd multiples of $\pi/2$.

$$\begin{aligned}
(\tan(x))' &= \left(\frac{\sin(x)}{\cos(x)}\right)'\\
&= \frac{\cos(x)(\sin(x))' - \sin(x)(\cos(x))'}{(\cos(x))^2} && \text{(quotient rule)}\\
&= \frac{(\cos(x))\cos(x) - \sin(x)(-\sin(x))}{(\cos(x))^2}\\
&= \frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)}\\
&= \frac{1}{\cos^2(x)} && \text{(} \sin^2(x)+\cos^2(x) = 1 \text{)}\\
&= \sec^2(x). && \text{(} \sec(x) = 1/\cos(x) \text{)}
\end{aligned}$$

This result is valid whenever x is not an odd multiple of $\pi/2$, and should be memorized. ♦

Formula 3.3.2: *Derivative of* tan(x)

$$D(\tan(x)) = \sec^2(x) \qquad (\text{ for all } x \text{ in the domain of } \tan(x)\ )$$

Because the numerator in the quotient rule is a difference, it is important to get the terms in the correct order. Here is an easy way to remember the quotient rule.

Observation 3.3.10: *How to Remember the Quotient Rule*

The following two steps can be helpful to remember the order of the terms in the numerator of the quotient rule for the derivative of f/g.

Step 1: Write down the parts where g^2 and g appear:

$$\frac{g\qquad\quad}{g^2}.$$

This ensures that the denominator is correct and one has a good start on the numerator.

Step 2: To complete the numerator, remember that it is the difference of two terms, each term a product of one function and one derivative:

$$\frac{gf' - fg'}{g^2}.$$

EXAMPLE 12. Compute $\left(\dfrac{x^2}{x^3+1}\right)'$, showing each step.

SOLUTION We obtain this derivative by following the two steps in the observation about how to remember the quotient rule:

$$\begin{aligned}
\left(\frac{x^2}{x^3+1}\right)' &= \frac{(x^3+1)\cdots\qquad\qquad}{(x^3+1)^2} && (\text{ Step 1: write denominator and start numerator })\\
&= \frac{(x^3+1)(x^2)' - (x^2)(x^3+1)'}{(x^3+1)^2} && (\text{ Step 2: complete numerator, remembering the minus sign })\\
&= \frac{(x^3+1)(2x) - (x^2)(3x^2)}{(x^3+1)^2} && (\text{ compute derivatives })\\
&= \frac{2x^4+2x-3x^4}{(x^3+1)^2} && (\text{ algebra })\\
&= \frac{2x-x^4}{(x^3+1)^2}. && (\text{ algebra: collecting })
\end{aligned}$$

◆

As Example 12 illustrates, the techniques for differentiating polynomials and quotients can be combined to differentiate any rational function, that is, any quotient of polynomials.

Summary

Let f and g be two differentiable functions and let C be a constant function. We obtained formulas for differentiating $f+g$, $f-g$, fg, Cf, $1/g$, and f/g.

With the aid of the formulas in Table 3.3.1, we can differentiate $\sec(x)$, $\csc(x)$, $\tan(x)$, and $\cot(x)$ using $(\sin(x))' = \cos(x)$ and $(\cos(x))' = -\sin(x)$. Exercises 17(a) and 17(c) concern the differentiation of $\cot(x)$ and $\csc(x)$. We also have shown that $D(x^a) = ax^{a-1}$ for any fixed rational number a.

Table 3.3.2 presents the current set of essential derivatives that must be learned and never forgotten. And, as previously noted, if one of these is forgotten it can always be recovered based on the definition involving limits of slopes of secant lines.

Rule	Formula	Comment
Constant Rule	$C' = 0$	C a constant
Sum Rule	$(f+g)' = f' + g'$	
Difference Rule	$(f-g)' = f' - g'$	
Product Rule	$(fg)' = f'g + fg'$	
Constant Multiple Rule	$(Cf)' = Cf'$	
Reciprocal Rule	$\left(\frac{1}{g}\right)' = \frac{-g'}{g^2}$	$g(x) \neq 0$
Quotient Rule	$\left(\frac{f}{g}\right)' = \frac{gf' - fg'}{g^2}$	$g(x) \neq 0$

Table 3.3.1

Function	Derivative	Comment
$f(x)$	$f'(x)$	
x^a	ax^{a-1}	a is a fixed number
e^x	e^x	
$\sin(x)$	$\cos(x)$	
$\cos(x)$	$-\sin(x)$	
$\tan(x)$	$\sec^2(x)$	for all x except odd multiples of $\pi/2$
$\sec(x)$	$\sec(x)\tan(x)$	for all x except odd multiples of $\pi/2$

Table 3.3.2

EXERCISES for Section 3.3

In Exercises 1 to 15 differentiate the given function. Use only the formulas presented in this and earlier sections.

1. $5x^3$ **2.** $5x^3 - 7x + 2^3$ **3.** $3\sqrt{x} - \sqrt[3]{x}$ **4.** $\frac{1}{\sqrt{x}}$ **5.** $(5+x)(x^2 - x + 7)$

6. $\sin(x)\cos(x)$ **7.** $3\tan(x)$ **8.** $3(\tan(x))^2$ **9.** $\frac{x^3-1}{2x+1}$ **10.** $\frac{\sin(x)}{e^x}$

11. $\frac{3x^2 + x + \sqrt{2}}{\cos(x)}$ **12.** $\frac{2}{x^3} + \frac{3}{x^4}$ **13.** $x^2\sin(x)e^x$ **14.** $\frac{\sqrt{x}}{e^x}$ **15.** $\sqrt{x}\sin(x)$

---

16. Differentiate the following functions: (a) $f(x) = \frac{(1+\sqrt{x})(x^3 + \sin(x))}{x^2 + 5x + 3e^x}$ and (b) $g(x) = \frac{(3+4x+5x^2)e^x}{\sin(x)}$.

17. Use the quotient rule to obtain the following derivatives.
(a) $D(\cot(x)) = -(\csc(x))^2$ (b) $D(\sec(x)) = \sec(x)\tan(x)$ (c) $D(\csc(x)) = -\csc(x)\cot(x)$

OBSERVATION: There is a pattern here: the minus sign goes with each "co" function (cos, cot, csc).

---

18. Find $\left(e^{2x}\right)'$ by writing e^{2x} as $e^x e^x$.

19. Find $\left(e^{3x}\right)'$ by writing e^{3x} as $e^x e^{2x}$.

20. Find $\left(e^{3x}\right)'$ by writing e^{3x} as $e^x e^x e^x$.

21. Find $(e^{-x})'$ by writing e^{-x} as $\frac{1}{e^x}$.

22. Find $\left(e^{-2x}\right)'$ by writing $e^{-2x} = e^{-x} \cdot e^{-x}$.

23. Find $\left(e^{-2x}\right)'$ by writing $e^{-2x} = \frac{1}{e^{2x}}$.

In Exercises 24 to 42 find the derivative of the function using formulas from this section. Simplify all answers.

24. $2^3 - \sqrt{\pi}$ **25.** $(x - x^{-1})^2$ **26.** $3\sin(x) - 5\cos(x)$ **27.** $5\tan(x)$

28. $u^5 - 6u^3 + u - 7$ **29.** $\frac{t^8}{8}$ **30.** $\frac{s^{-7}}{-7}$ **31.** $\sqrt{t}(t+4)$

32. $\frac{5}{u^5}$ **33.** $(x^3)^{1/2}$ **34.** $6\tan(x)$ **35.** $3\sec(x) - 4\cos(x)$

36. $\sec^2(\theta) - \tan^2(\theta)$ **37.** $(3x)^4$ **38.** u^2e^u **39.** $\frac{e^t\sin(t)}{\sqrt{t}}$

40. $(3+x^5)e^{-x}\tan(x)$ **41.** $(x-x^2)^3$ **42.** $\frac{\sqrt[3]{x}}{\sqrt[5]{x}}$

---

43. In Section 3.1 we showed that $D\left(\frac{1}{x}\right) = -\frac{1}{x^2}$. Obtain this formula by using the quotient rule.

44. At what point on the graph of $y = xe^{-x}$ is the tangent horizontal?

45. Using the formula for the derivative of a product, obtain the formula for $(fgh)'$.

46. If you had lots of time, how would you differentiate $(1+2x)^{100}$ using the formulas developed so far?

NOTE: In Section 3.5 we will obtain a shortcut for differentiating this function.

47. Obtain the formula for $(f-g)'$ by first writing $f-g$ as $f+(-1)g$.

48. Using the definition of the derivative as a limit, show that $(f-g)' = f' - g'$.

49. Using the definition of the derivative that makes use of both x and $x+h$, obtain the formula for differentiating the sum of two functions.

50. Using the definition of the derivative as $f'(a) = \lim_{x\to a} \frac{f(x)-f(a)}{x-a}$, obtain the formula for differentiating the product of two functions.

Exercises 51 to 53 are examples of **proof by mathematical induction**. In this technique the truth of the statement for n is used to prove the truth of the statement for $n+1$.

51. In Section 3.2 we showed that $D(x^n) = nx^{n-1}$, when n is a positive integer. Now that we have the formula for the derivative of a product of two functions we can obtain this result more easily.

(a) Show, using the definition of the derivative, that the formula $D(x^n) = nx^{n-1}$ holds when $n = 1$.
(b) Using (a) and the formula for the derivative of a product, show that the formula holds when $n = 2$.
(c) Using (b) and the formula for the derivative of a product, show that it holds when $n = 3$.
(d) Show that if it holds for some positive integer n, it also holds for $n+1$.
(e) Combine (c) and (d) to show that the formula holds for $n = 4$.
(f) Why must it hold for $n = 5$?
(g) Why must it hold for all positive integers?

52. Using induction, as in Exercise 51, show that for each positive integer n, $D(x^{-n}) = -nx^{-n-1}$.

53. Using induction, as in Exercise 51, show that for each positive integer n, $D(\sin^n(x)) = n\sin^{n-1}(x)\cos(x)$.

---

54. In Exercise 10 the formula for $\left(\frac{f}{g}\right)'$ was obtained by writing $\frac{f}{g}$ as the product of f and $\frac{1}{g}$. Obtain $\left(\frac{f}{g}\right)'$ directly from the definition of the derivative.

55. For every x in $[0.001, 0.002]$ the polynomial $P(x)$ has the value 0.
(a) Show that $P(x) = 0$ for all x in $(-\infty,\infty)$. (b) Justify the last step in Exercise 41(e) in Section 1.3.

3.4 The Chain Rule

We come now to the most important formula for computing derivatives. For example, it will help us to find the derivative of $(1+x^2)^{100}$ without having to multiply out one hundred copies of $(1+x^2)$. You might expect the derivative of $(1+x^2)^{100}$ to be $100(1+x^2)^{99}$. This cannot be right.

When you expand $(1+x^2)^{100}$ you get a polynomial of degree 200, so its derivative is a polynomial of degree 199. But when you expand $(1+x^2)^{99}$ you get a polynomial of degree 198. Something is wrong.

At this point we know the derivative of $\sin(x)$, but what is the derivative of $\sin(x^2)$? It is not the cosine of x^2. In this section we obtain a way to differentiate these functions easily.

The key is that both $(1+x^2)^{100}$ and $\sin(x^2)$ are composite functions. This section shows how to differentiate composite functions.

How to Differentiate a Composite Function

The composite function $y = (f\circ g)(x) = f(g(x))$ is built up by setting $u = g(x)$ and $y = f(u)$. The derivative of y with respect to x is the limit of $\Delta y/\Delta x$ as Δx approaches 0. The change in Δx causes a change Δu in u, which in turn causes the change Δy in y. (See Figure 3.4.1.) If Δu is not zero, then we may write

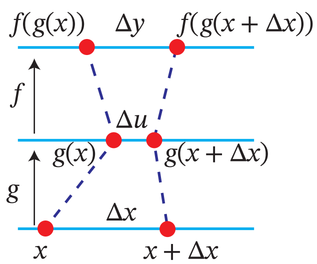


Figure 3.4.1

$$\frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta u}\frac{\Delta u}{\Delta x}.$$

Then

$$(f\circ g)'(x) = \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} = \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta u}\frac{\Delta u}{\Delta x} = \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta u}\lim_{\Delta x\to 0}\frac{\Delta u}{\Delta x}.$$

Since g is continuous, $\Delta u \to 0$ as $\Delta x \to 0$. So we have

$$(f\circ g)'(x) = \lim_{\Delta u\to 0}\frac{\Delta y}{\Delta u}\lim_{\Delta x\to 0}\frac{\Delta u}{\Delta x} = f'(u)g'(x).$$

Which gives us the following important general result.

Theorem 3.4.1: Chain Rule

Let g be differentiable at x and f be differentiable at $g(x)$. Then

$$(f\circ g)'(x) = f'(g(x))g'(x).$$

NOTE: It could happen that $\Delta u = 0$, as it would, for instance, if g were a constant function. This special case is treated in Exercise 75.

Algorithm: How To Differentiate a Composite Function

The chain rule tells us how to differentiate a composite function, $f\circ g$:

Step 1: Compute the derivative of the outer function f, evaluated at the inner function. This is $f'(g(x))$.

Step 2: Compute the derivative of the inner function, $g'(x)$.

Step 3: Multiply the derivatives found in Steps 1 and 2, obtaining $f'(g(x))g'(x)$.

In short, to differentiate $f(g(x))$, think of g as the inner function and f as the outer function. Then

$$\underbrace{f'(g(x))}_{\text{derivative of outer function evaluated at inner function}} \quad \text{times} \quad \underbrace{g'(x).}_{\text{derivative of inner function}}$$

The key to mastering the chain rule is practice. The five examples in this section are designed to help you to develop confidence and a solid routine.

EXAMPLE 1. Find $D\left((1+x^2)^{100}\right)$.

SOLUTION First, note that the derivative cannot be just $100(1+x^2)^{99}$. Why not? Because $(1+x^2)^{100}$ is a polynomial with degree 200, its derivative will be a polynomial with degree 199, but $100(1+x^2)^{99}$ has degree 198.

To use the chain rule to find the derivative of $(1+x^2)^{100}$, choose $g(x) = 1+x^2$ (the inner function) and $f(u) = u^{100}$ (the outer function). The first step is to compute $f'(u) = 100u^{99}$, which gives us $f'(g(x)) = 100(1+x^2)^{99}$. The second step is to find $g'(x) = 2x$. Then, by the chain rule,

$$D\left((1+x^2)^{100}\right) = (f\circ g)'(x) = f'(\underbrace{u}_{u=g(x)})g'(x) = \underbrace{100u^{99}}_{f'(u)}\cdot\underbrace{2x}_{g'(x)} = \underbrace{100(1+x^2)^{99}}_{f'(g(x))}\cdot\underbrace{2x}_{g'(x)} = 200x(1+x^2)^{99}.$$

Note that the result obtained by the chain rule, with the factor of $2x$ from the derivative of the inner function, is a polynomial with degree 199, as expected. ♦

The same example, done with Leibniz notation, looks like this: To find dy/dx, write

$$y = (1+x^2)^{100} = u^{100}, \quad \text{where} \quad u = 1+x^2.$$

Then the chain rule reads

$$\frac{dy}{dx} = \underbrace{\frac{dy}{du}\frac{du}{dx}}_{\text{chain rule}} = 100u^{99}\cdot 2x = \underbrace{100(1+x^2)^{99}}_{u=1+x^2}(2x) = 200x(1+x^2)^{99}.$$

Warning: $\frac{dy}{dx}$ is Not a Fraction

We avoided using Leibniz notation earlier, in particular, during the derivation of the chain rule, because it tempts the reader to cancel the du's in $\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$. The expressions dy, du, and dx will be given a meaning in Section 5.5, but even then it will be important to remember that $\frac{dy}{dx}$ is not a fraction.

In Leibniz's time, in the late seventeenth century, their meaning was fuzzy, standing for a quantity that was zero and also vanishingly small at the same time. Bishop Berkeley poked fun at this in his 1734 book *The Analyst: A Discourse Addressed to an Infidel Mathematician*, where he asks "may we not call them the ghosts of departed quantities?"

As you get more practice applying the chain rule — and the other differentiation rules — you will be able to find the derivative of almost any function without introducing extra symbols, such as u, which do not appear in the final answer. You will be writing just

$$D\left((1+x^2)^{100}\right) = 100(1+x^2)^{99}\cdot 2x = 200x(1+x^2)^{99}.$$

Developing this skill, like playing the guitar, takes more practice. We will provide four more examples, but then you should use the exercises at the end of this section (and chapter) to get more practice on your own.

Observation 3.4.2: *Multiple Roles of u*

When we write dy/du and du/dx, the u serves two roles. In dy/du it denotes an independent variable while in du/dx, u is a dependent variable. The different roles of the auxiliary variable in the chain rule usually causes no problems in computing derivatives.

EXAMPLE 2. If $y = \sin\left(x^2\right)$, find $\frac{dy}{dx}$.

SOLUTION The outer function is the sine and the inner function is x^2. So we have, in short,

$$(\underbrace{\sin}_{\text{outer}}\ \underbrace{(x^2)}_{\text{inner}})' = \underbrace{\cos(x^2)}_{\substack{\text{derivative of outer function}\\ \text{evaluated at inner function}}} \quad \text{times} \quad \underbrace{2x}_{\text{derivative of inner function}} = 2x\cos(x^2).$$

Alternatively, without explicitly identifying the outer and inner functions, write $y = \sin(u)$ and $u = x^2$. Then, by the chain rule,

$$\left(\sin(x^2)\right)' = \frac{dy}{dx} = \underbrace{\frac{dy}{du}\frac{du}{dx}}_{\text{chain rule}} = \cos(u)\cdot 2x = \cos(x^2)\cdot 2x = 2x\cos(x^2).$$

◆

The chain rule holds for compositions of more than two functions. We illustrate this in the next example.

EXAMPLE 3. Differentiate $y = \sqrt{\sin\left(x^2\right)}$.

SOLUTION This function is the composition of three functions:

$$u = x^2, \qquad v = \sin(u), \quad \text{and} \quad y = \sqrt{v}.$$

Then

$$\begin{aligned}
\frac{dy}{dx} &= \underbrace{\frac{dy}{dv}\frac{dv}{dx}}_{\text{chain rule}} \\
&= \frac{dy}{dv}\underbrace{\frac{dv}{du}\frac{du}{dx}}_{\substack{\text{chain rule,}\\ \text{again}}} \\
&= \underbrace{\frac{1}{2\sqrt{v}}}_{\substack{\text{derivative of}\\ \text{outer function}}} \cdot \underbrace{\cos(u)}_{\substack{\text{derivative of}\\ \text{middle function}}} \cdot \underbrace{2x}_{\substack{\text{derivative of}\\ \text{inner function}}} \\
&= \underbrace{\frac{1}{2\sqrt{\sin(x^2)}}}_{v=\sin(u)} \cdot \underbrace{\cos(x^2)}_{u=x^2} \cdot 2x \\
&= \frac{x\cos(x^2)}{\sqrt{\sin(x^2)}}.
\end{aligned}$$

♦

SUGGESTION: Before proceeding, it is highly recommended that you redo Example 3 on your own, this time without introducing any auxiliary symbols (u, v, and y).

EXAMPLE 4. Let $y = 2^x$. Find y'.

SOLUTION As it stands, 2^x is not a composite function. However, we can write $2 = e^{\ln(2)}$ and then 2^x equals $\left(e^{\ln(2)}\right)^x = e^{\ln(2)x}$, so 2^x can be expressed as the composite function

$$y = e^u, \quad \text{where} \quad u = (\ln(2))x.$$

Then

$$y' = \frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = e^u \cdot \ln(2) = e^{\ln(2)x}\ln(2) = 2^x \ln(2).$$

♦

Observation 3.4.3:

In Example 2 (Section 3.2), using a calculator, we found $D\left(2^x\right) \approx (0.693)2^x$. We have just seen that the exact formula is $D\left(2^x\right) = 2^x \ln(2)$. The fact that 0.693 is an approximation of $\ln(2)$ shows these results are consistent.

Example 5 shows how the chain rule is used in combination with other differentiation rules such as the product and quotient rules.

EXAMPLE 5. Find $D\left(x^3\tan(x^2)\right)$.

SOLUTION The function $x^3\tan(x^2)$ is the product of two functions. We first apply the product rule to obtain

$$D\left(x^3\tan(x^2)\right) = \left(x^3\right)'\tan(x^2) + x^3\left(\tan(x^2)\right)' = 3x^2\tan(x^2) + x^3\left(\tan(x^2)\right)'.$$

Since the derivative of the tangent is the square of the secant, the chain rule tells us that

$$\left(\tan(x^2)\right)' = \sec^2(x^2)(x^2)' = 2x\sec^2(x^2).$$

Thus,

$$D\left(x^3\tan(x^2)\right) = 3x^2\tan(x^2) + x^3\left(\tan(x^2)\right)' = 3x^2\tan(x^2) + x^3\left(2x\sec^2(x^2)\right) = 3x^2\tan(x^2) + 2x^4\sec^2(x^2).$$

◆

In the computation of $D(\tan(x^2))$ we did not introduce any new symbols. That is how your computations will look, once you become proficient using the chain rule.

Common Composite Functions

Certain types of composite functions occur so often that it is worthwhile memorizing their derivatives.

Function	Derivative	Example
$(g(x))^n$	$ng(x)^{n-1}g'(x)$	$\left((1+x^2)^{100}\right)' = 100(1+x^2)^{99}(2x)$
$\dfrac{1}{g(x)}$	$\dfrac{-g'(x)}{(g(x))^2}$	$\left(\dfrac{1}{\cos(x)}\right)' = \dfrac{-(-\sin(x))}{(\cos(x))^2}$
$\sqrt{g(x)}$	$\dfrac{g'(x)}{2\sqrt{g(x)}}$	$\left(\sqrt{\tan(x)}\right)' = \dfrac{(\sec(x))^2}{2\sqrt{\tan(x)}}$
$e^{g(x)}$	$e^{g(x)}g'(x)$	$\left(e^{x^2}\right)' = e^{x^2}(2x)$

Table 3.4.1

Summary

This section presented the single most important tool for computing derivatives: the chain rule, which says that the derivative of $f\circ g$ at x is

$$\left(f(g(x))\right)' = \underbrace{f'(g(x))}_{\substack{\text{derivative of outer function}\\ \text{evaluated at inner function}}} \quad \text{times} \quad \underbrace{g'(x)}_{\text{derivative of inner function}}$$

Introducing the symbol u, we described the chain rule for $y = f(u)$ and $u = g(x)$ with the brief notation

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}.$$

When the function is built up from more than two functions, such as $y = f(u)$, $u = g(v)$, and $v = h(x)$, we have

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dv}\frac{dv}{dx},$$

a chain of more derivatives.

With practice, applying the chain rule becomes second nature.

EXERCISES for Section 3.4

In Exercises 1 to 4, repeat the specified example from this section without introducing an auxiliary variable.

1. Example 2 **2.** Example 3 **3.** Example 4 **4.** Example 5

In Exercises 5 to 40 find the derivative of the function. Simplify your answers.

5. $(x^3+2)^5$ **6.** $(x^2+3x+1)^4$ **7.** $\sqrt{\cos(x^3)}$ **8.** $\sqrt{\tan(x^2)}$

9. $\left(\frac{1}{x}\right)^{10}$ **10.** $\cos(3x)\sin(2x)$ **11.** $x^2\tan(x^3)$ **12.** $(1+2x)\sin(x^4)$

13. $5(\tan(x^3))^2$ **14.** $\frac{\cos^3(2x)}{x^5}$ **15.** $\sin(2\exp(x))$ **16.** $e^{\cos(x)}$

17. $\frac{(1+2x)^2}{x^3}$ **18.** $\frac{\sec(5x)}{\cos(5x)}$ **19.** $(5x^2+3)^{10}$ **20.** $(\sin(3x))^3$

21. $\frac{1}{5t^2+t+2}$ **22.** $\frac{1}{e^{5s}+s}$ **23.** $\sqrt{4+u^2}$ **24.** $\left(\sqrt{\cos(2\theta)}\right)^3$

25. e^{5x^3} **26.** $\sin^2(3x)$ **27.** $e^{\tan(3t)}$ **28.** $\sqrt{\tan(2u)}$

29. $\sqrt[3]{\tan(s^2)}$ **30.** $v^3\tan(2v)$ **31.** $e^{2r}\sin(3r)$ **32.** $\frac{\sec(2x)}{x^2}$

33. $\exp(\sin(2x))$ **34.** $\frac{(3t+2)^4}{\sin(2t)}$ **35.** $e^{-5s}\tan(3s)$ **36.** e^{x^2}

37. $(\sin(2u))^5(\cos(3u))^6$ **38.** $\left(x+3^{3x}\right)^2\left(\sin(\sqrt{x})\right)^3$ **39.** $\frac{t^3}{t+\sin^2(3t)}$ **40.** $\frac{(3x+2)^4}{(x^3+x+1)^2}$

Learning to use the chain rule takes practice. Exercises 41 to 68 offer opportunities. They also show that sometimes the derivative of a function can be simpler than the function. In each case show that the derivative of the first expression is the second expression, the functions being separated by a semi-colon. The letters a, b, and c denote constants.

41. $\frac{b}{2a^2(ax+b)^2} - \frac{1}{a^2(ax+b)}$; $\frac{x}{(ax+b)^2}$

42. $\frac{-1}{2a(ax+b)^2}$; $\frac{1}{(ax+b)^3}$

43. $\frac{2}{3a}\sqrt{(ax+b)^3}$; $\sqrt{ax+b}$

44. $\frac{2(3ax-2b)}{15a^2}\sqrt{(ax+b)^3}$; $x\sqrt{ax+b}$

45. $\frac{-\sqrt{ax^2+c}}{cx}$; $\frac{1}{x^2\sqrt{ax^2+c}}$

46. $\frac{x}{c\sqrt{ax^2+c}}$; $(ax^2+c)^{-3/2}$

47. $\frac{1}{a}\sin(ax) - \frac{1}{3a}\sin^3(ax)$; $\cos^3(ax)$

48. $\frac{1}{a(n+1)}\sin^{n+1}(ax)$; $\sin^n(ax)\ \cos(ax)$

49. $\frac{2(ax-2b)}{3a^2}\sqrt{ax+b}$; $\frac{x}{\sqrt{ax+b}}$

50. $\frac{2(3a^2x^2-4abx+8b^2)}{15a^3}\sqrt{ax+b}$; $\frac{x^2}{\sqrt{ax+b}}$

51. $\frac{-\sqrt{ax^2+c}}{cx}$; $\frac{1}{x^2\sqrt{ax^2+c}}$

52. $\frac{-x^2}{a\sqrt{ax^2+c}}+\frac{2}{a^2}\sqrt{ax^2+c}$; $\frac{x^3}{(ax^2+c)^{3/2}}$

53. $\frac{x}{2}+\frac{\sin(2ax)}{4a}$; $\cos^2(ax)$

54. $\frac{3x}{8}-\frac{3\sin(2ax)}{16a}-\frac{\sin^3(ax)\cos(ax)}{4a}$; $\sin^4(ax)$

55. $\frac{1}{a}\tan(ax)$; $\frac{1}{\cos^2(ax)}$

56. $\frac{1}{a}\tan\left(\frac{ax}{2}\right)$; $\frac{1}{1+\cos(ax)}$

57. $\frac{\sin((a-b)x)}{2(a-b)}-\frac{\sin((a+b)x)}{2(a+b)}$; $\sin(ax)\sin(bx)$ $\quad (a^2\neq b^2)$

58. $\frac{\sin((a-b)x)}{2(a-b)}+\frac{\sin((a+b)x)}{2(a+b)}$; $\cos(ax)\cos(bx)$ $\quad (a^2\neq b^2)$

59. $\frac{1}{a}(\tan(ax)-\cot(ax))$; $\frac{1}{\sin^2(ax)\cos^2(ax)}$

60. $\frac{1}{a}\tan(ax)-x$; $\tan^2(ax)$

61. $\frac{\sec^n(ax)}{an}$; $\tan(ax)\sec^n(ax)$ $\quad (n\neq 0)$

62. $\frac{-1}{a}\cos(ax)+\frac{1}{3a}\cos^3(ax)$; $\sin^3(ax)$

63. $\frac{\sin(ax)}{a^2}-\frac{x\cos(ax)}{a}$; $x\sin(ax)$

64. $\frac{\cos(ax)}{a^2}+\frac{x\sin(ax)}{a}$; $x\cos(ax)$

65. $\frac{1}{a^2}e^{ax}(ax-1)$; xe^{ax}

66. $\frac{1}{a^3}e^{ax}(a^2x^2-2ax+2)$; x^2e^{ax}

67. $2\sqrt{2}\sin\left(\frac{x}{2}\right)$; $\sqrt{1+\cos(x)}$ $\quad (|x|\leq\pi)$

68. $\frac{e^{ax}(a\sin(bx)-b\cos(bx))}{a^2+b^2}$; $e^{ax}\sin(bx)$

Exercises 69 and 70 illustrate how differentiation can be used to obtain one trigonometry identity from another.

69. (a) Differentiate both sides of the identity $\sin^2(x)=\frac{1}{2}(1-\cos(2x))$. What trigonometric identity do you get?

(b) Differentiate the identity found in (a) to obtain another trigonometric identity. What identity is obtained?

(c) Does this process continued forever produce new identities?

70. Let k be a constant. Differentiate both sides of the identity $\sin(x+k)=\sin(x)\cos(k)+\cos(x)\sin(k)$ to obtain the corresponding identity for $\cos(x+k)$.

---

71. Differentiate $(e^x)^3$ by the following three ways. Which of these approaches do you prefer? Why?

(a) Directly, by the chain rule

(b) After writing it as $e^x\cdot e^x\cdot e^x$ and using the product rule

(c) After writing it as e^{3x} and using the chain rule

72. In Section 3.3 we obtained the derivative of $\frac{1}{g(x)}$ by using the definition of the derivative. Obtain that formula for the reciprocal rule by using the chain rule.

73. In our proof of the chain rule we had to assume that Δu is not 0 when Δx is sufficiently small. Show that if the derivative of g is not 0 at the argument x, then the proof is valid.

74. Here is an example of a differentiable function g not covered by the proof of the chain rule given in the text. Define $g(x)$ to be $x^2\sin\left(\frac{1}{x}\right)$ for x different from 0 and $g(0)$ to be 0.

(a) Sketch the part of the graph of g near the origin.

(b) Show that there are arbitrarily small values of Δx such that $\Delta u=g(\Delta x)-g(0)=0$.

(c) Show that g is differentiable at 0.

75. Here is a proof of the chain rule that manages to avoid division by $\Delta u = 0$. Let $f(u)$ be differentiable at $g(a)$, where g is differentiable at a. Let $\Delta f = f(g(a)+\Delta u) - f(g(a))$. Then $\frac{\Delta f}{\Delta u} - f'(g(a))$ is a function of Δu, which we call $p(\Delta u)$. It is defined for $\Delta u \neq 0$. By the definition of f', $p(\Delta u)$ tends to 0 as Δu approaches 0. Define $p(0)$ to be 0. Note that p is continuous at 0.

(a) Show that $\Delta f = f'(g(a))\Delta u + p(\Delta u)\Delta u$ when Δu is different than 0, and also when $\Delta u = 0$.

(b) Define $q(\Delta x) = \frac{\Delta u}{\Delta x} - g'(a)$. Observe that $q(\Delta x)$ approaches 0 as Δx approaches 0. Show that $\Delta u = g'(a)\Delta x + q(\Delta x)\Delta x$ when Δx is not 0.

(c) Combine (a) and (b) to show that $\Delta f = f'(g(a))\left(g'(a)\Delta x + q(\Delta x)\Delta x\right) + p(\Delta u)\Delta u$.

(d) Using (c), show that $\lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = f'(g(a))g'(a)$.

(e) Why did we have to define $p(0)$ but not $q(0)$?

3.5 Derivative of an Inverse Function

In this section we obtain the derivatives of the inverse functions of e^x and of the six trigonometric functions. This will complete the inventory of basic derivatives. The chain rule will be our main tool.

Differentiability of Inverse Functions

As mentioned in Section 1.1, the graph of an inverse function is a copy of the graph of the original function. One graph is obtained from the other by reflection across the line $y = x$. If the original function, f, is differentiable at a point (a, b), $b = f(a)$, then the graph of $y = f(x)$ has a tangent line at (a, b). The reflection across $y = x$ of the tangent line to the graph of f is the tangent line to the inverse function at (b, a). Thus, we expect that the inverse function, f^{-1}, is differentiable at (b, a), and we will assume it is.

RECALL: $a = f^{-1}(b)$ means $b = f(a)$.

We could find a general formula for the derivative of an inverse function, we find the derivative of the inverses of the exponential function (base e and base $b > 0$), the sine function, and the tangent function. The process used in these examples will be more useful than a general formula.

First, the chain rule will be used to find the derivative of $\log_e(x)$.

The Derivative of $\log_e(x)$

We want to find $y' = dy/dx$ where $y = \log_e(x)$. By the definition of logarithm as the inverse of the exponential function,

$$x = e^y. \tag{3.5.1}$$

We differentiate both sides of (3.5.1) with respect to x:

$$\begin{aligned} \frac{d(x)}{dx} &= \frac{d(e^y)}{dx} && (\ y \text{ is a function of } x\text{, so } e^y \text{ is a function of } x\ ) \\ 1 &= \frac{d\left(e^y\right)}{dx} && (\ dx/dx = 1\ ) \\ 1 &= e^y \frac{dy}{dx} && (\ \text{chain rule}\ ). \end{aligned}$$

Solving for dy/dx, we obtain

$$\frac{dy}{dx} = \frac{1}{e^y} = \frac{1}{x}.$$

This is another important differentiation rule.

Formula 3.5.1: *Derivative of* $\log_e(x)$

$$\left(\log_e(x)\right)' = \frac{1}{x}, \qquad x > 0$$

It may come as a surprise that such a complicated function has a simple derivative. It may also be a surprise that $\log_e(x)$ is one of the most important functions in calculus, principally because it has the derivative $1/x$.

EXAMPLE 1. Find $\left(\log_b(x)\right)'$ for any $b > 0$.

SOLUTION The function $\log_b x$ is a constant times $\log_e(x)$:

$$\log_b(x) = \left(\log_b(e)\right)\log_e(x).$$

Therefore

$$\left(\log_b(x)\right)' = \left(\log_b(e)\right)\frac{1}{x}.$$

If b is not e, then $\log_b(e)$ is not 1. If e is chosen as the base for logarithms, then the coefficient of $1/x$ becomes $\log_e(e) = 1$. That is another reason why we prefer e as the base for logarithms in calculus ♦

Recall from Section 2.2 that $\log_e(x)$ is the natural logarithm, which is denoted $\ln(x)$.

Warning: (Logarithm Notation)

$\ln(x)$ is often written simply as $\log(x)$, with the base understood to be e. All references in this book to the base-e logarigtm will use the notation ln and all references to the base-10 logarithm will use the notation $\log_{10}$.

The Derivative of $\arcsin(x)$

Inverse trigonometric functions were introduced in Section 1.2.

For x in $[-\pi/2, \pi/2]$, $\sin(x)$ is one-to-one and therefore has an inverse function, $\arcsin(x)$, which gives the angle, in radians, if you know the sine of the angle. For instance, $\arcsin(1) = \pi/2$, $\arcsin\left(\sqrt{2}/2\right) = \pi/4$, $\arcsin(-1/2) = -\pi/6$, and $\arcsin(-1) = -\pi/2$. The domain of $\arcsin(x)$ is $[-1, 1]$ and its range is $[-\pi/2, \pi/2]$.

For convenience, Figure 3.5.1 shows the graphs of $y = \sin(x)$ and $y = \arcsin(x)$. While these graphs will not be needed to find $D(\arcsin(x))$, they are helpful to verify some of the properties of the derivative formula we will obtain. For example, the derivative of $\arcsin(x)$ should be 1 when $x = 0$ and undefined at the endpoints, $x = \pi/2$ and $x = -\pi/2$.

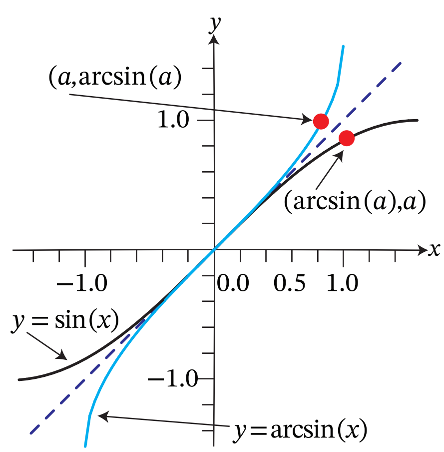


Figure 3.5.1

To find $D(\arcsin(x))$, we proceed as we did when finding the derivative of the logarithm function. Let $y = \arcsin(x)$, so $x = \sin(y)$. We then have

$$\begin{aligned}
x &= \sin(y) && \text{(inverse relation for } y = \sin(x)\text{)}\\
\frac{d(x)}{dx} &= \frac{d(\sin(y))}{dx} && \text{(differentiate with respect to } x\text{)}\\
1 &= \left(\cos(y)\right)y' && \text{(chain rule)}\\
y' &= \frac{1}{\cos(y)} && \text{(algebra).}
\end{aligned}$$

All that remains is to express $\cos(y)$ in terms of $x = \sin(y)$. From the relation $\cos^2(y) + \sin^2(y) = 1$, we conclude that $\cos(y) = \pm\sqrt{1-\sin^2(y)} = \pm\sqrt{1-x^2}$. We use the positive value: $\cos(y) = \sqrt{1-x^2}$ because arcsin is an increasing function. Consequently, we find

Formula 3.5.2: *Derivative of* arcsin(x)

$$\frac{d}{dx}(\arcsin(x)) = \frac{1}{\sqrt{1-x^2}}, \qquad |x| < 1$$

At $x = 1$ or at $x = -1$, the derivative is not defined. However, for x near 1 or -1 the derivative is very large (in absolute value), telling us that the graph of the arcsine function is very steep near its two ends. That is a reflection of the fact that the graph of $\sin(x)$ is horizontal at $x = -\pi/2$ and $x = \pi/2$. (See Figure 3.5.1.)

Functions such as $x^3 - x$, $x^{2/7}$, and $1/\sqrt{1-x^2}$ that can be written in terms of the algebraic operations of addition, subtraction, multiplication, division, raising to a power, and extracting a root are examples of algebraic functions. More precisely, an **algebraic function** is a function $y = f(x)$ defined implicitly as roots of an equation of the form $a_0(x) + a_1(x)y + \cdots + a_n(x)y^n = 0$, where each $a_i(x)$ is a polynomial in x. Functions that cannot be written in this way, including e^x, $\cos(x)$, and $\arcsin(x)$, are known as **transcendental functions**. The derivatives of $\ln(x)$ and $\arcsin(x)$ show that the derivative of a transcendental function can be an algebraic function. The derivative of an algebraic function will always be an algebraic function.

EXAMPLE 2. Differentiate $\arcsin\left(x^2\right)$.

SOLUTION By the chain rule,

$$\frac{d}{dx}\left(\arcsin\left(x^2\right)\right) = \frac{1}{\sqrt{1-\left(x^2\right)^2}} \cdot \frac{d}{dx}\left(x^2\right) = \frac{2x}{\sqrt{1-x^4}}.$$

♦

EXAMPLE 3. Differentiate $\frac{1}{2}\left(x\sqrt{a^2-x^2} + a^2 \arcsin\left(\frac{x}{a}\right)\right)$ where a is a positive constant, and $|x| < a$.

SOLUTION

$$\begin{aligned}
D\left(\frac{1}{2}\left(x\sqrt{a^2-x^2} + a^2\arcsin\left(\frac{x}{a}\right)\right)\right) &= D\left(\frac{x}{2}\sqrt{a^2-x^2}\right) + D\left(\frac{a^2}{2}\arcsin\left(\frac{x}{a}\right)\right) && (\text{ sum rule })\\
&= D\left(\frac{x}{2}\sqrt{a^2-x^2}\right) + \frac{a^2}{2}D\left(\arcsin\left(\frac{x}{a}\right)\right) && (\text{ constant multiple rule })\\
&= \left(\frac{1}{2}\sqrt{a^2-x^2} + \frac{x}{2}\left(\frac{\frac{1}{2}(-2x)}{\sqrt{a^2-x^2}}\right)\right) && (\text{ product and chain rules })\\
&\quad + \frac{a^2}{2}\left(\frac{1/a}{\sqrt{1-(x/a)^2}}\right) && (\text{ Formula 3.5.2; chain rule })\\
&= \frac{1}{2}\left(\sqrt{a^2-x^2} + \frac{-x^2}{\sqrt{a^2-x^2}} + \frac{a^2}{\sqrt{a^2-x^2}}\right) && (\text{ algebra })\\
&= \frac{1}{2}\left(\frac{a^2-x^2-x^2+a^2}{\sqrt{a^2-x^2}}\right) && (\text{ common denominator })\\
&= \sqrt{a^2-x^2}.
\end{aligned}$$

♦

We are quickly learning to not be surprised when a complicated function has a (relatively) simple derivative.

The Derivative of arctan(x)

For x in $(-\pi/2, \pi/2)$ the function $\tan(x)$ is one-to-one and has an inverse function, $\arctan(x)$, which tells us the angle, in radians, if we know the tangent of the angle. For instance, $\arctan(1) = \pi/4$, $\arctan(0) = 0$, and $\arctan(-1) = -\pi/4$. When x is a large positive number, $\arctan(x)$ is near, and smaller than, $\pi/2$. When x is a large negative number, $\arctan(x)$ is slightly larger than $-\pi/2$. Figure 3.5.2 shows the graph of $y = \arctan(x)$ and $y = \tan(x)$. We will not need this graph when differentiating $\arctan(x)$, but it serves as a check on the formula.

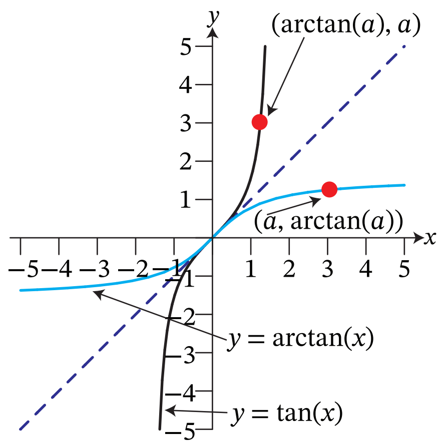


Figure 3.5.2

To find $(\arctan(x))'$, we again use the chain rule. Starting with $y = \arctan(x)$, we proceed as before:

$$\begin{aligned}
x &= \tan(y) \\
\frac{d(x)}{dx} &= \frac{d(\tan(y))}{dx} && \text{(differentiate with respect to } x \text{)} \\
1 &= \left(\sec^2(y)\right) y' && \text{(chain rule)} \\
y' &= \frac{1}{\sec^2(y)} && \text{(algebra)} \\
y' &= \frac{1}{1+\tan^2(y)} && \text{(trigonometric identity)} \\
y' &= \frac{1}{1+x^2} && \text{(} x = \tan(y) \text{).}
\end{aligned}$$

This derivation is summarized by a simple formula, which should be memorized.

Formula 3.5.3: *Derivative of* arctan(x)

$$D(\arctan(x)) = \frac{1}{1+x^2} \qquad \text{for all inputs } x.$$

EXAMPLE 4. Find $D(\arctan(3x))$.

SOLUTION By the chain rule

$$D(\arctan(3x)) = \frac{1}{1+(3x)^2}\frac{d(3x)}{dx} = \frac{3}{1+9x^2}.$$

◆

EXAMPLE 5. Find $D\left(x\tan^{-1}(x) - \frac{1}{2}\ln\left(1+x^2\right)\right)$.

SOLUTION

$$\begin{aligned}
D\left(x\tan^{-1}(x) - \frac{1}{2}\ln\left(1+x^2\right)\right) &= D\left(x\tan^{-1}(x)\right) - \frac{1}{2}D\left(\ln\left(1+x^2\right)\right) \\
&= \left(\tan^{-1}(x) + \frac{x}{1+x^2}\right) - \frac{1}{2}\frac{2x}{1+x^2} \\
&= \tan^{-1}(x).
\end{aligned}$$

◆

More on $\ln(x)$

An **antiderivative** of a function, $f(x)$, is another function, $F(x)$, whose derivative is equal to $f(x)$. That is, $F'(x) = f(x)$, and so $\ln(x)$ is an antiderivative of $1/x$. We showed that for $x > 0$, $\ln(x)$ is an antiderivative of $1/x$. But what if we needed an antiderivative of $1/x$ for negative x? The next example answers this question.

EXAMPLE 6. Show that for negative x, $\ln(-x)$ is an antiderivative of $\frac{1}{x}$.

SOLUTION Let $y = \ln(-x)$. By the chain rule,

$$\frac{dy}{dx} = \left(\frac{1}{-x}\right)\frac{d(-x)}{dx} = \frac{1}{-x}(-1) = \frac{1}{x}.$$

So $\ln(-x)$ is an antiderivative of $1/x$ when x is negative. ♦

In view of Example 6, $\ln|x|$ is an antiderivative of $1/x$, whether x is positive or negative.

Formula 3.5.4: *Derivative of* $\ln|x|$

$$D(\ln|x|) = \frac{1}{x} \qquad \text{for } x \neq 0.$$

We know the derivative of x^a for any rational number a. To extend this result to x^k for any number k, and positive x, we write x as $e^{\ln(x)}$.

EXAMPLE 7. Find $D(x^k)$ for $x > 0$ and any constant $k \neq 0$, rational or irrational.

SOLUTION For $x > 0$ we can write $x = e^{\ln(x)}$. Then

$$x^k = \left(e^{\ln(x)}\right)^k = e^{k\ln(x)}.$$

Looking at $y = e^{k\ln(x)}$ as a composite function, $y = e^u$ where $u = k\ln(x)$, we have

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = e^u\frac{k}{x} = x^k\frac{k}{x} = kx^{k-1}.$$

♦

The example shows that for positive x and any fixed exponent k, $\left(x^k\right)' = kx^{k-1}$. It probably does not come as a surprise. You may wonder why we worked so hard to get the derivative of x^a when a is an integer or rational number when this example covers all exponents. We had two reasons for treating the special cases. First, they include cases when x is negative. Second, they were simpler and helped introduce the derivative.

Observation 3.5.1: *Another View of* e

For each choice of the base b ($b > 0$), we obtain a value for $\lim_{x\to 0}\frac{b^x - 1}{x}$. We defined e to be the base for which the limit is as simple as possible, namely 1: $\lim_{x\to 0}\frac{e^x - 1}{x} = 1$.

Now that we know that the derivative of $\ln x = \log_e x$ is $1/x$, we can obtain a new view of e.

The derivative of $\ln(x)$ at 1 is $1/1 = 1$. By the definition of the derivative,

$$\lim_{h\to 0}\frac{\ln(1+h) - \ln(1)}{h} = 1.$$

Since $\ln(1) = 0$, we have

$$\lim_{h\to 0}\frac{\ln(1+h)}{h} = 1.$$

By a property of logarithms, we may rewrite the limit as

$$\lim_{h\to 0} \ln\left((1+h)^{1/h}\right) = 1.$$

Writing e^x as $\exp(x)$ for convenience, we conclude that

$$\exp\left(\lim_{h\to 0} \ln\left((1+h)^{1/h}\right)\right) = \exp(1) = e.$$

Since exp is a continuous function, we may switch exp and lim, getting

$$\lim_{h\to 0}\left(\exp\left(\ln\left((1+h)^{1/h}\right)\right)\right) = e.$$

But $\exp(\ln(p)) = p$ for any positive number, by the definition of a logarithm. That tells us that

$$\lim_{h\to 0}(1+h)^{1/h} = e.$$

This is a more direct view of e than the one in Section 2.2. As a check, let $h = 1/1000 = 0.001$. Then $(1+h)^{1/h} = (1+1/1000)^{1000} \approx 2.717$, and values of h closer to 0 give better estimates for e, whose decimal expansion begins 2.718.

The Derivatives of the Six Inverse Trigonometric Functions

Of the six inverse trigonometric functions, the most important are arcsin and arctan. The other four are treated in Exercises 70 to 73. Table 3.5.1 summarizes all six derivatives. There is no reason to memorize the formulas. If we need, say, an antiderivative of $-1/(1+x^2)$, we do not have to use $\text{arccot}(x)$. Instead, $-\arctan(x)$ would do. For finding antiderivatives, we don't need arccot, or any of the inverse co-functions. The formulas for the derivatives of arcsin, arctan, and arcsec suffice.

$D(\arcsin(x))$	$=$	$\dfrac{1}{\sqrt{1-x^2}}$	$(-1<x<1)$	$D(\arccos(x))$	$=$	$\dfrac{-1}{\sqrt{1-x^2}}$	$(-1<x<1)$
$D(\arctan(x))$	$=$	$\dfrac{1}{1+x^2}$	$(-\infty<x<\infty)$	$D(\text{arccot}(x))$	$=$	$\dfrac{-1}{1+x^2}$	$(-\infty<x<\infty)$
$D(\text{arcsec}(x))$	$=$	$\dfrac{1}{x\sqrt{x^2-1}}$	$(x>1 \text{ or } x<-1)$	$D(\text{arccsc}(x))$	$=$	$\dfrac{-1}{x\sqrt{x^2-1}}$	$(x>1 \text{ or } x<-1)$

Table 3.5.1

Summary

A geometric argument suggested that the inverse of every differentiable function is differentiable. The chain rule then helped find the derivatives of $\ln(x)$, $\arcsin(x)$, and $\arctan(x)$ and of the other four inverse trigonometric functions.

EXERCISES for Section 3.5

In Exercises 1 to 6 evaluate the function and its derivative at the given argument.

1. $\arcsin(x)$; $x = 1/2$

2. $\arcsin(t)$; $t = -1/2$

3. $\arctan(\theta)$; $\theta = -1$

4. $\arctan(u)$; $u = \sqrt{3}$

5. $\ln(s)$; $s = e$

6. $\ln(y)$; $y = 1$

In Exercises 7 to 28 differentiate the function. Simplify all answers.

7. $\arcsin(3x)\sin(3x)$

8. $\arctan(5x)\tan(5x)$

9. $e^{2x}\ln(3x)$

10. $e^{\left(\ln(3x)x^{\sqrt{2}}\right)}$

11. $x^2\arcsin(x^2)$

12. $(\arcsin(3x))^2$

13. $\dfrac{\arctan(2x)}{1+x^2}$

14. $\dfrac{x^3}{\arctan(6x)}$

15. $\log_{10}(x)$

16. $\log_x(10)$

17. $\arcsin\left(x^3\right)$

18. $\arctan\left(x^2\right)$

19. $(\arctan(3x))^2$

20. $(\arccos(5x))$

21. $\dfrac{\arcsin\left(1+x^2\right)}{1+3x}$

22. $\operatorname{arcsec}\left(x^3\right)$

23. $x^2\arcsin(3x)$

24. $\dfrac{\arctan(3x)}{\tan(2x)}$

25. $\dfrac{\arctan\left(x^3\right)}{\arctan(x)}$

26. $\ln(\sin(3x))$

27. $\ln\left(\sin(x)^3\right)$

28. $\ln(\exp(4x))$

In Exercise 29 to 64 check that the derivative of the first expression is the second expression, separated by a semicolon. The letters a, b, and c denote constants.

29. $x(\ln(ax))^2 - 2x\ln(ax) + 2x$; $(\ln(ax))^2$

30. $-\dfrac{1}{2}\ln\left(\dfrac{1+\cos(x)}{1-\cos(x)}\right)$; $\dfrac{1}{\sin(x)} = \csc(x)$

31. $x\arctan(ax) - \dfrac{1}{2a}\ln\left(1+a^2x^2\right)$; $\arctan(ax)$

32. $x\arccos(ax) - \dfrac{1}{a}\sqrt{1-a^2x^2}$; $\arccos(ax)$

33. $\dfrac{1}{a}\ln\left(\tan\left(\dfrac{ax}{2}\right)\right)$; $\dfrac{1}{\sin(ax)}$

34. $\ln(\ln(ax))$; $\dfrac{1}{x\ln(ax)}$

35. $\dfrac{-1}{(n-1)(\ln(ax))^{n-1}}$; $\dfrac{1}{x(\ln(ax))^n}$

36. $x\arcsin(ax) + \dfrac{1}{a}\sqrt{1-a^2x^2}$; $\arcsin(ax)$

37. $\dfrac{1}{ab}\left(ax - \ln\left(b+ce^{ax}\right)\right)$; $\dfrac{1}{b+ce^{ax}}$

38. $\dfrac{1}{ab}\arctan\left(\dfrac{bx}{a}\right)$; $\dfrac{1}{a^2+b^2x^2}$

39. $\dfrac{x}{2a^2(a^2+x^2)} + \dfrac{1}{2a^3}\arctan\left(\dfrac{x}{a}\right)$; $\dfrac{1}{(a^2+x^2)^2}$

40. $\dfrac{1}{2a^2}\arctan\left(\dfrac{x^2}{a^2}\right)$; $\dfrac{x}{a^4+x^4}$

41. $\dfrac{2\sqrt{x}}{b^2} - 2\dfrac{a}{b^3}\arctan\left(\dfrac{b\sqrt{x}}{a}\right)$; $\dfrac{\sqrt{x}}{a^2+b^2x}$

42. $\dfrac{1}{a^2}\cos(ax) + \dfrac{x}{a}\sin(ax)$; $x\cos(ax)$

43. $\dfrac{1}{a^3}e^{ax}\left(a^2x^2 - 2ax + 2\right)$; x^2e^{ax}

44. $\arccos\left(\dfrac{a-x}{a}\right)$; $\dfrac{1}{\sqrt{2ax-x^2}}$ $\quad (0 \le x \le 2a)$

45. $\arcsin(x) - \sqrt{1-x^2}$; $\sqrt{\dfrac{1+x}{1-x}}$

46. $\dfrac{1}{ab}\left(ax - \ln\left(b+ce^{ax}\right)\right)$; $\dfrac{1}{b+ce^{ax}}$

47. $\dfrac{1}{a^2+b^2}e^{ax}(a\sin(bx) - b\cos(bx))$; $e^{ax}\sin(bx)$

48. $\ln(\sec(x)+\tan(x))$; $\sec(x)$

49. $\dfrac{1}{cn}\ln\left(\dfrac{x^n}{ax^n+c}\right)$; $\dfrac{1}{x(ax^n+c)}$

50. $\dfrac{1}{a\sqrt{bc}}\arctan\left(e^{ax}\sqrt{\dfrac{b}{c}}\right)$; $\dfrac{1}{be^{ax}+ce^{-ax}}$ $(b, c>0)$

51. $2\arcsin\left(\sqrt{\dfrac{x-b}{a-b}}\right)$; $\dfrac{1}{\sqrt{(a-x)(x-b)}}$ $(b<x<a)$

52. $\dfrac{2}{n\sqrt{-c}}\operatorname{arcsec}\left(\sqrt{\dfrac{ax^n}{-c}}\right)$; $\dfrac{1}{x\sqrt{ax^n+c}}$ $(c<0)$

53. $\dfrac{1}{b^2}(a+bx-a\ln(a+bx))$; $\dfrac{x}{a+bx}$ $(a+bx>0)$

54. $x(\arcsin(ax))^2-2x+\dfrac{2}{a}\sqrt{1-a^2x^2}\arcsin(ax)$; $(\arcsin(ax))^2$

55. $\dfrac{1}{n\sqrt{c}}\ln\left(\dfrac{\sqrt{ax^n+c}-\sqrt{c}}{\sqrt{ax^n+c}+\sqrt{c}}\right)$; $\dfrac{1}{x\sqrt{ax^n+c}}$ $(c>0)$

56. $\sqrt{ax^2+c}+\sqrt{c}\ln\left(\dfrac{\sqrt{ax^2+c}-\sqrt{c}}{x}\right)$; $\dfrac{\sqrt{ax^2+c}}{x}$ $(c>0)$

57. $\sqrt{ax^2+c}-\sqrt{-c}\arctan\left(\dfrac{\sqrt{ax^2+c}}{\sqrt{-c}}\right)$; $\dfrac{\sqrt{ax^2+c}}{x}$ $(c<0)$

58. $\dfrac{2}{\sqrt{4ac-b^2}}\arctan\left(\dfrac{2ax+b}{\sqrt{4ac-b^2}}\right)$; $\dfrac{1}{ax^2+bx+c}$ $(b^2<4ac)$

59. $\dfrac{1}{\sqrt{b^2-4ac}}\ln\left(\dfrac{2ax+b-\sqrt{b^2-4ac}}{2ax+b+\sqrt{b^2-4ac}}\right)$; $\dfrac{1}{ax^2+bx+c}$

60. $\dfrac{1}{2}\left((x-a)\sqrt{2ax-x^2}+a^2\arcsin\left(\dfrac{x-a}{a}\right)\right)$; $\sqrt{2ax-x^2}$ $(0\le x\le 2a)$

61. $\dfrac{1}{b^3}\left(a+bx-2a\ln(a+bx)-\dfrac{a^2}{a+bx}\right)$; $\dfrac{x^2}{(a+bx)^2}$, $(a+bx>0)$

62. $x\operatorname{arcsec}(ax)-\dfrac{1}{a}\ln\left(ax+\sqrt{a^2x^2-1}\right)$; $\operatorname{arcsec}(ax)$ $(ax\ge 1)$

63. $\dfrac{x^2}{2}\arcsin(ax)-\dfrac{1}{4a^2}\arcsin(ax)+\dfrac{x}{4a}\sqrt{1-a^2x^2}$; $x\arcsin(ax)$

64. $x(\arcsin(ax))^2-2x+\dfrac{2}{a}\sqrt{1-a^2x^2}\arcsin(ax)$; $(\arcsin(ax))^2$

---

65. Find $D(\ln^3(x))$
(a) by the chain rule. (b) by writing $\ln^3(x)$ as $\ln(x)\cdot\ln(x)\cdot\ln(x)$. (c) Which method do you prefer? Why?

66. We have used the equation $\sec^2(x)=1+\tan^2(x)$. (a) Derive it from the equation $\cos^2(x)+\sin^2(x)=1$. (b) Derive $\cos^2(x)+\sin^2(x)=1$ from the Pythagorean Theorem.

67. Find two antiderivatives of (a) $2x$, (b) x^2, (c) $1/x$, and (d) $\sqrt{x}$.

68. Find two antiderivatives of (a) e^{3x}, (b) $\cos(x)$, (c) $\sin(x)$, and (d) $1/(1+x^2)$.

69. This problem provides some additional experience with the development of the formula $\log_b(x)=\log_b(e)\log_e(x)$. Let $b>0$. Recall that $\log_b(a)=\dfrac{\log_e(a)}{\log_e(b)}$.

NOTE: This formula was used previously in Example 1 in Section 3.5.

(a) Show that $\log_b(e)=\dfrac{1}{\log_e(b)}$. (b) Conclude that $\log_b(x)=\dfrac{\log_b(e)}{\log_e(x)}$.

In Exercises 70 to 73 use the chain rule to verify each each expression has the indicated derivative.

70. $(\arccos(x))' = \dfrac{-1}{\sqrt{1-x^2}}$ **71.** $(\operatorname{arcsec}(x))' = \dfrac{1}{|x|\sqrt{x^2-1}}$ **72.** $(\operatorname{arccot}(x))' = \dfrac{-1}{1+x^2}$ **73.** $(\operatorname{arccsc}(x))' = \dfrac{-1}{|x|\sqrt{x^2-1}}$

---

74. Verify that $D\left(2(\sqrt{x}-1)e^{\sqrt{x}}\right) = e^{\sqrt{x}}$.

75. **SAM:** I say that $D(\log_b(x)) = \dfrac{1}{x\ln(b)}$. It's simple. Let $y = \log_b(x)$. That tells me $x = b^y$. I differentiate both sides of that, getting $1 = b^y(\ln(b))y'$. So $y' = \dfrac{1}{b^y\ln(b)} = \dfrac{1}{x\ln(b)}$.

JANE: Well, not so fast. I start with the equation $\log_b(x) = (\log_b(e))\ln(x)$. So $D(\log_b(x)) = \dfrac{\log_b(e)}{x}$.

SAM: Something is wrong. Where did you get that equation you started with?

JANE: Just take $\log_b$ of both sides of $x = e^{\ln(x)}$.

SAM: I hope this won't be on the next midterm.

Settle this argument.

We did not need the chain rule to find the derivatives of inverse functions. Instead, we could have taken a geometric approach, using the interpretation of the derivative of the slope of the tangent line. When we reflect the graph of f around the line $y = x$ to obtain the graph of f^{-1}, the reflection of the tangent line to the graph of f with slope m is the tangent line to the graph of f^{-1} with slope $1/m$. (See Section 1.1.) Exercises 76 to 80 use this approach to develop formulas obtained in this section.

76. Let $f(x) = \ln(x)$. The slope of the graph of $y = \ln(x)$ at $(a, \ln(a))$, $a > 0$, is the reciprocal of the slope of the graph of $y = e^x$ at $(\ln(a), a)$. Use this to show that the slope of the graph of $y = \ln(x)$ when $x = a$ is $1/a$.

In Exercises 77 to 80 use the technique illustrated in Exercise 76 to differentiate the function.

77. $f(x) = \arctan(x)$ **78.** $f(x) = \arcsin(x)$ **79.** $f(x) = \operatorname{arcsec}(x)$ **80.** $f(x) = \arccos(x)$

---

81. (a) Evaluate $\lim\limits_{x\to\infty} \dfrac{1}{1+x^2}$ and $\lim\limits_{x\to-\infty} \dfrac{1}{1+x^2}$.

(b) What do these results tell you about the graph of the arctangent function?

82. **SAM:** I can get the formula for $(fg)'$ real easy if I assume fg is differentiable when f and g are.

JANE: How?

SAM: Start with $\ln(fg) = \ln(f) + \ln(g)$, which is OK if $f(x)$ and $g(x)$ are positive.

JANE: I am with you, so far.

SAM: Then differentiate like mad, using the chain rule for each of the three terms — and resisting the urge to expand $(fg)'$: $\dfrac{1}{fg}(fg)' = \dfrac{f'}{f} + \dfrac{g'}{g}$.

JANE: So?

SAM: Then solve for $(fg)'$ and out pops $(fg)' = f'g + fg'$.

JANE: I wonder why the book used all those Δs instead.

Why didn't the book use Sam's approach?

83. Use the assumptions and methods in Exercise 82 to find $D\left(\dfrac{f}{g}\right)$.

84. **SAM:** In Exercise 82 I assumed that fg is differentiable if f and g are. I can get around that by using the differentiability of exp and ln.

JANE: How so?

SAM: Again I assume $f(x)$ and $g(x)$ are positive and I write fg as $\exp(\ln(fg))$.

JANE: So?

SAM: But $\ln(fg) = \ln(f) + \ln(g)$, and that does it.

JANE: I'm lost.

SAM: Well, $fg = \exp(\ln(f) + \ln(g))$ and use the chain rule. It's good for more than grinding out derivatives. In fact, when you differentiate both sides of my equation, you get that fg is differentiable and $(fg)'$ is $f'g + fg'$.

JANE: Why wouldn't the authors use this approach?

SAM: It would make things too easy and reveal that calculus is all about e, exponentials, and logarithms. I peeked at Chapter 12 and saw that you can even get sine and cosine out of e^x.

Is Sam's argument correct?

3.6 Antiderivatives and Slope Fields

So far in this chapter we have started with a function and found its derivative. In this section we will go in the opposite direction: given a function f, we will be interested in finding a function F whose derivative is f. Why? Because going from the derivative back to the function plays a central role in **integral calculus**, as we will see in Chapter 5. Any function, F, with the property that $F'(x) = f(x)$, is called an **antiderivative** of f. Chapter 6 describes several ways to find antiderivatives.

Some Antiderivatives

EXAMPLE 1. Find an antiderivative of x^6.

SOLUTION When we differentiate x^a we get ax^{a-1}. The exponent in the derivative, $a-1$, is one less than the original exponent, a. So we expect an antiderivative of x^6 to involve x^7.

Because $\left(x^7\right)' = 7x^6$, x^7 is an antiderivative of $7x^6$, not of x^6. We want to get rid of that coefficient 7 in front of x^6. If we divide x^7 by 7 we have

$$\left(\frac{x^7}{7}\right)' = \frac{7x^6}{7} = x^6$$

We conclude that $x^7/7$ is an antiderivative of x^6.

However, $x^7/7$ is not the only antiderivative of x^6. For instance,

$$\left(\frac{1}{7}x^7 + 2024\right)' = \frac{1}{7}7x^6 + 0 = x^6.$$

In fact, we can add any constant to $x^7/7$ and the result is always an antiderivative of x^6. ◆

Observation 3.6.1: *Antiderivatives are Not Unique*

A constant added to any antiderivative of a function f gives another antiderivative of f. If $F(x)$ is an antiderivative of $f(x)$ so is $F(x) + C$ for any constant C.

The reasoning in this example suggests that $x^{a+1}/(a+1)$ is an antiderivative of x^a. This formula is meaningless when $a+1=0$. We have to expect a different formula for antiderivatives of $x^{-1} = 1/x$. In Section 3.5 we saw that $(\ln(x))' = 1/x$. That's one reason the function $\ln(x)$ is so important: it provides an antiderivative for $1/x$.

Formula 3.6.1: *Antiderivatives of x^a*

For any number a, except -1, the antiderivatives of x^a are

$$\frac{1}{a+1}x^{a+1}+C \quad \text{for any constant } C.$$

The antiderivatives of $x^{-1}=\dfrac{1}{x}$ are, when $x>0$,

$$\ln(x)+C \quad \text{for any constant } C.$$

Every time you compute a derivative, you are also finding an antiderivative. For instance, since $D(\sin(x)) = \cos(x)$, $\sin(x)$ is an antiderivative of $\cos(x)$. So is $\sin(x)+C$ for any constant C.

There are tables of antiderivatives that go on for hundreds of pages. Here is a small table with entries corresponding to the derivatives that we have found so far. Table 3.6.1 summarizes the basic antiderivatives that we currently know from our earlier discussion of derivatives.

NOTE: For more extensive lists of antiderivatives, search online for "antiderivative table".

Function (f)	Antiderivative (F)	Comment
x^a	$\frac{1}{a+1}x^{a+1}$	for $a \neq -1$
$x^{-1}=\frac{1}{x}$	$\ln(x)$	for $x>0$
e^x	e^x	
$\cos(x)$	$\sin(x)$	
$\sin(x)$	$-\cos(x)$	
$\sec^2(x)$	$\tan(x)$	Example 8 in Section 3.3
$\sec(x)\tan(x)$	$\sec(x)$	Example 11 in Section 3.3
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin(x)$	Section 3.4
$\frac{1}{1+x^2}$	$\arctan(x)$	Section 3.4

Table 3.6.1

An **elementary function** is a function that can be expressed in terms of polynomials, powers, trigonometric functions, exponentials, logarithms, and the functions obtained from them by algebra and by composition of functions. The derivative of an elementary function is elementary. We might expect that every elementary function would have an elementary antiderivative.

In 1833, the French mathematician and engineer Joseph Liouville (1809–1882) proved that there are elementary functions that do not have elementary antiderivatives. Here are five examples of such functions:

$$e^{-x^2}, \qquad \frac{\sin(x)}{x}, \qquad x\tan(x), \qquad \sqrt{x}\sqrt[3]{1+x}, \quad \text{and} \quad \sqrt[4]{1+x^2}.$$

The four operations of algebra are $+$, $-$, $\times$ and $/$.

There are two types of elementary functions: **algebraic** and **transcendental**. Algebraic functions, defined in Section 3.5, consist of polynomials, quotients of polynomials (the rational functions), and functions that can be built up by the four operations of algebra and taking roots. For instance, $\left(\sqrt{x+\sqrt[3]{x}+x^2}\right)/(1+2x)^5$ is algebraic while functions such as $\sin(x)$ and 2^x are not algebraic; these functions are called **transcendental**.

It is difficult to tell whether a given elementary function has an elementary antiderivative. For instance, $x\sin(x)$ does, namely $-x\cos(x)+\sin(x)$, as may be checked, but $x\tan(x)$ does not. The function e^{x^2} does not, as mentioned earlier. However, $e^{\sqrt{x}}$, which looks more complicated, has an elementary antiderivative. (See Exercise 74 in Section 3.5.)

The importance of antiderivatives will be revealed in Chapter 5. Some techniques for finding antiderivatives are developed in Chapter 8.

Picturing Antiderivatives

If it is not possible to find an explicit formula for the antiderivative of many elementary functions, why do we believe that they have antiderivatives? The following shows why.

The **slope field** for a function $f(x)$ is made of short line segments with slope $f(x)$ at a few points whose x-coordinate is x. By drawing a slope field you can convince yourself that an antiderivative exists, and see the shape of its graph.

Imagine that we are looking for an antiderivative $F(x)$ of $\sqrt{1+x^3}$. We want $F'(x)$ to be $\sqrt{1+x^3}$. Or, to put it geometrically, we want the slope of the curve $y=F(x)$ to be $\sqrt{1+x^3}$. For instance, when $x=2$, we want the slope to be $\sqrt{1+x^3}=3$. We do not know what $F(2)$ is, but at least we can draw a short piece of the tangent line at all points for which $x=2$: they all have slope 3. (See Figure 3.6.1(a).)

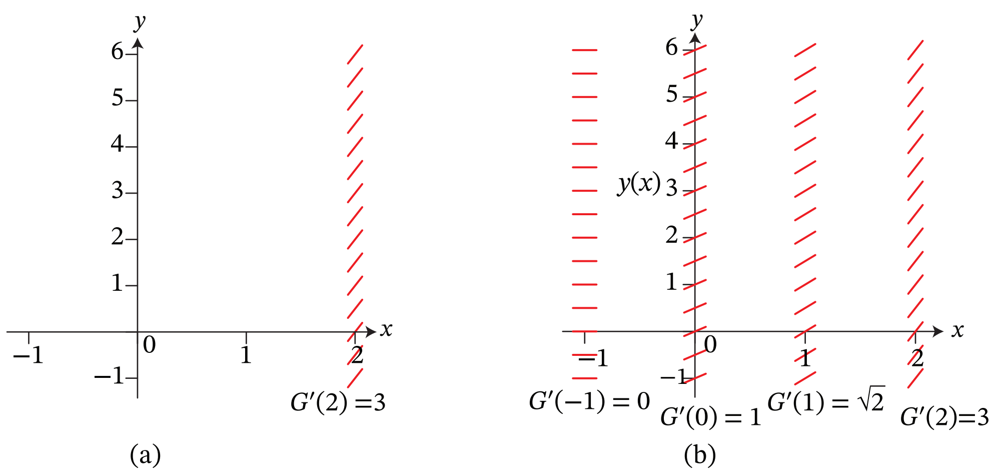


Figure 3.6.1

When $x=1$, $\sqrt{1+x^3}=\sqrt{2}\approx 1.4$. So we draw short lines with slope $\sqrt{2}$ on the vertical line $x=1$. When $x=0$, $\sqrt{1+x^3}=1$: the tangent lines for $x=0$ all have slope 1. When $x=-1$, the slopes are $\sqrt{1+x^3}=0$ so the tangent lines are all horizontal. (See Figure 3.6.1(b).)

The plot of a slope field is most commonly made with the aid of specialized software on a graphing calculator or computer.

For a sample of available resources, search online for "slope field plot".

A typical slope field, showing more segments of tangent lines than we have the patience to draw by hand, is in Figure 3.6.2(a) which shows a computer-generated direction field for $f(x)=\sqrt{1+x^3}$ that has many more segments of tangent lines than Figure 3.6.1(a).

We can now visualize the curves that follow the slope field for $f(x)=\sqrt{1+x^3}$. Start at a point, say $(-1,0)$. There the slope is $F'(-1)=f(-1)=0$, and the curve starts moving horizontally to the right. As soon as the curve leaves this

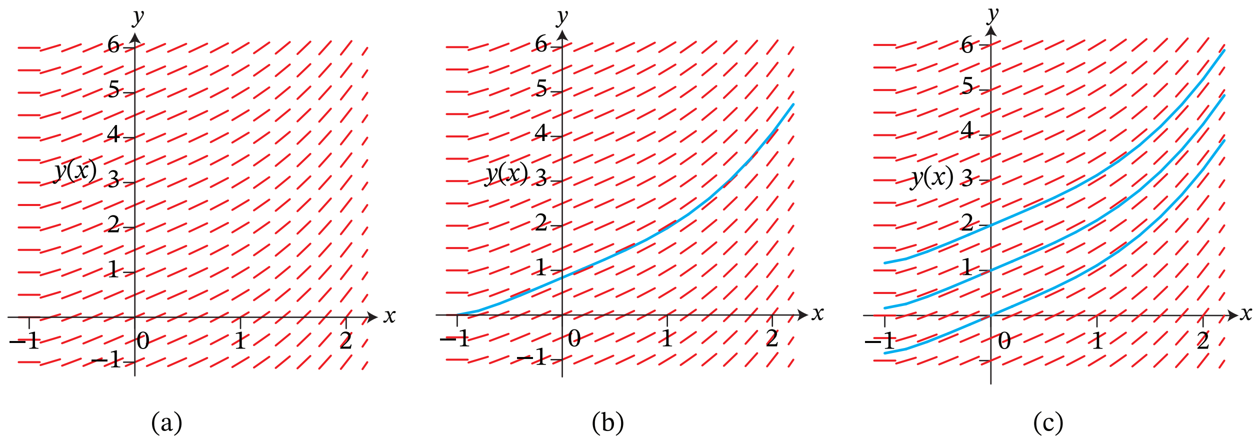


Figure 3.6.2

initial point the slope, as given by $F'(x) = f(x)$, becomes slightly positive. This pushes the curve upward. The slope continues to increase as x increases. The curve in Figure 3.6.2(b) is the graph of the antiderivative of $f(x) = \sqrt{1+x^3}$ that equals 0 when x is -1.

Starting at a different initial point will produce a different antiderivative. Three antiderivatives are shown in Figure 3.6.2(c). Many other antiderivatives for $f(x) = \sqrt{1+x^3}$ are visible in the slope field. None are elementary.

This suggests that different antiderivatives of a function differ by a constant: the graph of one is the graph of the other raised or lowered by their constant difference. The next example suggests that the constant functions are the only antiderivatives of the zero function. Both suggestions are correct, as will be shown in Section 4.1.

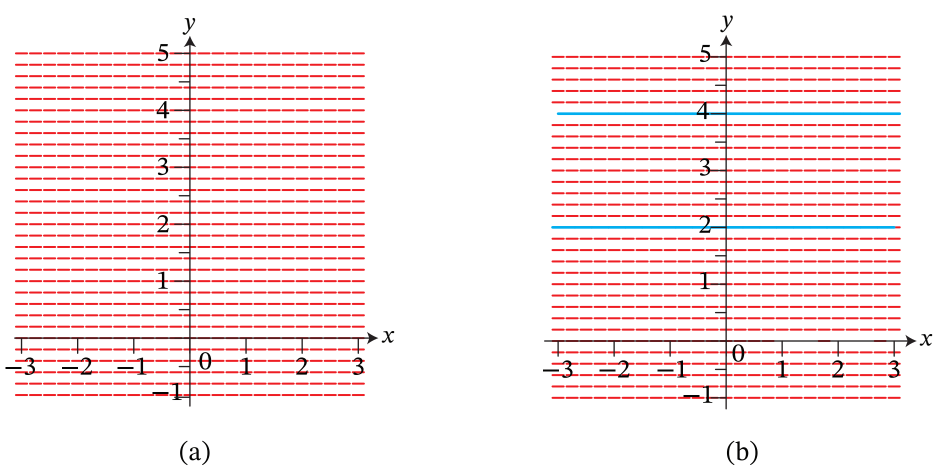


Figure 3.6.3

EXAMPLE 2. Draw the slope field for $\dfrac{dy}{dx} = 0$. Also, graph two antiderivatives of 0.

SOLUTION Since the slope is 0 everywhere, each of the tangent lines is represented by a horizontal line segment, as in Figure 3.6.3(a).

Figure 3.6.3(b) shows two possible antiderivatives of 0, namely the constant functions $f(x) = 2$ and $g(x) = 4$. ♦

The following results will be established using the definitions and theorems of calculus in Section 4.1.

Observation 3.6.2: *Constants are the Antiderivatives of the Zero Function*

- Every antiderivative of the zero function on an interval is constant. That is, if $f'(x) = 0$ for all x in an interval, then $f(x) = C$ for some constant C.
- Two antiderivatives of a function on an interval differ by a constant. That is, if $F'(x) = G'(x)$ for all x in an interval, then $F(x) = G(x) + C$ for some constant C.

Summary

If $F' = f$, then F is an antiderivative of f; so is $F + C$ for any constant C.

We introduced the notion of an elementary function. Such a function is built up from polynomials, logarithms, exponentials, and the trigonometric functions by algebraic operations and the most important operation, composition. While the derivative of an elementary function is elementary, its antiderivative does not need to be. Each elementary function is either algebraic or transcendental.

We showed how a slope field can help analyze an antiderivative even though we may not know a formula for it. Slope fields will be used later for other purposes.

EXERCISES for Section 3.6

1. (a) Verify that $-x\cos(x) + \sin(x)$ is an antiderivative of $x\sin(x)$.
(b) Spend at least one minute and at most ten minutes trying to find an antiderivative of $x\tan(x)$.

In Exercises 2 to 11 give two antiderivatives for each function.

2. x^3 **3.** x^4 **4.** x^{-2} **5.** $\dfrac{1}{x^3}$ **6.** $\sqrt[3]{x}$

7. $\dfrac{2}{x}$ **8.** $\sec(x)\tan(x)$ **9.** $\sin(x)$ **10.** e^{-x} **11.** $\sin(2x)$

In Exercises 12 to 20
(a) Draw the slope field for the given derivative,
(b) Use it to draw the graphs of two possible antiderivatives $F(x)$.

12. $F'(x) = 2$ **13.** $F'(x) = x$ **14.** $F'(x) = \dfrac{-x}{2}$

15. $F'(x) = \dfrac{1}{x}, x > 0$ **16.** $F'(x) = \cos(x)$ **17.** $F'(x) = \sqrt{x}$

18. $F'(x) = e^{-x}, x > 0$ **19.** $F'(x) = \dfrac{1}{x^2}, x > 0$ **20.** $F'(x) = \dfrac{1}{x-1}, x > 1$

In Exercises 21 to 30 use differentiation to check that the first expression is an antiderivative of the second expression.

21. $2x\sin(x) - (x^2-2)\cos(x)$; $x^2\sin(x)$

22. $(4x^3 - 24x)\sin(x) - (x^4 - 12x^2 + 24)\cos(x)$; $x^4\sin(x)$

23. $\dfrac{-1}{2x^2}$; $\dfrac{1}{x^3}$

24. $\dfrac{-2}{\sqrt{x}}$; $\dfrac{1}{x^{3/2}}$

25. $(x-1)e^x$; xe^x

26. $(x^2-2x+2)e^x$; x^2e^x

27. $\frac{1}{2}e^u(\sin(u)-\cos(u))$; $e^u\sin(u)$

28. $\frac{1}{2}e^u(\sin(u)+\cos(u))$; $e^u\cos(u)$

29. $\frac{x}{2}-\frac{\sin(x)\cos(x)}{2}$; $\sin^2(x)$

30. $2x\cos(x)+(x^2-2)\sin(x)$; $x^2\cos(x)$

---

31. (a) Draw the slope field for $\frac{dy}{dx}=e^{-x^2}$.

(b) Draw the graph of the antiderivative of e^{-x^2} that passes through the point $(0,1)$.

32. (a) Draw the slope field for $\frac{dy}{dx}=f(x)$ where $f(x)=\frac{\sin(x)}{x}$, when $x\neq 0$, and $f(0)=1$.

(b) What is the slope for any point on the y-axis?

(c) Draw the graph of the antiderivative of $f(x)$ that passes through the point $(0,1)$.

33. Two tables of antiderivatives list the following two antiderivatives of $\frac{1}{x^2(a+bx)}$, where a and b are constants and $\frac{a+bx}{x}>0$:

$$\frac{-1}{a^2}\left(\frac{a+bx}{x}-b\ln\left(\frac{a+bx}{x}\right)\right) \quad \text{and} \quad -\frac{1}{ax}+\frac{b}{a^2}\ln\left(\frac{a+bx}{x}\right).$$

(a) By differentiating the two antiderivatives, show that both are correct.

(b) Show that the two antiderivatives differ by a constant by finding their difference.

34. If $F(x)$ is an antiderivative of $f(x)$, find an antiderivative of (a) $g(x)=2f(x)$ and (b) $h(x)=f(2x)$.

35. (a) Draw the slope field for $\frac{dy}{dx}=-y$.

(b) Draw the graph of the function $y=F(x)$ such that $F(0)=1$ and $F'(x)=-F(x)$.

(c) Do you think $\lim_{x\to\infty}F(x)$ exists? If so, what is its value?

Historical Note: How Computers Find Antiderivatives

There are algorithms implemented in software on computers, hand-held devices, and calculators that determine if a given elementary function has an elementary antiderivative. The most well-known is the **Risch algorithm**, developed in 1968, based on differential equations and abstract algebra. An online search for "Risch antiderivative elementary symbolic" produces links with more information about the Risch algorithm.

3.7 Motion and the Second Derivative

In a drag race Melanie Troxel reached a speed of 324 miles per hour, which is about 475 feet per second, in a mere 4.539 seconds. By comparison, a 1968 Fiat 850 Idromatic could reach a speed of 60 miles per hour in 25 seconds, a 1997 Porsche 911 Turbo S in 3.6 seconds, and a 2021 Tesla Model S P100D (in ludicrous plus mode) in 2.3 seconds.

Since Troxel increased her speed from 0 feet per second to 475 feet per second in 4.539 seconds her speed was increasing at the rate of $475/4.539 \approx 105$ feet per second per second, assuming she kept the motor at constant power throughout the time interval. That acceleration is more than three times the acceleration due to gravity at sea level (32 feet per second per second). Ms. Troxel must have felt quite a force as her seat pressed against her back. Extensive acceleration data can be found with an online search for "automobile acceleration rates".

Before beginning the discussion of **acceleration**, we need to remember that the sign of velocity indicates the direction of motion. By contrast, **speed**, the absolute value of velocity, does not indicate direction.

This brings us to the definition of **acceleration** and an introduction to higher derivatives.

Acceleration

Velocity is the rate at which the position of an object moving on a line changes. The rate at which velocity changes is called **acceleration**, denoted a. Thus if $y = f(t)$ denotes position on the y-axis at time t, then the derivative $\frac{dy}{dt}$ equals the velocity, and the derivative of the derivative equals the acceleration. That is,

$$v = \frac{dy}{dt} \quad \text{and} \quad a = \frac{dv}{dt} = \frac{d}{dt}\left(\frac{dy}{dt}\right).$$

The derivative of the derivative of a function $y = f(x)$ is called the **second derivative**. It is denoted in many different ways, including

$$\frac{d^2y}{dx^2}, \qquad D^2y, \qquad y'', \qquad f'', \qquad D^2f, \qquad f^{(2)}, \quad \text{and} \quad \frac{d^2f}{dx^2}.$$

If $y = f(t)$, where t denotes time, the first and second derivatives dy/dt and d^2y/dt^2 are sometimes denoted $\dot{y}$ and $\ddot{y}$, respectively.

For instance, if $y = x^3$,

$$\frac{dy}{dx} = 3x^2 \quad \text{and} \quad \frac{d^2y}{dx^2} = 6x.$$

Other ways of denoting the second derivative are

$$D^2(x^3) = 6x, \ \frac{d^2(x^3)}{dx^2} = 6x, \quad \text{and} \quad (x^3)'' = 6x.$$

Table 3.7.1 lists four functions and their first and second derivatives.

Most functions met in applications of calculus can be differentiated repeatedly in the sense that Df exists, the derivative of Df, namely, D^2f, exists, the derivative of D^2f exists, and so on.

The derivative of the second derivative is called the **third derivative** and is denoted many ways, such as

$$\frac{d^3y}{dx^3}, \qquad D^3y, \qquad y''', \qquad f''', \qquad f^{(3)}, \quad \text{and} \quad \frac{d^3f}{dx^3}.$$

The fourth derivative is defined similarly, as the derivative of the third derivative. In the same way the n^{th} derivative can be defined for any positive integer n and denote this by such symbols as

$$\frac{d^ny}{dx^n}, \qquad D^ny, \qquad f^{(n)}, \quad \text{and} \quad \frac{d^nf}{dx^n}.$$

y	$\frac{dy}{dx}$	$\frac{d^2y}{dx^2}$
x^3	$3x^2$	$6x$
$\frac{1}{x} = x^{-1}$	$\frac{-1}{x^2} = -x^{-2}$	$\frac{2}{x^3} = 2x^{-3}$
e^x	e^x	e^x
$\sin(5x)$	$5\cos(5x)$	$-25\sin(5x)$

Table 3.7.1

Each of these is read as "the n^{th} derivative of y with respect to x." For instance, if $f(x) = 2x^3 + x^2 - x + 5$, we have

$$f^{(1)}(x) = 6x^2 + 2x - 1, \qquad f^{(2)}(x) = 12x + 2, \qquad f^{(3)}(x) = 12, \quad \text{and} \quad f^{(4)}(x) = 0.$$

And, because $f^{(4)}(x) = 0$ for all x, all subsequent derivatives are also zero: $f^{(n)}(x) = 0$ for all $n \geq 5$.

EXAMPLE 1. Find $D^n(e^{-2x})$ for each positive integer n.

SOLUTION $D^1\left(e^{-2x}\right) = D\left(e^{-2x}\right) = -2e^{-2x}$, $D^2\left(e^{-2x}\right) = D\left(-2e^{-2x}\right) = (-2)^2e^{-2x}$, and $D^3\left(e^{-2x}\right) = D\left((-2)^2e^{-2x}\right) = (-2)^3e^{-2x}$. At each differentiation another (-2) becomes part of the coefficient. Thus $D^n\left(e^{-2x}\right) = (-2)^ne^{-2x}$. This can also be written

$$D^n\left(e^{-2x}\right) = (-1)^n2^ne^{-2x},$$

because the power $(-1)^n$ records a plus if n is even and a minus if n is odd. ◆

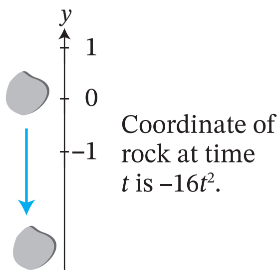


Figure 3.7.1

Finding Velocity and Acceleration from Position

EXAMPLE 2. A falling rock drops $16t^2$ feet in the first t seconds. Find its velocity and acceleration.

SOLUTION Place the y-axis in the usual position, with 0 at the beginning of the fall and the part with positive values above 0, as in Figure 3.7.1. At time t the object has the y coordinate

$$y = -16t^2.$$

The velocity is $v = (-16t^2)' = -32t$ feet per second, and the acceleration is $a = (-32t)' = -32$ feet per second per second. The velocity changes at a constant rate. That is, the acceleration is constant. ♦

Finding Position from Velocity and Acceleration

To calculate the position of a moving object it is enough to know the object's acceleration, its initial position, and its initial velocity. This will be demonstrated in the next two examples in the special case that the acceleration is constant. In the first example, the acceleration is 0.

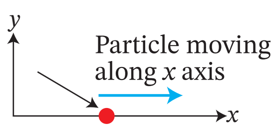


Figure 3.7.2

EXAMPLE 3.

In the simplest motion, no forces act on a moving particle, hence its acceleration is 0. Assume that a particle is moving on the x-axis and no forces act on it. Let its location at time t seconds be $x = f(t)$ feet. See Figure 3.7.2. If at time $t = 0$, $x = 3$ feet and the velocity is 5 feet per second, determine $f(t)$.

SOLUTION The assumption that no force operates on the particle means that there is no acceleration: $d^2x/dt^2 = 0$. Call the velocity v. Then

$$\frac{dv}{dt} = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d^2x}{dt^2} = 0$$

That is, v is a function of time whose derivative is 0. At the end of Section 3.6 we saw that constant functions are the antiderivatives of 0. Thus, v must be constant:

$$v(t) = C \qquad \text{for some constant } C.$$

Since $v(0) = 5$, the constant C must be 5.

We know that the derivative of the position x is the velocity v. Hence

$$\frac{dx}{dt} = 5.$$

Similar reasoning tells us that $x = f(t)$ has the form

$$x = 5t + K \qquad \text{for some constant } K.$$

When $t = 0$, $x = 3$, so $K = 3$. Thus at time t seconds the particle is at $x = 5t + 3$ feet. ♦

In the next example the acceleration is constant, but not zero.

EXAMPLE 4. A ball is thrown straight up, with an initial speed of 64 feet per second, from a cliff 96 feet above a beach.

(a) Where is the ball t seconds later? (b) When does it reach its maximum height? (c) How high above the beach does the ball rise? (d) When does the ball hit the beach? Assume that there is no air resistance and that the acceleration due to gravity is constant.

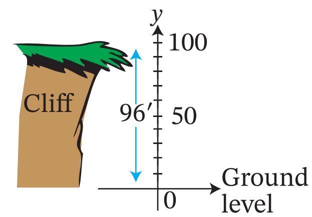


Figure 3.7.3

SOLUTION Introduce a vertical coordinate axis to describe the position of the ball. It is more natural to call it the y-axis, and so the velocity and acceleration are dy/dt and d^2y/dt^2, respectively. Place the origin at ground level and let the positive part of the y-axis be above the ground, as in Figure 3.7.3. At time $t=0$, the velocity dy/dt is 64, since the ball is thrown up at a speed of 64 feet per second. As time increases, dy/dt decreases from 64 to 0 (when the ball reaches the top of its path and begins its descent) and continues to decrease through larger and larger negative values as the ball falls to the ground. Since v is decreasing, the acceleration dv/dt is negative. The constant value of dv/dt, gravitational acceleration, is approximately -32 feet per second per second.

(a) From the fact that velocity is an antiderivative of acceleration and the equation for acceleration:

$$a = \frac{dv}{dt} = -32,$$

it follows that

$$v = -32t + C,$$

where C is some constant. To find C, we use the fact that $v = 64$ when $t = 0$ so $64 = -32 \cdot 0 + C$ and so $C = 64$. Hence $v = -32t + 64$ for any time t until the ball hits the beach.
So we have

$$\frac{dy}{dt} = v = -32t + 64.$$

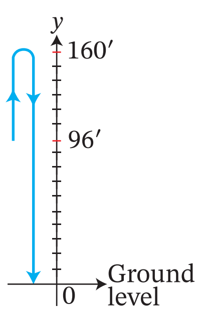


Figure 3.7.4

Since the position function y is an antiderivative of the velocity, $-32t+64$, we have

$$y(t) = -16t^2 + 64t + K,$$

where K is a constant. To find K, recall that $y = 96$ when $t = 0$. Thus $96 = -16 \cdot 0^2 + 64 \cdot 0 + K$, and $K = 96$.

We have obtained a complete description of the position of the ball at any time t while it is in the air:

$$y = -16t^2 + 64t + 96.$$

This, together with $v = -32t + 64$, provides answers to many questions about the ball's flight. (As a check, note that when $t = 0$, $y = 96$, the initial height.)

(b) When does it reach its maximum height? When it is neither rising nor falling. That is, the velocity is neither positive nor negative, and so must be 0. The velocity is zero when $-32t + 64 = 0$, which is when $t = 2$ seconds.

(c) How high above the ground does the ball rise? Compute y when $t = 2$. This gives $-16 \cdot 2^2 + 64 \cdot 2 + 96 = 160$ feet. (See Figure 3.7.4.)

(d) When does the ball hit the beach? When $y = 0$. Find t such that

$$y = -16t^2 + 64t + 96 = 0.$$

Division by -16 yields $t^2 - 4t - 6 = 0$, which has the solutions

$$t = \frac{4 \pm \sqrt{16+24}}{2} = 2 \pm \sqrt{10}.$$

Since $2 - \sqrt{10}$ is negative and the ball cannot hit the beach before it is thrown, the only physically meaningful solution is $2 + \sqrt{10}$. The ball lands $2 + \sqrt{10}$ seconds after it is thrown, so it is in the air for about 5.2 seconds.

The graphs of position, velocity, and acceleration as functions of time provide another perspective on the motion of the ball, as shown in Figure 3.7.5(a), (b), and (c), respectively. ◆

Reasoning like that in Examples 3 and 4 establishes the following description of motion in all cases where the acceleration is constant.

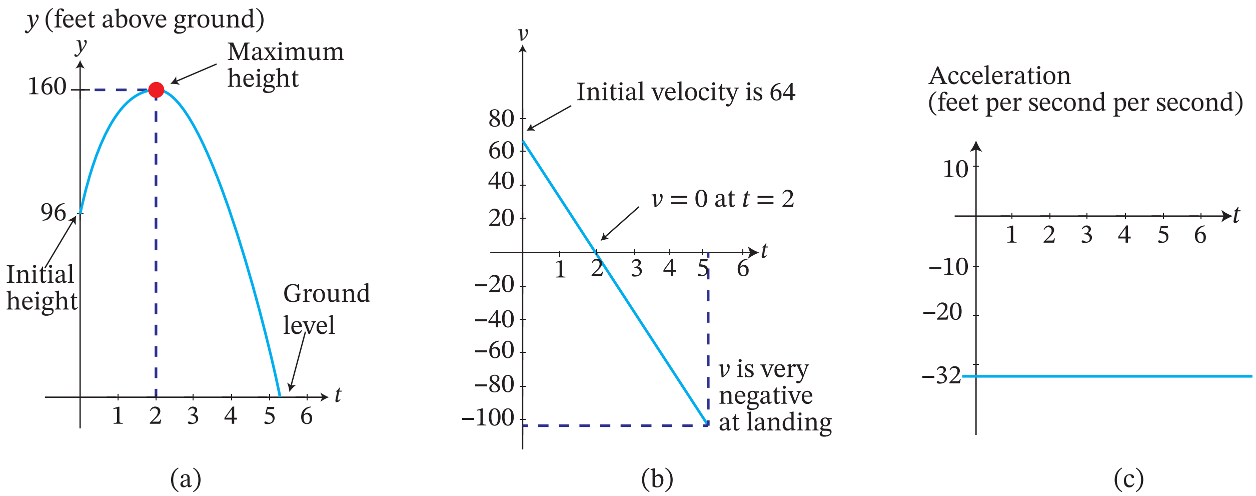


Figure 3.7.5

Observation 3.7.1: *Motion Under Constant Acceleration*

Assume that a particle moving on the y-axis has a constant acceleration a. Assume also that at time $t = 0$ it has the initial velocity v_0 and has the initial y-coordinate y_0. Then at any time $t \ge 0$ its y-coordinate is

$$y = \frac{a}{2}t^2 + v_0 t + y_0.$$

In Example 3, $a = 0$, $v_0 = 5$, and $y_0 = 3$ and in Example 4, $a = -32$ $v_0 = 64$, and $y_0 = 96$. The data must be given in consistent units, for instance, all in meters or all in feet.

Summary

We defined the higher derivatives of a function. They are obtained by repeatedly differentiating. The second derivative is the derivative of the derivative, the third derivative is the derivative of the second derivative, and so on. The first and second derivatives, $D(f)$ and $D^2(f)$, are used in many applications. We used them to analyze motion under constant acceleration.

EXERCISES for Section 3.7

In Exercises 1 to 16 find the first and second derivatives of the functions.

1. $y = 2x + 3$

2. $y = e^{-x^3}$

3. $y = x^5$

4. $y = \ln(6x + 1)$

5. $y = \sin(\pi x)$

6. $y = 4x^3 - x^2 + x$

7. $y = \dfrac{x}{x+1}$

8. $y = \dfrac{x^2}{x-1}$

9. $y = x\cos(x^2)$

10. $y = \dfrac{x}{\tan(3x)}$

11. $y = (x-2)^4$

12. $y = (x+1)^3$

13. $y = e^{3x}$

14. $y = \tan(x^2)$

15. $y = x^2 \arctan(3x)$

16. $y = -\dfrac{\arcsin(2x)}{x^2}$

Exercises 17 to 19 concern Example 4.

17. (a) How long after the ball is thrown does it pass by the top of the cliff? (b) What are its speed and velocity at this instant?

18. Suppose the ball in Example 4 had simply been dropped from the cliff. Find the position y as a function of time. How long would it take the ball to reach the beach?

19. In view of the result of Exercise 18, provide a physical interpretation of the three terms on the right-hand side of the formula $y = -16t^2 + 64t + 96$.

20. At time $t = 0$ a particle is at $y = 3$ feet and has a velocity of -3 feet per second; it has a constant acceleration of 6 feet per second per second. Find its position at any time t.

21. At time $t = 0$ a particle is at $y = 10$ feet and has a velocity of 8 feet per second and has a constant acceleration of -8 feet per second per second. (a) Find its position at any time t. (b) What is its maximum y coordinate?

22. At time $t = 0$ a particle is at $y = 0$ feet and has a velocity of 0 feet per second. Find its position at any time t if its acceleration is always -32 feet per second per second.

23. At time $t = 0$ a particle is at $y = -4$ feet and has a velocity of 6 feet per second and it has a constant acceleration of -32 feet per second per second. (a) Find its position at any time t. (b) What is its largest y coordinate?

24. The (Leaning) Tower of Pisa started to lean even before it was completed in 1350. By 1990, the tilt was 5.5°. After almost two decades of work to stabilize the Tower the tilt was reduced to just under 4° in 2001. Use calculus, specifically derivatives, to restate the following report about the Tower of Pisa:

> **Historical Note: Until 2001, the Tower of Pisa's angle from the vertical was increasing more rapidly**
>
> The tower, begun in 1174 and completed in 1350, is 179 feet tall and leans about 14 feet from the vertical. Each day it leaned on the average, another 1/5000 inch until the tower was propped up in 2001.

25. Find all functions f such that $D^2(f) = 0$ for all x.

26. Find all functions f such that $D^3(f) = 0$ for all x.

In Exercises 27 to 36 find the given derivatives.

27. $D^3\left(5x^2 - 2x + 7\right)$

28. $D^4\left(\sin(2x)\right)$

29. $D^n\left(e^x\right)$

30. $D(\sin(x))$, $D^2(\sin(x))$, $D^3(\sin(x))$, and $D^4(\sin(x))$

31. $D(\cos(x))$, $D^2(\cos(x))$, $D^3(\cos(x))$, and $D^4(\cos(x))$

32. $D(\ln(x))$, $D^2(\ln(x))$, $D^3(\ln(x))$, and $D^4(\ln(x))$

33. $D^4(x^4)$ and $D^5(x^4)$

34. $D^{200}(\sin(x))$

35. $D^{200}(e^x)$

36. $D^2(5^x)$

37. A jet with velocity 500 miles per hour begins its descent 120 miles from an airport. Its landing velocity is 180 miles per hour. Assuming a constant deceleration, how long does the descent take?

38. Verify Observation 3.7.1. That is, let $y = f(t)$ describe the motion on the y-axis of an object whose acceleration has the constant value a, and show that $y = \frac{a}{2}t^2 + v_0 t + y_0$, where v_0 is the velocity when $t = 0$ and y_0 is the position when $t = 0$.

39. Which has the highest acceleration? Melanie Troxel's dragster, a 1997 Porsche 911 Turbo S, a Tesla Model 3 (dual motor), or an airplane being launched from an aircraft carrier? The plane reaches a velocity of 180 miles per hour in 2.5 seconds, within a distance of 300 feet.

40. Why do engineers call the third derivative of position with respect to time the **jerk**? What do they call the fourth, fifth, and sixth derivatives of position?

41. Give two functions f such that $D^2(f) = 9f$, neither a constant multiple of the other.

42. Give two functions f such that $D^2(f) = -4f$, neither a constant multiple of the other.

43. A car accelerates with constant acceleration from 0 (rest) to 60 miles per hour in 15 seconds. How far does it travel in this period? Be sure to do your computations either all in seconds, or all in hours.

60 mph = 88 fps

44. Show that a ball thrown straight up from the ground takes as long to rise as to fall back to its initial position. Disregard air resistance. How does the velocity with which it strikes the ground compare with its initial velocity? How do the initial and landing speeds compare?

3.8 The Precise Definition of Limits at Infinity: $\lim_{x\to\infty} f(x) = L$

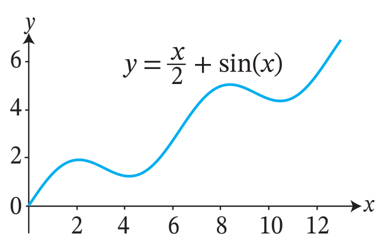


Figure 3.8.1

One day a teacher drew on the board the graph of $y = x/2 + \sin(x)$, shown in Figure 3.8.1.

Then the class was asked whether they thought that $\lim_{x\to\infty} f(x) = \infty$. A third of the class voted "No" because "it keeps going up and down." A third voted "Yes" because "the function tends to get very large as x increases." A third didn't vote. Such a variety of views on such a fundamental concept suggests that we need a more precise definition of a limit than the ones developed in Sections 2.2 and 2.3.

How would you vote?

The definitions of the limits considered in Chapter 2 used such phrases as "x approaches a," "$f(x)$ approaches a specific number," "as x gets larger," and "$f(x)$ becomes and remains arbitrarily large." Such phrases, although appealing to the intuition and conveying the sense of a limit, are not precise. The definitions seem to suggest moving objects and call to mind the motion of a pencil as it traces the graph of a function.

The informal approach was adequate during the early development of calculus, from Leibniz and Newton in the seventeenth century through the Bernoullis, Euler, and Gauss in the eighteenth and early nineteenth centuries. By the mid-nineteenth century, mathematicians, facing more complicated functions and more difficult theorems, no longer could depend solely on intuition. They realized that glancing at a graph was no longer adequate to understand the behavior of functions, especially if theorems covering a broad class of functions were needed.

It was Weierstrass who developed, over the period 1841–1856, a way to define limits without any reference to motion or pencils tracing out graphs. His approach, on which he lectured after joining the faculty at the University of Berlin in 1859, has since been followed by pure and applied mathematicians throughout the world. Even an undergraduate advanced calculus course depends on Weierstrass's approach.

In this section we examine how Weierstrass would define infinite and finite limits at infinity:

$$\lim_{x\to\infty} f(x) = \infty \qquad \text{and} \qquad \lim_{x\to\infty} f(x) = L.$$

In the next section we consider finite limits at finite points: $\lim_{x\to a} f(x) = L$. Other limits, including one-sided limits, are considered in the Exercises of Sections 3.8 and 3.9.

The Precise Definition of $\lim_{x\to\infty} f(x) = \infty$

First we treat the case in which the limit is infinite, which includes the example that introduces this section. We had a definition of $\lim_{x\to\infty} f(x) = \infty$ in Section 2.2.

Informal definition of $\lim_{x\to\infty} f(x) = \infty$

1. $f(x)$ is defined for all x beyond some number.
2. As x gets large through positive values, $f(x)$ becomes and remains arbitrarily large.

To take us part way to the precise definition, let us reword the informal definition, paraphrasing it in the following definition, which is still informal.

Reworded informal definition of $\lim_{x\to\infty} f(x) = \infty$

Let $f(x)$ be a function.

1. Assume that $f(x)$ is defined for all x greater than some number c.
2. If x is sufficiently large and positive, then $f(x)$ is necessarily large (and positive).

The precise definition parallels the reworded definition.

Definition: *Precise Definition of* $\lim\limits_{x\to\infty} f(x) = \infty$

The statement $\lim_{x\to\infty} f(x) = \infty$ means the following two conditions are satisfied:

1. $f(x)$ is defined for all x greater than some number c.
2. For each number E there is a number D such that for all $x > D$ it is true that $f(x) > E$.

Think of the number E as a challenge and D as the reply. The larger E is, the larger D must usually be. Only if a number D (which depends on E) can he found for every number E can we make the claim that $\lim_{x\to\infty} f(x) = \infty$.

Observation 3.8.1: *One Way to Think about* E *and* D

In this challenge and reply approach to limits we think of E as the "enemy" and D as the "defense". The limit exists only when it can be shown that there is an appropriate "defense" for any possible "enemy".

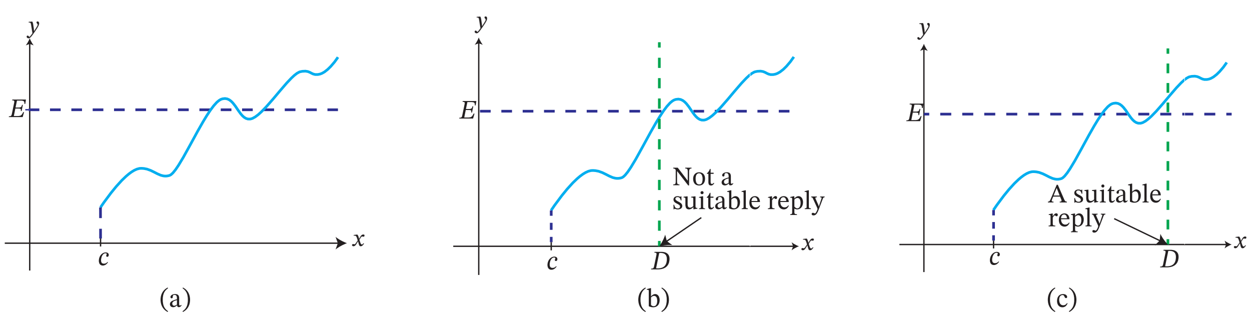


Figure 3.8.2

In other words, D could be expressed as a function of E. To picture the idea behind the precise definition, consider the graph in Figure 3.8.2(a) of a function f for which $\lim_{x\to\infty} f(x) = \infty$. For each possible choice of a horizontal line, say at height E, if you are far enough to the right on the graph of f, you stay above that horizontal line. That is, there is a number D such that if $x > D$, then $f(x) > E$.

The number D in Figure 3.8.2(b) is not a suitable reply. It is too small since there are some values of $x > D$ such that $f(x) \le E$.

The number D in Figure 3.8.2(c) does fulfill the second part of the definition. For every value of $x > D$, $f(x) > E$.

Examples 1 and 2 illustrate how the precise definition is used.

EXAMPLE 1. Using the precise definition of $\lim_{x\to\infty} f(x) = \infty$, show that $\lim\limits_{x\to\infty} 2x = \infty$.

SOLUTION Let E be any positive number. We must show that there is a number D such that whenever $x > D$ it follows that $2x > E$. (For example, if $E = 100$, then $D = 50$ would do because if $x > 50$, then $2x > 100$.) The number D will depend on E. Our goal is to find a formula for D for any value of E.

REMINDER: It is important to remember that your choice for D typically depends on the value of E. That is, the defense you choose depends on the enemy you are facing.

The inequality $2x > E$ is equivalent to $x > E/2$. So if $x > E/2$, then $2x > E$. Choosing $D = E/2$ will suffice. To verify this, when $x > D$ $(= E/2)$, $2x > 2D = 2(E/2) = E$. This allows us to conclude that $\lim_{x\to\infty} 2x = \infty$. ♦

In Example 1 a formula was provided for a suitable D in terms of E, namely, $D = E/2$ (see Figure 3.8.3). When challenged with $E = 1000$, the response $D = 500$ suffices. Any larger value of D also is suitable. If $x > 600$, it is still the case that $2x > 1000$ (since $2x > 1200$). If one value of D is a satisfactory response to a given challenge E, then any larger value of D also is a satisfactory response.

Now that we have a precise definition of $\lim_{x\to\infty} f(x) = \infty$ we can settle the question, is $\lim_{x\to\infty}(x/2 + \sin(x)) = \infty$?

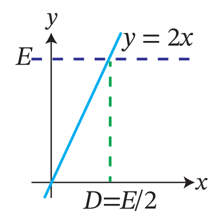


Figure 3.8.3

EXAMPLE 2. Using the precise definition of $\lim_{x\to\infty} f(x) = \infty$, show that $\lim\limits_{x\to\infty} \frac{x}{2} + \sin(x) = \infty$.

SOLUTION Let E be any number. We need to exhibit a number D, depending on E, such that $x > D$ implies

$$\frac{x}{2} + \sin(x) > E. \tag{3.8.1}$$

Now, $\sin(x) \ge -1$ for all x. So, if we can force

$$\frac{x}{2} + (-1) > E$$

then it will follow that

$$\frac{x}{2} + \sin(x) > E.$$

The smallest value of x that satisfies (3.8.1) can be found as follows:

$$\begin{aligned} \frac{x}{2} &> E + 1 && \text{(add 1 to both sides of (3.8.1))} \\ x &> 2(E+1) && \text{(multiply both sides of (3.8.1) by a positive constant).} \end{aligned}$$

Thus $D = 2(E + 1)$ will suffice. That is,

$$\text{If } x > 2(E+1)\text{, then } \frac{x}{2} + \sin(x) > E.$$

To verify this we check that $D = 2(E + 1)$ is a satisfactory reply to E.

Assume that $x > D = 2(E + 1)$. Then

$$\frac{x}{2} > E + 1 \qquad \text{and} \qquad \sin(x) \ge -1.$$

If $a > b$ and $c \geq d$, then $a+c > b+d$.

Adding these inequalities gives

$$\frac{x}{2} + \sin(x) > (E+1) + (-1) = E,$$

which is inequality (3.8.1). Therefore we can conclude that

$$\lim_{x\to\infty} \left(\frac{x}{2} + \sin(x)\right) = \infty.$$

As x increases, the function does become and remain large, despite the small dips downward. ♦

Limits of the form $\lim_{x\to\infty} f(x) = -\infty$, $\lim_{x\to-\infty} f(x) = \infty$, and $\lim_{x\to-\infty} f(x) = -\infty$ can be treated similarly to limits of the form $\lim_{x\to\infty} f(x) = \infty$. (see Exercises 27, 28, and 29, respectively).

We now turn our attention to the second type of limit considered in this section. In these limits the independent variable (x) still grows without bound but the the value of the limit is now a finite real number, L, not infinity (∞) or negative infinity ($-\infty$).

The Precise Definition of $\lim_{x\to\infty} f(x) = L$

We defined $\lim_{x\to\infty} f(x) = L$ informally in Section 2.2. Remember that in this discussion the limit exists only when L is a finite number. When the values of $f(x)$ increase or decrease with any bound, the limit does not exist.

Informal definition of $\lim_{x\to\infty} f(x) = L$

1. $f(x)$ is defined for all x beyond some number.
2. As x gets large through positive values, $f(x)$ approaches L.

Again, we restate this definition before offering the precise definition.

Reworded informal definition of $\lim_{x\to\infty} f(x) = L$

Let $f(x)$ be a function.

1. Assume that $f(x)$ is defined for all x greater than some number c.
2. If x is sufficiently large, then $f(x)$ is necessarily near L.

As before, the precise definition parallels the reworded definition. In order to make precise the phrase "$f(x)$ is necessarily near L," we shall use the absolute value of $f(x) - L$ to measure the distance from $f(x)$ to L. The following definition says that if x is large enough, then $|f(x) - L|$ is as small as we please.

Definition: *Precise Definition of* $\lim\limits_{x\to\infty} f(x) = L$

The statement $\lim_{x\to\infty} f(x) = L$, with L a finite number, means the following two conditions are satisfied:

1. $f(x)$ is defined for all x greater than some number c.
2. For each positive number ϵ there is a number D such that for all $x > D$ it is true that

$$|f(x) - L| < \epsilon.$$

This definition can also be interpreted graphically. Draw two lines parallel to the x-axis, one of height $L+\epsilon$ and one of height $L-\epsilon$. They are the two edges of a band of width 2ϵ centered at $y = L$. Assume that for each positive ϵ, a number D can be found such that the part of the graph to the right of $x = D$ lies within the band. Then we say that as x approaches ∞, $f(x)$ approaches L and write

The Greek letter "ϵ" (epsilon) corresponds to the English letter "e".

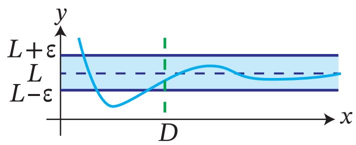


Figure 3.8.4

$$\lim_{x\to\infty} f(x) = L.$$

The positive number ϵ is the challenge, and D is a reply. The smaller ϵ is, the narrower the band is, and the larger D usually must be chosen. The geometric meaning of the precise definition of $\lim_{x\to\infty} f(x) = L$ is shown in Figure 3.8.4.

EXAMPLE 3. Use the precise definition of $\lim_{x\to\infty} f(x) = L$ to show that $\lim_{x\to\infty}\left(1+\frac{1}{x}\right) = 1$.

SOLUTION Here $f(x) = 1 + 1/x$, which is defined for all $x \neq 0$. The number L is 1. We must show that for each positive number ϵ, however small, there is a number D such that, for all $x > D$,

$$\left|\left(1+\frac{1}{x}\right)-1\right| < \epsilon. \tag{3.8.2}$$

Inequality (3.8.2) reduces to

Again, D is expected to depend on ϵ.

$$\left|\frac{1}{x}\right| < \epsilon.$$

Since we may consider only $x > 0$, it is equivalent to

$$\frac{1}{x} < \epsilon. \tag{3.8.3}$$

Multiplying inequality (3.8.3) by the positive number x yields the equivalent inequality

$$1 < x\epsilon. \tag{3.8.4}$$

Dividing inequality (3.8.4) by the positive number ϵ yields

$$\frac{1}{\epsilon} < x \qquad \text{or} \qquad x > \frac{1}{\epsilon}.$$

These steps are reversible, which means that $D = 1/\epsilon$ is a suitable reply to the challenge ϵ. If $x > D = 1/\epsilon$, then

$$\left|\left(1+\frac{1}{x}\right)-1\right| < \epsilon.$$

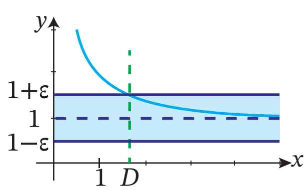


Figure 3.8.5

That is, (3.8.2) is satisfied.

According to the precise definition of $\lim_{x\to\infty} f(x) = L$, we conclude that

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right) = 1.$$

◆

The graph of $f(x) = 1 + 1/x$, shown in Figure 3.8.5, reinforces the argument presented in Example 3. It seems plausible that no matter how narrow a band someone may place around the line $y = 1$, it will always be possible to find a number D such that the part of the graph to the right of $x = D$ stays within it. In Figure 3.8.5 a typical band is shown shaded in cyan.

The precise definitions can also be used to show that some claims about an alleged limit are false. The next example illustrates this.

EXAMPLE 4. Show that the claim that $\lim_{x\to\infty} \sin(x) = 0$ is false.

SOLUTION To show that the claim is false, we must exhibit a challenge $\epsilon > 0$ for which no response D can be found. That is, we exhibit a positive number ϵ such that no D exists for which $|\sin(x) - 0| < \epsilon$ for all $x > D$.

Because $\sin(x) = 1$ whenever $x = \pi/2 + 2n\pi$ for any integer n, there are arbitrarily large values of x for which $\sin(x) = 1$. This suggests how to exhibit an $\epsilon > 0$ for which no response D can be found. Pick ϵ to be a positive number less than or equal to 1. For instance, $\epsilon = 0.7$ will do.

For any number D there is always $x^* > D$ such that we have $\sin(x^*) = 1$. This means that $|\sin(x^*) - 0| = 1 > 0.7$. Hence no response can he found for $\epsilon = 0.7$. Thus the claim that $\lim_{x\to\infty} \sin(x) = 0$ is false. ♦

To conclude this section, we show how the precise definition of the limit can be used to obtain information about new limits.

EXAMPLE 5. Use the precise definition of $\lim\limits_{x\to\infty} f(x) = L$ to show that if f and g are defined everywhere, $\lim\limits_{x\to\infty} f(x) = 2$, and $\lim\limits_{x\to\infty} g(x) = 3$, then $\lim\limits_{x\to\infty} (f(x) + g(x)) = 5$.

SOLUTION The objective is to show that for each positive number ϵ, however small, there is a number D such that, for all $x > D$,

$$|(f(x) + g(x)) - 5| < \epsilon.$$

Because $|(f(x) + g(x)) - 5|$ can be written as $|(f(x) - 2) + (g(x) - 3))|$ it is no larger than $|f(x) - 2| + |g(x) - 3|$. If we can show that for all x sufficiently large both $|f(x) - 2| < \epsilon/2$ and $|g(x) - 3| < \epsilon/2$, then their sum will be no larger than $\epsilon/2 + \epsilon/2 = \epsilon$.

Here is how we can do this.

Because $\lim_{x\to\infty} f(x) = 2$ we know that for any $\epsilon > 0$ there exists a number D_1 with the property that $|f(x) - 2| < \epsilon/2$ for all $x > D_1$. (In this case $\epsilon/2$ is the challenge and D_1 is the response.) Likewise, because $\lim_{c\to\infty} g(x) = 3$ we know that for any $\epsilon > 0$ there exists a number D_2 with the property that $|g(x) - 2| < \epsilon/2$ for all $x > D_2$.

Let D be the larger of D_1 and D_2. For any x greater than D we know that

$$|f(x) + g(x) - 5| < |f(x) - 2| + |g(x) - 3| < \epsilon/2 + \epsilon/2 = \epsilon.$$

According to the precise definition of a limit at infinity, we conclude that $\lim_{x\to\infty}(f(x) + g(x)) = 2 + 3 = 5$. ♦

Summary

We developed a precise definition of the limit of a function as the argument becomes arbitrarily large: $\lim_{x\to\infty} f(x)$. The definition involves being able to respond to a challenge. In the case of an infinite limit, the challenge is a large number. In the case of a finite limit, the challenge is a small number used to describe a narrow horizontal band on both sides of the value of the limit.

EXERCISES for Section 3.8

1. Let $f(x) = 3x$.
- (a) Find a number D such that $x > D$ implies $f(x) > 600$.
- (b) Find another number D such that $x > D$ implies $f(x) > 600$.
- (c) What is the smallest number D such that $x > D$ implies $f(x) > 600$?

2. Let $f(x) = 4x$.
- (a) Find a number D such that $x > D$ implies $f(x) > 1000$.
- (b) Find another number D such that $x > D$ implies $f(x) > 1000$.
- (c) What is the smallest number D such that $x > D$ implies $f(x) > 1000$?

3. Let $f(x) = 5x$. Find a number D such that, for all $x > D$, (a) $f(x) > 2000$ and (b) $f(x) > 10,000$.

4. Let $f(x) = 6x$. Find a number D such that, for all $x > D$, (a) $f(x) > 1200$ and (b) $f(x) > 1800$.

In Exercises 5 to 12 use the precise definition of $\lim_{x\to\infty} f(x) = \infty$ to establish

5. $\lim_{x\to\infty} 3x = \infty$

6. $\lim_{x\to\infty} 4x = \infty$

7. $\lim_{x\to\infty} (x+5) = \infty$

8. $\lim_{x\to\infty} (x-600) = \infty$

9. $\lim_{x\to\infty} (2x+4) = \infty$

10. $\lim_{x\to\infty} (3x-1200) = \infty$

11. $\lim_{x\to\infty} (4x+100\cos(x)) = \infty$

12. $\lim_{x\to\infty} (2x-300\cos(x)) = \infty$

---

13. Let $f(x) = x^2$.

(a) Find a number D such that, for all $x > D$, $f(x) > 100$.

(b) When E is a nonnegative number. Find a number D such that, for all $x > D$, it follows that $f(x) > E$.

(c) When E is a negative number. Find a number D such that, for all $x > D$, it follows that $f(x) > E$.

(d) Using the precise definition of $\lim_{x\to\infty} f(x) = \infty$, show that $\lim_{x\to\infty} x^2 = \infty$.

14. Using the precise definition of $\lim_{x\to\infty} f(x) = \infty$, show that $\lim_{x\to\infty} x^3 = \infty$.

Exercises 15 to 26 concern the precise definition of $\lim_{x\to\infty} f(x) = L$.

15. Let $f(x) = 3 + \frac{1}{x}$ if $x \neq 0$.

(a) Find a number D such that $x > D$ implies $|f(x) - 3| < \frac{1}{10}$.

(b) Find another number D such that $x > D$ implies $|f(x) - 3| < \frac{1}{10}$.

(c) What is the smallest number D such that $x > D$ implies $|f(x) - 3| < \frac{1}{10}$?

(d) Using the precise definition of $\lim_{x\to\infty} f(x) = L$, show that $\lim_{x\to\infty} \left(3 + \frac{1}{x}\right) = 3$.

16. Let $f(x) = \frac{2}{x}$ if $x \neq 0$.

(a) Find a number D such that $x > D$ implies $|f(x) - 0| < \frac{1}{100}$.

(b) Find another number D such that $x > D$ implies $|f(x) - 0| < \frac{1}{100}$.

(c) What is the smallest number D such that $x > D$ implies $|f(x) - 0| < \frac{1}{100}$?

(d) Using the precise definition of $\lim_{x\to\infty} f(x) = L$, show that $\lim_{x\to\infty} \frac{2}{x} = 0$.

In Exercises 17 to 22 use the precise definition of $\lim_{x\to\infty} f(x) = L$ to establish the indicated limits.

17. $\lim_{x\to\infty} \frac{\sin(x)}{x} = 0$

18. $\lim_{x\to\infty} \frac{x+\cos(x)}{x} = 1$

19. $\lim_{x\to\infty} \frac{4}{x^2} = 0$

20. $\lim_{x\to\infty} \frac{2x+3}{x} = 2$

21. $\lim_{x\to\infty} \frac{1}{x-100} = 0$

22. $\lim_{x\to\infty} \frac{2x+10}{3x-5} = \frac{2}{3}$

In Exercises 23 to 26 use the precise definition to prove each statement is false.

23. $\lim_{x\to\infty} \frac{x}{x+1} = \infty$

24. $\lim_{x\to\infty} \sin(x) = \frac{1}{2}$

25. $\lim_{x\to\infty} 3x = 6$

26. $\lim_{x\to\infty} 2x = -7$

Exercises 27 to 30 develop precise definitions of the limit. Phrase them in terms of a challenge number E or ϵ and a reply D. Show the geometric meaning of your definition on a graph.

27. $\lim_{x\to\infty} f(x) = -\infty$

28. $\lim_{x\to-\infty} f(x) = \infty$

29. $\lim_{x\to-\infty} f(x) = -\infty$

30. $\lim_{x\to-\infty} f(x) = L$

---

31. Let $f(x) = 5$ for all x. Using a precise definition, show that (a) $\lim_{x\to\infty} f(x) = 5$ and (b) $\lim_{x\to-\infty} f(x) = 5$.

32. Is this argument correct?

> I will prove that $\lim_{x\to\infty}(2x+\cos(x)) = \infty$. Let E be given. I want $2x+\cos(x) > E$ for all $x > D$, for some number D. So, $2x > E-\cos(x)$. Then, $2x > E-\cos(x)$. And, finally, $x > \frac{1}{2}(E-\cos(x))$. Thus, if $D = \frac{1}{2}(E-\cos(x))$, then $2x+\cos(x) > E$ for all $x > D$.

33. Use the precise definition of $\lim_{x\to\infty} f(x) = L$, to prove this version of the sum law for limits: if $\lim_{x\to\infty} f(x) = A$ and $\lim_{x\to\infty} g(x) = B$, then $\lim_{x\to\infty}(f(x)+g(x)) = A+B$.

34. Use the precise definition of $\lim_{x\to\infty} f(x) = L$, to prove this version of the product law for limits: if $\lim_{x\to\infty} f(x) = A$, then $\lim_{x\to\infty}(f(x)^2) = A^2$.

35. Use the precise definition of $\lim_{x\to\infty} f(x) = L$, to prove this version of the product law for limits: if $\lim_{x\to\infty} f(x) = A$ and $\lim_{x\to\infty} g(x) = B$, then $\lim_{x\to\infty}(f(x)g(x)) = AB$.

36. Assume that $\lim_{x\to\infty} f(x) = 5$. Is there necessarily a number c such that for $x > c$, $f(x)$ stays in the closed interval $[4.5, 5]$? Explain in detail.

37. Assume that $\lim_{x\to\infty} f(x) = 5$. Is there necessarily a number c such that for $x > c$, $f(x)$ stays in the open interval $(4.9, 5.3)$? Explain in detail.

38.

SAM:	I got lost in Example 5 when $\epsilon/2$ came out of nowhere.
JANE:	It's just another small number. They were looking ahead to what they needed.
SAM:	Why must the two numbers add up to ϵ?
JANE:	They don't have to. They could add up to ϵ divided by 12 for instance.
SAM:	What if they added up to 12ϵ? Would that work too?
JANE:	No.
SAM:	I'm getting a headache.

Explain Jane's explanation for Sam's benefit.

3.9 The Precise Definition of Limits at a Finite Point: $\lim_{x\to a} f(x) = L$

To conclude the discussion of limits, we extend the ideas developed in Section 3.8 to the limit of a function at a number a.

Informal definition of $\lim_{x\to a} f(x) = L$

Let f be a function and a some fixed number.

1. Assume that the domain of f contains open intervals (c, a) and (a, b) for some $c < a$ and $b > a$.
2. If, as x approaches a, either from the left or from the right, $f(x)$ approaches a number L, then L is called the **limit** of $f(x)$ as x approaches a. This is written

$$\lim_{x\to a} f(x) = L.$$

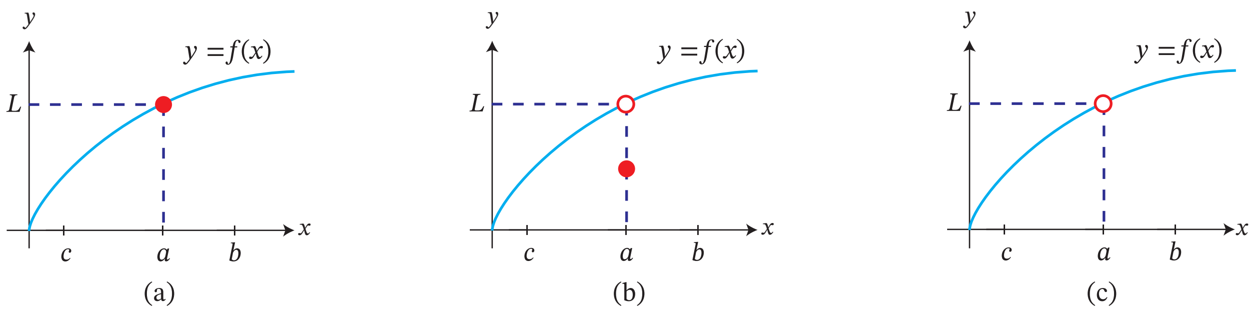


Figure 3.9.1

Figure 3.9.1 illustrates three possibilities for $\lim_{x\to a} f(x) = L$: (a) a is in the domain of f and $f(a) = L$, (b) a is in the domain of f and $f(a) \neq L$, and (c) a is not in the domain of f. These remind us that a need not be in the domain of f. And, even if it is, the value of $f(a)$ plays no role in determining whether $\lim_{x\to a} f(x)$ exists.

Reworded informal definition of $\lim_{x\to a} f(x) = L$

Let f be a function and assume a is some number.

1. Assume that the domain of f contains open intervals (c, a) and (a, b) for some $c < a$ and $b > a$.
2. If x is is sufficiently close to a but not equal to a, then $f(x)$ is necessarily near L.

The following precise definition parallels the reworded informal definition.

Definition: *Precise definition of* $\lim_{x\to a} f(x) = L$

Let f be a function and assume a is some fixed number.

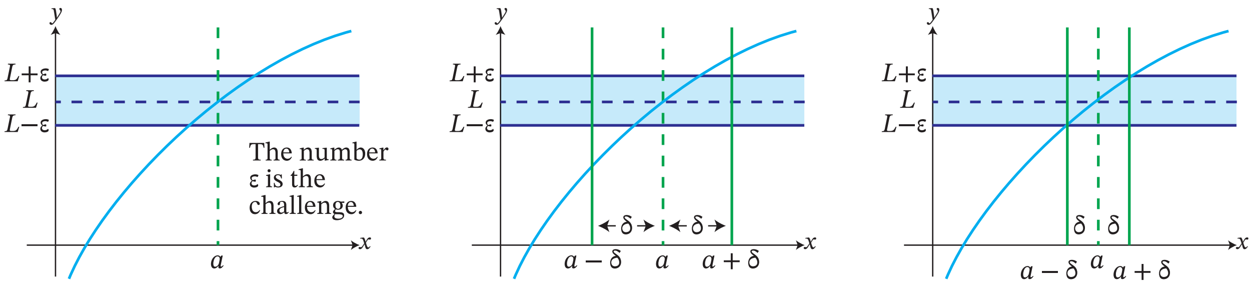
1. Assume that the domain of f contains open intervals (c, a) and (a, b) for some $c < a$ and some $b > a$.
2. For each positive number ϵ there is a positive number δ such that

for all x that satisfy the inequality	$0 < \lvert x - a\rvert < \delta$
it is true that	$\lvert f(x) - L\rvert < \epsilon$.

The Greek letters "δ" (delta) and "Δ" (delta) correspond to the lower case and upper case English letters "d" and "D".

The inequality $0 < |x-a|$ that appears in the definition is a way of saying "x is not a." The inequality $|x-a| < \delta$ asserts that x is within a distance δ of a. The two inequalities are combined as the single statement $0 < |x-a| < \delta$, which describes the open interval $(a-\delta, a+\delta)$ from which a is deleted. This deletion is made since $f(a)$ plays no role in the definition of $\lim_{x\to a} f(x)$.

Once again ϵ is the challenge. The reply is δ. Usually, the smaller ϵ is, the smaller δ will have to be.



(a) (b) (c)

Figure 3.9.2

The geometric significance of the precise definition of $\lim_{x\to a} f(x) = L$ is shown in Figure 3.9.2. The narrow horizontal band of width 2ϵ is the challenge (see Figure 3.9.2(a)). The desired response is a sufficiently narrow vertical band, of width 2δ, such that the part of the graph within the vertical band (except perhaps at $x = a$) also lies in the horizontal band of width 2ϵ. In Figure 3.9.2(b) the vertical band is not narrow enough to meet the challenge of the horizontal band, but the vertical band in Figure 3.9.2(c) is narrow enough.

Assume that for each positive number ϵ it is possible to find a positive number δ such that the parts of the graph between $x = a-\delta$ and $x = a$ and between $x = a$ and $x = a+\delta$ lie within the given horizontal band. Then we say that "as x approaches a, $f(x)$ approaches L". The narrower the horizontal band around the line $y = L$, the smaller δ usually must be.

EXAMPLE 1. Use the precise definition of $\lim\limits_{x\to a} f(x) = L$ to show that $\lim\limits_{x\to 2}(3x+5) = 11$.

SOLUTION Here $f(x) = 3x+5$, $a = 2$, and $L = 11$. Let ϵ be a positive number. We wish to find a number $\delta > 0$ such that for $0 < |x-2| < \delta$ we have $|(3x+5)-11| < \epsilon$.

Let us find out for which x it is true that $|(3x+5)-11| < \epsilon$. This is equivalent to

$$|3x-6| < \epsilon$$

or

$$3|x-2| < \epsilon$$

or

$$|x-2| < \frac{\epsilon}{3}.$$

Thus $\delta = \epsilon/3$ is a suitable response. If $0 < |x-2| < \epsilon/3$, then $|(3x+5)-11| < \epsilon$. ◆

OBSERVATION: Any positive number less than $\epsilon/3$ is also a suitable response.

The algebra of finding a response δ can be more involved for other functions, such as $f(x) = x^2$. The precise definition of limit can actually be easier to apply in more general situations where f and a are not given explicitly. To illustrate, we present a proof of the permanence property.

When the permanence property was introduced in Section 2.5, the only justification we provided was a picture and an appeal to intuition that a continuous function cannot jump instantaneously from a positive value to zero or a negative value — the function has to remain positive on some open interval. Mathematicians call this a "proof by handwaving." We can prove without the use of intuition or handwaving that there must be an open interval around a given input, a, such that for any x in that interval $f(x)$ stays near $f(a)$.

EXAMPLE 2. Prove the permanence property: Assume that f is continuous in an open interval that contains a and that $f(a) = p > 0$. Then for any $q < p$ there is an open interval I containing a such that $f(x) > q$ for all x in I.

SOLUTION Let $p = f(a) > 0$ and let q be a number less than p. Pick $\epsilon = p - q$. (The reason for this choice for ϵ will become clear in a moment.) Because f is continuous at a, $\lim_{x\to a} f(x) = f(a)$. By the precise definition of $\lim_{x\to a} f(x) = L$, when $\epsilon = p - q$ there is a positive number δ such that

$$|f(a) - f(x)| < p - q \qquad (\text{ for } a-\delta < x < a+\delta\,).$$

Thus

$$-(p-q) < f(a) - f(x) < p - q.$$

In particular,

$$f(a) - f(x) < p - q \tag{3.9.1}$$

Because $f(a) = p$, (3.9.1) can be rewritten as

$$p - f(x) < p - q$$

or

$$f(x) > q.$$

Thus $f(x)$ is greater than q if x is in the interval $I = (a-\delta, a+\delta)$. ♦

Observation 3.9.1: *A Common Use of the Permanence Property*

A very common instance of the permanence property is the case of Example 2 with $p = f(a) > 0$ and $q = 0$:

> If a continuous function $f(x)$ is positive for one input number a, so $f(a) > 0$,
> then $f(x)$ must be positive on an interval containing a.

Summary

This section developed a precise definition of the limit of a function as the argument approaches a fixed number: $\lim_{x\to a} f(x)$. It involves being able to respond to an arbitrary challenge number. For a finite limit, the challenge is a small positive number. The smaller that number, the harder it is to meet the challenge.

In addition, it also gave a rigorous proof of the permanence property. That we could deduce it from the precise definition of a limit reassures us that the precise definition expresses what we feel the word "limit" means.

EXERCISES for Section 3.9

In Exercises 1 to 4 use the precise definition of $\lim_{x\to a} f(x) = L$ to justify the indicated limit.

1. $\lim_{x\to 2} 3x = 6$
2. $\lim_{x\to 3} (4x-1) = 11$
3. $\lim_{x\to 1} (x+2) = 3$
4. $\lim_{x\to 5} (2x-3) = 7$

In Exercises 5 to 8 find a positive number δ such that the point $(x, f(x))$ lies in the shaded band for all x in the interval $(a-\delta, a+\delta)$.

5. $f(x) = \dfrac{x}{2} + 1$, $a = 2$, $\epsilon = 0.1$ See Figure 3.9.3(a).
6. $f(x) = \dfrac{3}{2}x + 6$, $a = 2$, $\epsilon = 0.4$ See Figure 3.9.3(b).
7. $f(x) = 2x^2$, $a = 1$, $\epsilon = 0.4$ See Figure 3.9.3(c).

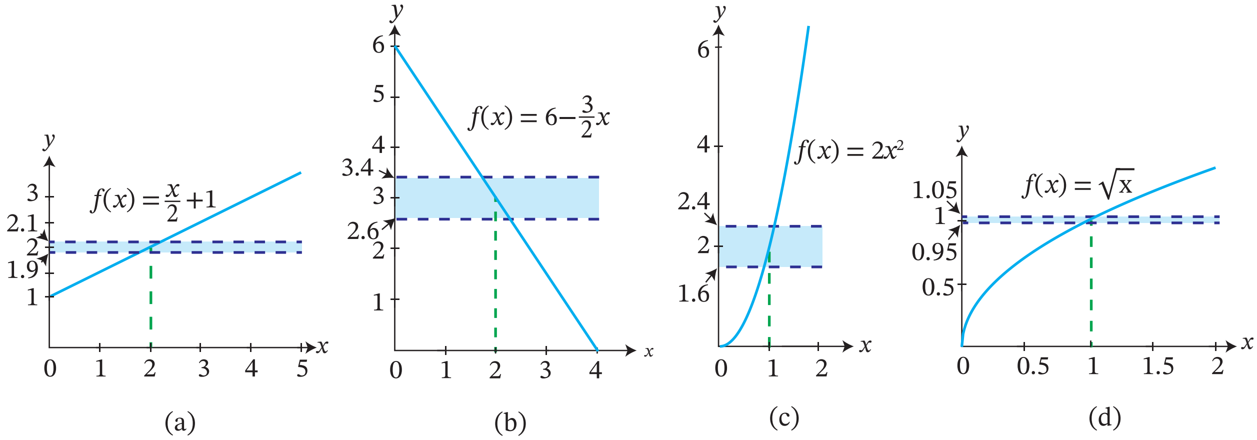


Figure 3.9.3

8. $f(x)=\sqrt{x}$, $a=1$, $\epsilon=0.05$ See Figure 3.9.3(d).

In Exercises 9 to 12 use the precise definition of $\lim_{x\to a} f(x)=L$ to justify the indicated limit.

9. $\lim_{x\to 1}(3x+5)=8$ **10.** $\lim_{x\to 1}\frac{5x+3}{4}=2$ **11.** $\lim_{x\to 0}\frac{x^2}{4}=0$ **12.** $\lim_{x\to 0}4x^2=0$

---

13. Give an example of a number $\delta>0$ such that $|x^2-4|<1$ if $0<|x-2|<\delta$.

14. Give an example of a number $\delta>0$ such that $|x^2+x-2|<0.5$ if $0<|x-1|<\delta$.

15. Let $f(x)=9x^2$.

(a) Find $\delta>0$ such that, for $0<|x-0|<\delta$, it follows that $|9x^2-0|<\frac{1}{100}$.

(b) Let ϵ be any positive number. Find a positive number δ such that, for $0<|x-0|<\delta$ we have $|9x^2-0|<\epsilon$.

(c) Show that $\lim_{x\to 0}9x^2=0$.

16. Let $f(x)=x^3$.

(a) Find $\delta>0$ such that, for $0<|x-0|<\delta$, it follows that $|x^3-0|<\frac{1}{1000}$.

(b) Show that $\lim_{x\to 0}x^3=0$.

17. Show that the assertion "$\lim_{x\to 2}3x=5$" is false. To do this, it is necessary to exhibit a positive number ϵ such that there is no response number $\delta>0$.

18. Show that the assertion"$\lim_{x\to 2}x^2=3$" is false.

Develop precise definitions of limits in the forms given in Exercise 19 to 24. Phrase your definitions in terms of a challenge, E or ϵ, and a response, δ.

19. $\lim_{x\to a^+}f(x)=L$ **20.** $\lim_{x\to a^-}f(x)=L$ **21.** $\lim_{x\to a}f(x)=\infty$

22. $\lim_{x\to a}f(x)=-\infty$ **23.** $\lim_{x\to a^+}f(x)=\infty$ **24.** $\lim_{x\to a^-}f(x)=\infty$

---

25. In the proof of the permanence property given in Example 2, $p = f(a) > 0$ and $q < p$.
(a) Would the argument have worked if we had used $\epsilon = 2(p-q)$?
(b) Would the argument have worked if we had used $\epsilon = \frac{1}{2}(p-q)$?
(c) Would the argument have worked if we had used $\epsilon = q$?
(d) What is the largest value of ϵ for which the proof of the permanence property works? $\epsilon = p - q$.

26. The permanence property discussed in Example 2 and Exercise 25 pertains to limits at a finite point a. State, and prove, a version of the permanence property that is valid when a is replaced by ∞.

27. (a) Show that if $0 < \delta < 1$ and $|x-3| < \delta$ then $|x^2 - 9| < 7\delta$.
(b) Use (a) to deduce that $\lim_{x\to 3} x^2 = 9$.

28. (a) Show that if $0 < \delta < 1$ and $|x-4| < \delta$ then $|\sqrt{x} - 2| < \frac{\delta}{\sqrt{3}+2}$.
(b) Use (a) to deduce that $\lim_{x\to 4} \sqrt{x} = 2$.

29. (a) Show that if $0 < \delta < 1$ and $|x-3| < \delta$ then $|x^2 + 5x - 24| < 12\delta$.
(b) Use (a) to deduce that $\lim_{x\to 3} (x^2 + 5x) = 24$.

30. (a) Show that if $0 < \delta < 1$ and $|x-2| < \delta$ then $\left|\frac{1}{x} - \frac{1}{2}\right| < \frac{\delta}{2}$.
(b) Use (a) to deduce that $\lim_{x\to 2} \frac{1}{x} = \frac{1}{2}$.

31. Use a precise definition of limit to prove: if f is defined in an open interval including a and f is continuous at a, then $3f$ is also continuous at a.

32. Use a precise definition of limit to prove: if f and g are both defined in an open interval including a and both are continuous at a, then $f+g$ is also continuous at a.

33. Use a precise definition of limit to prove: if f and g are both continuous at a, then their product, fg, is also continuous at a. Assume that both functions are defined at least in an open interval around a.

34. Assume that $f(x)$ is continuous at a and is defined on an open interval containing a. Assume that $f(x) = p > 0$ and that q is a number greater than p. Using the precise definition of a limit, show that there is an open interval, I, containing a such that $f(x) < q$ for all x in I.

3.S Chapter Summary

In this chapter we defined the derivative of a function, developed ways to compute derivatives, and applied them to graphs and motion.

The derivative of a function f at a number $x = a$ is defined as the limit of the slopes of secant lines through the points $(a, f(a))$ and $(b, f(b))$ as the input b is taken closer and closer to a.

Algebraically, the derivative is the limit of a quotient, the change in the output divided by the change in the input. It is usually written as

$$\lim_{x\to a} \frac{f(x)-f(a)}{x-a}, \quad \text{or} \quad \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}, \quad \text{or} \quad \lim_{\Delta x\to 0} \frac{\Delta y}{\Delta x}.$$

The derivative is denoted in several ways, such as f', $f'(x)$, df/dx, dy/dx, or $D(f)$.

The list of derivatives of basic functions in Table 3.S.1 should be memorized.

In addition to memorizing the basic derivatives in Table 3.S.1, it is also important to understand that functions most frequently encountered in applications are differentiable, that is, they have derivatives. Geometrically, the derivative exists whenever the graph of the function on a small interval looks almost like a straight line.

The derivative records how fast something changes. The velocity of a moving object is defined as the derivative of the object's position. Also, the derivative gives the slope of the tangent line to the graph of a function.

Function	Derivative	Note	Function	Derivative	Note
x^a	ax^{a-1}	a is any constant: positive, negative, zero, rational, or irrational	$\sqrt{x}$	$\frac{1}{2\sqrt{x}} = \frac{1}{2}x^{-1/2}$	$x > 0$
			$\frac{1}{x} = x^{-1}$	$\frac{-1}{x^2} = -x^{-2}$	$x \neq 0$
$\sin(x)$	$\cos(x)$		$\cos(x)$	$-\sin(x)$	
e^x	e^x		a^x	$a^x \ln(a)$	$a > 0$ $(a^x = e^{x(\ln(a))})$
$\ln(x)$	$\frac{1}{x}$	$x > 0$	$\ln\|x\|$	$\frac{1}{x}$	$x \neq 0$
$\tan(x)$	$\sec^2(x)$	$x \neq \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots$	$\sec(x)$	$\sec(x)\tan(x)$	$x \neq \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots$
$\arcsin(x)$	$\frac{1}{\sqrt{1-x^2}}$	$-1 < x < 1$	$\arctan(x)$	$\frac{1}{1+x^2}$	

Table 3.S.1

We then developed ways to compute the derivative of functions expressible in terms of the functions met in algebra and trigonometry, as well as exponentials with a fixed base and logarithms, the elementary functions. They were based on three limits:

$$\lim_{x\to a} \frac{x^n - a^n}{x - a} = na^{n-1} (n \text{ a positive integer}), \qquad \lim_{x\to 0} \frac{e^x - 1}{x} = 1, \text{and} \qquad \lim_{x\to 0} \frac{\sin(x)}{x} = 1.$$

Using them, we obtained the derivatives of x^n, e^x, and $\sin(x)$. We showed, if we knew the derivatives of two functions, how to compute the derivatives of their sum, difference, product, and quotient.

The next step was the development of a most important computational tool: the chain rule. It enables us to differentiate a composite function, such as $\cos^3(x^2)$, telling us that its derivative is $3\cos^2(x^2)(-\sin(x^2))(2x)$.

Differentiating inverse functions enabled us to show that the derivative of $\ln(x)$ is $1/x$ for $x > 0$ and the derivative of $\arcsin(x)$ is $1/\sqrt{1-x^2}$ for $-1 < x < 1$.

As you work with derivatives you may begin to think of them as slope or velocity or rate of change, and forget their underlying definition as limits. However, we will return to the definition in terms of limits as we develop more applications of the derivative.

We also introduced the antiderivative and, related to it, the slope field. While the derivative of an elementary function is again elementary, an antiderivative often is not. For instance, $\sqrt{1+x^3}$ does not have an elementary antiderivative. However, as we will see in Chapter 6, it does have an antiderivative. Chapter 8 will present different ways to find antiderivatives.

The derivative of the derivative is the second derivative. For motion, the second derivative describes acceleration. It can be denoted several ways, such as f'', $f''(x)$, d^2f/dx^2, d^2y/dx^2, and $D^2(f)$. While the first and second derivatives suffice for most applications, higher derivatives of all orders are used in Chapter 5 to estimate the error when approximating a function by a polynomial.

The final two sections provided precise definitions of limits and a proof of the permanence property.

EXERCISES for Section 3.S

In Exercises 1 to 19 differentiate the expression.

1. $\exp\left(x^2\right)$
2. 2^{x^2}
3. $x^3 \sin(4x)$
4. $\dfrac{1+x^2}{1+x^3}$
5. $\ln\left(x^3\right)$
6. $\ln\left(x^3+1\right)$
7. $\cos^4\left(x^2\right)\tan(2x)$
8. $\sqrt{5x^2+x}$
9. $\arcsin\left(\sqrt{3+2x}\right)$
10. $x^2\arctan(2x)e^{3x}$
11. $\sec^2(3x)$
12. $\sec^2(3x)-\tan^2(3x)$
13. $\left(\dfrac{3+2x}{4+5x}\right)^3$
14. $\dfrac{1}{1+2e^{-x}}$
15. $\dfrac{x}{\sqrt{x^2+1}}$
16. $(\arcsin(3x))^2$
17. $x^2\arctan(3x)$
18. $\sin^5(3x^2)$
19. $\dfrac{1}{(2^x+3^x)^{20}}$

In Exercises 20 to 29 give an antiderivative of the expression. Use differentiation to check each answer.

20. $4x^3$
21. x^3
22. $\dfrac{3}{x^2}$
23. $\cos(x)$
24. $\cos(2x)$
25. $\sin^{100}(x)\cos(x)$
26. $\dfrac{1}{x+1}$
27. $5e^{4x}$
28. $\dfrac{1}{e^x}$
29. 2^x

In Exercises 30 to 33, sketch the slope field and draw the solution curve through the point $(0,1)$.

30. $\dfrac{dy}{dx}=\dfrac{1}{x+1}$
31. $\dfrac{dy}{dx}=e^{-x^2}$
32. $\dfrac{dy}{dx}=-y$
33. $\dfrac{dy}{dx}=y-x$

In Exercises 34 to 55 evaluate the derivative to verify each equation. Determine all values of x for which each equation is valid. The letters a, b, c, and d denote constants. These exercises, based on tables of antiderivatives, provide practice in differentiation and algebra.

34. $\dfrac{d}{dx}\left(\dfrac{1}{a}\tan^{-1}\left(\dfrac{x}{a}\right)\right)=\dfrac{1}{a^2+x^2}$
35. $D\left(\dfrac{1}{2a}\ln\left(\dfrac{a+x}{a-x}\right)\right)=\dfrac{1}{a^2-x^2}$
36. $\left(\ln\left(x+\sqrt{a^2+x^2}\right)\right)'=\dfrac{1}{\sqrt{a^2+x^2}}$
37. $\dfrac{d}{dx}\left(\dfrac{1}{a}\ln\left(\dfrac{\sqrt{a^2+x^2}-a}{x}\right)\right)=\dfrac{1}{x\sqrt{a^2+x^2}}$
38. $D\left(\dfrac{-1}{b(a+bx)}\right)=\dfrac{1}{(a+bx)^2}$
39. $\left(\dfrac{1}{b^2}(a+bx-a\ln(a+bx))\right)'=\dfrac{x}{a+bx}$
40. $\dfrac{d}{dx}\left(\dfrac{1}{b^2}\left(\dfrac{a}{2(a+bx)^2}-\dfrac{1}{a+bx}\right)\right)=\dfrac{x}{(a+bx)^3}$
41. $D\left(\dfrac{1}{ad-bc}\ln\left(\dfrac{c+dx}{a+bx}\right)\right)=\dfrac{1}{(a+bx)(c+dx)}$
42. $D\left(\dfrac{1}{a}\arccos\left(\dfrac{a}{x}\right)\right)=\dfrac{1}{|x|\sqrt{x^2-a^2}}$
43. $\left(\dfrac{x}{2}\sqrt{a^2-x^2}+\dfrac{a^2}{2}\arcsin\left(\dfrac{x}{a}\right)\right)'=\sqrt{a^2-x^2}$
44. $\left(\dfrac{-x}{2}\sqrt{a^2-x^2}+\dfrac{a^2}{2}\arcsin\left(\dfrac{x}{a}\right)\right)'=\dfrac{x^2}{\sqrt{a^2-x^2}}$
45. $D\left(-\dfrac{\sqrt{a^2-x^2}}{x}-\arcsin\left(\dfrac{x}{a}\right)\right)=\dfrac{\sqrt{a^2-x^2}}{x^2}$
46. $\left(\arcsin(x)-\sqrt{1-x^2}\right)'=\sqrt{\dfrac{1+x}{1-x}}$
47. $\dfrac{d}{dx}\left(\dfrac{x}{2}-\dfrac{1}{2}\cos(x)\sin(x)\right)=\sin^2(x)$
48. $D\left(x\arcsin(x)+\sqrt{1-x^2}\right)=\arcsin(x)$
49. $\left(x\tan^{-1}(x)-\dfrac{1}{2}\ln(1+x^2)\right)'=\arctan(x)$

50. $\dfrac{d}{dx}\left(\dfrac{e^{ax}}{a^2}(ax-1)\right) = xe^{ax}$

51. $D\left(x - \ln(1+e^x)\right) = \dfrac{1}{1+e^x}$

52. $\left(\dfrac{x}{2}(\sin(\ln(ax)) - \cos(\ln(ax)))\right)' = \sin(\ln(ax))$

53. $\left(\dfrac{e^{ax}(a\sin(bx) - b\cos(bx))}{a^2+b^2}\right)' = e^{ax}\sin(bx)$

54. $\left(\dfrac{2}{\sqrt{4ac-b^2}}\arctan\left(\dfrac{2cx+b}{\sqrt{4ac-b^2}}\right)\right)' = \dfrac{1}{a+bx+cx^2} \qquad (4ac > b^2)$

55. $\left(\dfrac{1}{\sqrt{b^2-4ac}}\ln\left(\dfrac{2cx+b-\sqrt{b^2-4ac}}{2cx+b+\sqrt{b^2-4ac}}\right)\right)' = \dfrac{1}{a+bx+cx^2} \qquad (4ac < b^2)$

In Exercises 56 to 59 give two antiderivatives for each expression.

56. xe^{x^2}

57. $(x^2+x)e^{x^3+3x}$

58. $\cos^3(x)\sin(x)$

59. $\sin(2x)$

---

60. Verify that $2(\sqrt{x}-1)e^{\sqrt{x}}$ is an antiderivative of $e^{\sqrt{x}}$.

61. Sam threw a baseball straight up and caught it 6 seconds later. (Assume the ball is released and caught at the top of Sam's head.)

(a) How high above his head did it rise?
(b) How fast was it going as it left his hand?
(c) How fast was it going when he caught it?
(d) Translate the answers in (b) and (c) to miles per hour. (60 mph = 88 fps)

62. Assuming that $D(x^4) = 4x^3$ and $D(x^7) = 7x^6$, you could find $D(x^3)$ from them by viewing x^3 as x^7/x^4 and using the formula for differentiating a quotient. Show how you could use them to find (a) $D(x^{11})$, (b) $D(x^{-4})$, (c) $D(x^{28})$, and (d) $D(x^8)$.

63. Let $y = x^{m/n}$, where $x > 0$ and m and $n \neq 0$ are integers. Assuming that y is differentiable, show that $\dfrac{dy}{dx} = \dfrac{m}{n}x^{\frac{m}{n}-1}$ by starting with $y^n = x^m$ and differentiating both sides with respect to x.

64. A spherical balloon is being filled with helium at the rate of 3 cubic feet per minute. At what rate is the radius increasing when the radius is (a) 2 feet? (b) 3 feet?

NOTE: The volume of a ball of radius r is $\frac{4}{3}\pi r^3$.

65. An object at the end of a vertical spring is at rest. When pulled down and released, it goes up and down for a while. With the origin of the y-axis at the original rest position, and downward motion corresponding to $y > 0$, the position of the object t seconds later is $y = 3e^{-2t}\cos(2\pi t)$ inches.

(a) What is the physical significance of 3 in the formula for the position?
(b) What does e^{-2t} tell us?
(c) What does $\cos(2\pi t)$ tell us?
(d) How long does it take the object to complete a full cycle (go from $y = 0$, down, up, then down to $y = 0$)?
(e) What happens to the object after a long time?

66. The motor on a moving motor boat is turned off. It then coasts along the x-axis. Its position, in meters, at time t (seconds) is $x = 500 - 50e^{-3t}$. (This means the force of the water slowing the boat is proportional to the velocity of the boat.) (a) Where is it at time $t = 0$? (b) What is its velocity at time t? (c) What is its acceleration at time t? (d) How far does it coast? (e) Show that its acceleration is proportional to its velocity.

See also Exercise 79.

67. It is safe to switch the "sin" and "lim" in $\sin\left(\lim_{x\to 0}\frac{e^x-1}{x}\right)=\lim_{x\to 0}\left(\sin\left(\frac{e^x-1}{x}\right)\right)$. However, such a switch sometimes is not correct. Let f be defined by $f(x)=2$ for $x\neq 1$ and $f(1)=0$.

(a) Show that $f\left(\lim_{x\to 0}\frac{e^x-1}{x}\right)$ is not equal to $\lim_{x\to 0} f\left(\frac{e^x-1}{x}\right)$.

(b) What property of $\sin(x)$ permits us to switch it with lim?

It is important to keep in mind the definition of a derivative as a limit. Exercises 68 to 72 are intended to reinforce the definition.

68. Define the derivative of the function $g(x)$ at $x=a$ in
(a) the x and $x+h$ notation, (b) the x and a notation, and (c) the Δy and Δx notation.

69. We obtained the derivative of $\sin(x)$ using the x and $x+h$ notation and the addition identity for $\sin(x+h)$. Instead, obtain the derivative of $\sin(x)$ using the x and a notation. That is, find $\lim_{x\to a}\frac{\sin(x)-\sin(a)}{x-a}$.

(a) Show that $\sin(x)-\sin(y)=2\sin\left(\frac{1}{2}(x-y)\right)\cos\left(\frac{1}{2}(x+y)\right)$. (b) Use the identity in (a) to find $\lim_{x\to a}\frac{\sin(x)-\sin(a)}{x-a}$.

70. We obtained the derivative for $\tan(x)$ by writing it as $\sin(x)/\cos(x)$. Instead, obtain it directly by finding $\lim_{h\to 0}\frac{\tan(x+h)-\tan(x)}{h}$.

71. Show that $\frac{\tan(a)}{\tan(b)}>\frac{a}{b}>\frac{\sin(a)}{\sin(b)}$ for all angles a and b in the first quadrant with $a>b$.

72. We obtained the derivative of $\ln(x)$, $x>0$, by viewing it as the inverse of $\exp(x)$. Instead, find it directly from the definition of the derivative.

Exercises 73 and 75 show how we could have predicted that $\ln(x)$ would provide an antiderivative for $1/x$.

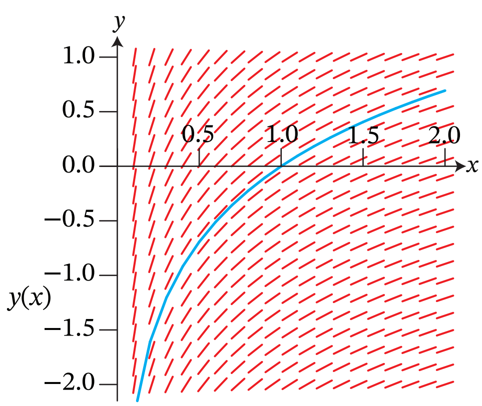


Figure 3.S.1

73. The slope field for $\frac{1}{x}$ and the antiderivative of $\frac{1}{x}$ passing through $(1,0)$ are shown in Figure 3.S.1. The antiderivative of $\frac{1}{x}$ that passes through $(1,0)$ is $\ln(x)$. One would expect that for t near 1, the antiderivative of $\frac{1}{x^t}$ that passes through $(1,0)$ would look much like $\ln(x)$ when x is near 1. To verify that this is true

(a) with $t=1.1$, graph the slope field for $\frac{1}{x^t}$ and the antiderivative of $\frac{1}{x^t}$ that passes through $(1,0)$

(b) repeat (a) for $t=0.9$

(c) repeat (a) for $t=1.01$

(d) repeat (a) for $t=0.99$

74.
JANE: I can show $\lim_{x\to 0}\frac{2^x-1}{x}=\ln(2)$ more easily than the book does.
SAM: I need it.
JANE: I just rewrite the limit like this: $\lim_{x\to 0}\frac{2^x-1}{x}=\lim_{x\to 0}\frac{\left(e^{\ln(2)}\right)^x-1}{x}=\lim_{x\to 0}\frac{e^{\ln(2)x}-1}{x}=\lim_{x\to 0}\frac{\ln(2)\left(e^{\ln(2)x}-1\right)}{\ln(2)x}$.
SAM: It looks like you just made it more complicated. Why did you do that?
JANE: Because it's obvious that $\lim_{x\to 0}\frac{\left(e^{\ln(2)x}-1\right)}{\ln(2)x}\cdot\ln(2)=1\cdot\ln(2)=\ln(2)$.
SAM: Not to me.

Explain Jane's argument.

75. (a) Verify that for $t \neq 1$ the antiderivative of $\frac{1}{x^t}$ that passes through $(1,0)$ is $\frac{x^{1-t}-1}{1-t}$.

(b) Holding x fixed and letting t approach 1, show that $\lim_{t\to 1} \frac{x^{1-t}-1}{1-t} = \ln(x)$.

See also Exercise 73.

---

76. Define f as $f(x) = \begin{cases} x & \text{if } x \text{ is rational} \\ -x & \text{if } x \text{ is irrational.} \end{cases}$

Use a dotted curve to indicate that points are missing.

(a) What does the graph of f look like? (b) Does $\lim_{x\to 0} f(x)$ exist? (c) Does $\lim_{x\to 1} f(x)$ exist? (d) Does $\lim_{x\to\sqrt{2}} f(x)$ exist? (e) For which numbers a does $\lim_{x\to a} f(x)$ exist?

77. Define f as $f(x) = \begin{cases} x^2 & \text{if } x \text{ is rational} \\ x^3 & \text{if } x \text{ is irrational.} \end{cases}$

Use a dotted curve to indicate that points are missing.

(a) What does the graph of f look like? (b) Does $\lim_{x\to 0} f(x)$ exist? (c) Does $\lim_{x\to 1} f(x)$ exist? (d) Does $\lim_{x\to\sqrt{2}} f(x)$ exist? (e) For which numbers a does $\lim_{x\to a} f(x)$ exist?

78. A heavy block rests on a horizontal table covered with thick oil. The block, which is at the origin of the x-axis, is given an initial velocity v_0 at time $t = 0$. It then coasts along the positive x-axis.

Assume that its acceleration is $-k\sqrt{v(t)}$, where $v(t)$ is the velocity at time t and k is a constant. (That means it meets a resistance force proportional to the square root of its velocity.)

(a) Show that $\frac{dv}{dt} = -kv^{1/2}$.
(b) Is k positive or negative? Explain.
(c) Show that $2v^{1/2}$ and $-kt$ have the same derivative with respect to t.
(d) Show that $2v^{1/2} = -kt + 2v_0^{1/2}$.
(e) When, in terms of v_0 and k, does the block come to a rest?
(f) How far, in terms of v_0 and k, does the block slide?

79. A motorboat traveling along the x-axis at the speed v_0 stops its motor at time $t = 0$ when it is at the origin. It then coasts along the positive x-axis. Assume the resistance force of the water is proportional to the velocity. That implies the acceleration of the boat is proportional to its velocity, $v(t)$.

See also Exercise 66.

(a) Show that there is a constant k such that $\frac{dv}{dt} = -kv(t)$.
(b) Is k positive or negative? Explain.
(c) Deduce that $\ln(v)$ and $-kt$ have the same derivative with respect to t.
(d) Deduce that $\ln(v(t)) = -kt + \ln(v_0)$.
(e) Deduce that $v(t) = v_0 e^{-kt}$.
(f) According to (e), how long, in terms of v_0 and k, does the boat continue to move?
(g) How far, in terms of v_0 and k, does it move during that time?

80. Archimedes used the following property of a parabola in his study of the equilibrium of floating bodies. Let P be a point on the parabola $y = x^2$ other than the origin. The line perpendicular to the parabola at P meets the y-axis in a point Q. The line through P and parallel to the x-axis meets the y-axis in a point R. Show that the length of QR is constant, independent of the choice of P.

This problem introduces the **subnormal** of the graph; compare this with Exercises 24 and 25 in Section 3.2.

Calculus is Everywhere # 4

Reflections on Reflections: Ellipses, Parabolas, and a Solar Cooker

The shiny surface behind a flashlight bulb has a remarkable property. Light from the bulb, after bouncing off the reflector, moves parallel to the axis of the reflector, whose shape is formed by rotating a parabola about its axis, as shown in Figure C.4.1. A satellite dish is also parabolic, but is used in reverse. All the radio waves arriving parallel to the axis are reflected by the surface and pass through a single point on the axis, called the **focus**. Because all parabolas are similar we lose no generality by assuming the one we will use is the graph of $y = x^2$.

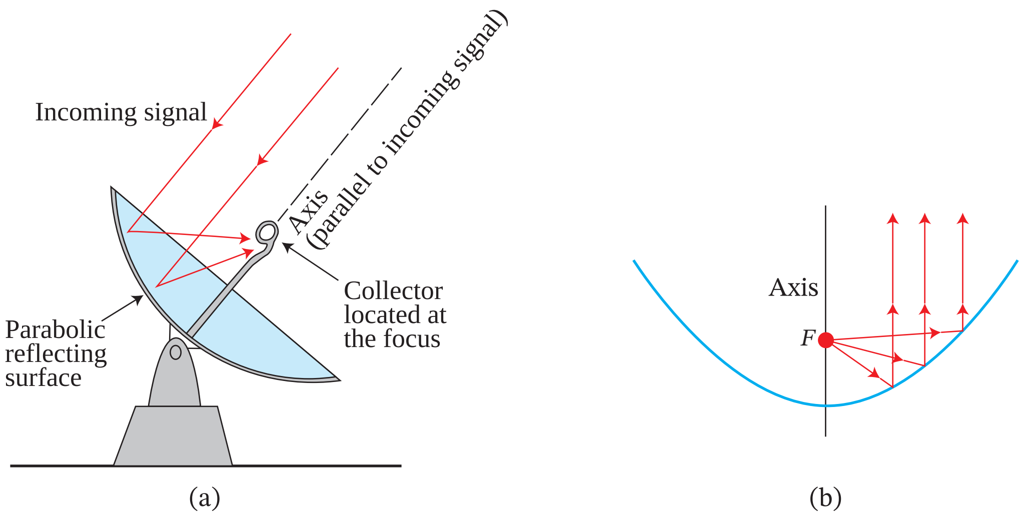


Figure C.4.1

There is another bounded surface with a similar but different reflecting property. This surface is formed by rotating an ellipse about one of its axes. To describe this property we need a certain property of an ellipse, a property often used to define and to draw an ellipse.

Put two tacks in a piece of cardboard and tie the ends of a string to the tacks. The length of the string must be greater than the distance between the tacks. A pencil held firmly against the string traces out half of an ellipse. At the tacks are the **foci** (fo-si; singular: focus) of the ellipse, labeled F and F' in Figure C.4.2(a).

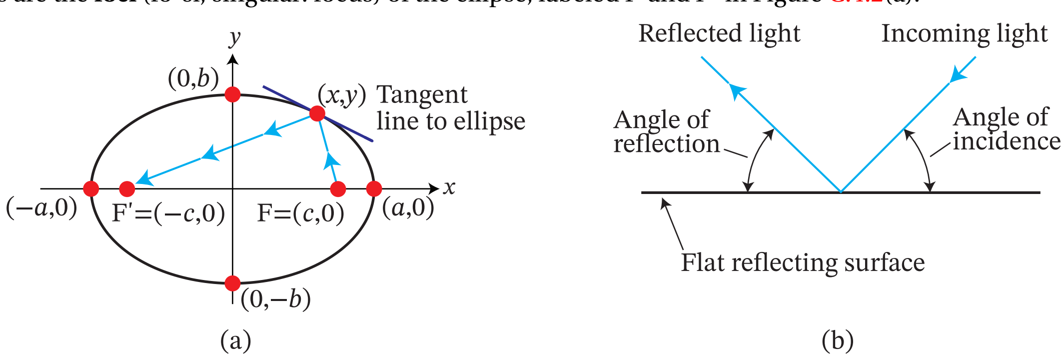


Figure C.4.2

Sound produced at one focus, after reflection by the ellipse, then travels to the other focus. This is the basis for the whispering room in the United States Capitol. Whispers at one focus can be heard easily at the other focus. Luckily, sounds of all pitches travel at the same speed.

Before we delve into the mathematics we begin with a brief description of the physics on which the analysis will be based.

The Physics

Light or sound, after striking a smooth flat surface at an angle, bounces off at the same angle, as shown in Figure C.4.2(b).

The Mathematics

We start with a curve C and a point P on it where light, say, strikes the curve. Then we draw the tangent to the curve at P, as shown in Figure C.4.3(a). Rather than viewing the light as striking the curve at P, we view it as striking the tangent at P, as shown in Figure C.4.3(b).

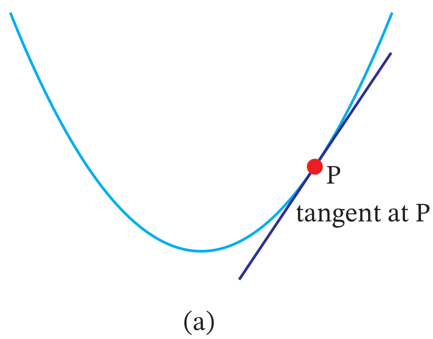


Figure C.4.3

We will also need the tangent of the angle formed by two intersecting lines, L with slope m and L' with slope m'.

The Angle Between Two Lines

To establish the reflection properties we will use the principle that the angle of reflection equals the angle of incidence, as in Figure C.4.2, and work with the angle between two lines, given their slopes.

A line L in the xy-plane forms an **angle of inclination** α, $0 \leq \alpha < \pi$, with the x-axis. The slope of L is $\tan(\alpha)$. If $\alpha = \pi/2$, the slope is not defined.

Why is $\tan(\alpha)$ equal to the slope of L? See Figure C.4.4(a).

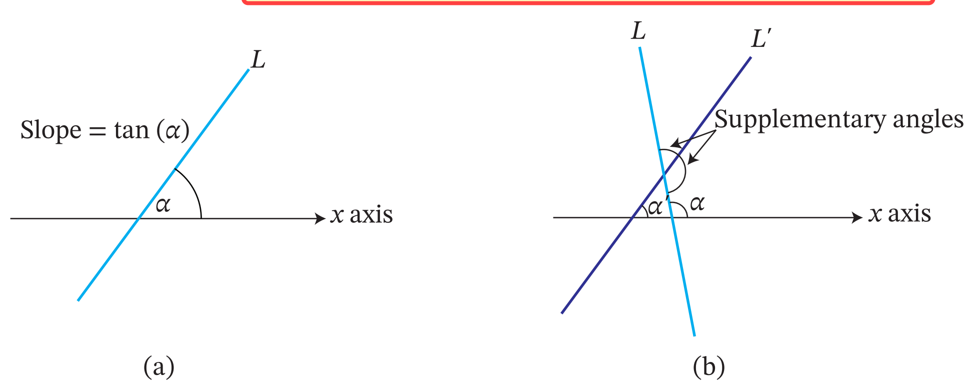


Figure C.4.4

Two lines L and L' with angles of inclination α and α' and slopes m and m', respectively, as in Figure C.4.4(b) intersect and form two (supplementary) angles. The following definition distinguishes one as the angle between L and L'.

Definition: *Angle between two lines*

Given two lines L and L' in the xy-plane, named so that L has the larger angle of inclination, $\alpha > \alpha'$. The **angle** θ **between** L **and** L' is defined to be

$$\theta = \alpha - \alpha'.$$

If L and L' are parallel, θ is defined to be 0.

So, θ is the counterclockwise angle from L' to L and $0 \le \theta < \pi$. The tangent of θ can be expressed in terms of the slopes m of L and m' of L':

$$\begin{aligned}
\tan(\theta) &= \tan(\alpha - \alpha') && (\text{ definition of } \theta \text{ }) \\
&= \frac{\tan(\alpha) - \tan(\alpha')}{1 + \tan(\alpha)\tan(\alpha')} && (\text{ by the identity for } \tan(A - B) \text{ }) \\
&= \frac{m - m'}{1 + mm'}. && (\text{ definition of } m \text{ and } m' \text{ })
\end{aligned}$$

Thus

$$\tan(\theta) = \frac{m - m'}{1 + mm'}.$$

This equation is the main tool in establishing the reflection property of both parabolas and ellipses.

The Reflection Property of a Parabola

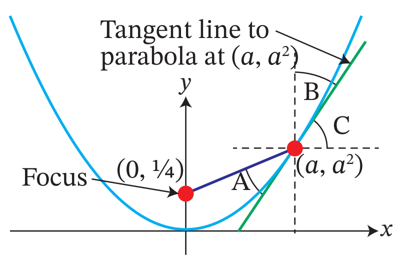


Figure C.4.5

Figure C.4.5 shows a graph of the parabola $y = x^2$ and it's focus $(0, 1/4)$. We will show that the angles A and B at the typical point (a, a^2) on the parabola are equal. We will do this by showing that $\tan(A) = \tan(B)$.

First of all, recall that the slope of the parabola at (a, a^2) is $2a$. Since B is the complement of C,

$$\tan(B) = \tan\left(\frac{\pi}{2} - C\right) = \frac{1}{\tan(C)} = \frac{1}{2a}.$$

The slope of the line through the focus $(0, 1/4)$ and a point on the parabola $\left(a, a^2\right)$ is

$$\frac{a^2 - \frac{1}{4}}{a - 0} = \frac{4a^2 - 1}{4a}.$$

Therefore, as the formula $\tan(\theta) = (m - m')/(1 + mm')$ promises, the tangent of the angle between these two lines is

$$\tan(A) = \frac{2a - \frac{4a^2 - 1}{4a}}{1 + 2a\left(\frac{4a^2 - 1}{4a}\right)}.$$

Exercise 1 asks you to supply the algebra to complete the proof that $\tan(B) = \tan(A)$, namely that both sides equal $1/(2a)$.

The Reflection Property of an Ellipse

Now let's look at the reflecting property of an ellipse.

An ellipse consists of points such that the sum of the distances from a point to two fixed points is constant. Assume that the foci of the ellipse are a distance $2c$ apart, and the sum of the distances is $2a$, where $a > c$. If the foci are at $(c,0)$ and $(-c,0)$ and $b^2 = a^2 - c^2$, the equation of the ellipse is

See Figure C.4.6.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad \text{hence,} \quad \frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1.$$

As with the parabola, our goal is to show that $\tan(A) = \tan(B)$. Exercise 3 asks you to carry out the calculation, which uses the same approach as was used for the parabola. One reason to do Exercise 3 is to appreciate more fully the power of vector calculus, developed in Chapter 15, for with that tool you can establish the reflection property of either the parabola or the ellipse with a lot less work.

Why is $b^2 = a^2 - c^2$? See the right triangle with vertices $(0,b)$, $(c,0)$, and $(0,0)$.

See Exercises 10 and 11 in Section 15.S.

The reflection property of an ellipse is used in wind tunnel tests of aircraft noise. The test is run in an elliptical chamber, with the aircraft model at one focus and a microphone at the other focus.

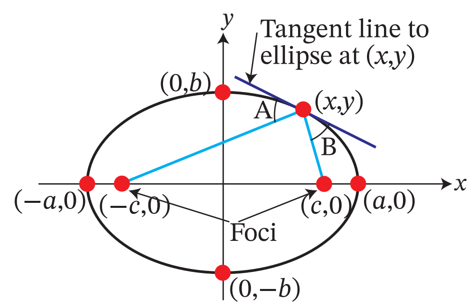


Figure C.4.6

Whispering rooms, such as National Statuary Hall in the Capitol in Washington, D.C., are based on the same principle. A person talking quietly at one focus can be heard easily at the other focus but not at points between the foci. (The whisper would be unintelligible except for the additional property that all the paths of the sound from one focus to the other have the same length.)

An ellipsoidal reflector cup is used for crushing kidney stones. An electrode is placed at one focus and an ellipsoid positioned so that the stone is at the other focus. Shock waves generated at the electrode bounce off the ellipsoid, concentrate on the other focus, and pulverize the stones without damaging other parts of the body. The patient recovers in three to four days instead of the two to three weeks required after surgery. This advance also reduced the mortality rate from kidney stones by a factor of 200, from 1 in 50 to 1 in 10,000.

The reflecting property of the ellipse also is used in the study of air pollution. One way to detect air pollution is by light scattering. A laser is aimed through one focus of a shiny ellipsoid. When a particle passes through this focus, the light is reflected to the other focus where a light detector is located. The number of particles detected is used to determine the amount of pollution in the air.

Solar Cookers

Diocles, in his book *On Burning Mirrors*, written around 190 B.C., studied spherical and parabolic reflectors, both of which had been considered by earlier writers. At that time, some had thought that a spherical reflector focuses incoming light to a single point. This is false.

Diocles showed that a spherical reflector subtending an angle of 60° reflects light that is parallel to its axis of symmetry to points on the axis that occupy about one-thirteenth of the radius. He proposed an experiment, "Perhaps you would like to make two examples of a burning-mirror, one spherical, one parabolic, so that you can measure the burning power of each."' Though the reflection property of a parabola was already known, *On Burning Mirrors* contains the first known proof.

Exercise 4 shows that a spherical oven is pretty effective, since a potato or hamburger is not a point.

EXERCISES for CIE C.4

1. Do the algebra to complete the proof that $\tan(A) = \tan(B)$ in the case of the parabola.

2. Show that if $0 < C < \frac{\pi}{2}$, then $\tan\left(\frac{\pi}{2} - C\right) = \frac{1}{\tan(C)}$.

3. This exercise establishes the reflection property of an ellipse. Refer to Figure C.4.6 for the meaning of the notation.

(a) Find the slope of the tangent line at (x, y).
(b) Find the slope of the line through $F = (c, 0)$ and (x, y).
(c) Find $\tan(B)$.
(d) Find the slope of the line through $F' = (c', 0)$ and (x, y).
(e) Find $\tan(A)$.
(f) Check that $\tan(A) = \tan(B)$.

4. Use trigonometry to show that a spherical mirror of radius r and subtending an angle of 60° causes light parallel to its axis of symmetry to reflect and meet the axis in an interval of length $\left(\frac{1}{\sqrt{3}} - \frac{1}{2}\right) r \approx \frac{r}{12.9}$. See Figure C.4.7.

NOTE: This question is also asked in Section 1.S as Exercise 40 .

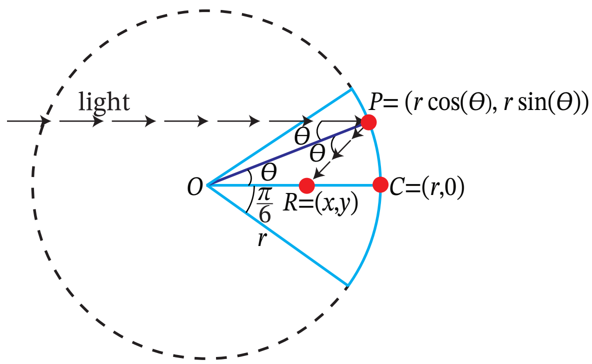


Figure C.4.7

Chapter 4

Derivatives and Curve Sketching

When you graph a function you typically plot a few points and connect them with straight line segments or a curve. Most electronic graphing devices use the same approach and obtain better results by plotting more points and using shorter segments. The more points used, the smoother the graph will appear. This chapter will show how to choose the key points.

Three properties of the derivative developed in Section 4.1 and proved in Section 4.4 will be used in Section 4.2 to help graph a function. In Section 4.3 we will see what the second derivative tells about the graph of a function.

4.1 Three Theorems about the Derivative

This section is based on plausible observations about the graphs of differentiable functions, which we restate as theorems. The ideas will be applied in Section 4.2 to sketch graphs of functions.

Theorem of the Interior Extremum

Observation 4.1.1: ***Slope of Tangent Line at an Extremum***

Suppose that $f(c)$ is the largest value of $f(x)$ over an open interval that contains c, as shown in Figure 4.1.1. The maximum occurs at a point $(c, f(c))$, which we call P. If $f(x)$ is differentiable at c, then the tangent line at P will exist. What can we say about this tangent line?

For a differentiable function, the tangent line at an extreme value is horizontal.

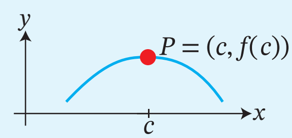


Figure 4.1.1

If the tangent at P were not horizontal (that is, not parallel to the x-axis), then it would be tilted. A small piece of the graph around P that appears to be almost straight would look as shown in Figure 4.1.2(a) or (b).

In the first case P could not be the highest point on the curve because there would be higher points to the right of P. In the second case P could not be the highest point because there would be higher points to the left of P. Therefore the tangent at P must be horizontal, as shown in Figure 4.1.2(c). That is, $f'(c) = 0$.

This suggests the following criterion for identifying local extrema.

Theorem 4.1.2: Theorem of the Interior Extremum

Let f be a function defined on the open interval (a, b). If f takes on an extreme value at c in this interval, then either (i) $f'(c) = 0$ or (ii) $f'(c)$ does not exist.

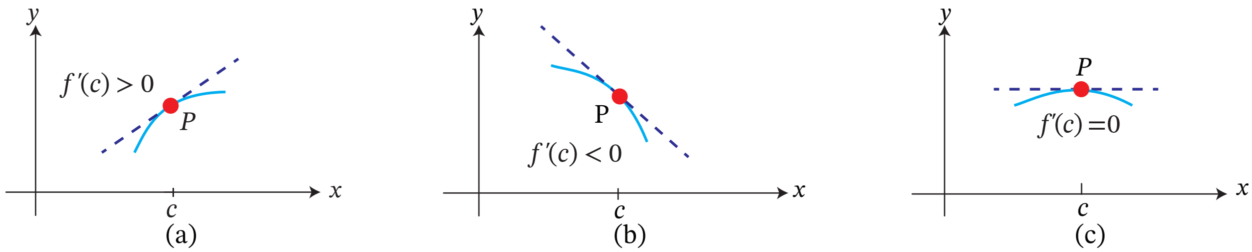


Figure 4.1.2

The basic message of Theorem 4.1.2 is this: if an extreme value occurs within an open interval and the derivative exists there, the derivative must be 0 there. While this makes intuitive sense, we have to be careful not to misapply it in either of the following ways.

1. If in Theorem 4.1.2 the open interval (a, b) is replaced by a closed interval $[a, b]$ the conclusion may not hold. A glance at Figure 4.1.3(a) shows why — the extreme value could occur at an endpoint ($x = a$ or $x = b$).
2. The converse of Theorem 4.1.2 is not true. Having the derivative equal to 0 at a point does not guarantee that there is an extremum at this point. The graph of $y = x^3$, Figure 4.1.3(b), shows why. Since $f'(x) = 3x^2$, $f'(0) = 0$. While the tangent line is indeed horizontal at $(0, 0)$, it crosses the curve at this point. The graph has neither a maximum nor a minimum at the origin.

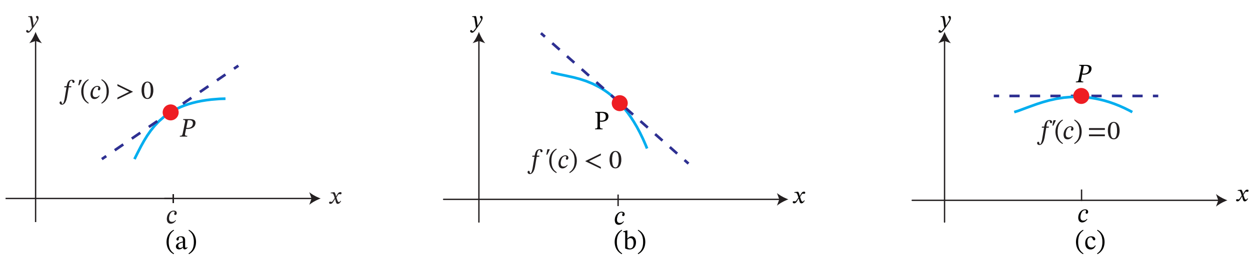


Figure 4.1.3

Rolle's Theorem

Though the next observation is phrased in terms of slopes we will see that it has implications for velocity and any changing quantity.

Observation 4.1.3: *Motivation for Rolle's Theorem*

For a smooth curve on a closed and bounded interval, there must be a point between the endpoints where the tangent line has the same slope as the chord between the endpoints.

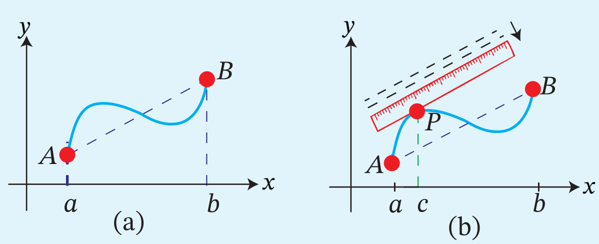


Figure 4.1.4

RECALL: A **chord** of f is a line segment that joins two points on the graph of $y = f(x)$.

Let $A = (a, f(a))$ and $B = (b, f(b))$ be two points on the graph of a differentiable function f defined on the interval $[a, b]$, as shown in Figure 4.1.4(a). Draw the chord AB joining A and B. Assume part of the graph lies above that line. Imagine holding a ruler parallel to AB and lowering it until it just touches the graph of $y = f(x)$, as in Figure 4.1.4(b).

The ruler touches the curve at a point P and lies along the tangent at P. At P the slope of AB is $f'(c)$. (In Figure 4.1.4(b) there is a second number between a and b where the tangent line to $y = f(x)$ is parallel to the chord between A and B.)

It is customary to state two separate theorems based on the observation about chords and tangent lines. The first, Rolle's theorem, is a special case of the second, the mean value theorem.

The next theorem is suggested by a special case of Observation 4.1.3. When the points A and B have the same y-coordinate, the chord AB has slope 0. (See Figure 4.1.5.) In this case, the observation tells us there must be a horizontal tangent to the graph. Expressed in terms of derivatives, this suggests Rolle's theorem.

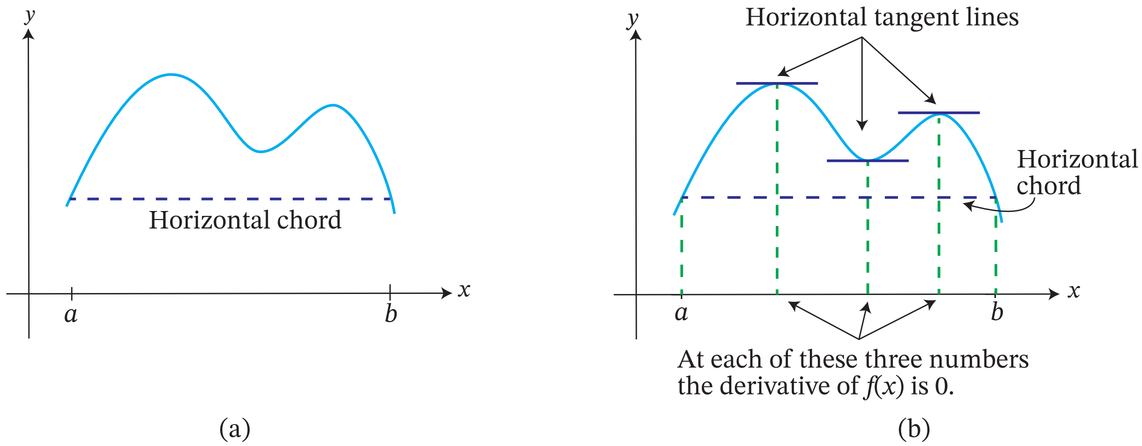


Figure 4.1.5

Theorem 4.1.4: Rolle's Theorem

Let f be a continuous function on the closed interval $[a, b]$ and have a derivative at all x in the open interval (a, b). If $f(a) = f(b)$, then there is at least one number c in (a, b) such that $f'(c) = 0$.

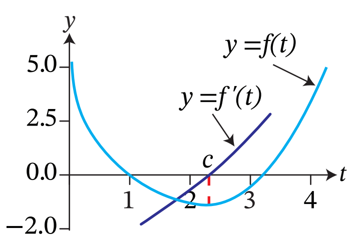


Figure 4.1.6

In the first example we will rely upon the graph of $y = f'(x)$ to conclude that there is only one c satisfying Rolle's theorem. Then, Example 2 describes the basic technique for using Rolle's theorem to show that a function has at most one x-intercept.

EXAMPLE 1. Verify Rolle's theorem for $f(t) = (t^2 - 1)\ln\left(\frac{t}{\pi}\right)$ on $[1, \pi]$.

SOLUTION The function $f(t)$ is defined and differentiable for $t > 0$, in particular, on the closed interval $[1, \pi]$. We see that $f(1) = 0$ and, because $\ln(1) = 0$, $f(\pi) = 0$. Therefore, by Rolle's theorem, there must be a value of c between 1 and π where $f'(c) = 0$.

The derivative $f'(t) = 2t\ln(t/\pi) + (t^2 - 1)/t$ is complicated. Though it is not possible to find the exact value of c with $f'(c) = 0$, Rolle's theorem guarantees that it exists. Figure 4.1.6 confirms that there is only one solution to $f'(c) = 0$ on $[1, \pi]$. ♦

Assume that $f(x)$ is a differentiable function such that $f'(x)$ is never 0 for x in an interval. Then the equation $f(x) = 0$ can have at most one solution in that interval. See Figure 4.1.6. (If it had two solutions, a and b, then $f(a) = 0$ and $f(b) = 0$, and we could apply Rolle's theorem on $[a, b]$.)

EXAMPLE 2. Use Rolle's theorem to determine the number of real roots for the equation $x^3 - 6x^2 + 15x + 3 = 0$.

SOLUTION Define $f(x) = x^3 - 6x^2 + 15x + 3$. In Section 2.5 it was observed that the intermediate value theorem guarantees that every odd-degree polynomial has at least one x-intercept. Let r be one root of $f(x)$, that is, $f(r) = 0$.

Could there be another x-intercept, s?

If so, by Rolle's theorem, there would be a number c between r and s at which $f'(c) = 0$.

The plan is compute the derivative of $f(x)$ and see if it is ever 0. We have $f'(x) = 3x^2 - 12x + 15 = 3(x^2 - 4x + 5)$. To find when $f'(x)$ is 0, we solve the equation $x^2 - 4x + 5 = 0$ by the quadratic formula, obtaining

$$x = \frac{-(-12) \pm \sqrt{(-4)^2 - 4(1)(5)}}{2}$$
$$= \frac{4 \pm \sqrt{-4}}{2} = 2 \pm \sqrt{-1} = 2 \pm i.$$

Because the two roots of the $f'(x)$ are complex-valued, there are no points where $f'(x) = 0$ and so there cannot be a second point where $f(x) = 0$.

The equation $x^3 - 6x^2 + 15x + 3 = 0$ has only one real root, which turns out to be approximately -0.186. ♦

Historical Note: Michel Rolle

Michel Rolle (1652–1719) was a French mathematician and an early critic of calculus before later changing his opinion. He is also the first person known to have placed the index in the opening of a radical to denote the n^{th} root of a number: $\sqrt[n]{x}$.

The ideas used in Example 2 can be generalized as follows:

Observation 4.1.5: *Monotonic Functions have At Most One x-Intercept*

More generally, in any interval in which the derivative $f'(x)$ exists and is never 0, the graph of $y = f(x)$ can have no more than one x-intercept. The graph of a monotonic differentiable function has at most one x-intercept.

Note that in Example 2, $f'(x) = 3x^2 - 12x + 15 = 3((x-2)^2 + 1) \geq 3$ for all x, so the graph of $y = f(x)$ is monotonic on the real line. So, by Observation 4.1.5, it has at most one x-intercept.

Mean Value Theorem

The mean value theorem is a generalization of Rolle's theorem in that it applies to any chord, not just horizontal ones.

In geometric terms, the theorem asserts that for any chord of the graph of a well-behaved function, somewhere above or below it the graph has at least one tangent line parallel to the chord. Let us translate this geometric statement into the language of functions.

Call the ends of the chord $(a, f(a))$ and $(b, f(b))$. The slope of the chord is

$$\frac{f(b) - f(a)}{b - a}.$$

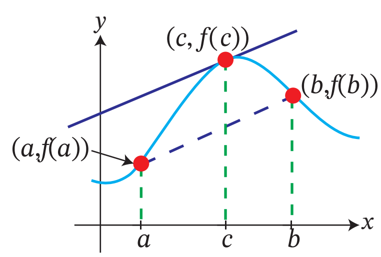


Figure 4.1.7

Since the tangent line and the chord are parallel, they have the same slopes. If the tangent line is tangent at the point $(c, f(c))$, then

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

This observation is the basis behind the next theorem.

Theorem 4.1.6: Mean Value Theorem

Let f be a continuous function on the closed interval $[a,b]$ and have a derivative at every x in the open interval (a,b). Then there is at least one number c in (a,b) such that

$$f'(c) = \frac{f(b)-f(a)}{b-a}.$$

EXAMPLE 3. Verify the mean value theorem for $f(t) = \sqrt{4-t^2}$ on the interval $[0,2]$.

SOLUTION Because $4-t^2 \geq 0$ for t between -2 and 2 f is continuous on $[0,2]$ and differentiable on $(0,2)$. The slope of the chord through $(a, f(a)) = (0,2)$ and $(b, f(b)) = (2,0)$ is

$$\frac{f(b)-f(a)}{b-a} = \frac{0-2}{2-0} = -1.$$

According to the mean value theorem, there is at least one number c between 0 and 2 where $f'(c)$ is -1.

Let us try to find c. Since $f'(t) = -t/\sqrt{4-t^2}$, we need to solve the equation $-c/\sqrt{4-c^2} = -1$. We get

$$\begin{aligned}
\frac{-c}{\sqrt{4-c^2}} &= -1 \\
-c &= -\sqrt{4-c^2} && (\text{ multiply both sides by } \sqrt{4-c^2}\ ) \\
c^2 &= 4-c^2 && (\text{ square both sides }) \\
2c^2 &= 4 && (\text{ add } c^2 \text{ to both sides}) \\
c^2 &= 2 && (\text{ divide both sides by 2 }).
\end{aligned}$$

While both $c = \sqrt{2}$ and $c = -\sqrt{2}$ satisfy $c^2 = 2$, only $\sqrt{2}$ lies in $[0,2]$. The other root, $-\sqrt{2}$, is of no interest because it is not in $[0,2]$. Thus $c = \sqrt{2}$ is the only number in $[0,2]$ whose existence is assured by the mean value theorem. ♦

The interpretation of the derivative as slope suggested the mean value theorem. What does the mean value theorem say when the function describes the position of a moving object, and the derivative, its velocity? This is answered in Example 4.

EXAMPLE 4. A car moving on the x-axis has x-coordinate $x = f(t)$ at time t. At time a its position is $f(a)$. At some later time b its position is $f(b)$. What does the mean value theorem assert for this car?

SOLUTION The quotient

$$\frac{f(b)-f(a)}{b-a} \quad \text{equals} \quad \frac{\text{Change in position}}{\text{Change in time}}.$$

The mean value theorem asserts that at some time c, $f'(c)$ is equal to the quotient $(f(b)-f(a))/(b-a)$. This says that the velocity at time c is the same as the average velocity during the time interval $[a,b]$. For example, if a car travels 210 miles in 5 hours, then at some time its speedometer must read 42 miles per hour. ♦

Consequences of the Mean Value Theorem

There are several ways of writing the mean value theorem. The equation

$$f'(c) = \frac{f(b)-f(a)}{b-a}$$

is equivalent to

$$f(b) - f(a) = (b-a)f'(c)$$

and hence to

$$f(b) = f(a) + (b-a)f'(c). \tag{4.1.1}$$

In this last form, the mean value theorem asserts that $f(b)$ is equal to $f(a)$ plus a quantity that involves the derivative f' at some number c between a and b. The following important corollaries are based on this form.

Corollary 4.1.7: Only Constant Functions Have Slope Zero on an Interval

If the derivative of a function is 0 *throughout an interval, then the function is constant on the interval.*

Corollary 4.1.7 is seen to be plausible when interpreted in terms of motion. It asserts that if an object has zero velocity for a period of time, then it does not move during that time. The following proof makes use of the mean value theorem.

Proof of Corollary 4.1.7

Let a and b be two numbers in the interval I and denote the function by f. To prove the corollary, it suffices to prove that $f(a) = f(b)$, for that is the defining property of a constant function.

By the mean value theorem in the form (4.1.1), there is a number c between a and b such that

$$f(b) = f(a) + (b-a)f'(c).$$

But $f'(c) = 0$, since $f'(x) = 0$ for all x in I. Hence

$$f(b) = f(a) + (b-a)(0)$$

which proves that $f(b) = f(a)$ for all values of a and b. That is, f is a constant function. ●

EXAMPLE 5. Use calculus to show that $f(x) = (e^x + e^{-x})^2 - e^{2x} - e^{-2x}$ is a constant. Find the constant.

SOLUTION The function f is differentiable for all numbers x. Its derivative is

$$f'(x) = 2(e^x + e^{-x})(e^x - e^{-x}) - 2e^{2x} + 2e^{-2x} = 2(e^{2x} - e^{-2x}) - 2e^{2x} + 2e^{-2x} = 0.$$

Because $f'(x)$ is always zero, f must be a constant.

To find the constant, evaluate $f(x)$ for any convenient value of x. If we choose $x = 0$ we see $f(0) = (e^0 + e^0)^2 - e^0 - e^0 = 2^2 - 2 = 2$. Thus, $(e^x + e^{-x})^2 - e^{2x} - e^{-2x} = 2$ for all numbers x. ♦

NOTE: The identity found in Example 5 can be checked by squaring $e^x + e^{-x}$ and simplifying.

We now turn our attention to understanding the relationship between two functions with the same derivative.

Corollary 4.1.8: Two Functions with the Same Derivative Differ by a Constant

If two functions have the same derivative throughout an interval, then they differ by a constant. That is, if $F'(x) = G'(x)$ for all x in an interval, then there is a constant C such that $F(x) = G(x) + C$.

Proof of Corollary 4.1.8

Define h as $h(x) = F(x) - G(x)$. Then, because $F'(x) = G'(x)$,

$$h'(x) = F'(x) - G'(x) = 0.$$

Since the derivative of h is 0, Corollary 4.1.7 implies that h is constant, that is, $h(x) = C$ for some fixed number C. Thus

$$F(x) - G(x) = C \quad \text{or} \quad F(x) = G(x) + C,$$

and Corollary 4.1.8 is proved. ●

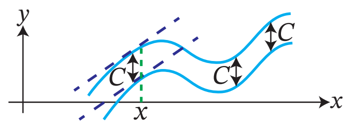


Figure 4.1.8

Is Corollary 4.1.8 plausible when the derivative is interpreted as slope? In this case, it asserts that if the graphs of two functions have parallel tangent lines for every x, then one graph can be obtained from the other by raising (or lowering) it by a constant amount C. If you sketch two such graphs (as in Figure 4.1.8), you will see that the corollary is reasonable.

EXAMPLE 6. What functions have a derivative equal to $2x$ everywhere?

SOLUTION One such solution is x^2; another is x^2+25. For a constant C, $D(x^2+C)=2x$. Are there any other possibilities? Corollary 4.1.8 tells us there are not, for if F is a function such that $F'(x)=2x$, then $F'(x)=(x^2)'$ for all x. Thus the functions F and x^2 differ by a constant, say C,

$$F(x)=x^2+C.$$

The only antiderivatives of $2x$ are of the form x^2+C. ♦

Corollary 4.1.7 asserts that if $f'(x)=0$ for all x, then f is a constant. What can be said about f if $f'(x)$ is positive for all x in an interval? In terms of the graph of f this implies that all tangent lines slope upward. It is reasonable to expect that as we move from left to right on the graph in Figure 4.1.9, the y-coordinate increases, that is, the function is increasing. (See Section 1.1.)

Corollary 4.1.9: Using the Derivative to Characterize Increasing and Decreasing Functions

(a) *If f is continuous on the closed interval $[a,b]$ and has a positive derivative on the open interval (a,b), then f is increasing on $[a,b]$.*
(b) *If f is continuous on the closed interval $[a,b]$ and has a negative derivative on the open interval (a,b), then f is decreasing on $[a,b]$.*

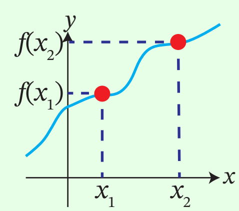


Figure 4.1.9

Proof of Corollary 4.1.9
We prove the increasing case; the other case is handled in Exercise 43. Take two numbers x_1 and x_2 such that

$$a\le x_1<x_2\le b.$$

The goal is to show that $f(x_2)>f(x_1)$.
By the mean value theorem, there is a number c between x_1 and x_2 such that

$$f(x_2)=f(x_1)+(x_2-x_1)f'(c).$$

Because $x_2>x_1$ we know x_2-x_1 is positive. Since $f'(c)$ is assumed to be positive, and the product of two positive numbers is positive, it follows that

$$(x_2-x_1)f'(c)>0.$$

Thus, $f(x_2)>f(x_1)$, which implies that $f(x)$ is an increasing function. ●

EXAMPLE 7. Determine whether $2x+\sin(x)$ is an increasing function, a decreasing function, or neither.

SOLUTION The function $2x+\sin(x)$ is the sum of two simpler functions: $2x$ and $\sin(x)$. The first part is an increasing function. The second increases for x between 0 and $\pi/2$ and decreases for x between $\pi/2$ and π. It is not clear what type of function you will get when you add $2x$ and $\sin(x)$. Let us see what Corollary 4.1.9 tells us.

The derivative of $2x+\sin(x)$ is $2+\cos(x)$. Since $\cos(x)\ge -1$ for all x, $(2x+\sin(x))'=2+\cos(x)\ge 2+(-1)=1$. Because $(2x+\sin(x))'$ is positive for all numbers x, $2x+\sin(x)$ is an increasing function. Figure 4.1.10 shows its graph together with the graphs of $2x$ and $\sin(x)$. ♦

While Corollary 4.1.9 and the definitions of increasing and decreasing are stated in terms of intervals, there are times when we will talk about a function being increasing or decreasing at a point. When we say a function is **increasing at** c we mean the function is "increasing in an interval that contains c."

In particular, if f' is continuous and $f'(c) > 0$, the permanence property in Section 2.5 tells us there is an interval (a,b) containing c where $f'(x)$ remains positive for all numbers x in (a,b). Thus, f is increasing on (a,b), and hence increasing at c. The meaning of **decreasing at** c is analogous.

More generally, if $f'(x)$ is never negative, that is $f'(x) \geq 0$ for all x, then f is nondecreasing. In the same manner, if $f'(x) \leq 0$ for all x, then f is nonincreasing.

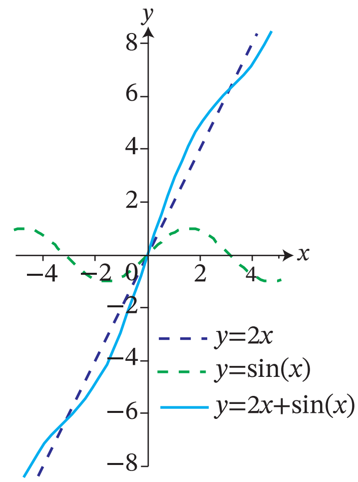


Figure 4.1.10

Summary

This section focused on three theorems, which we state informally.

The theorem of the interior extremum says that at a local extreme the derivative must be zero. (The converse is not true.)

Rolle's theorem asserts that if a function has equal values at two inputs, its derivative must equal zero at least at one number between these inputs. The mean value theorem, a generalization of Rolle's theorem, asserts that for any chord on the graph of a function there is a tangent line parallel to it. This means that for $a < b$ there is c in (a,b) such that $f'(c) = (f(b) - f(a))/(b-a)$, or in a more useful form, $f(b) = f(a) + f'(c)(b-a)$.

From the mean value theorem it follows that where a derivative is positive, a function is increasing, where it is negative it is decreasing, and where it stays at the value zero it is constant. The last assertion implies that two antiderivatives of the same function differ by a constant (which may be zero).

EXERCISES for Section 4.1

1. State Rolle's theorem in words, using as few mathematical symbols as you can.

2. Draw a graph illustrating Rolle's theorem. Identify its critical features.

3. Draw a graph illustrating the mean value theorem. Identify its critical features.

4. Express the mean value theorem in words, using no symbols to denote the function or the interval.

5. Express the mean value theorem in symbols, where the function is denoted g and the interval is $[e, f]$.

6. Which of the corollaries to the mean value theorem implies that
  (a) if two cars on a straight road have the same velocity at every instant, they remain a fixed distance apart?
  (b) if all tangents to a curve are horizontal, the curve is a horizontal line?
Explain each answer in terms of theorems in this section.

Exercises 7 to 12 concern the theorem of the interior extremum.

7. Let the function $f(x) = x^2$ be defined for x in $[-1,2]$.
  (a) Graph $f(x)$ for x in $[-1,2]$.
  (b) What is the maximum value of $f(x)$ for x in $[-1,2]$?
  (c) Does $f'(x)$ exist at the maximum?
  (d) Does $f'(x)$ equal zero at the maximum?
  (e) What is the minimum value of $f(x)$ for x in $[-1,2]$?
  (f) Does $f'(x)$ equal zero at the minimum?

8. Let the function $f(x) = \sin(x)$ be defined for x in $[0,\pi]$.
  (a) Graph $f(x)$ for x in $[0,\pi]$.

(b) What is the maximum value of $f(x)$ for x in $[0,\pi]$?
(c) Does $f'(x)$ exist at the maximum?
(d) Does $f'(x)$ equal zero at the maximum?
(e) What is the minimum value of $f(x)$ for x in $[0,\pi]$?
(f) Does $f'(x)$ equal zero at the minimum?

9. Repeat Exercise 7 on the intervals: (a) $[1,2]$, (b) $(-1,2)$, (c) $(1,2)$, (d) $[0,2\pi]$, (e) $(0,\pi)$, and (f) $(0,2\pi)$.

10. (a) Graph $y=-x^2+3x+2$ for x in $[0,2]$.
(b) Looking at the graph, estimate the x-coordinate where the maximum value of y occurs for x in $[0,2]$.
(c) Find where $\frac{dy}{dx}=0$.
(d) Using (c), determine exactly where the maximum occurs.

11. (a) Graph $y=2x^2-3x+1$ for x in $[0,1]$.
(b) Looking at the graph, estimate the x-coordinate where the maximum value of y occurs for x in $[0,1]$. At which value of x does it occur?
(c) Looking at the graph, estimate the x-coordinate where the minimum value of y occurs for x in $[0,1]$.
(d) Find where $\frac{dy}{dx}=0$.
(e) Using (d), determine exactly where the minimum occurs.

---

12. For the following functions,
(i) show that the derivative of the function is 0 when $x=0$ and
(ii) use a graph of $y=f(x)$ to decide whether the function has an extremum at $x=0$.
(a) $f(x)=1-\cos(x)$, (b) $f(x)=x-e^x$, (c) $f(x)=x^2\sin(x)$, (d) $f(x)=x^2-x^3$, and (e) $f(x)=x^3-x^4$.

In Exercises 13 to 18, verify that the function satisfies the hypotheses of Rolle's theorem for the given interval. Find all numbers c that satisfy the conclusion of the theorem.

13. $f(x)=x^2-2x-3$, $[0,2]$
14. $f(x)=x^3-x$, $[-1,1]$
15. $f(x)=x^4-2x^2+1$, $[-2,2]$
16. $f(x)=\sin(x)+\cos(x)$, $[0,4\pi]$
17. $f(x)=e^x+e^{-x}$, $[-2,2]$
18. $f(x)=x^2e^{-x^2}$, $[-2,2]$

Exercises 19 to 21 also concern Rolle's theorem.

19. (a) Graph $f(x)=x^{2/3}$ for x in $[-1,1]$.
(b) Show that $f(-1)=f(1)$.
(c) Is there a number c in $(-1,1)$ such that $f'(c)=0$?
(d) Why does this not contradict Rolle's theorem?

20. (a) Graph $f(x)=1/x^2$ for x in $[-1,1]$.
(b) Show that $f(-1)=f(1)$.
(c) Is there a number c in $(-1,1)$ such that $f'(c)=0$?
(d) Why does this not contradict Rolle's theorem?

21. Let $f(x)=\ln(x^2)$. Note that $f(-1)=f(1)$. Is there a number c in $(-1,1)$ such that $f'(c)=0$? If so, find at least one such number. If not, does this contradict Rolle's theorem?

Exercises 22 to 27 concern the mean value theorem. In Exercises 22 to 25, find all values of c that satisfy the mean value theorem for the given functions and intervals.

22. $f(x)=x^2-3x$, $[1,4]$
23. $f(x)=2x^2+x+1$, $[-2,3]$
24. $f(x)=3x+5$, $[1,3]$
25. $f(x)=5x-7$, $[0,4]$

26. (a) Graph $y = \sin(x)$ for x in $\left[\frac{\pi}{2}, \frac{7\pi}{2}\right]$.

(b) Draw the chord joining $\left(\frac{\pi}{2}, f\left(\frac{\pi}{2}\right)\right)$ and $\left(\frac{7\pi}{2}, f\left(\frac{7\pi}{2}\right)\right)$.

(c) Draw all tangents to the graph parallel to the chord drawn in (b).

(d) Using (c), determine how many numbers c there are in $\left(\frac{\pi}{2}, \frac{7\pi}{2}\right)$ such that $f'(c) = \dfrac{f\left(\frac{7\pi}{2}\right) - f\left(\frac{\pi}{2}\right)}{\frac{7\pi}{2} - \frac{\pi}{2}}$.

(e) Use the graph to estimate the values of the c's.

27. (a) Graph $y = \cos(x)$ for x in $\left[\frac{\pi}{2}, \frac{9\pi}{2}\right]$.

(b) Draw the chord joining $(0, f(0))$ and $\left(\frac{9\pi}{2}, f\left(\frac{9\pi}{2}\right)\right)$.

(c) Draw all tangents to the graph parallel to the chord drawn in (b).

(d) Using (c), determine how many numbers c there are in $\left(0, \frac{9\pi}{2}\right)$ such that $f'(c) = \dfrac{f\left(\frac{9\pi}{2}\right) - f(0)}{\frac{7\pi}{2} - 0}$.

(e) Use the graph to estimate the values of the c's.

---

28. At time t seconds a ball thrown upwards is at a height of $f(t) = -16t^2 + 32t + 40$ feet.

(a) What is the initial height? That is, the height when t is zero.

(b) Show that after 2 seconds it returns to its initial height.

(c) What does Rolle's theorem imply about the velocity of the ball?

(d) Verify Rolle's theorem in this case by computing the numbers c that it asserts exist.

29. Find all points where $f(x) = 2x^3(x-1)$ has an extreme value on the following intervals (a) $\left(\frac{-1}{2}, 1\right)$, (b) $\left[\frac{-1}{2}, 1\right]$, (c) $\left[\frac{-1}{2}, \frac{1}{2}\right]$, and (d) $\left(\frac{-1}{2}, \frac{1}{2}\right)$.

30. Let $f(x) = |2x - 1|$.

(a) Explain why $f'\left(\frac{1}{2}\right)$ does not exist.

(b) Find $f'(x)$. (Treat the cases $x < \frac{1}{2}$ and $x > \frac{1}{2}$ separately.)

(c) Does the mean value theorem apply for this function and the interval $[-1, 2]$?

31. The year is 2030. Because a gallon of gas costs ten dollars and Highway 80 is full of tire-wrecking potholes, the California Highway Patrol no longer patrols the 77 miles between Sacramento and Berkeley. Instead it uses two cameras. One, in Sacramento, records the license number and time of a car on the freeway, and another does the same in Berkeley. A computer processes the data instantly. Assume that the two cameras show that a car that was in Sacramento at 10:45 reached Berkeley at 11:40. Show that the mean value theorem justifies giving the driver a ticket for exceeding the 70 mile-per-hour speed limit.

Historical Note: Legality of Unmanned Speed Traps (in California)

While it makes a nice story, reality is that the California Vehicle Code forbids this way to catch speeders. It reads, "No speed trap shall be used in securing evidence as to the speed of any vehicle. A 'speed trap' is a particular section of highway measured as to distance in order that the speed of a vehicle may be calculated by securing the time it takes the vehicle to travel the known distance."

32. Verify the mean value theorem for $f(x) = x^2 e^{-x/3}$ on $[1, 10]$. See Example 1.

33. Find all antiderivatives of the following functions. Check your answer by differentiation.
(a) $f(x) = 3x^2$, (b) $f(x) = \sin(x)$, (c) $f(x) = \frac{1}{1+x^2}$, and (d) $f(x) = e^x$.

34. Find all antiderivatives of each of the following functions. Check your answer by differentiation.
(a) $f(x) = \cos(x)$, (b) $f(x) = \sec(x)\tan(x)$, (c) $f(x) = \frac{1}{x}$ $(x > 0)$, and (d) $f(x) = \sqrt{x}$ $(x > 0)$.

35. (a) Differentiate $\sec^2(x)$ and $\tan^2(x)$.
(b) The derivatives in (a) are equal. Corollary 4.1.8 then asserts that there exists a constant C such that $\sec^2(x) = \tan^2(x) + C$. Find the constant.

36. Show by differentiation that $f(x) = \ln\left(\frac{x}{5}\right) - \ln(5x)$ is constant for all positive x. Find the constant.

37. Find all functions whose second derivatives are 0 for all x in $(-\infty, \infty)$.

38. Use Rolle's theorem to determine how many real roots there are for the equation $x^3 - 6x^2 + 15x + 3 = 0$.

39. Use Rolle's theorem to determine how many real roots there are for the equation $3x^4 + 4x^3 - 12x^2 + 4 = 0$. For each root give an interval that contains that root and no other root.

40. Use Rolle's theorem to determine how many real roots there are for the polynomial $f(x) = 3x^4 + 4x^3 - 12x^2 + A$. The number may depend on A. For which A is there exactly one root? Are there any values of A for which there is an odd number of real roots? Exercise 39 uses this equation with $A = 4$.

41. The number of real roots to $x^3 - ax^2 + 15x + 3 = 0$ depends on the value of a.
(a) Find all values of a for which the equation has 3 real roots.
(b) Find all values of a for which the equation has 1 real root.
(c) Are there any values of a for which the equation has exactly two real roots?

Exercise 38 uses this equation with $a = 6$.

42. If f is differentiable for all real numbers and $f'(x) = 0$ has three solutions, what can be said about the number of solutions of $f(x) = 0$? of $f(x) = 5$?

43. Prove the decreasing case of Corollary 4.1.9.

44. For which values of the constant k is the function $7x + k\sin(2x)$ always increasing?

45. If two functions have the same second derivative for all x in $(-\infty, \infty)$, what can be said about their difference?

46. If a function f is differentiable for all x and c is a number, is there necessarily a chord of the graph of f that is parallel to the tangent line at $(c, f(c))$? Explain.

47. Sketch a graph of a differentiable function $f(x)$ such that $f'(1)$ is 2, yet there is no open interval around 1 on which f is increasing.

48. Establish that for x in $\left[0, \frac{\pi}{2}\right)$, $\tan(x)$ is greater than x by first showing by
(a) showing that $f(x) = \tan(x) - x$ is increasing, and (b) showing that $f(x) = \frac{\tan(x)}{x}$ is increasing.

Exercises 49 and 50 are related.

The authors thought twice about including Exercise 49 because this is a calculus text, not a trigonometry text. It is good preparation for Exercise 50 (which does use calculus ideas), so it made the final cut.

49. Using trigonometric identities but no calculus show that
(a) for θ in $\left(0, \frac{\pi}{4}\right)$, $\frac{\sin(2\theta)}{\sin(\theta)}$ is decreasing, and (b) for θ in $(0, \pi/6)$, $\frac{\sin(3\theta)}{\sin(\theta)}$ is decreasing.

Observation 4.1.10:

Trigonometric identities can be used to show that for each positive integer n, $\sin(n\theta)/\sin(\theta)$ is decreasing for θ in $\left(0, \frac{\pi}{2n}\right)$.

50. Using calculus, show that for any positive number $n > 1$, $\frac{\sin(n\theta)}{\sin(\theta)}$ is decreasing for θ in $\left(0, \frac{\pi}{n-1}\right)$.

NOTE: This result is true on a larger interval, but the exact interval is much harder to determine.

4.2 The First Derivative and Graphing

Section 4.1 showed the connection between extrema and the places where the derivative is zero. In this section we use this connection to find high and low points on a graph.

The graph of a differentiable function f is shown in Figure 4.2.1. The points P, Q, R, and S are of special interest. S is the highest point on the graph for all x in the domain. We call it a **global maximum** or **absolute maximum**. The point P is higher than all points near it on the graph; it is called a **local maximum** or **relative maximum**. Similarly, Q is called a **local minimum** or **relative minimum**. The point R is neither a relative maximum nor a relative minimum.

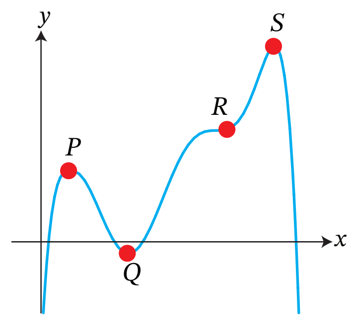


Figure 4.2.1

A point that is either a maximum or minimum is called an **extremum**.

If you were to walk left to right along the graph in Figure 4.2.1, you would call P the top of a hill, Q the bottom of a valley, and S the highest point on your walk (it is also a top of a hill). You might notice R, for you get a momentary break from climbing from Q to S. For just this one instant it would be like walking along a horizontal path.

These aspects of a function and its graph are made precise in definitions phrased in terms of a general domain. In most cases the domain of the function will be an interval — open, closed, or half-open.

Definition: ***Relative Maximum and Minimum (Local Maximum and Minimum)***

(i) The function f has a **relative maximum** (or **local maximum**) at c if there is an open interval around c such that $f(c) \geq f(x)$ for all x in that interval that lie in the domain of f.

(ii) The function f has a **relative minimum** (or **local minimum**) at c if there is an open interval around c such that $f(c) \leq f(x)$ for all x in that interval that lie in the domain of f.

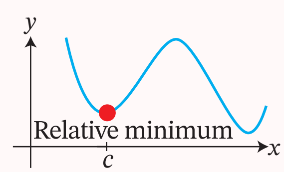


Figure 4.2.2

Definition: ***Absolute Maximum and Minimum (Global Maximum and Minimum)***

(i) The function f has an **absolute maximum** (or **global maximum**) at c if $f(c) \geq f(x)$ for all x in the domain of f.

(ii) The function f has an **absolute minimum** (or **global minimum**) at c if $f(c) \leq f(x)$ for all x in the domain of f.

A local minimum is like the summit of a single mountain or the lowest point in a valley. (See Figure 4.2.2.) A global maximum corresponds to Mt. Everest at more than 29,000 feet above sea level and a global minimum corresponds to the Mariana Trench in the Pacific Ocean 36,000 feet below sea level, the lowest point on Earth's crust. REMINDER: Every global extremum is also a local extremum.

In this section it is assumed that functions are differentiable. If a function is not differentiable at an isolated point, the point will need to be considered separately.

Definition: *Critical Number and Critical Point.*

A number c at which $f'(c) = 0$ is called a **critical number** for the function f. The corresponding point $(c, f(c))$ on the graph of f is a **critical point**.

The theorem of the interior extremum in Section 4.1 says that every local maximum and minimum of a function f occurs where the tangent line to the curve either is horizontal or does not exist.

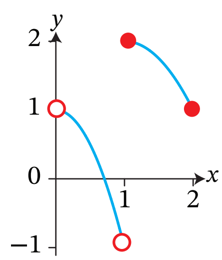
Some functions have extreme values and others do not. The next theorem gives conditions under which both a global maximum and a global minimum are guaranteed to exist. To convince yourself that this is plausible, imagine drawing the graph of the function. Somewhere your pencil reaches a highest point and elsewhere a lowest point.

Theorem 4.2.1: Extreme Value Theorem

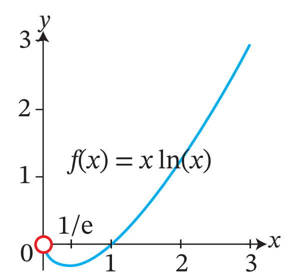
Let f be a continuous function on a closed interval $[a, b]$. Then f attains an absolute maximum value $M = f(c)$ and an absolute minimum value $m = f(d)$ at some c and d in $[a, b]$.

EXAMPLE 1. Find the absolute extrema on the interval $[0, 2]$ of the function whose graph is shown in Figure 4.2.3(a).

SOLUTION The function has an absolute maximum value of 2 but no absolute minimum value. The range is $(-1, 2]$. This function takes on values that are arbitrarily close to -1, but -1 is not in the range of this function. This can occur because the function is not continuous at $x = 1$. ◆



(a)



(b)

Figure 4.2.3

Corollary 4.1.9 provides a convenient test to determine if a function is increasing or decreasing at a point: if $f'(c) > 0$ then f is increasing at $x = c$ and if $f'(c) < 0$ then f is decreasing at $x = c$.

EXAMPLE 2. Let $f(x) = x\ln(x)$ for all $x > 0$. Find the intervals on which f is increasing, decreasing, or neither.

SOLUTION The function is increasing at numbers x where $f'(x) > 0$ and decreasing where $f'(x) < 0$. More effort is needed to determine the behavior at points where $f'(x) = 0$ (or does not exist). (The domain of f is $(0, \infty)$.) By the Product Rule,

$$f'(x) = \ln(x) + x\left(\frac{1}{x}\right) = \ln(x) + 1.$$

To find where $f'(x)$ is positive or is negative we first find where it is zero. At such numbers the derivative may switch sign, and the function may switch between increasing and decreasing. So we solve the equation $f'(x) = 0$. From $\ln(x) + 1 = 0$ we find $\ln(x) = -1$, so $e^{\ln(x)} = e^{-}1$ or $x = e^{-1}$.

When x is larger than $e^{-1} \approx 0.3679$, $\ln(x)$ is larger than -1 so that $f'(x) = \ln(x) + 1$ is positive and f is increasing. Finally, f is decreasing when x is between 0 and e^{-1} because $\ln(x) < -1$, which makes $f'(x) = \ln(x) + 1$ negative. The graph of $y = x\ln(x)$ in Figure 4.2.3(b) confirms these findings.

In addition, observe that $x = e^{-1}$ provides a minimum value $f(e^{-1}) = e^{-1}\ln\left(e^{-1}\right) = -1/e$. ♦

Using Critical Numbers to Identify Local Extrema

Previous examples suggest there is a connection between critical points and local extrema. Generally, just to the left of a local maximum the function is increasing, while just to the right it is decreasing. The opposite holds for a local minimum. The first-derivative test for a local extreme value at $x = c$ is a precise statement of this observation.

Theorem 4.2.2: First Derivative Test for a Local Extreme Value

Let f be a function and let c be a number in its domain. Suppose f is continuous on an open interval that contains c and is differentiable there, except possibly at c. Then

(i) If f' changes from positive to negative as x moves from left to right through the value c, then f has a local maximum at c.

(ii) If f' changes from negative to positive as x moves from left to right through the value c, then f has a local minimum at c.

(iii) If f' does not change sign at c, then f does not have a local extremum at c.

EXAMPLE 3. Classify each critical number of $f(x) = 3x^5 - 20x^3 + 10$ as a local maximum, as a local minimum, or as neither. Use this information to sketch the graph of $y = f(x)$.

SOLUTION To identify the critical numbers of f we find and factor the derivative:

$$f'(x) = 15x^4 - 60x^2 = 15x^2(x^2 - 4) = 15x^2(x-2)(x+2).$$

The critical numbers of f are 0, 2, and -2. To determine if they provide local extrema it is necessary to know where f is increasing and where it is decreasing.

Because f' is continuous the critical numbers are the only places the sign of f' can possibly change. As a result, on each of the intervals $(-\infty, -2)$, $(-2, 0)$, $(0, 2)$, and $(2, \infty)$, f is either increasing or decreasing; all that remains is to determine which. This is easily determined from the table of function values shown in the first two rows of Table 4.2.1.

x	$\to -\infty$	-2	0	2	$\to \infty$
$f(x)$	$\to -\infty$	74	10	-54	$\to \infty$
$f'(x)$		0	0	0	

Table 4.2.1

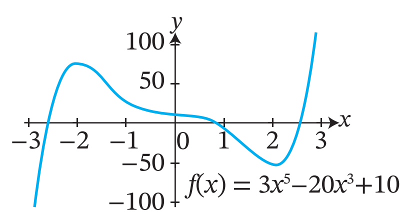


Figure 4.2.4

From $f(-2) = 74 > 10 = f(0)$ we conclude f is decreasing on $(-2, 0)$. Likewise, f must be decreasing on $(0, 2)$ because $f(0) = 10 > -54 = f(2)$. For the two unbounded intervals, limits at $\pm\infty$ must be used but the idea is the same. Since $\lim_{x\to-\infty} f(x) = -\infty$, the function must be increasing on $(-\infty, -2)$. Likewise, in order to have $\lim_{x\to\infty} f(x) = +\infty$, f must be increasing on $(2, \infty)$. (See Figure 4.2.4.)

Because the graph of f changes from increasing to decreasing at $x = -2$, there is a local maximum at $(-2, 74)$. At $x = 2$ the graph changes from decreasing to increasing, so a local minimum occurs at $(2, -54)$. Because the derivative does not change sign at $x = 0$, this critical number does not provide a local extremum. ♦

EXAMPLE 4. Find all local extrema of $f(x) = (x+1)^{2/7}e^{-x}$.

SOLUTION (The domain of f is $(-\infty, \infty)$.) The product and chain rules for derivatives can be used to obtain

$$\begin{aligned} f'(x) &= \frac{2}{7}(x+1)^{-5/7}e^{-x} + (x+1)^{2/7}e^{-x}(-1) && = \frac{2}{7}(x+1)^{-5/7}e^{-x} - (x+1)^{2/7}e^{-x} \\ &= (x+1)^{-5/7}e^{-x}\left(\frac{2}{7} - (x+1)\right) && = (x+1)^{-5/7}e^{-x}\left(-x - \frac{5}{7}\right) \\ &= \frac{-x - \frac{5}{7}}{(x+1)^{5/7}e^{x}}. \end{aligned}$$

The only solution to $f'(x) = 0$ is $x = -5/7$; so $c = -5/7$ is the only critical number. In addition, because the denominator of $f'(x)$ is zero when $x = -1$, f is not differentiable for $x = -1$.

x	$\to -\infty$	-1	$-5/7$	$\to \infty$
$f(x)$	$\to \infty$	0	$(2/7)^{2/7}e^{5/7} \approx 1.43$	$\to 0$
$f'(x)$		dne	0	

Table 4.2.2

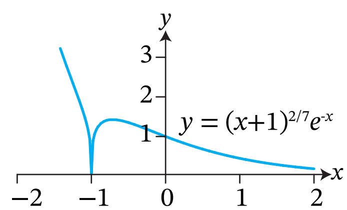


Figure 4.2.5

NOTE: In Table 4.2.2, "dne" means the limit does not exist.

Using the information in Table 4.2.2, we conclude f is decreasing on $(-\infty, -1)$, increasing on $(-1, -5/7)$, and decreasing on $(-5/7, \infty)$.

In addition, by the first-derivative test, f has a local maximum at $(-5/7, (2/7)^{2/7}e^{5/7}) \approx (-0.71, 1.43)$ and a local minimum at $(-1, 0)$. The first-derivative test applies at $x = -1$ even though f is not differentiable at -1. A graph is shown in Figure 4.2.5. ♦

Extreme Values on a Closed Interval

Many applied problems involve a continuous function whose domain is a closed interval $[a, b]$.

Observation 4.2.3: *Extreme Values on a Closed Interval*

The extreme value theorem guarantees the function attains both a maximum and a minimum at some numbers in the interval. The extreme values occur either at

(i) an endpoint ($x = a$ or $x = b$),
(ii) a critical number ($x = c$ where $f'(c) = 0$), or
(iii) where f is not differentiable ($x = c$ where $f'(c)$ is not defined).

EXAMPLE 5. Find the absolute maximum and minimum values of $f(x) = x^4 - 8x^2 + 1$ on the interval $[-1, 3]$.

SOLUTION The function is continuous on a closed and bounded interval. The absolute maximum and minimum values occur either at a critical number or at an end of the interval. The ends are $x = -1$ and $x = 3$. To find the critical numbers we solve $f'(x) = 0$:

$$f'(x) = 4x^3 - 16x = 4x(x^2 - 4) = 4x(x-2)(x+2) = 0.$$

There are three critical numbers, $x = 0$, 2, and -2, but only $x = 0$ and $x = 2$ are in the interval.

x	-1	0	2	3
$f(x)$	-6	1	-15	10
$f'(x)$	12	0	0	0

Table 4.2.3

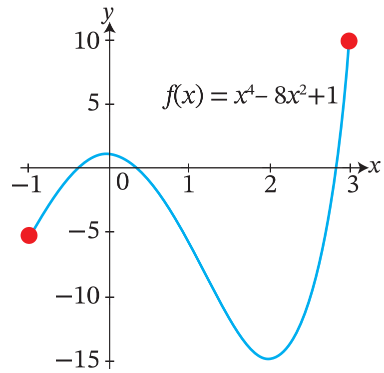


Figure 4.2.6

We simply scan the list of four function values at the ends and at the critical numbers – row 2 of Table 4.2.3 – for the largest and smallest values of $f(x)$. The largest value is 10: the global maximum occurs at $x = 3$. The smallest value is -15: the global minimum occurs at $x = 2$. See Figure 4.2.6. ♦

Summary

This section showed how to use the first derivative to find extreme values of a function. Namely, identify when the derivative is zero, positive, and negative, and where it changes sign.

A continuous function on a closed interval $[a, b]$ always has a maximum and a minimum. All extrema occur either where $f'(c) = 0$, at a or b, or where f is not differentiable.

EXERCISES for Section 4.2

In Exercises 1 to 28, sketch the graph of the function. Find all intercepts and critical points, determine the intervals where the function is increasing and where it is decreasing, and identify all local extreme values.

1. $f(x) = x^5$
2. $f(x) = (x-1)^4$
3. $f(x) = 3x^4 + x^3$
4. $f(x) = 2x^3 + 3x^2$
5. $f(x) = x^4 - 8x^2 + 1$
6. $f(x) = x^3 - 3x^2 + 3x$
7. $f(x) = x^4 - 4x + 3$
8. $f(x) = 2x^2 + 3x + 5$
9. $f(x) = x^4 + 2x^3 - 3x^2$
10. $f(x) = 2x^3 + 3x^2 - 6x$
11. $f(x) = xe^{-x/2}$
12. $f(x) = xe^{x/3}$
13. $f(x) = e^{-x^2}$
14. $f(x) = xe^{-x^2/2}$
15. $f(x) = x\sin(x) + \cos(x)$
16. $f(x) = x\cos(x) - \sin(x)$
17. $f(x) = \dfrac{\cos(x) - 1}{x^2}$
18. $f(x) = x\ln(x)$
19. $f(x) = \dfrac{\ln(x)}{x}$
20. $f(x) = \dfrac{e^x - 1}{x}$
21. $f(x) = \dfrac{e^{-x}}{x}$
22. $f(x) = \dfrac{x - \arctan(x)}{x^3}$
23. $f(x) = \dfrac{3x+1}{3x-1}$
24. $f(x) = \dfrac{x}{x^2+1}$
25. $f(x) = \dfrac{x}{x^2-1}$
26. $f(x) = \dfrac{1}{2x^2 - x}$
27. $f(x) = \dfrac{1}{x^2 - 3x + 2}$
28. $f(x) = \dfrac{\sqrt{x^2+1}}{x}$

In Exercises 29 to 36 sketch the graph, using the given information. Assume the function and its derivative are defined for all x and are continuous. Explain your reasoning.

29. Critical point $(1,2)$, $f'(x) < 0$ for $x < 1$, and $f'(x) > 0$ for $x > 1$.

30. Critical point $(1,2)$, and $f'(x) < 0$ for all x except $x = 1$.

31. x intercept -1, critical points $(1,3)$ and $(2,1)$, $\lim_{x\to\infty} f(x) = 4$, and $\lim_{x\to-\infty} f(x) = -1$.

32. y intercept 3, critical point $(1,2)$, $\lim_{x\to\infty} f(x) = \infty$, and $\lim_{x\to-\infty} f(x) = 4$.

33. x intercept -1, critical points $(1,5)$ and $(2,4)$, $\lim_{x\to\infty} f(x) = 5$, and $\lim_{x\to-\infty} f(x) = -\infty$.

34. x intercept 1, y intercept 2, critical points $(1,0)$ and $(4,4)$, $\lim_{x\to\infty} f(x) = 3$, and $\lim_{x\to-\infty} f(x) = \infty$.

35. x intercepts 2 and 4, y intercept 2, critical points $(1,3)$ and $(3,-1)$, $\lim_{x\to\infty} f(x) = \infty$, and $\lim_{x\to-\infty} f(x) = 1$.

36. No x intercepts, y intercept 1, no critical points, $\lim_{x\to\infty} f(x) = 2$, and $\lim_{x\to-\infty} f(x) = 0$.

Exercises 37 to 52 concern functions whose domains are restricted to closed intervals. Find the maximum and minimum values for the given function on the interval.

37. $f(x) = x^2 - x^4$ on $[0,1]$

38. $f(x) = 4x - x^2$ on $[0,5]$

39. $f(x) = 2x^2 - 5x$ on $[-1,1]$

40. $f(x) = x^3 - 2x^2 + 5x$ on $[-1,3]$

41. $f(x) = \dfrac{x}{x^2+1}$ on $[0,3]$

42. $f(x) = \dfrac{x+1}{\sqrt{x^2+1}}$ on $[0,3]$

43. $f(x) = x^2 + x^4$ on $[0,1]$

44. $f(x) = \sin(x) + \cos(x)$ on $[0,\pi]$

45. $f(x) = \sin(x) - \cos(x)$ on $[0,\pi]$

46. $f(x) = x + \sin(x)$ on $[-\pi/2, \pi/2]$

47. $f(x) = x + \sin(x)$ on $[-\pi, 2\pi]$

48. $f(x) = \dfrac{x}{2} + \sin(x)$ on $[-\pi, 2\pi]$

49. $f(x) = 2\sin(x) - \sin(2x)$ on $[-\pi,\pi]$

50. $f(x) = \sin\left(x^2\right) + \cos\left(x^2\right)$ on $[0, \sqrt{2\pi}]$

51. $f(x) = \sin(x) - \cos(x)$ on $[-2\pi, 2\pi]$

52. $f(x) = \sin^2(x) - \cos^2(x)$ on $[-2\pi, 2\pi]$

In Exercises 53 to 59 graph the function.

53. $f(x) = \dfrac{\sin(x)}{1+2\cos(x)}$

54. $f(x) = \dfrac{\sqrt{x^2-1}}{x}$

55. $f(x) = \dfrac{3x^2+5}{x^2-1}$

56. $f(x) = \dfrac{3x^2+5}{x^2+1}$

57. $f(x) = 2x^{1/3} + x^{4/3}$

58. $f(x) = \sqrt{3}\sin(x) + \cos(x)$

59. $f(x) = \dfrac{1}{(x-1)^2(x-2)}$

---

60. Graph $f(x) = (x^2-9)^{1/3}e^{-x}$.

61. A differentiable function has $f'(x) < 0$ for $x < 1$ and $f'(x) > 0$ for $x > 1$. Also, $f(0) = 3$, $f(1) = 1$, and $f(2) = 2$.

(a) What is the minimum value of $f(x)$ for x in $[0,2]$? Why?

(b) What is the maximum value of $f(x)$ for x in $[0,2]$? Why?

In Exercises 62 to 65 decide if there is a differentiable function that meets all the conditions. If you think there is, sketch its possible graph. Otherwise, explain why a function cannot meet all of the conditions.

62. $f(x) > 0$ for all x and $f'(x) < 0$ for all x.

63. $f(3) = 1$, $f(5) = 1$, and $f'(x) > 0$ for x in $[3,5]$.

64. $f'(3) = f'(5) = 0$ and $f'(x) \neq 0$ for all other x and $f(-2) = f(4) = f(5) = 0$ and $f(x) \neq 0$ for all other x.

65. What is the minimum value of $y = (x^3 - x)/(x^2 - 4)$ for $x > 2$?

4.3 The Second Derivative and Graphing

The sign of the first derivative tells whether a function is increasing or decreasing. In this section we examine what the sign of the second derivative tells us about a function and its graph. This information will be used to help graph functions and also to provide an additional way to test whether a critical point is a maximum or minimum.

Concavity and Points of Inflection

The second derivative is the derivative of the first derivative. Thus, the sign of the second derivative determines if the first derivative is increasing or decreasing. If $f''(x)$ is positive for all x in an interval (a, b), then f' is an increasing function throughout the interval — the slope of the graph of $y = f(x)$ increases as x increases from left to right.

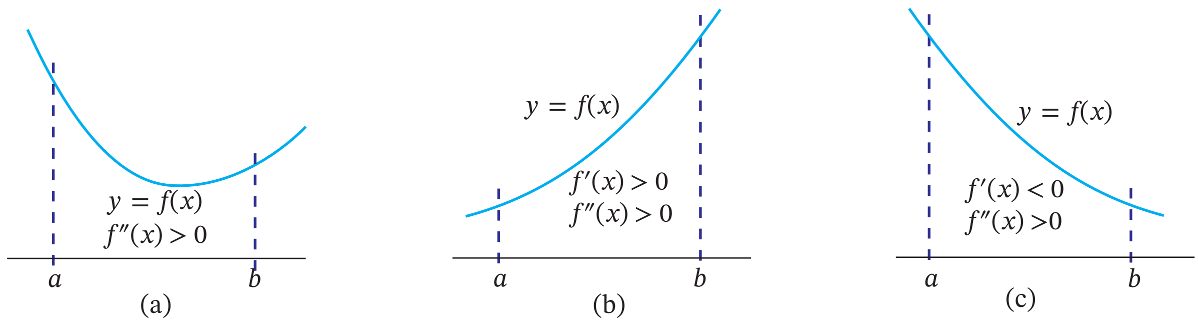


Figure 4.3.1

The slope may increase from negative values to zero to positive values, as in Figure 4.3.1(a). It may be positive throughout (a, b), as in Figure 4.3.1(b), or it may be negative throughout (a, b), as in Figure 4.3.1(c).

If $f''(x)$ is negative on the interval (a, b) then f' is decreasing on (a, b). The slope of the graph of $y = f(x)$ decreases as x increases from left to right on that part of the graph corresponding to (a, b).

Definition: *Concave Up and Concave Down.*

(i) A function whose first derivative is increasing on the open interval is called **concave up** in that interval.
(ii) A function whose first derivative is decreasing on the open interval is called **concave down** in that interval.

Observation 4.3.1: *Concavity and Chords*

(i) When a curve is concave up, it lies above its tangent lines and below its chords. The graph of a concave up function is shaped like a cup. See Figure 4.3.2(a).
(ii) When a curve is concave down, it lies below its tangent lines and above its chords. The graph of a concave down function is shaped like a frown. See Figure 4.3.2(b).

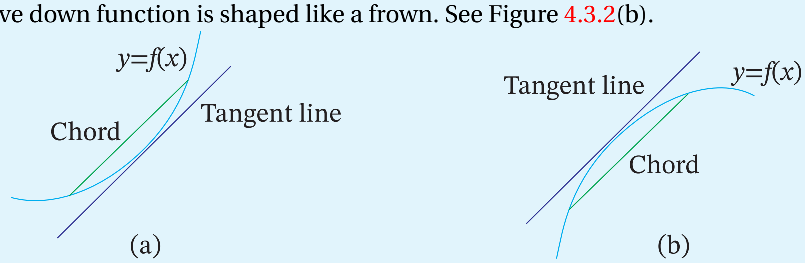


Figure 4.3.2

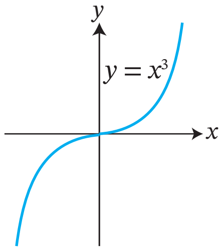


Figure 4.3.3

EXAMPLE 1. Where is the graph of $f(x) = x^3$ concave up? concave down?

SOLUTION First, compute the first and second derivatives of $f(x)$: $f'(x) = 3x^2$ and $f''(x) = 6x$. The second derivative is positive when x is positive and negative when x is negative. Thus, the graph is concave up for $x > 0$ and is concave down for $x < 0$. The sense of concavity changes at $x = 0$, where $f''(x) = 0$. (See Figure 4.3.3.) ◆

In an interval where $f''(x)$ is positive, the function $f'(x)$ is increasing, and so the function f is concave up. However, if a function is concave up, $f''(x)$ need not be positive for all x in the interval. For instance, $y = x^4$ has a second derivative $12x^2$ that is zero for $x = 0$, but the first derivative $4x^3$ is increasing on any interval; so the graph is concave up over any interval.

A point where the graph of a function changes concavity is important.

Definition: *Inflection Number and Inflection Point.*

Let f be a function and let a be a number. Assume there are numbers b and c such that $b < a < c$ and

1. f is continuous on the open interval (b, c)
2. f is concave up on (b, a) and concave down on (a, c)
   or
   f is concave down on (b, a) and concave up on (a, c).

Then the point $(a, f(a))$ is called an **inflection point** or **point of inflection** of f and the number a is called an **inflection number** of f.

CAUTION: If the second derivative changes sign at the number a, then a is an inflection number. If the second derivative exists at an inflection number, it must be 0. There can be an inflection point at a if $f''(a)$ is not defined, as is illustrated in the next example.

EXAMPLE 2. Examine the concavity of the graph of $y = x^{1/3}$.

SOLUTION Here $y' = x^{-2/3}/3$ and $y'' = -2x^{-5/3}/9$. Although $x = 0$ is in the domain of y, neither y' nor y'' is defined for $x = 0$. When x is negative, y'' is positive; when x is positive, y'' is negative. Thus, the concavity changes from concave up to concave down at $x = 0$. This means $x = 0$ is an inflection number and $(0, 0)$ is an inflection point. See Figure 4.3.4. ◆

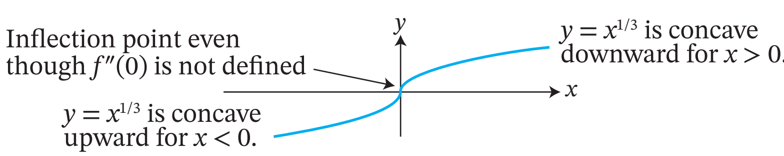


Figure 4.3.4

To find inflection points look for numbers where the second derivative changes sign.

Observation 4.3.2: *Identifying Inflection Points of* $y = f(x)$

Inflection points of $y = f(x)$ correspond to sign changes of f''. To find the sign changes of $f''(x)$:

1. Compute $f'(x)$ and $f''(x)$.
2. Look for numbers a such that $f''(a) = 0$ or f'' is not defined at a.
3. For each interval defined by the numbers found in Step 2, determine the sign of $f''(x)$.

REMINDER: Just because $f''(a) = 0$ does not automatically make a an inflection number of f. To be an inflection number, concavity has to change at a.

This process can be done using the same ideas used to identify critical points, as Example 3 shows.

EXAMPLE 3. Find the inflection point(s) of $f(x) = x^4 - 8x^3 + 18x^2$.

SOLUTION We have $f'(x) = 4x^3 - 24x^2 + 36x$ and

$$f''(x) = 12x^2 - 48x + 36 = 12(x^2 - 4x + 3) = 12(x-1)(x-3).$$

Because f'' is defined for all real numbers, the only candidates for inflection numbers are the solutions to $f''(x) = 0$, that is, the solutions to

$$12(x-1)(x-3) = 0.$$

The solutions are 1 and 3.

To decide whether 1 and 3 are inflection numbers, look at the sign of $f''(x) = 12(x-1)(x-3)$. For $x > 3$ both $x-1$ and $x-3$ are positive; so $f''(x)$ is positive. For x in $(1,3)$, $x-1$ is positive and $x-3$ is negative; so $f''(x)$ is negative. For $x < 1$, both $x-1$ and $x-3$ are negative; so $f''(x)$ is positive. This is recorded in Table 4.3.1.

x	$(-\infty,1)$	1	$(1,3)$	3	$(3,\infty)$
$f''(x)$	+	0	−	0	+

Table 4.3.1

Since sign changes in $f''(x)$ correspond to changes in concavity of the graph of f, the function has two inflection points: $(1,11)$ and $(3,27)$.

◆

Using Concavity in Graphing

The second derivative, together with the first derivative and the other tools of graphing, can help us sketch the graph of a function. Example 4 continues Example 3.

EXAMPLE 4. Graph $f(x) = x^4 - 8x^3 + 18x^2$.

SOLUTION As a nonconstant polynomial, the graph of $y = f(x)$ has no asymptotes. Its y-intercept is 0 since $f(0) = 0^4 - 8(0^3) + 18(0^2) = 0$. To find its x-intercepts look for solutions of $f(x) = 0$:

$$x^4 - 8x^3 + 18x^2 = 0$$
$$x^2(x^2 - 8x + 18) = 0.$$

The **discriminant** of ax^2+bx+c is b^2-4ac.

Thus $x = 0$ or $x^2 - 8x + 18 = 0$, which can be solved by the quadratic formula. The discriminant is $(-8)^2 - 4(1)(18) = -8$ which is negative, so there are no real solutions of $x^2 - 8x + 18 = 0$. The only x-intercept is $x = 0$.

In Example 3 we found

$$f'(x) = 4x^3 - 24x^2 + 36x = 4x(x^2 - 6x + 9) = 4x(x-3)^2,$$

which is 0 only when $x = 0$ and $x = 3$. The two critical points are $(0, f(0)) = (0,0)$ and $(3, f(3)) = (3,27)$. The information in Table 4.3.2 allows us to conclude that f is decreasing on $(-\infty,0)$ and increasing on $(0,\infty)$ with a local minimum at $(0,0)$.

x	$(-\infty,0)$	0	$(0,3)$	3	$(3,\infty)$
$f'(x)$	$-$	0	$+$	0	$+$

Table 4.3.2

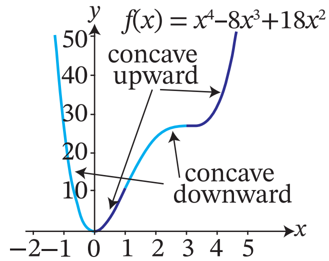


Figure 4.3.5

By Example 3, the graph of $y = x^4 - 8x^3 + 18x^2$ is concave up on $(-\infty, 1)$ and $(3, \infty)$ and concave down on $(1, 3)$.

To begin to sketch the graph of $y = f(x)$, plot the three points $(0, f(0)) = (0, 0)$, $(1, f(1)) = (1, 11)$, and $(3, f(3)) = (3, 27)$. They divide the domain into four intervals. On $(-\infty, 0)$ the function is decreasing and concave up, on $(0, 1)$ it is increasing and concave up, on $(1, 3)$ it is increasing and concave down, and on $(3, \infty)$ it is once again increasing and concave up. The final graph is shown in Figure 4.3.5. Each colored part indicates a section where both f' and f'' do not change sign. ♦

The procedure in Example 4 has several advantages. It was necessary to evaluate $f(x)$ only at a few important inputs x, which cut the domain into intervals where neither the first derivative nor the second derivative changes sign. On each of these intervals the graph of the function will have one of the four shapes shown in Figure 4.3.6. A graph usually is made up of these four shapes.

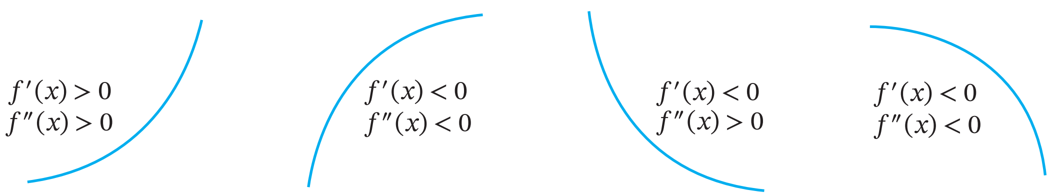


Figure 4.3.6

Local Extrema and the Second-Derivative Test

The second derivative is also useful in testing whether a critical number corresponds to a relative minimum or relative maximum. For this, we will use the relationships between concavity and tangent lines shown in Figure 4.3.2.

Let a be a critical number for the function f. Assume, for instance, that $f''(a)$ is negative. If f'' is continuous in some open interval that contains a, then (by the permanence property) $f''(x)$ remains negative for a small open interval that contains a. This means the graph of f is concave down near $(a, f(a))$: it lies below its tangent lines. In particular, it lies below the horizontal tangent line at the critical point $(a, f(a))$, as illustrated in Figure 4.3.7. Thus the function f has a relative maximum at the critical number a. Similarly, if $f'(a) = 0$ and $f''(a) > 0$, the critical point $(a, f(a))$ is a relative minimum because the graph of f is concave up and lies above the horizontal tangent line at $(a, f(a))$. These observations justify the following test for a relative extremum.

Theorem 4.3.3: Second-Derivative Test for a Relative Extremum

Let f be a function such that $f'(x)$ and $f''(x)$ are defined on some open interval containing the number a. Assume that $f''(x)$ is continuous.

(i) *If $f'(a) = 0$ and $f''(a) > 0$, then f has a relative minimum at $(a, f(a))$.*

(ii) *If $f'(a) = 0$ and $f''(a) < 0$, then f has a relative maximum at $(a, f(a))$.*

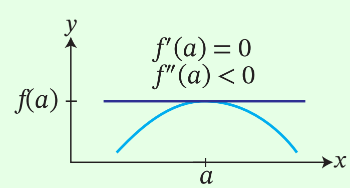


Figure 4.3.7

EXAMPLE 5. Use the second derivative test to classify all local extrema of $f(x) = x^4 - 8x^3 + 18x^2$.

SOLUTION This is the same function as in Examples 3 and 4. The two critical points are $(0,0)$ and $(3,27)$. The second derivative is $f''(x) = 12x^2 - 48x + 36$. At $x = 0$ we have

$$f''(0) = 12(0^2) - 48(0) + 36 = 36,$$

which is positive. Since $f'(0) = 0$ and $f''(0) > 0$, f has a local minimum at $(0,0)$. At $x = 3$ we have

$$f''(3) = 12(3^2) - 48(3) + 36 = 0.$$

Since $f''(3) = 0$, the second derivative test tells us nothing about the critical number 3.

Because $f''(x)$ changes sign at $x = 3$, the point $(3,27)$ is an inflection point and not a local extreme point.

These conclusions are consistent with our previous findings in Examples 3 and 4. ◆

Summary

Table 4.3.3 shows the meaning of the signs of $f(x)$, $f'(x)$, and $f''(x)$ in terms of the graph of $y = f(x)$.

Where the	is positive (>0)	is negative (<0)	changes sign	is zero ($=0$)
ordinate $f(x)$	the graph is above the x-axis.	the graph is below the x-axis.	the graph crosses the x-axis.	there is an x intercept.
slope $f'(x)$	the graph slopes upward.	the graph slopes downward.	the graph has a horizontal tangent and a relative extremum.	there is a critical point.
$f''(x)$	the graph is concave up (like a cup).	the graph is concave down (like a frown).	the graph has an inflection point.	there may be an inflection point.

Table 4.3.3

EXERCISES for Section 4.3

In Exercises 1 to 16 describe the intervals where the function is concave up and concave down and list all inflection points.

1. $f(x) = x^3 - 3x^2 + 2$

2. $f(x) = x^3 - 6x^2 + 1$

3. $f(x) = x^2 + x + 1$

4. $f(x) = 2x^2 - 5x$

5. $f(x) = x^4 - 4x^3$

6. $f(x) = 3x^5 - 5x^4$

7. $f(x) = \dfrac{1}{1+x^2}$

8. $f(x) = \dfrac{1}{1+x^4}$

9. $f(x) = x^3 + 6x^2 - 15x$

10. $f(x) = \dfrac{x^2}{2} + \dfrac{1}{x}$

11. $f(x) = e^{-x^2}$

12. $f(x) = xe^x$

13. $f(x) = \tan(x)$

14. $f(x) = \sin(x) + \sqrt{3}\cos(x)$

15. $f(x) = \cos(x)$

16. $f(x) = \cos(x) + \sin(x)$

In Exercises 17 to 30 graph the functions, showing critical points, inflection points, and intercepts.

17. $f(x) = x^3 + 3x^2$

18. $f(x) = 2x^3 + 9x^2$

19. $f(x) = x^4 - 4x^3 + 6x^2$

20. $f(x) = x^4 + 4x^3 + 6x^2 - 2$

21. $f(x) = x^4 - 6x^3 + 12x^2$

22. $f(x) = 2x^6 - 10x^4 + 10$

23. $f(x) = 2x^6 + 3x^5 - 10x^4$

24. $f(x) = 3x^4 + 4x^3 - 12x^2 + 4$

25. $f(x) = 3x^4 + 4x^3 - 12x^2 + 4$

26. $f(x) = xe^{-x}$

27. $f(x) = e^{-x^2/2}$

28. $f(x) = e^{x^3}$

29. $f(x) = 3x^5 - 20x^3 + 10$ See Example 3 in Section 4.2.

30. $f(x) = 2x^6 - 15x^4 + 20x^3 - 20x + 10$

In Exercises 31 to 38 sketch the general appearance of the graph of the function near $(1,1)$ on the basis of the information given. Assume that f, f', and f'' are continuous.

31. $f(1) = 1$, $f'(1) = 0$, and $f''(1) = -1$

32. $f(1) = 1$, $f'(1) = 0$, and $f''(1) = 1$

33. $f(1) = 1$, $f'(1) = 0$, and $f''(1) = 0$ Sketch five quite different possibilities.

34. $f(1) = 1$, $f'(1) = 0$, $f''(1) = 0$, $f''(x) < 0$ for $x < 1$, and $f''(x) > 0$ for $x > 1$

35. $f(1) = 1$, $f'(1) = 0$, $f''(1) = 0$, and $f''(x) < 0$ for x near 1

36. $f(1) = 1$, $f'(1) = 1$, $f''(1) = 0$, and $f''(x) > 0$ for x near 1

37. $f(1) = 1$, $f'(1) = 1$, $f''(1) = 0$, $f''(x) < 0$ for $x < 1$, and $f''(x) > 0$ for $x > 1$

38. $f(1) = 1$, $f'(1) = 1$, and $f''(1) = -1$

---

39. Find all inflection points of $f(x) = x\ln(x)$. On what intervals is the graph of $y = f(x)$ concave up? concave down? Graph $y = f(x)$ on an interval large enough to show all interesting features of the graph. On what intervals is the graph increasing? decreasing? See Example 2 of Section 4.2.

40. Find all inflection points of $f(x) = x + \ln(x)$. On what intervals is the graph of $y = f(x)$ concave up? concave down? Graph $y = f(x)$ on an interval large enough to show all interesting features of the graph. On what intervals is the function increasing? decreasing?

41. Find all inflection points of $f(x) = (x+1)^{2/7}e^{-x}$. On what intervals is the graph of $y = f(x)$ concave up? concave down? On what intervals is the function increasing? decreasing? See Example 4 of Section 4.2.

42. Find the critical points and inflection points of $f(x) = x^2 e^{-x/3}$.

In Exercises 43 to 44 sketch a graph of a function that meets the conditions. Assume f' and f'' are continuous. Explain your reasoning.

43. Critical point $(2,4)$, inflection points $(3,1)$ and $(1,1)$, $\lim_{x\to\infty} f(x) = 0$, and $\lim_{x\to-\infty} f(x) = 0$.

44. Critical points $(-1,1)$ and $(3,2)$, inflection point $(4,1)$, $\lim_{x\to 0^+} f(x) = -\infty$, $\lim_{x\to 0^-} f(x) = \infty$, $\lim_{x\to\infty} f(x) = 0$, and $\lim_{x\to-\infty} f(x) = \infty$.

---

45. Graph $y = 2(x-1)^{5/3} + 5(x-1)^{2/3}$, paying attention to points where y' does not exist.

46. Graph $y = x + (x+1)^{1/3}$.

47. Find the critical points and inflection points in $[0, 2\pi]$ of $f(x) = \sin^2(x)\cos(x)$.

48. Can a polynomial of degree 6 have (a) no inflection points? (b) exactly one inflection point? Explain.

49. Can a polynomial of degree 5 have (a) no inflection points? (b) exactly one inflection point? Explain.

50. Let f be a function such that $f''(x) = (x-1)(x-2)$. (a) For which x is f concave up? (b) For which x is f concave down? (c) List its inflection number(s). (d) Find a function f whose second derivative is $(x-1)(x-2)$.

51. Figure 4.3.8 is the graph of the derivative of a function f that is continuous for all x and differentiable for all x other than 0. CONTRIBUTED BY: David Hayes

(a) Answer the following questions about f (not about f'). Where is f increasing? decreasing? concave up? concave down? What are the critical numbers? Where do any relative extrema occur? Explain.

(b) Assuming that $f(0) = 1$, graph a hypothetical function f that satisfies the conditions.

(c) Sketch $f''(x)$.

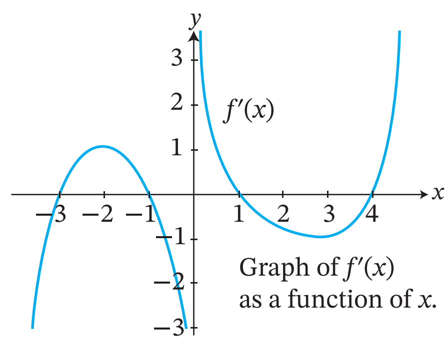


Figure 4.3.8

52. In the theory of **inhibited growth** it is assumed that the growing quantity y approaches a limiting size M. Specifically, one assumes that the rate of growth is proportional to both the quantity present and the amount left to grow: $\frac{dy}{dt} = ky(M-y)$, where k is a positive number. Prove that the graph of y as a function of time has an inflection point when the amount y is exactly half the limiting amount M.

53. A function $y = f(x)$ has the property that $y' = \sin(y) + 2y + x$. Show that at a critical number it has a local minimum.

54. Assume that the domain of $f(x)$ is the entire x-axis and $f'(x)$ and $f''(x)$ are continuous. Also, $(1,1)$ is the only critical point and $\lim_{x\to\infty} f(x) = 0$. (a) Can $f(x)$ be negative for some $x > 1$? (b) Must $f(x)$ be decreasing for $x > 1$? (c) Must $f(x)$ have an inflection point?

4.4 Proofs of the Three Theorems in Section 4.1

In Section 4.1 two observations about tangent lines led to the theorem of the interior extremum, Rolle's theorem, and the mean value theorem. Now, using the definition of the derivative and no pictures, we prove them. That the proofs need only the definition of the derivative as a limit reassures us that the definition is suitable to serve as part of the foundation of calculus.

Proof of the Theorem of the Interior Extremum (Theorem 4.1.2)

Suppose the maximum of f on the open interval (a,b) occurs at c. This means that $f(c) \geq f(x)$ for each x between a and b.

Our challenge is to use only this information and the definition of the derivative as a limit to show that either $f'(c) = 0$ or f is not differentiable at c.

Assume that f is differentiable at c. We will show that $f'(c) \geq 0$ and $f'(c) \leq 0$, forcing $f'(c)$ to be zero.

By the definition of the derivative,

$$f'(c) = \lim_{\Delta x \to 0} \frac{f(c+\Delta x) - f(c)}{\Delta x}.$$

Theorem 4.1.2: Let f be a function defined on (a,b). If f takes on an extreme value at c in this interval, then $f'(c) = 0$ or $f'(c)$ does not exist.

The assumption that f is differentiable on (a,b) implies that $f'(c)$ exists. In the difference quotient

$$\frac{f(c+\Delta x)-f(c)}{\Delta x} \tag{4.4.1}$$

take Δx so small that $c+\Delta x$ is in the interval (a,b). Then $f(c+\Delta x) \le f(c)$. Hence $f(c+\Delta x)-f(c) \le 0$. Therefore, when Δx is positive, the difference quotient in (4.4.1) will be negative or zero. Consequently, as $\Delta x \to 0$ through positive values,

$$f'(c) = \lim_{\Delta x \to 0^+} \frac{f(c+\Delta x)-f(c)}{\Delta x} \le 0.$$

If, on the other hand, Δx is negative, then the difference quotient in (4.4.1) will be positive or zero. Hence, as $\Delta x \to 0$ through negative values,

$$f'(c) = \lim_{\Delta x \to 0^-} \frac{f(c+\Delta x)-f(c)}{\Delta x} \ge 0.$$

The only way $f'(c) \le 0$ and $f'(c) \ge 0$ can both hold is when $f'(c) = 0$. This proves that if f has a maximum on (a,b), then $f'(c) = 0$.

The proof when f has a minimum on (a,b) is essentially the same. (See Exercise 16.) ●

The proofs of Rolle's theorem and the mean value theorem are related. Suppose f is continuous on $[a,b]$ and differentiable on (a,b).

Proof of Rolle's Theorem (Theorem 4.1.4)

The goal is to use the facts that f is continuous on $[a,b]$, differentiable on (a,b), and $f(a) = f(b)$ to conclude that there must be a number c in (a,b) with $f'(c) = 0$.

Since f is continuous on the closed interval $[a,b]$, it has a maximum value M and a minimum value m on the interval. There are two cases: $m = M$ and $m < M$.

Case 1: If $m = M$, f is constant and $f'(x) = 0$ for all x in $[a,b]$. Then any number in (a,b) will serve as c.

Case 2: Suppose $m < M$. Because $f(a) = f(b)$ the minimum and maximum cannot both occur at the ends of the interval. At least one of the extrema occurs at a number c strictly between a and b. By assumption, f is differentiable at c, so $f'(c)$ exists. By the theorem of the interior extremum, $f'(c) = 0$.

This completes the proof of Rolle's theorem. ●

> Theorem 4.1.4:
> Let f be continuous on $[a,b]$ and differentiable on (a,b). If $f(a) = f(b)$, then there is at least one c in (a,b) where $f'(c) = 0$.

The idea behind the proof of the mean value theorem is to define a function to which Rolle's theorem can be applied.

Proof of the Mean Value Theorem (Theorem 4.1.6)

Let $y = L(x)$ be the equation of the chord through $(a, f(a))$ and $(b, f(b))$. Its slope is $L'(x) = (f(b)-f(a))/(b-a)$. Define $h(x) = f(x) - L(x)$. Note that $h(a) = h(b) = 0$ because $f(a) = L(a)$ and $f(b) = L(b)$.

By assumption, f is continuous on the closed interval $[a,b]$ and differentiable on the open interval (a,b). So h, being the difference of f and L, is also continuous on $[a,b]$ and differentiable on (a,b).

Rolle's theorem applies to h on $[a,b]$. Therefore, there is at least one c in (a,b) where $h'(c) = 0$. Because $h'(c) = f'(c) - L'(c)$ we have $f'(c) = L'(c)$ and therefore

$$f'(c) = \frac{f(b)-f(a)}{b-a}.$$ ●

> Theorem 4.1.6:
> Let f be continuous on $[a,b]$ and differentiable on (a,b). Then there is at least one c in (a,b) where $f'(c) = \frac{f(b)-f(a)}{b-a}$.

Summary

Using only the definition of the derivative and the assumption that a continuous function defined on a closed interval assumes maximum and minimum values, we proved the theorem of the interior extremum, Rolle's theorem, and the mean value theorem. We did not appeal to any pictures or to geometric intuition.

EXERCISES for Section 4.4

In each of Exercises 1 to 3 sketch a graph of a differentiable function that meets the given conditions. (There is no need to give a formula for the function.)

1. $f'(x) < 0$ for all x

2. $f'(3) = 0$ and $f'(x) < 0$ for x not equal to 3

3. $f'(x) = 0$ only when $x = 1$ or 4; $f(1) = 3$, $f(4) = 1$; $f'(x) > 0$ for $x < 1$ and for $x > 4$

In Exercises 4 and 5 explain why no differentiable function satisfies all the conditions.

4. $f(1) = 3$, $f(2) = 4$, $f'(x) < 0$ for all x

5. $f(x) = 2$ only when $x = 0$, 1, and 3; $f'(x) = 0$ only when $x = \frac{1}{4}, \frac{3}{4}$, and 4.

---

6. In *Surely You're Joking, Mr. Feynman*, Norton, New York, 1985, Nobel laureate Richard P. Feynman writes:

> I often liked to play tricks on people when I was at MIT. One time, in mechanical drawing class, some joker picked up a French curve (a piece of plastic for drawing smooth curves — a curly funny-looking thing) and said, "I wonder if the curves on that thing have some special formula?"
>
> I thought for a moment and said, "Sure they do. The curves are very special curves. Lemme show ya," and I picked up my French curve and began to turn it slowly. "The French curve is made so that at the lowest point on each curve, no matter how you turn it, the tangent is horizontal."
>
> All the guys in the class were holding their French curve up at different angles, holding their pencil up to it at the lowest point and laying it down, and discovering that, sure enough, the tangent is horizontal.

How was Feynman playing a trick on his classmates?

7. Let f be a differentiable function. What can be said about the number of solutions of $f(x) = 3$
(a) when $f'(x) > 0$ for all x?, and (b) when $f'(x) > 0$ for $x < 7$ and $f'(x) < 0$ for $x > 7$?.

8. For $f(x) = x^3 + ax^2 + c$, show that if $a < 0$ and $c > 0$, then $f(x) = 0$ has exactly one negative solution.

9. Obtain the mean value theorem from Rolle's theorem.

> CHALLENGE: Answer this question without looking anything up in this book — or any other reference.

10. (a) Using the definition of $L(x)$ in the proof of the mean value theorem, show that $L(x) = f(a) + \frac{x-a}{b-a}\left(f(b) - f(a)\right)$.

(b) Using (a), confirm that $L'(x) = \frac{f(b) - f(a)}{b-a}$.

11. Show that Rolle's theorem is a special case of the mean value theorem.

12. Is this argument a proof of the mean value theorem?

Alleged Proof of Theorem 4.1.6

Tilt the x- and y-axes and the graph of the function until the x-axis is parallel to the given chord. The chord is now horizontal, and we may apply Rolle's theorem.

13. This exercise shows that a polynomial $f(x)$ of degree n, $n \geq 1$, can have at most n distinct real roots, that is, solutions to $f(x) = 0$.

(a) Use algebra to show that the statement holds for $n = 1$ and $n = 2$.

(b) Use calculus to show that the statement then holds for $n = 3$.
(c) Use calculus to show that the statement continues to hold for $n = 4$ and $n = 5$.
(d) Why does it hold for all positive integers n?

14. Is there a differentiable function f whose domain is the x-axis such that f is increasing and yet the derivative is not positive for all x?

15. Prove that if f is continuous on $[a, b]$ and $f'(x) < 0$ on (a, b) then f is decreasing on the interval $[a, b]$.

16. Prove the theorem of the interior extremum when the minimum of f on (a, b) occurs at c.

This Exercise provides an analytic justification for the first part of the statement, in Section 4.3, that when a curve is concave up, it lies above its tangent lines and below its chords. (The second part is proved in Exercise 52 in Section 4.S.)

17. Show that in an open interval in which f'' is positive, tangents to the graph of f lie below the curve.

---

18. We stated in Section 4.3 that if $f(x)$ is defined in an open interval around a critical number a and $f''(a)$ is negative, then $f(x)$ has a relative maximum at a. Explain why this is so, following these steps.
(a) Why is $\lim_{\Delta x \to 0} \frac{f'(a + \Delta x) - f'(a)}{\Delta x}$ negative?
(b) Deduce that if Δx is small and positive, then $f'(a + \Delta x)$ is negative.
(c) Show that if Δx is small and negative, then $f'(a + \Delta x)$ is positive.
(d) Show that $f'(x)$ changes sign from positive to negative at a. By the first-derivative test for a relative maximum, $f(x)$ has a relative maximum at a.

19. To keep differentiation skills sharp, differentiate each of the following expressions
(a) $\sqrt{1 - x^2}\sin(3x)$ (b) $\frac{\sqrt[3]{x}}{x^2 + 1}$ (c) $\tan\left(\frac{1}{(2x + 1)^2}\right)$ (d) $\ln\left(\frac{(x^2 + 1)^3\sqrt{1 - x^2}}{\sec^2(x)}\right)$ (e) e^{x^4}

4.S Chapter Summary

In this chapter we saw that the signs of the function and of its first and second derivatives influence the shape of its graph. In particular the derivatives show where the function is increasing or decreasing and where it is concave up or down. That enabled us to find extreme points and inflection points.

We state here the main ideas informally for a function, f, with continuous first and second derivatives.

If a function f has an extremum at a number c, then either $f'(c) = 0$, or $f'(c)$ is not defined, or c may be the endpoint of the domain. This narrows the search for extrema. If $f'(c) = 0$ and $f''(c)$ is not zero, the function has an extremum at c.

Rolle's theorem asserts that if a differentiable function has the same value at two inputs on an interval in its domain, its derivative must be zero somewhere between them. This helps find the number of solutions to $f(x) = 0$, and thus the number of x-intercepts of the graph of f.

The Mean Value Theorem generalizes Rolle's theorem. It says that between any two points on the graph of a differentiable function f there is a point on the graph where the tangent is parallel to the chord through the two points. We use this to show that if a and b are numbers, then $f(b) = f(a) + f'(c)(b - a)$ for some c between a and b.

If $f'(a)$ is positive and f' is continuous on an open interval containing a, then, by the permanence property, $f'(x)$ remains positive for some open interval containing a. This implies that if the derivative is positive at some number, then the function is increasing for inputs near that number. A similar statement holds when $f'(a)$ is negative.

We close with a conversation between our ever-present students, Sam and Jane.

SAM: Why bother me with limits? The authors say we need them to define derivatives.
JANE: Aren't you curious about why the formula for the derivative of a product is what it is?

Sam: No. It's been true for over three centuries. Just tell me what it is. If someone says the speed of light is 186,000 miles per second am I supposed to find a meter stick and clock and check it out?

Jane: But what if you forget the formula during a test?

Sam: That's not much of a reason.

Jane: I agree. But avoiding limits and proofs based on them is like building a brick wall without mortar. At any moment you may collapse, especially during an exam.

Sam: Don't mention exams.

Jane: Well, my physics class uses limits all the time. Density of mass, density of electric charge, for instance.

Sam: All right, I'll go review the whole course.

Jane: You'll see that once you delete the exercises and examples, there aren't many pages to read. We really haven't covered much. Just master the proofs and you'll be as confident as our instructor.

Sam: That cheers me on.

EXERCISES for Section 4.S

In each of Exercises 1 to 13 decide if it is possible for a differentiable function to have all the properties listed. If it is possible, sketch a graph of a differentiable function that meets the conditions. (There is no need to try to find a formula for it.) If it is not possible, explain why.

1. $f(0)=1$, $f(x)>0$ for all x, and $f'(x)<0$ for all positive x
2. $f(0)=-1$, $f'(x)<0$ for all x in $[0,2]$, and $f(2)=0$
3. x-intercepts at 1 and 5, y intercept at 2, $f'(x)<0$ for $x<4$ and $f'(x)>0$ for $x>4$
4. x-intercepts at 2 and 5, y intercept at 3, $f'(x)>0$ for $x<1$ and for $x>3$, and $f'(x)<0$ for x in $(1,3)$
5. $f(0)=1$, $f'(x)<0$ for all positive x, and $\lim_{x\to\infty} f(x)=1/2$
6. $f(2)=5$, $f(3)=-1$, and $f'(x)\geq 0$ for all x
7. x-intercepts only at 1 and 2 and $f(3)=-1$, $f(4)=2$
8. $f'(x)=0$ only when $x=1$ or 4, $f(1)=3$, $f(4)=1$, $f'(x)<0$ for $x<1$, and $f'(x)>0$ for $x>4$
9. $f(0)=f(1)=1$ and $f'(0)=f'(1)=1$
10. $f(0)=f(1)=1$, $f'(0)=f'(1)=1$, and $f(x)\neq 0$ for all x in $[0,1]$
11. $f(0)=f(1)=1$, $f'(0)=f'(1)=1$, and $f(x)=0$ for exactly one number x in $[0,1]0$
12. $f(0)=f(1)=1$, $f'(0)=f'(1)=1$, and $f(x)$ has exactly two inflection numbers in $[0,1]$
13. $f(0)=f(1)=1$, $f'(0)=f'(1)=1$, and $f(x)$ has exactly two extrema in $[0,1]$

Figure 4.S.1

14. State the assumptions and conclusions of the theorem of the interior extremum for a function F defined on (a,b).

15. State the assumptions and conclusions of the mean value theorem for a function g defined on $[c,d]$.

16. Find all functions $f(x)$ such that $f'(x)=2$ for all x and $f(1)=4$.

17. Find all twice differentiable functions such that $f(1)=3$, $f'(1)=-1$, and $f''(1)=e^x$.

18. The newspaper headline shown in Figure 4.S.1 appeared in February 2012. Let $f(t)$ be the average home price at time t. Translate this headline into a sentence about calculus, that is, about the derivatives of f.

19. At high tide and at low tide, the height of the tide changes very slowly. The same holds for an outdoor thermometer: the temperature changes the slowest when it is at its highest or at its lowest. Why is that?

20. The following discussion on higher derivatives in economics appears on page 124 of the *College Mathematics Journal* **37** (2006):

Historical Note: Higher Derivatives and Economics

Charlie Marion of Shrub Oak, NY, submitted this excerpt from "Curses! The Second Derivative" by Jeremy J. Siegel in the October 2004 issue of *Kiplinger's* (p. 73):

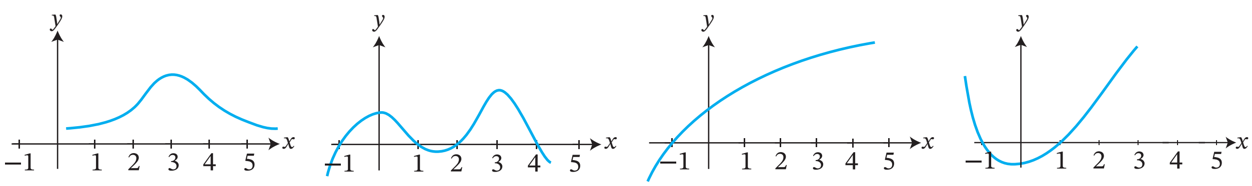
> "... I think what is bugging the market is something that I have seen happen many times before: the Curse of the Second Derivative. The second derivative, for all those readers who are a few years away from their college calculus class, is the rate of change of the rate of change — or, in this case, whether corporate earnings, which are still rising, are rising at a faster or slower pace."

In the October 1996 issue of the *Notices of the American Mathematical Society*, Hugo Rossi wrote, "In the fall of 1972 President Nixon announced that the rate of increase of inflation was decreasing. This was the first time a sitting president used the third derivative to advance his case for reelection."

Explain why the third derivative is involved in President Nixon's statement.

21. (a) Graph $y = \sin^2(2\theta)\cos(2\theta)$ for θ in $[-\pi/2, \pi/2]$. (b) What is the maximum value of y?

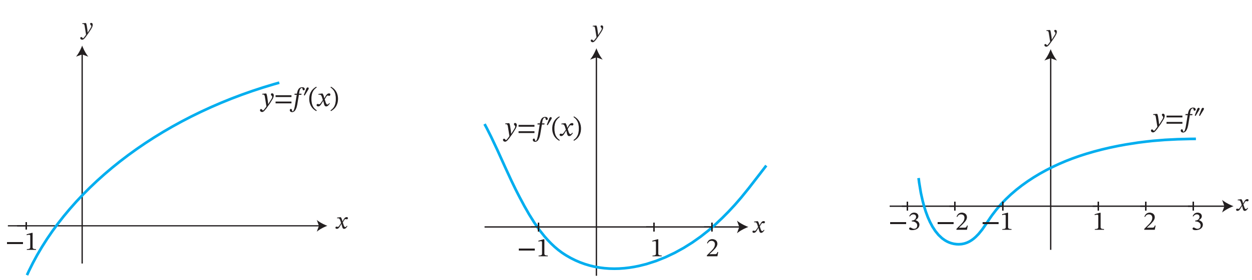
In Exercises 22 to 25, from the graph of function f with continuous f' and f'', sketch a possible graph of f' and a possible graph of f''.



(a) (b) (c) (d)

Figure 4.S.2

22. Figure 4.S.2(a) **23.** Figure 4.S.2(b) **24.** Figure 4.S.2(c) **25.** Figure 4.S.2(d)



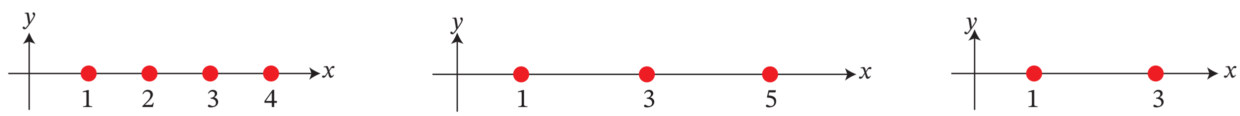
(a) (b) (c)

Figure 4.S.3

26. Sketch the graphs of two possible functions f whose first derivative f' is graphed in Figure 4.S.3(a).

27. Sketch the graphs of two possible functions f whose first derivative f' is graphed in Figure 4.S.3(b).

28. Sketch the graph of a function f whose second derivative f'' is graphed in Figure 4.S.3(c).



(a) (b) (c)

Figure 4.S.4

29. Figure 4.S.4(a) shows the only x-intercepts of a function f. Sketch three possible graphs of f' and f''.

30. Figure 4.S.4(b) shows the only arguments at which $f'(x) = 0$. Sketch three possible graphs of f' and f''.

31. Figure 4.S.4(c) shows the only arguments at which $f''(x) = 0$. Sketch three possible graphs of f' and f''.

In Exercises 32 to 39 graph the functions, showing extrema, inflection points, and asymptotes.

32. $e^{-2x}\sin(x)$ on $[0, 4\pi]$ **33.** $\dfrac{e^x}{1-e^x}$ **34.** $x^3 - 9x^2$ **35.** $x\sqrt{3-x}$

36. $\dfrac{x-1}{x-2}$ **37.** $\cos(x) - \sin(x)$ on $[0, 2\pi]$ **38.** $x^{1/2} - x^{1/4}$ **39.** $\dfrac{x}{4-x^2}$

---

40. Figure 4.S.5 shows the graph of f. Estimate where
(a) f changes sign , (b) f' changes sign, and (c) f'' changes sign.

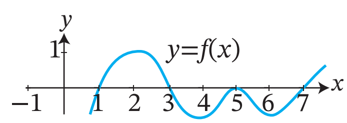


Figure 4.S.5

41. Find the maximum value of $e^{2\sqrt{3}x}\cos(2x)$ for x in $[0, \pi/4]$.

42. (a) Show that the equation $5x - \cos(x) = 0$ has exactly one solution.
(b) Find an interval of length less than 0.1 that contains the solution.

43. Assume f has continuous first and second derivatives defined on an open interval.
(a) If $f'(a) = 0$ and $f''(a) = 0$, does f necessarily have an extremum at a? Explain.
(b) If $f''(a) = 0$, does f necessarily have an inflection point at $x = a$?
(c) If $f'(a) = 0$ and $f''(a) = 3$, does f necessarily have an extremum at a?

44. Define $f(x) = x^3 - 3x$.
(a) Solve $f'(x) = 0$.
(b) Use the theorem of the interior extremum to show that the maximum value of $x^3 - 3x$ for x in $[1, 5]$ occurs either at 1 or at 5.

45. Let f and g be polynomials without a common root.
(a) Show that if the degree of g is odd, the graph of $\dfrac{f}{g}$ has a vertical asymptote.
(b) Show that if the degree of f is less than or equal to the degree of g, then $\dfrac{f}{g}$ has a horizontal asymptote.

46. If $\lim_{x\to\infty} f'(x) = 0$, does it follow that the graph of $y = f(x)$ has a horizontal asymptote? Explain.

47. Let f be a positive function on $(0, \infty)$ with f' and f'' both continuous. Let $g(x) = (f(x))^2$.
(a) If f is increasing, is g? (b) If f is concave up, is g?

48. Give an example of a positive function f on $(0, \infty)$ that is concave down but f^2 is concave up.

49. Graph $\cos(2\theta) + 4\sin(\theta)$ for θ in $[0, 2\pi]$.

50. Graph $\cos(2\theta) + 2\sin(\theta)$ for θ in $[0, 2\pi]$.

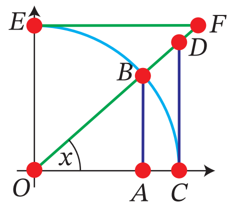


Figure 4.S.6

51. Figure 4.S.6 shows part of a unit circle. The line segment CD is tangent to the circle and has length x. This exercise uses calculus to show that $|AB| < |\widehat{BC}| < |CD|$.

NOTE: $|\widehat{BC}|$ is the length of the arc joining B and C.

(a) Express $|AB|$ and $|BC|$ in terms of x.
(b) Using (a) and calculus, show that for $x > 0$, $|AB| < |BC| < |CD|$.

52. Assume that $f''(x)$ is positive for x in an open interval. Let a and b, $a < b$, be in the interval. This exercise shows that the chord joining $(a, f(a))$ to $(b, f(b))$ lies above the graph of f. ("A concave up curve has chords that lie above the curve.") Compare with Exercise 17 in Section 4.4.

(a) Why does one want to prove that

$$f(a) + \frac{f(b) - f(a)}{b - a}(x - a) > f(x), \qquad \text{for } a < x < b?$$

(b) Why does one want to prove that

$$\frac{f(b) - f(a)}{b - a} > \frac{f(x) - f(a)}{x - a}? \tag{4.S.1}$$

(c) Show that right-hand side of (4.S.1) is increasing for $a < x < b$. Why does this show that the chords lie above the curve?

53. He's at it again. He's been thinking, and leaves the following message for Jane:

SAM: I can do Exercise 52 more easily. I'll show that (4.S.1) is true. By the mean value theorem, I can write the left side as $f'(c)$ where c is in $[a, b]$ and the right side as $f'(d)$ where d is in $[a, x]$. Since $b > x$, I know $c > d$, hence $f'(c) > f'(d)$. Nothing to it.

Is Sam's reasoning correct?

54. (a) Graph $y = \dfrac{\sin(x)}{x}$ showing intercepts and asymptotes.
(b) Graph $y = x$ and $y = \tan(x)$ on the same axes.
(c) Use (b) to find how many solutions there are to $x = \tan(x)$.
(d) Write a short commentary on the critical points of $\sin(x)/x$.
(e) Refine the graph in (a) to show several critical points.

55. Let $f(x) = ax^3 + bx^2 + cx + d$, where $a \neq 0$.
(a) Show that the graph of $y = f(x)$ always has exactly one inflection point.
(b) Show that the inflection point separates the graph into two parts that are congruent.

56. Assume that a_i and b_i, $0 \le i \le n$, are positive and the ratios a_i/b_i increase as a function of the index i. (That is, $a_0/b_0 < a_1/b_1 < \cdots < a_n/b_n$.) Then it is known that $f(x) = \dfrac{\sum_{i=0}^{n} a_i x^i}{\sum_{i=0}^{n} b_i x^i}$ is an increasing function for $x > 0$. This is used in the statistical theory of reliability. Verify the assertion for (a) $n = 1$ and (b) $n = 2$.

CONTRIBUTED BY: Frank Saminiego

57. (a) Graph $y = \dfrac{1}{1 + 2^{-x}}$.
(b) The point $\left(0, \dfrac{1}{2}\right)$ is on the graph and divides it into two pieces. Are the pieces congruent?

NOTE: Curves of this type model the depletion of a finite resource; x is time and y is the fraction of the resource consumed up to time x.

58. (a) If the graph of f has a horizontal asymptote (say $\lim_{x\to\infty} f(x) = L$), does it follow that $\lim_{x\to\infty} f'(x)$ exists?
(b) If $\lim_{x\to\infty} f'(x)$ exists in (a), must it be 0?

59. Assume that f is continuous on $[1,3]$, $f(1)=5$, $f(2)=4$, and $f(3)=5$. Show that the graph of f has a horizontal chord of length 1.

60. Can a straight line meet the curve $y=x^5$ four times? No. Three times is possible, but not four.

61. A polynomial function f defined on the whole x-axis has exactly one inflection point. In at most how many points can a straight line intersect the graph of f? Explain. (Examples include x^n, n an odd integer greater than 1.)

62. Let f be an increasing function with continuous f' and f''. What, if anything, can be said about the concavity of the composite function $f\circ f$ if (a) f is concave up? (b) f is concave down?

63. Assume f has continuous first and second derivatives. Show that if f and $g=f^2$ have inflection points at the same argument a, then $f'(a)=0$.

64. Assume $\lim\limits_{x\to\infty} f'(x)=3$. Show that for x sufficiently large, $f(x)$ is greater than $2x$.

65. Assume that f is differentiable for all numbers x.

(a) If f is an even function, what, if anything, can be said about $f'(0)$?

(b) If f is an odd function, what, if anything, can be said about $f'(0)$?

Explain your answers.

66. Graph $y=\sin(x^2)$ on the interval $[-\sqrt{\pi},\sqrt{\pi}]$. Identify the extreme points and the inflection points.

67. Assume that $f(x)$ is a continuous function not identically 0 defined on $(-\infty,\infty)$ and that $f(x+y)=f(x)\cdot f(y)$ for all x and y.

(a) Show that $f(0)=1$.

(b) Show that $f(x)$ is never 0.

(c) Show that $f(x)$ is positive for all x.

(d) Letting $f(1)=a$, find $f(2)$, $f(1/2)$, and $f(-1)$, in terms of a.

(e) Show that $f(x)=a^x$ for all x.

68. Assume $y=f(x)$ is a twice differentiable function with $f(0)=1$ and $f''(x)<-1$ for all x. Is it possible that $f(x)>0$ for all x in $(1,\infty)$? No. Eventually $f'(x)<0$ and f stays below its tangents.

69. If $\lim\limits_{x\to\infty} f'(x)=3$, does it follow that the graph of $y=f(x)$ is asymptotic to some line of the form $y=a+3x$ for some constant a? No. $f(x)=3x+ln(x)$.

70. Assume that $f(x)$ is defined for all real numbers and has a continuous derivative. Assume that $f'(x)$ is positive for all x other than c and that $f'(c)=0$.

(a) Give two examples of functions with these properties.

(b) Must any function with these properties be increasing?

Exercises 71 to 74 involve the **hyperbolic functions**. The **hyperbolic sine** function is $\sinh(x)=\dfrac{e^x-e^{-x}}{2}$ and the **hyperbolic cosine** function is $\cosh(x)=\dfrac{e^x+e^{-x}}{2}$. Hyperbolic functions are discussed in greater detail in Section 5.8.

71. (a) Show that $\dfrac{d}{dx}\sinh(x)=\cosh(x)$. (b) Show that $\dfrac{d}{dx}\cosh(x)=\sinh(x)$.

72. Define $\operatorname{sech}(x)=\dfrac{1}{\cosh(x)}=\dfrac{2}{e^x+e^{-x}}$ and $\tanh(x)=\dfrac{\sinh(x)}{\cosh(x)}=\dfrac{e^x-e^{-x}}{e^x+e^{-x}}$.

(a) Show that $\dfrac{d}{dx}\tanh(x)=(\operatorname{sech}(x))^2$. (b) Show that $\dfrac{d}{dx}\operatorname{sech}(x)=-\operatorname{sech}(x)\tanh(x)$.

73. Use calculus to show that $(\cosh(x))^2-(\sinh(x))^2$ is a constant. Find the constant.

74. Use calculus to show that $(\operatorname{sech}(x))^2+(\tanh(x))^2$ is a constant. Find the constant.

Calculus is Everywhere # 5

Calculus Reassures a Bicyclist

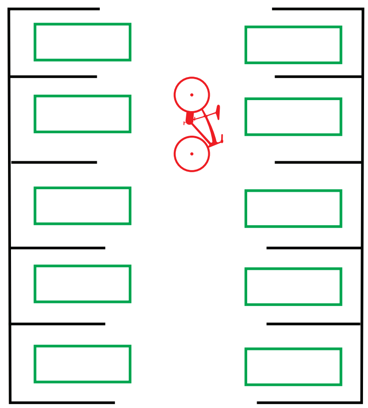


Figure C.5.1

Each author enjoys bicycling for pleasure and running errands in a flat town. Some of these trips across campus involve navigating through a parking lot. On each side of the route is a row of parked cars. (See Figure C.5.1.) At any moment a car can back into his path. Wanting to avoid a collision, one of the authors wonders where he should ride. The farther he rides from a row, the safer he is with respect to that row. However, the farther he rides from one row, the closer he is to the other row. Where should he ride?

Instinct tells him to ride midway between the two rows, an equal distance from both. But he has second thoughts. Maybe it is best to ride, say, one-third of the way from one row to the other, which is the same as two-thirds of the way from the other row. That would mean he has two safest routes, depending on which row he is nearer. Wanting a definite answer, he resorts to calculus.

He introduces a function, $f(x)$, which is the probability that he gets through safely when his distance from one row is x, disregarding cars in the other row. Then he let d represent the distance between the two rows. When he is at a distance x from one row, he is at a distance $d-x$ from the other row.

When he is very close to a row, x is small and $f(x)$ is near 0 because there is a high probability that a car backs out and hits the bicycle (or the bicycle hits the car). When he is far from a row, x is a little less than d and $f(x)$ is near 1 because he has a high chance of getting through safely.

The probability that he does not collide with a car backing out from either row is then the product, $f(x)f(d-x)$. His intuition says that this is maximized when $x = d/2$, putting him midway between the two rows.

What does he know about f? First of all, the farther he rides from one line of cars, the less likely he collides with a car backing out from that line; thus f is an increasing function. He also assumes f is differentiable, so $f'(x)$ is positive. Moreover, when he is very far from the cars, the probability of riding safely through the lot approaches 1. So he assumes $\lim_{x\to\infty} f(x) = 1$ (which it turned out he does not need).

Though $f(x)$ may be defined for all positive x it is of interest only for x in the interval $[0, d]$ because the distance between the two lines of cars is d.

The derivative of f' measures the rate at which he gains safety as he increases his distance from the cars. When x is small and he rides near the cars, $f'(x)$ is large: he gains a great deal of safety by increasing x. However, when he is far from the cars, he gains very little. That means that f' was a decreasing function. In other words, and assuming that f is twice differentiable, f'' is negative.

Does this information about f imply that midway is the safest route?

In other words, does the maximum of $f(x)f(d-x)$ occur when $x = d/2$? Symbolically, is

$$f\left(\frac{d}{2}\right) f\left(\frac{d}{2}\right) \geq f(x)f(d-x)?$$

He takes the logarithm of that expression, in order to replace a product by something easier, a sum. He wants to see if

$$2\ln\left(f\left(\frac{d}{2}\right)\right) \geq \ln(f(x)) + \ln(f(d-x)).$$

Denoting the composite function $\ln(f(x))$ as $g(x)$, he faces the inequality,

$$2g\left(\frac{d}{2}\right) \geq g(x) + g(d-x),$$

or

$$g\left(\frac{d}{2}\right) \geq \frac{1}{2}(g(x) + g(d-x)).$$

This inequality asserts that the point $\left(d/2, g(d/2)\right)$ on the graph of g is at least as high as the midpoint of the chord joining $(x, g(x))$ to $(d-x, g(d-x))$. This would be the case if the second derivative of g were negative, and the graph of g were concave down. He had to compute g'' and hope it was negative. Because $g(x) = \ln(f(x))$, $g'(x)$ is $f'(x)/f(x)$ and one more differentiation gives

$$g''(x) = \frac{f(x)f''(x) - (f'(x))^2}{f(x)^2}.$$

The denominator is positive. Because $f(x)$ is positive and $f''(x)$ is negative, the numerator is negative; so the quotient is negative. That means that the safest path is midway between the two rows. The bicyclist continues to follow that route but, after these calculations, with more confidence that it is indeed the safest way.

EXERCISES for CIE C.5

1. The reasoning of the bicyclist is expressed in the language of calculus. Express the assumptions informally, in everyday language, and show that midway between the two rows is the safest place to ride. Your reasoning should persuade someone who knows no mathematics.

Calculus is Everywhere # 6

Graphs in Economics

Elementary economics texts are full of graphs. They provide visual images of a variety of concepts, such as production, revenue, cost, supply, and demand. Here we show how economists use graphs to help analyze production as a function of the amount of labor, that is, the number of workers.

Assume that $P(L)$ is the amount of some product, such as cell phones, produced by a firm employing L workers. Since both workers and cell phones come in integer amounts, the graph of $P(L)$ is a bunch of dots. In practice, they suggest a curve, and economists use it in their analysis. So $P(L)$ is viewed as a differentiable function defined for some interval of the form $[0,b]$.

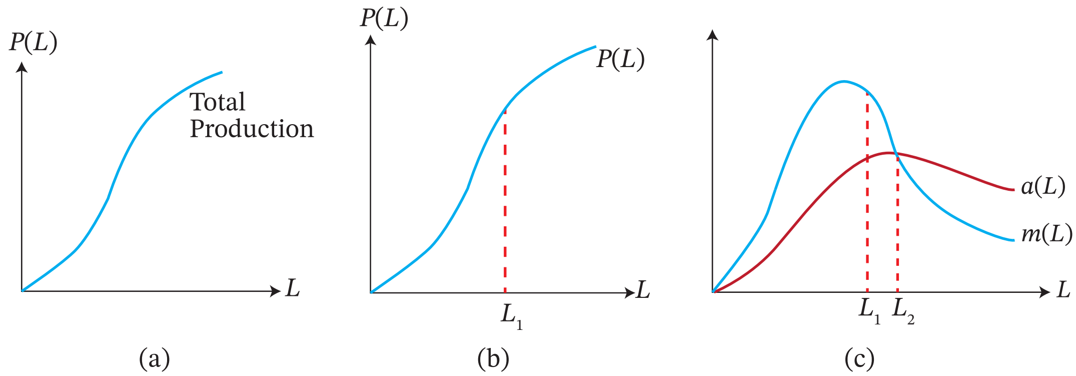


Figure C.6.1

If there are no workers, there is no production, so $P(0) = 0$. When the first few workers are added, production may increase rapidly, but as more are hired, production may still increase but not as rapidly. Figure C.6.1(a) is a typical **production curve**. It seems to have an inflection point when the gain from adding more workers begins to decline. The inflection point of $P(L)$ occurs at L_2 in Figure C.6.1(b).

When the firm employs L workers and adds one more, production increases by $P(L+1) - P(L)$, the marginal production. Economists relate this to the derivative by writing:

$$P(L+1) - P(L) = \frac{P(L+1) - P(L)}{(L+1) - L} \tag{C.6.1}$$

The right-hand side of (C.6.1) is change in output divided by change in input, which is, by the definition of the derivative, an approximation to the derivative, $P'(L)$. For this reason economists define the **marginal production** as $P'(L)$, and think of it as the extra production produced by the "L plus first" worker. We denote the marginal production as $m(L)$, that is, $m(L) = P'(L)$.

The **average production** per worker when there are L workers is defined as the quotient $P(L)/L$, which we denote $a(L)$. We have three functions: $P(L)$, $m(L) = P'(L)$, and $a(L) = P(L)/L$.

Now the fun begins.

QUESTION 1: At what point on the graph of the production function is the average production a maximum?

Since $a(L) = P(L)/L$, it is the slope of the line from the origin to the point $(L, P(L))$ on the graph. Therefore we are looking for the point on the graph where the slope is a maximum. One way to find it is to rotate a straightedge around the origin clockwise, starting at the vertical axis until it meets the graph, as in Figure C.6.1(b). Call the point of tangency $(L_1, P(L_1))$. For L less than L_1 or greater than L_1, average productivity is less than $a(L_1)$.

At L_1, the average production is the same as the marginal production, for the slope of the tangent at $(L_1, P(L_1))$ is both the quotient $P(L_1)/L_1$ and the derivative $P'(L_1)$.

We can use calculus to obtain the same conclusion: Since $a(L)$ has a maximum when the input is L_1, its derivative is 0 then. The derivative of $a(L)$ is

$$\frac{d}{dL}\left(\frac{P(L)}{L}\right) = \frac{LP'(L) - P(L)}{L^2}. \tag{C.6.2}$$

At L_1 the quotient in (C.6.2) is 0. Therefore, its numerator is 0, from which it follows that $P'(L_1) = P(L_1)/L_1$. (You might take a few minutes to see why this equation should hold, without using graphs or calculus.)

The graphs of $m(L)$ and $a(L)$ cross when L is L_1. For smaller values of L, the graph of $m(L)$ is above that of $a(L)$, and for larger values it is below, as shown in Figure C.6.1(c).

QUESTION 2: What does the maximum point on the marginal production graph tell about the production graph?

Assume that $m(L)$ has a maximum at L_2. For L smaller than L_2 the derivative of $m(L)$ is positive. For L larger than L_2 the derivative of $m(L)$ is negative. Since $m(L)$ is defined as $P'(L)$, the second derivative of $P(L)$ switches from positive to negative at L_2, showing that the production curve has an inflection point at $(L_2, P(L_2))$.

Economists use similar techniques to deal with a variety of concepts, such as marginal and average cost or marginal and average revenue, viewed as functions of labor or of capital.

Chapter 5

More Applications of Derivatives

Chapter 2 constructed the foundation for derivatives, namely the concept of a limit. Chapters 3 and 4 developed the derivative and applied it to graphs of functions. The present chapter will apply the derivative in a variety of ways, such as finding the most efficient method to accomplish a task (Section 5.1), connecting the rate one variable changes to the rate another changes (Sections 5.2 and 5.3), approximating functions such as e^x by polynomials (Sections 5.4 and 5.5), evaluating limits (Section 5.6), describing natural growth and decay (Section 5.7), and defining certain special functions (Section 5.8).

5.1 Applied Maximum and Minimum Problems

In Chapter 4 we saw how the first and second derivatives are of use in finding the maxima and minima of a function – the locally high and low points on its graph. Now we will use these techniques to find extrema in applied problems. Though the examples will be drawn mainly from geometry they all follow the same general four-step procedure.

Algorithm: Solving Applied Optimization Problems

Step 1: Develop a feel for the problem by experimenting with particular cases.

Step 2: Devise a formula for the function whose maximum or minimum you want to find.

Step 3: Determine the domain of the function – that is, the inputs that make sense in the application.

Step 4: Find the maximum or minimum of the function found in Step 2 for inputs that are in the domain identified in Step 3.

The main challenge in these situations is figuring out the formula for the function that describes the quantity to be maximized (or minimized). To become skillful at doing this takes practice.

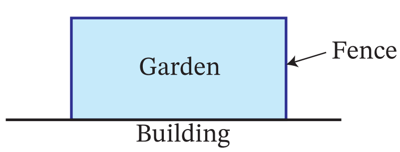


Figure 5.1.1

EXAMPLE 1. Enclosing a Large Garden. A couple have enough wire to construct 100 feet of fence. They wish to use it to form three sides of a rectangular garden, one side of which is along a building, as shown in Figure 5.1.1. What shape garden should they choose in order to enclose the largest possible area?

SOLUTION

Step 1: First make a few experiments. Figure 5.1.2 shows three ways of laying out the 100 feet of fence. In the first the side parallel to the building is very long, in an attempt to make a large area. However, doing this forces the other sides of the garden to be short. The area is $90 \times 5 = 450$ square feet. In the second, the garden has a larger area, $60 \times 20 = 1200$ square feet. In the third case, the side parallel to the building is only 20 feet long, but the other sides are longer. The area is $20 \times 40 = 800$ square feet.

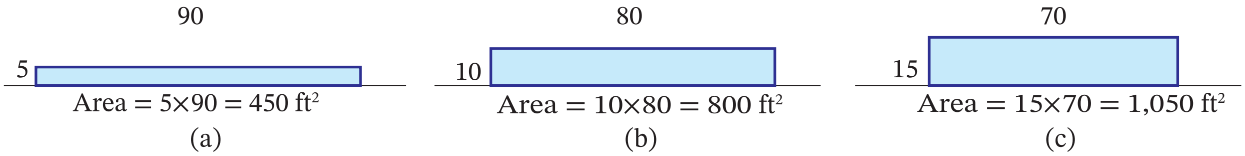


Figure 5.1.2

In all cases, once the length of the side parallel to the building is set, the other side lengths are known and the area can be computed. The area of the garden as a function of the length of the side parallel to the building.

Step 2: Let $A(x)$ be the area of the garden when the length of the side parallel to the building is x feet, as in Figure 5.1.3. The other sides of the garden have length y. Since the total length of the fence is 100 feet, y is determined by x:

$$x + 2y = 100.$$

Thus $y = (100 - x)/2$.

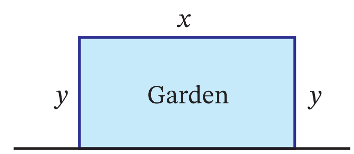


Figure 5.1.3

Since the area of a rectangle is its length times its width,

$$A(x) = xy = x\left(\frac{100 - x}{2}\right) = 50x - \frac{x^2}{2}. \tag{5.1.1}$$

(See Figure 5.1.4.) We now have the function to be optimized.

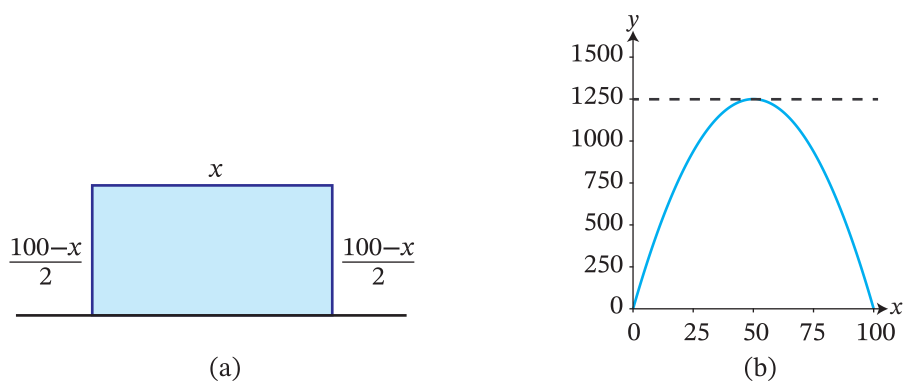


Figure 5.1.4

Step 3: Which values of x in (5.1.1) correspond to possible gardens?

Since there is only 100 feet of fence, $x \leq 100$. Furthermore, it makes no sense to have a negative amount of fence; so $x \geq 0$. Therefore the domain on which we wish to consider the function (5.1.1) is the closed interval $[0, 100]$.

Step 4: To maximize $A(x) = 50x - x^2/2$ on $[0, 100]$ we compare $A(0)$, $A(100)$, and the value of $A(x)$ at any critical numbers.

To find critical numbers, differentiate $A(x) = 50x - x^2/2$: $A'(x) = 50 - x$. Then solve $A'(x) = 0$ to find $0 = 50 - x$, from which we conclude $x = 50$.

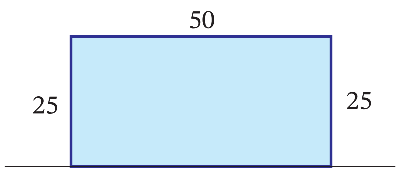


Figure 5.1.5

There is one critical number, $x = 50$. To find the largest of $A(0)$, $A(100)$, and $A(50)$, compute $A(0) = 50 \cdot 0 - 0^2/2 = 0$, $A(100) = 50 \cdot 100 - 100^2/2 = 0$, and $A(50) = 50 \cdot 50 - 50^2/2 = 1250$.

The maximum possible area is 1250 square feet; the fence should be laid out as shown in Figure 5.1.5. ♦

EXAMPLE 2. Designing A Tray With Maximum Volume. We start with a square piece of cardboard 12 inches on each side (see Figure 5.1.6(a)). Four congruent squares are cut out of the corners, as in Figure 5.1.6(b). The four flaps are folded up to obtain a tray without a top (see Figure 5.1.6 (c)). What size squares should be cut in order to maximize the volume of the tray?

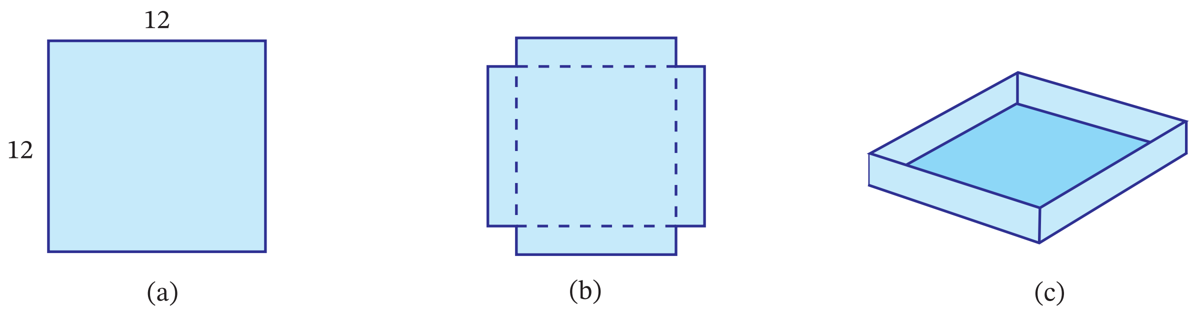


Figure 5.1.6

SOLUTION

Step 1: To get a feel for the problem we consider some special cases.

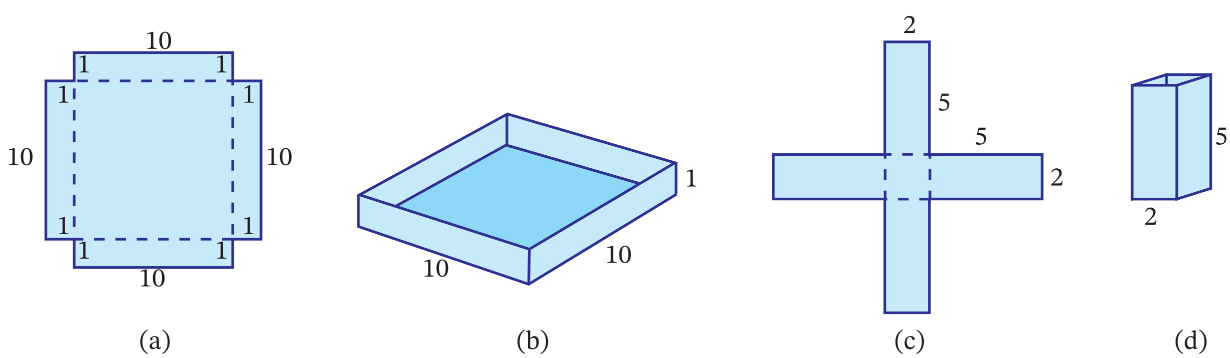


Figure 5.1.7

Say that we remove small squares that are 1 inch by 1 inch, as in Figure 5.1.7(a). When we fold up the flaps we obtain a tray whose base is a 10-inch by 10-inch square and whose height is 1 inch, as in Figure 5.1.7(b). The volume of the tray is

$$\text{Area of base} \times \text{height} = \underbrace{10 \times 10}_{\text{base area}} \times \underbrace{1}_{\text{height}} = 100 \text{ cubic inches.}$$

For our second experiment, cut out a large square from each corner, say 5 inches by 5 inches, as in Figure 5.1.7(c). When we fold up the flaps, we get a very tall tray with a very small base, as in Figure 5.1.7(d). Its volume is

$$\text{Area of base} \times \text{height} = 2 \times 2 \times 5 = 20 \text{ cubic inches.}$$

The volume depends on the size of the cut-out squares.

The function we will investigate is $V(x)$, the volume of the tray formed by removing four squares whose sides all have length x.

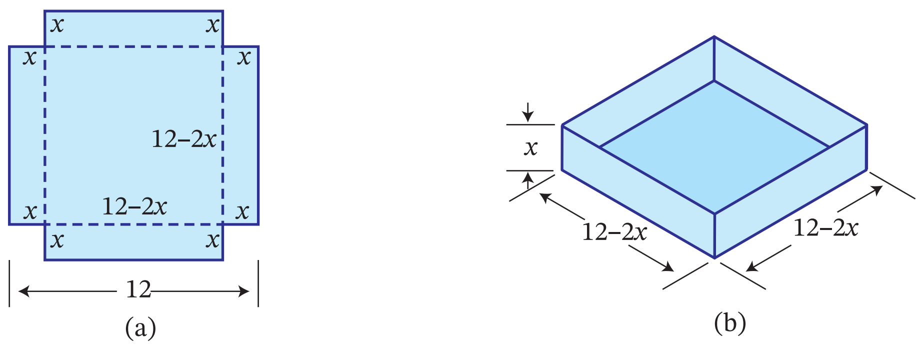


Figure 5.1.8

Step 2: To find the formula for $V(x)$ we make a large, clear diagram of the typical case, as in Figure 5.1.8(a) and Figure 5.1.8(b). We see that

$$\begin{aligned}\text{Volume of tray} &= \underbrace{(12-2x)}_{\text{length}}\underbrace{(12-2x)}_{\text{width}}\underbrace{x}_{\text{height}}\\ &= (12-2x)^2 x,\end{aligned}$$

hence

$$V(x) = (12-2x)^2 x = 4x^3 - 48x^2 + 144x.$$

We have obtained a formula for volume as a function of the length of the sides of the cut-out squares.

Step 3: Next determine the domain of the function $V(x)$ that is meaningful in the problem.

The smallest that x can be is 0. Then the tray has height 0 and is just a flat piece of cardboard, with volume 0. The size of the cut cannot be more than 6 inches, since the cardboard has sides of length 12 inches. The cut can be as near 6 inches as we please, and the nearer it is to 6 inches, the smaller is the base of the tray. For convenience, we allow cuts with $x = 6$, when the area of the base is 0 square inches and the height is 6 inches. The volume is again 0 cubic inches. Therefore the domain of the volume function $V(x)$ is the closed interval $[0,6]$.

Step 4: To maximize $V(x) = 4x^3 - 48x^2 + 144x$ on $[0,6]$, evaluate $V(x)$ at critical numbers in $[0,6]$ and at the endpoints of $[0,6]$. We have

$$V'(x) = 12x^2 - 96x + 144 = 12(x^2 - 8x + 12) = 12(x-2)(x-6).$$

A critical number is a solution to the equation

$$0 = 12(x-2)(x-6).$$

Hence $x - 2 = 0$ or $x - 6 = 0$. The critical numbers are 2 and 6.

The endpoints of the interval $[0,6]$ are 0 and 6. Therefore the maximum value of $V(x)$ for x in $[0,6]$ is the largest of $V(0)$, $V(2)$, and $V(6)$. Since $V(0) = 0$ and $V(6) = 0$, the largest value is

$$V(2) = 4(2^3) - 48(2^2) + 144 \cdot 2 = 128 \text{ cubic inches.}$$

The cut that produces the tray with the largest volume is $x = 2$ inches. ♦

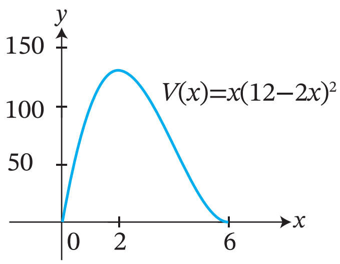


Figure 5.1.9

Figure 5.1.9 shows the graph of the volume, $V(x)$, as a function of the side length, x, of the squares removed from each corner. At $x = 2$ and $x = 6$ the tangent line is horizontal. In Example 2 one might say $x = 0$ and $x = 6$ do not correspond to what would be called a tray. That would restrict the domain of $V(x)$ to the open interval $(0,6)$. It would be necessary to examine the behavior of $V(x)$ for x near 0 and for x near 6. Making the domain $[0,6]$ from the start avoids the extra work of examining $V(x)$ for x near the ends of the interval.

We pause now to re-emphasize that the key step in these two examples, and in any applied problem, is Step 2: finding a formula for the quantity whose extremum is to be found. If the problem is geometrical, the following chart may help.

Algorithm: Setting Up the Objective Function for a Max/Min Problem

Step 1: Draw and label the appropriate diagrams.
(Make them large enough so that there is room for labels.)

Step 2: Label the various quantities by letters, such as x, y, A, V.

Step 3: Identify the quantity to be maximized (or minimized).

Step 4: Express the quantity to be maximized (or minimized) in terms of one or more of the other variables.

Step 5: Finally, express the objective function in terms of only one variable.

EXAMPLE 3. Designing A Tin Can With Maximum Volume. Of all the tin cans that enclose a volume of 100π cubic centimeters, which requires the least metal?

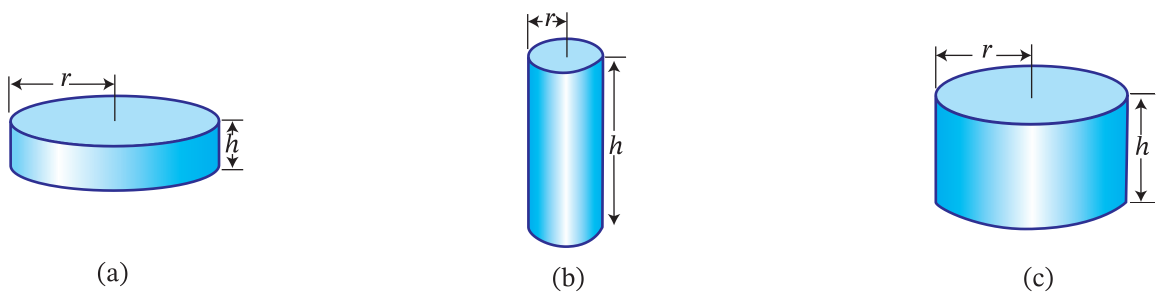


Figure 5.1.10

SOLUTION

Step 1: The can may be flat or tall. If it is flat, as in Figure 5.1.10(a), the side uses little metal, but then the top and bottom bases are large. If it is shaped like a mailing tube, as in Figure 5.1.10(b), then the two bases require little metal, but the curved side requires a great deal of metal. A third can is shown in Figure 5.1.10(c). In each case r denotes the radius and h is the height of the can. What is the ideal compromise between these extremes?

Step 2: The surface area S of the can is the sum of the area of the top, side, and bottom. The top and bottom are disks with radius r. Their total area is $2\pi r^2$. Figure 5.1.11 shows why the area of the side is $2\pi rh$. The total surface area of the can is

$$S = 2\pi r^2 + 2\pi rh. \tag{5.1.2}$$

Since the amount of metal in the can is proportional to S, it suffices to minimize S.

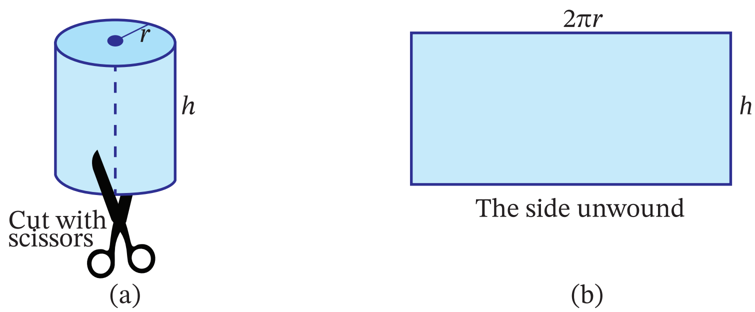


Figure 5.1.11

Equation (5.1.2) gives S as a function of two variables, but we can express one in terms of the other. The radius and height are related by

$$V = \pi r^2 h = 100\pi \tag{5.1.3}$$

since the volume is 100π cubic centimeters. To express S as a function of one variable, use (5.1.3) to eliminate either r or h. If we choose to eliminate h, we solve (5.1.3) for h,

$$h = \frac{100}{r^2}.$$

Substitution into (5.1.2) yields

$$S = 2\pi r^2 + 2\pi r \frac{100}{r^2} \qquad \text{or} \qquad S = 2\pi r^2 + \frac{200}{r}\pi. \tag{5.1.4}$$

Equation (5.1.4) expresses S as a function of just one variable, r.

There is no upper limit on the radius.

Step 3: The function $S(r)$ is continuous and differentiable on $(0, \infty)$.

Step 4: Compute dS/dr:

$$\frac{dS}{dr} = 4\pi r - \frac{200\pi}{r^2} = \frac{4\pi r^3 - 200\pi}{r^2}. \tag{5.1.5}$$

Set the derivative equal to 0 to find critical numbers. We have

$$0 = \frac{4\pi r^3 - 200\pi}{r^2},$$

hence

$$0 = 4\pi r^3 - 200\pi$$

or, solving for r, $4\pi r^3 = 200\pi$, so $r^3 = 200/4 = 50$, and, finally, $r = \sqrt[3]{50} \approx 3.684$.

When r is near 0 or very large, S is large. Thus there will be no maximum, but there will be a minimum. Because there is only one critical number a minimum must occur there. But we will check that this is the case in two ways: first by the first-derivative test, then by the second-derivative test.

The first derivative is

$$\frac{dS}{dr} = \frac{4\pi r^3 - 200\pi}{r^2}. \tag{5.1.6}$$

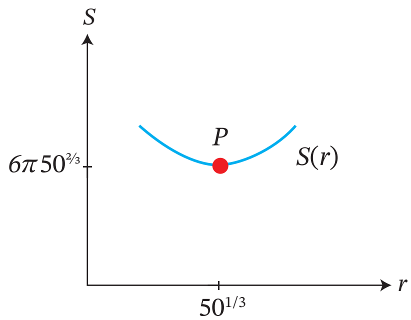


Figure 5.1.12

When $r = \sqrt[3]{50}$, the numerator in (5.1.6) is 0. When $r < \sqrt[3]{50}$ the numerator is negative and when $r > \sqrt[3]{50}$ it is positive. (The denominator is always positive.) Since $dS/dr < 0$ for $r < \sqrt[3]{50}$ and $dS/dr > 0$ for $r > \sqrt[3]{50}$, $S(r)$ decreases for $r < \sqrt[3]{50}$ and increases for $r > \sqrt[3]{50}$. That shows that a local minimum occurs at $\sqrt[3]{50}$.

This local minimum is also the global minimum. While Figure 5.1.12 provides visual confirmation of this fact, it is also easily verified this by the second-derivative test. Differentiation of (5.1.5) gives

$$\frac{d^2S}{dr^2} = 4\pi + \frac{400}{r^3}\pi. \tag{5.1.7}$$

which shows that for all positive values of r, d^2S/dr^2 is positive. We found not only a relative minimum but a global minimum, Observe also that the graph of $S(r)$ lies above its tangents, in particular, the horizontal tangent at the location of the critical point: $r = \sqrt[3]{50/\pi}$.

To find the height of the most economical can, solve (5.1.5) for h

$$\begin{aligned} h = \frac{100}{r^2} &= \frac{100}{(\sqrt[3]{50})^2} \\ &= \frac{100}{(\sqrt[3]{50})^2}\frac{\sqrt[3]{50}}{\sqrt[3]{50}} \quad \text{(rationalize the denominator)} \\ &= \frac{100}{50}\sqrt[3]{50} \\ &= 2\sqrt[3]{50}. \end{aligned}$$

The height of the can is equal to twice its radius, that is, its diameter. The total surface area is

$$S = 2\pi\left(50^{1/3}\right)^2 + \frac{200\pi}{50^{1/3}} = (2\pi + 4\pi)50^{2/3} = 6\pi 50^{2/3} \approx 255.83 \text{ square centimeters.}$$

◆

Summary

We showed how to use calculus to solve applied problems. Experiment, set up a function, find its domain and its critical points. Then test the critical points and endpoints of the domain to determine the extrema.

If the domain is a closed interval, the maximum or minimum will occur at a critical point or an endpoint. If the interval is not closed, more care is needed to confirm that a critical number provides an extremum.

EXERCISES for Section 5.1

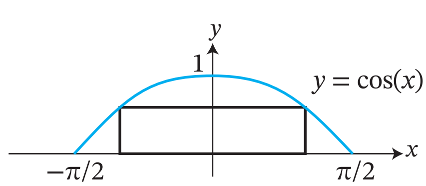


Figure 5.1.13

1. A gardener wants to make a rectangular garden with 100 feet of fence. What is the largest area the fence can enclose?

2. Of all rectangles with area 100 square feet, find the one with the shortest perimeter.

3. Solve Example 1, expressing A in terms of y instead of x.

4. Solve Example 3, expressing S in terms of h instead of r.

5. A gardener is going to put a rectangular garden inside one arch of the cosine curve, as shown in Figure 5.1.13. What is the garden with the largest area?

Exercises 6 to 9 are related to Example 2. In each find the length of the cut that maximizes the volume of the tray. The dimensions of the cardboard are given.

6. 5 inches by 5 inches **7.** 7 inches by 15 inches **8.** 5 inches by 8 inches, **9.** 3 inches by 8 inches,

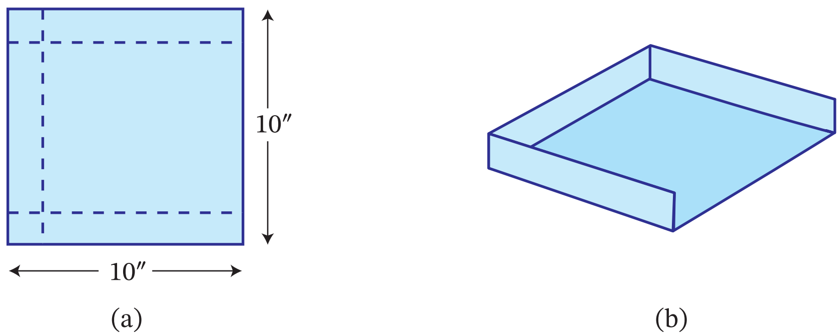


Figure 5.1.14

10. Starting with a square piece of paper 10″ on a side, Sam wants to make a paper holder with three sides. The pattern he will use is shown in Figure 5.1.14(a) along with the tray. He will remove two squares and fold up three flaps (as shown in Figure 5.1.14(b)).

(a) What size of the removed squares maximizes the volume of the tray? (b) What is that volume?

11. Of all cylindrical tin cans without a top that contain 100 cubic inches, which requires the least material?

12. Of all enclosed rectangular boxes with square bases that have a volume of 1000 cubic inches, which uses the least material?

13. Of all topless rectangular boxes with square bases that have a volume of 1000 cubic inches, which uses the least material?

14. Find the dimensions of the rectangle of largest area that can be inscribed in a circle of radius a. Start by expressing the area in terms of the angle θ.

The typical rectangle is shown in Figure 5.1.15.

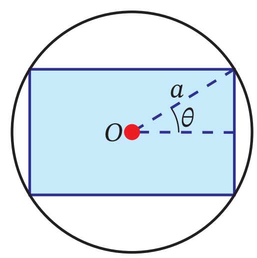


Figure 5.1.15

15. Repeat Exercise 14, this time expressing the area in terms of half the width of the rectangle, x.

16. Find the dimensions of the rectangle of largest perimeter that can be inscribed in a circle of radius a.

17. A chef wants to make a cake pan out of a circular piece of aluminum of radius 12 inches. To do this he plans to cut the circular base from the center of the piece and then cut the side from the remaining ring. What should the radius and height be to maximize the volume of the pan?

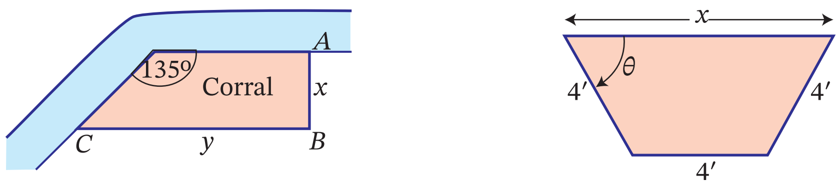
18. Show that of all rectangles of a given area, the square has the shortest perimeter.

Call the fixed area A and keep in mind that this area is a constant.

19. A rancher wants to construct a rectangular corral. He also wants to divide it by a fence parallel to one of the sides. He has 240 feet of fence. What are the dimensions of the corral of largest area he can enclose?

20. A river has a 45° turn, as indicated in Figure 5.1.16(a). A rancher wants to construct a corral bound on two sides by the river and on two sides by 1 mile of fence ABC. Find the dimensions of the corral of largest area.

21. An irrigation channel is to have a cross section in the form of an isosceles trapezoid, three of whose sides are 4 feet long. Find the length of the fourth side that maximizes the area of the trapezoid. (See Figure 5.1.16(b).)



(a) (b)

Figure 5.1.16

22. (a) Repeat Exercise 21, expressing the area as a function of θ instead of x. (See Figure 5.1.16(b).)
(b) Do the answers in (a) and Exercise 21 agree? Explain.

23. (a) How should one choose two nonnegative numbers whose sum is 1 in order to maximize the sum of their squares?
(b) To minimize the sum of their squares?

24. How should one choose two nonnegative numbers whose sum is 1 in order to maximize the product of the square of one of them and the cube of the other?

In Exercises 25 to 28 use the fact that the combined length and girth (distance around) of a package to be sent through the mail by the United States Postal Service (USPS) cannot exceed 108 inches.

NOTE: The length is always the longest dimension of a package, and height and width are the lengths of the sides perpendicular to the length. Girth is the distance around the thickest part of the remaining non-length sides. For rectangular packages, the girth is twice the sum of the height and width and, for cylindrical packages, the girth is the circumference of the cylinder.

The combined length and girth of a package sent by the US Postal Service (USPS) cannot exceed 130 inches. The United Parcel Service (UPS) limit is 165 inches for combined length and girth with the length not exceeding 108 inches.

25. Find the dimensions of the right circular cylinder of largest volume that USPS will accept.

26. Find the dimensions of the right circular cylinder of largest surface area that the USPS will accept.

27. Find the dimensions of the rectangular box with square base of largest volume that the USPS will accept.

28. Find the dimensions of the rectangular box with square base of largest surface area that the USPS will accept.

---

29. (a) Repeat Exercise 25 for a package sent by UPS.
(b) Generalize your solutions to Exercise 25 for a package subject to a combined length and girth that does not exceed M inches.

30. (a) Repeat Exercise 26 for a package sent by UPS.
(b) Generalize your solutions to Exercise 26 for a package subject to a combined length and girth that does not exceed M inches.

31. A cylindrical can is to be made to hold 100 cubic inches. The material for its top and bottom costs twice as much per square inch as the material for its side. Find the radius and height of the most economical can. (This is not the same as Example 3.)

(a) Would you expect the most economical can in this problem to be taller or shorter than the solution to Example 3? Use common sense, not calculus.

(b) Call the cost of 1 square inch of the material for the side k cents. Thus the cost of 1 square inch of the material for the top and bottom is $2k$ cents. (The value of k will not affect the answer.) Show that a can of radius r and height h costs $C = 4k\pi r^2 + 2k\pi rh$ cents.

(c) Find r that minimizes the function C in (b). Keep in mind that k is constant.

(d) Find the corresponding h.

32. A camper at A will walk to the river in Figure 5.1.17, put some water in a pail at P, and take it to the campsite at B.

(a) Express $AP + PB$ as a function of x.

(b) Use calculus to decide where P should be located to minimize the length of the walk, $AP + PB$?

NOTE: A similar exercise was first encountered as Exercise 34 in Section 1.1, where it was solved by geometry.

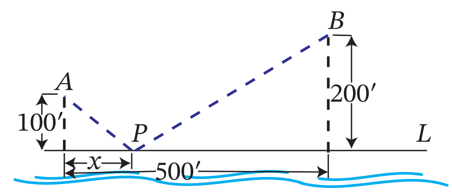


Figure 5.1.17

33. Sam is at the edge of a circular lake of radius one mile and Jane is at the edge, directly opposite. Sam wants to visit Jane. He can walk 3 miles per hour (mph) and he has a canoe. What mix of paddling and walking should Sam use to minimize the time needed to reach Jane (a) if he paddles at 1.5 mph?, (b) if he paddles at 2 mph?, and (c) if he paddles at least 3 mph?.

34. Let $\triangle ABC$ be a right triangle with C being at the right angle. There are two routes from A to B. One is direct, along the hypotenuse. The other is along the two legs, from A to C and then to B. What is the largest percentage saving possible by walking along the hypotenuse instead of along the two legs? For which shape right triangle does the maximum occur?

35. A rectangular box with a square base is to hold 100 cubic inches. Material for the top of the box costs 2 cents per square inch, material for the sides costs 3 cents per square inch, and material for the bottom costs 5 cents per square inch. Find the dimensions of the most economical box.

36. The cost of operating a truck (for gasoline, oil, and depreciation) is $(20 + s/2)$ cents per mile when it travels at a speed of s miles per hour. A truck driver earns \$18 per hour. What is the most economical speed at which to operate the truck during a 600-mile trip?

(a) If you considered only the truck, would you want s to be small or large?

(b) If you, the employer, considered only the expense of the driver's wages, would you want s to be small or large?

(c) Express cost as a function of s and solve. (Be sure to put all costs in terms of cents or in terms of dollars.)

(d) Would the answer be different for a 1000-mile trip?

37. A government contractor who is removing earth from a large excavation can route trucks over either of two roads. There are 10,000 cubic yards of earth to move. Each truck holds 10 cubic yards. On one road the cost per truckload is $1 + 2x^2$ cents, when x trucks use that road. (The function records the cost of congestion.) On the other road the cost is $2 + x^2$ cents per truckload when x trucks use it. How many trucks should be dispatched to each of the two roads?

38. A text on the dynamics of airplanes states "Recalling that $I = A\cos^2(\theta) + C\sin^2(\theta) - 2E\cos(\theta)\sin(\theta)$, we wish to find θ when I is a maximum or a minimum." Show that at an extremum of I, $\tan 2\theta = \frac{2E}{C-A}$. (Assume $A \neq C$.)

REFERENCE: *Dynamics of Airplanes and Airplane Structures*, J.E. Younger and B.M. Woods, Wiley, 1931.

39. A physics text states "By differentiating the equation for the horizontal range, $R = \frac{v_0^2 \sin(2\theta)}{g}$, show that the initial elevation angle θ for maximum range is 45°." In the formula for R, v_0 and g are constants.

(a) Using calculus, show that the maximum range occurs when $\theta = 45°$. (b) Solve the same problem without calculus.

REFERENCE: *University Physics*, A. Hudson and R. Nelson, Harcourt, 1982.

NOTE: R is the horizontal distance traveled by a projectile thrown at an angle θ with speed v_0, disregarding air resistance.

40. On one side of a river 1 mile wide is an electric power station and on the other side, s miles upstream, is a factory. (See Figure 5.1.18.) It costs 3 dollars per foot to run cable over land and 5 dollars per foot under water. What is the most economical way to run cable from the station to the factory?

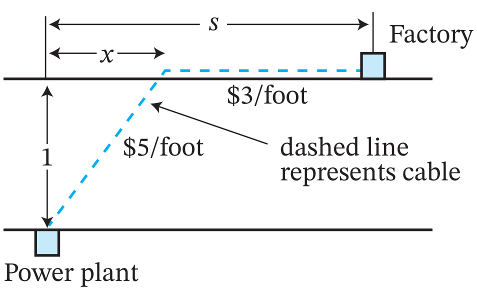


Figure 5.1.18

(a) Using no calculus, what do you think would be (approximately) the best route if s were very small? if s were very large?
(b) Use calculus to find the optimal path when $s = \frac{1}{2}$.
(c) Repeat (b) for $s = \frac{3}{4}$, 1, and 2.
(d) Draw the optimal paths for the cases considered in parts (b) and (c).
(e) Find the optimal path for an arbitrary s.

NOTE: Minimizing the length of cable is not the same as minimizing its cost.

41. A gardener has 10 feet of fence and wishes to make a triangular garden next to two buildings, as in Figure 5.1.19(a). How should he place the fence to enclose the maximum area?

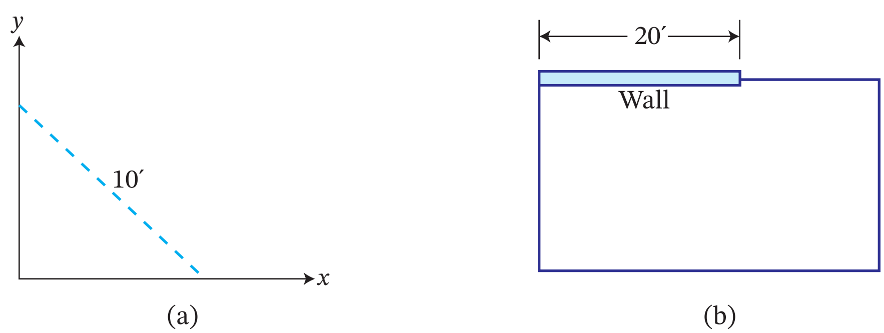


Figure 5.1.19

42. Fencing is to be added to an existing wall of length 20 feet, as shown in Figure 5.1.19(b). How should the extra fence be added to maximize the area of the enclosed rectangle if the additional fence

(a) is 40 feet long? (b) is 80 feet long? (c) is 60 feet long?

43. Let A and B be constants. Find the maximum and minimum values of $A\cos(t) + B\sin(t)$.

44. A spider at corner S of a cube of side 1 inch wishes to capture a fly at the opposite corner F. The spider, who must walk on the surface of the solid cube, wishes to find the shortest path. (See Figure 5.1.20(a).)
(a) Find a shortest path without the aid of calculus. (b) Find a shortest path with calculus.

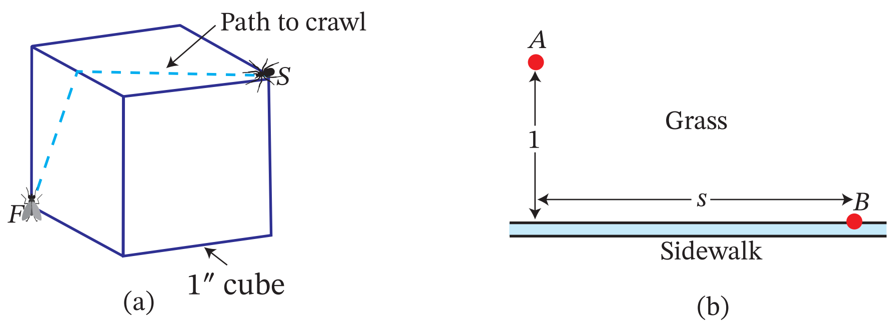


Figure 5.1.20

45. A woman walks 3 miles per hour on grass and 5 miles per hour on sidewalk. What route should she follow to walk from point A to point B, shown in Figure 5.1.20(b), in the least time (a) if $s = \frac{1}{2}$? (b) if $s = \frac{3}{4}$? (c) if $s = 1$?

46. A ladder of length b leans against a wall of height a, $a < b$. What is the maximum horizontal distance that the ladder can extend beyond the wall if its base rests on the horizontal ground?

47. The potential energy, U, in a diatomic molecule is given by the formula $U(r) = U_0\left(\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right)$, where U_0 and r_0 are constants and r is the distance between the atoms. For which value of r is $U(r)$ a minimum?

48. Find the dimensions of the right circular cylinder with largest volume that can be inscribed in a sphere of radius a?

49. The stiffness of a rectangular beam is proportional to the product of the width and the cube of the height of its cross section. What shape beam should be cut from a log in the form of a right circular cylinder of radius r in order to maximize its stiffness?

50. A rectangular box-shaped house is to have a square floor. Three times as much heat per square foot enters through the roof as through the walls. What shape should the house be if it is to enclose a volume of 12,000 cubic feet and minimize heat entry? (Assume no heat enters through the floor.)

51. (See Figure 5.1.21(a).) Find the coordinates of the points $P = (x, y)$, with $y \le 1$, on the parabola $y = x^2$, that
(a) minimize $|PA|^2 + |PB|^2$ and (b) maximize $|PA|^2 + |PB|^2$.

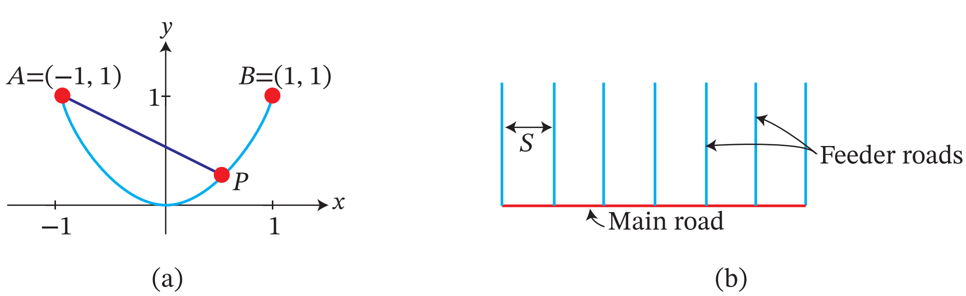


Figure 5.1.21

52. The speed of traffic through the Lincoln Tunnel in New York City depends on the amount of traffic. Let S be the speed in miles per hours at which the traffic moves and let D be the amount of traffic measured in vehicles per mile. The relation between S and D is approximated closely, for $D \le 100$, by $S = 42 - \frac{D}{3}$. (a) Express the total number of vehicles entering the tunnel per hour in terms of D. (b) What value of D maximizes the flow in (a)?

53. When a tract of timber is to be logged, a main logging road is built from which small roads branch off as feeders. If too many feeders are built, the cost of construction would be prohibitive. If too few feeders are built, the time spent moving the logs to the roads would be prohibitive. The formula for total cost, $y = \frac{CS}{4} + \frac{R}{VS}$, is used in a logger's manual to find how many feeder roads are to be built. R, C, and V are constants: R is the cost of road at unit spacing, C is the cost of moving a log a unit distance, and V is the value of timber per acre. S denotes the distance between the regularly spaced feeder roads. (See Figure 5.1.21(b).) Thus the cost y is a function of S, and the object is to find the value of S that minimizes y. The manual says, "To find the desired S set the two summands equal to each other and solve $\frac{CS}{4} = \frac{r}{VS}$." Show that the method is valid.

54. A delivery service is deciding how many warehouses to build in a large city. The warehouses will serve similarly shaped regions of equal area A and, let us assume, an equal number of people.

(a) Why would transportation costs per item presumably be proportional to $\sqrt{A}$?

(b) Assuming that the warehouse cost per item is inversely proportional to A, show that C, the cost of transportation and storage per item, is of the form $t\sqrt{A} + \frac{w}{A}$, where t and w are constants.

(c) Show that C is a minimum when $A = \left(\frac{2w}{t}\right)^{2/3}$.

Exercises 55 and 56 are related.

55. A pipe of length b is carried down a long corridor of width $a < b$ and then around corner C. (See Figure 5.1.22.) During the turn y starts at 0, reaches a maximum, and then returns to 0. (Try this with a short stick.) Find the maximum in terms of a and b. (Express y in terms of a, b, and θ; θ is a variable, while a and b are constants.)

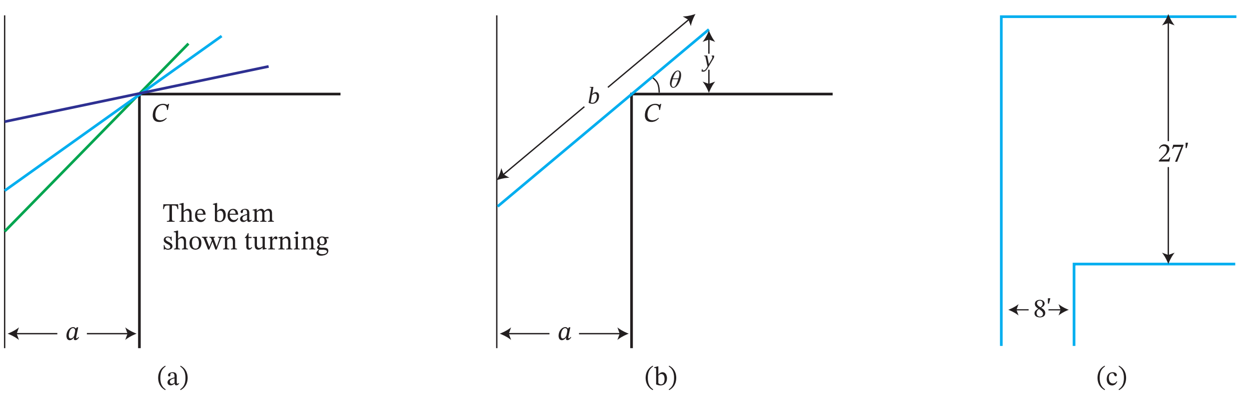


Figure 5.1.22

56. (Do Exercise 55 first.) Figure 5.1.22(c) shows two corridors meeting at right angle. One has width 8 feet and the other, width 27 feet. Find the length of the longest pipe that can be carried horizontally from one hall, around the corner and into the other hall.

---

57. The base of a painting on a wall is a feet above the eye of an observer, as shown in Figure 5.1.23(a). The vertical side of the painting is b feet long. How far from the wall should the observer stand to maximize the angle that the painting subtends?

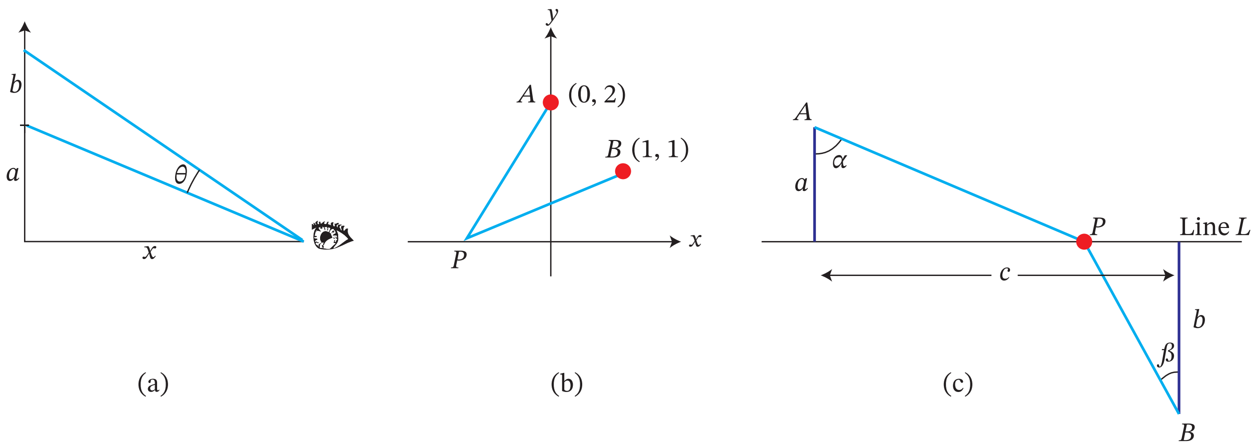


Figure 5.1.23

58. Find the point P on the x-axis such that maximizes angle APB in Figure 5.1.23(b). (See the hint in Exercise 57.)

59. *(Economics)* Let p denote the price in dollars of a commodity and y the number of units sold at that price. Assume that $y = 250 - p$ for $0 \le p \le 250$ and that it costs the producer $100 + 10y$ dollars to manufacture y units. What price p should the producer choose in order to maximize total profit, that is, revenue minus cost?

60. *(Leibniz on light)* A ray of light travels from point A to point B in Figure 5.1.23(c) in minimal time. The point A is in one medium, such as air or a vacuum. The point B is in another medium, such as water or glass. In the first medium, light travels at velocity v_1 and in the second at velocity v_2. The media are separated by line L. Show that for the path APB of minimal time, $\dfrac{\sin\alpha}{v_1} = \dfrac{\sin(\beta)}{v_2}$. Leibniz solved this problem with calculus in a paper published in 1684. (The result is called **Snell's law of refraction**.)

Leibniz wrote, "Other very learned men have sought in many devious ways what someone versed in this calculus can accomplish in these lines as by magic."

REFERENCE: C. H. Edwards Jr., *The Historical Development of the Calculus*, p. 259, Springer-Verlag, 1979.

Exercises 61 to 64 concern the intensity of light.

61. Why is it reasonable to assume that the intensity of light from a lamp is inversely proportional to the square of the distance from the lamp? (Imagine the light spreading out in all directions.)

62. A solar panel perpendicular to the sun's rays catches more light than when it is tilted at any other angle, as shown in Figure 5.1.24(a). Let θ be the angle the panel is tilted, as in Figure 5.1.24(b). Show that it then receives $\cos(\theta)$ times the light the panel would receive when perpendicular to the sun's rays.

63. In view of the preceding two exercises, the intensity of light on a small (flat) surface is inversely proportional to the square of the distance from the source and proportional to the angle between the surface and a surface perpendicular to the source.

(a) A person wants to put a light at a horizontal distance of ten feet from his address, which is on a wall (a vertical surface). At what height should the lamp be placed to maximize the intensity of light at the address?

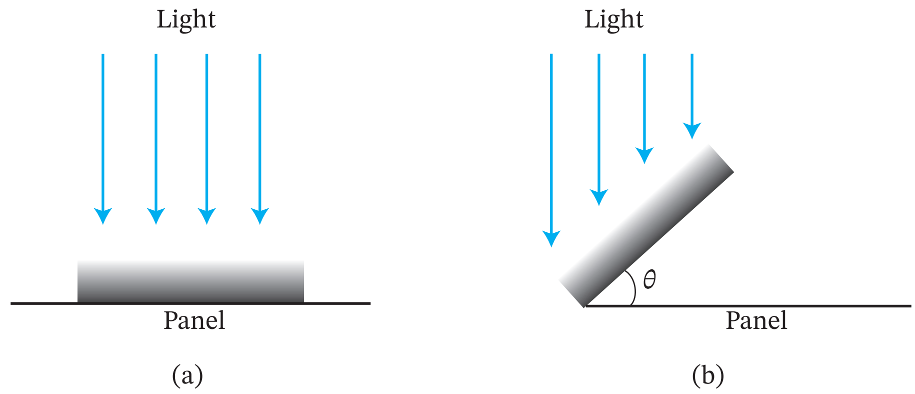


Figure 5.1.24

(b) Suppose the address is painted on the horizontal surface of the curb. The lamp will be placed at a horizontal distance of ten feet from the address. Without doing any calculations sketch what the graph of "intensity of light on the address versus height of lamp" might look like.

(c) Find the height the lamp should have to maximize the light on the address. (Use height as the independent variable.) $10/sqrt2$ feet

64. Solve Exercise 63(c) using an angle as the independent variable.

---

65. The following calculation occurs in an article concerning the optimum size of new cities: "The net utility to the total client-centered system is $U = \frac{RLv}{A} n^{1/2} - nK - \frac{ALc}{v} n^{-1/2}$. All symbols except U and n are constant; n is a measure of decentralization. Regarding U as a differentiable function of n, we can determine when $dU/dn = 0$. This occurs when $\frac{RLv}{2A} n^{-1/2} - K + \frac{ALc}{2v} n^{-3/2} = 0$. This is a cubic equation for $n^{-1/2}$."

(a) Check that the differentiation is correct. (b) Of what cubic polynomial is $n^{-1/2}$ a root?

66. A tangent to the curve $y = x^{-2}$ at a point in the first quadrant, together with the x- and y-axes, determine a triangle. (a) What is the largest area of such a triangle? (b) What is the smallest area of such a triangle?

67. Let f be a differentiable function that is never zero on its domain. Let $g(x) = (f(x))^2$. Show that the functions f and g have the same critical numbers. This is useful for avoiding square roots.

68. In which of these cases must the continuous function $f(x)$ have a global minimum? In each case explain your answer. (a) Domain: $[0,1]$ (b) Domain: $(0,1)$ (c) Domain: $(0,\infty)$ (d) Domain: $(0,\infty)$ with $\lim_{x\to 0} f(x) = \lim_{x\to\infty} f(x) = \infty$

69. The differentiable function $f(x)$ has domain $(0,\infty)$ with $\lim_{x\to 0} f(x) = \infty$ and $\lim_{x\to\infty} f(x) = \infty$. The only number where $f'(x) = 0$ is $x = a$. (a) Must $f(x)$ have a relative minimum at a? (b) Must $f(x)$ have a global minimum at a?

5.2 Implicit Differentiation

Sometimes a function $y = f(x)$ is given indirectly by an equation that links y and x. This section shows how to differentiate the function without solving for it in terms of x.

A Function Given Implicitly

The equation

$$x^2 + y^2 = 25 \tag{5.2.1}$$

describes a circle of radius 5 and center at the origin, as in Figure 5.2.1(a).

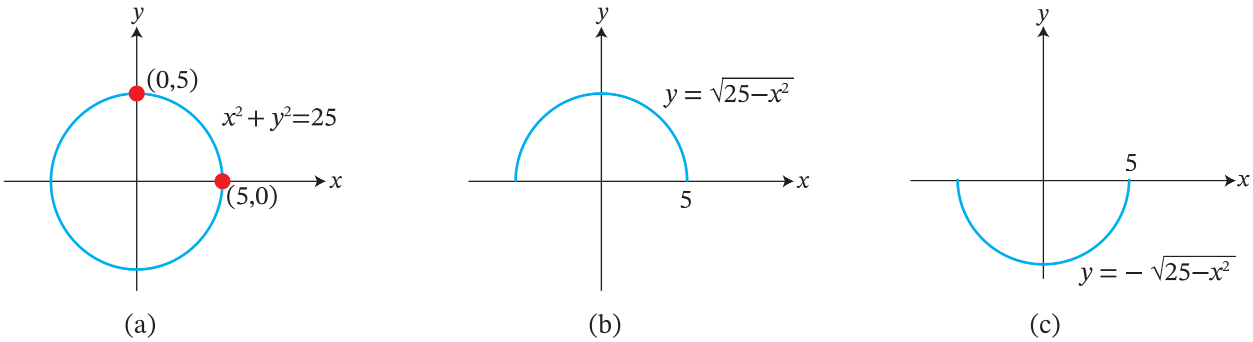


Figure 5.2.1

This circle is not the graph of a function since some vertical lines meet the circle in more than one point. However, the top half is the graph of a function and so is the bottom half. To find these functions explicitly, solve (5.2.1) for y:

$$y^2 = 25 - x^2 \qquad \text{so} \qquad y = \pm\sqrt{25 - x^2}.$$

The graph of $y = \sqrt{25 - x^2}$ is the top semicircle (see Figure 5.2.1(b)) and the graph of $y = -\sqrt{25 - x^2}$ is the bottom semicircle (see Figure 5.2.1(c)). There are two continuous functions that satisfy (5.2.1).

The equation $x^2 + y^2 = 25$ is said to describe the function $y = f(x)$ **implicitly**. The equations

$$y = \sqrt{25 - x^2} \qquad \text{and} \qquad y = -\sqrt{25 - x^2}$$

describe the function $y = f(x)$ **explicitly**. Separately, each explicit equation describes only one part (branch) of the graph. The implicit definition includes both parts of the curve.

Differentiating an Implicitly-Defined Function

It is possible to differentiate a function given implicitly without having to express it explicitly. An example will illustrate the method, which is referred to as **implicit differentiation**. The basic idea is to differentiate both sides of the equation that defines the function implicitly, remembering that some variables depend (implicitly) on other variables.

EXAMPLE 1. Let $y = f(x)$ be the continuous function that satisfies the equation $x^2 + y^2 = 25$ such that $y = -4$ when $x = 3$. Find $\frac{dy}{dx}$ when $x = 3$ and $y = -4$.

SOLUTION In this case we can solve for y explicitly, $y = \sqrt{25 - x^2}$ or $y = -\sqrt{25 - x^2}$. Because y equals -4 when x is 3, we are involved with $y = -\sqrt{25 - x^2}$, not $\sqrt{25 - x^2}$. We could then find the derivative by direct differentiation. However, the square roots complicate the algebra.

Instead, we differentiate both sides of the implicit equation (5.2.1) with respect to x:

$$\frac{d}{dx}(x^2 + y^2) = \frac{d}{dx}(25) \qquad (\text{ derivative of (5.2.1) with respect to } x\text{ })$$

$$2x + \frac{d(y^2)}{dx} = 0 \qquad (\text{ } y \text{ depends on } x\text{ }).$$

To differentiate y^2 with respect to x, write $w = y^2$, where y is a function of x. By the chain rule,

$$\frac{d(y^2)}{dx} = \frac{dw}{dx} = \frac{dw}{dy}\frac{dy}{dx} = 2y\frac{dy}{dx}.$$

Thus, the differentiated implicit equation is:

$$2x + 2y\frac{dy}{dx} = 0,$$

or

$$x + y\frac{dy}{dx} = 0.$$

In particular, when $x = 3$ and $y = -4$ are substituted into the previous equation,

$$3 + (-4)\frac{dy}{dx} = 0.$$

Now it is straightforward to solve for dy/dx (when $x = 3$ and $y = -4$):

$$\frac{dy}{dx} = \frac{3}{4}.$$

This answers the original question. ♦

Observation 5.2.1: *Working with Implicit Functions has its Benefits*

It is worthwhile to notice that at no point in Example 1 did we have to solve for y as a explicit function of x or worry about choosing the right sign of a square root.

Observation 5.2.2: *Implicit Differentiation was Used in Section 3.5*

We used implicit differentiation in Section 3.5 when finding the derivatives of $\ln(x)$, $\arcsin(x)$, and $\arctan(x)$. For instance, $y(x) = \ln(x)$ was defined implicitly by $e^{y(x)} = x$. We differentiated both sides of the equation with respect to x to find $y'(x)$.

Now, we see that implicit differentiation is useful in other situations where it is difficult or otherwise inconvenient to express one vaeriable explicitly in terms of one or more other variables.

In the next example implicit differentiation is the only way to find the derivative, for in this case there is no explicit formula expressible in terms of trigonometric and algebraic functions giving y explicitly in terms of x.

EXAMPLE 2. Assume that the equation $2xy+\pi\sin(y) = 2\pi$ defines a relationship between x and y, that is, $y = f(x)$ for some unknown (and unknowable) function f. Find $\frac{dy}{dx}$ when $x = 1$ and $y = \frac{\pi}{2}$.

SOLUTION First, verify that the point $(1, \pi/2)$ is on the graph of $y = f(x)$ by checking that the implicit equation is satisfied when $x = 1$ and $y = \pi/2$: $2(1)(\pi/2) + \pi\sin(\pi/2) = \pi + \pi = 2\pi$.

The use of implicit differentiation to determine dy/dx when $x = 1$ and $y = \pi/2$ involves the following steps:

$$\begin{aligned}
\frac{d}{dx}(2xy + \pi\sin(y)) &= \frac{d(2\pi)}{dx} && \text{(differentiate the implicit equation relating } x \text{ and } y\text{)}\\
\left(2\frac{dx}{dx}y + 2x\frac{dy}{dx}\right) + \pi\cos(y)\frac{dy}{dx} &= 0 && \text{(remembering that } y = y(x)\text{, use the chain rule)}\\
2\left(\frac{\pi}{2}\right) + 2(1)\frac{dy}{dx} + \pi\cos(\left(\frac{\pi}{2}\right))\frac{dy}{dx} &= 0 && \text{(substitute given values into the differentiated equation)}\\
\pi + 2\frac{dy}{dx} &= 0 && \text{(simplify in preparation of solving for unknown quantity)}\\
\frac{dy}{dx} &= -\frac{\pi}{2} && \text{(solve for the unknown quantity, } dy/dx\text{)}
\end{aligned}$$

So, $dy/dx = -\pi/2$ when $x = 1$ and $y = \pi/2$. ♦

Warning: Do Not Substitute Values Until Differentiation is Complete

The last two steps, substituting given values into the differentiated equation and solving for the unknown quantity, could be performed in either order. The one temptation that must be avoided is substituting values before completing all differentiation.

Implicit Differentiation and Extrema

Example 3 of Section 5.1 answered the question, "Of all the tin cans that enclose a volume of 100π cubic inches, which requires the least metal?" The radius of the most economical can is $\sqrt[3]{50}$. From this and the fact that its volume is 100π cubic inches, its height was found to be $2\sqrt[3]{50}$, exactly twice the radius. In the next example implicit differentiation is used to answer the same question. Not only will the algebra be simpler but it will provide the shape – the proportion between height and radius – easily.

EXAMPLE 3. Of all the tin cans that enclose a volume of 100π cubic inches, which requires the least metal?

SOLUTION Let the height and radius of the can be represented by h and r, respectively. The objective function is the surface area of the can:

$$S = 2\pi r^2 + 2\pi r h. \tag{5.2.2}$$

The constraint is that the volume is 100π cubic inches: $\pi r^2 h = 100\pi$.

While h, and hence S, are functions of r, it is not necessary to find explicit formulas for h and S in terms of r. Differentiation of the constraint equation and the objective function with respect to r yields

$$2\pi r h + \pi r^2 \frac{dh}{dr} = \frac{d(100\pi)}{dr} = 0 \quad \text{and} \quad \frac{dS}{dr} = 4\pi r + 2\pi h + 2\pi r \frac{dh}{dr}. \tag{5.2.3}$$

When S is a minimum, $dS/dr = 0$, so we have

$$0 = 4\pi r + 2\pi h + 2\pi r \frac{dh}{dr}. \tag{5.2.4}$$

To obtain a relation between h and r, factor πr out of the first equation in (5.2.3) and 2π out of (5.2.4):

$$2h + r\frac{dh}{dr} = 0 \qquad \text{and} \qquad 2r + h + r\frac{dh}{dr} = 0. \tag{5.2.5}$$

Elimination of dh/dr from (5.2.5) yields $2r + h + r(-2h/r) = 0$, which simplifies to

$$2r = h. \tag{5.2.6}$$

We have obtained the shape before finding the specific dimensions. Equation (5.2.6) asserts that the height of the most economical can is the same as its diameter. Moreover, this is the ideal shape no matter what the prescribed volume happens to be.

The dimensions of the most economical can with a volume of 100π are found by substituting $h = 2r$ into the constraint:

$$\pi r^2(2r) = 100\pi \qquad \text{or} \qquad r^3 = 50.$$

The tin can with $r = \sqrt[3]{50} \approx 3.684$ in and $h = 2r = 2\sqrt[3]{50} \approx 7.368$ in has volume 100π cubic inches and has the smallest surface area. ♦

The procedure illustrated in Example 3 is general. It may be of use when maximizing (or minimizing) a quantity (the **objective function**) that at first is expressed as a function of two variables that are linked by an equation. The equation that links them is called the **constraint**. In Example 3, the constraint is $\pi r^2 h = 100\pi$.

Algorithm: Using Implicit Differentiation in Max/Min Problems

1. Name the various quantities in the problem by letters, such as x, y, h, r, A, V.
2. Identify the **objective function**, the quantity to be maximized (or minimized).
3. Express it in terms of other quantities, such as x and y.
4. Identify the **constraint equation(s)** that provide relationships between x and y.
5. Differentiate implicitly both the constraint and the quantity to be maximized (or minimized), interpreting all quantities as functions of a single variable (which you choose).
6. Set the derivative of the quantity to be maximized (or minimized) equal to 0 and use the derivative of the constraint to obtain an equation relating x and y at a maximum (or minimum).
7. Step 6 gives only a relation between x and y at an extremum. If the explicit values of x and y are needed, find them by noting that x and y also satisfy the constraint.

Warning

Sometimes an extremum occurs where a derivative, such as dy/dx, is not defined, as Exercise 38 illustrates.

Implicit Differentiation and the Second Derivative

As the next example shows, second derivatives can also be found by implicit differentiation.

EXAMPLE 4. The function $y = y(x)$ is given implicitly by $2xy + \pi\sin(y) = 2\pi$. Find y' and y''.

SOLUTION This implicit equation was introduced in Example 2. There, after differentiating but before substituting the given values of x and y, we found that $2y+2xy'+\pi\cos(y)y' = 0$. Solving for y' yields $y' = -2y/\left(2x+\pi\cos(y)\right)$. Now, differentiating once again:

$$\begin{aligned}
y'' = (y')' &= \left(\frac{-2y}{2x+\pi\cos(y)}\right)' \\
&= \frac{\left(2x+\pi\cos(y)\right)\left(-2y'\right)-\left(-2y\right)\left(2-\pi\sin(y)y'\right)}{\left(2x+\pi\cos(y)\right)^2} && \text{(quotient and chain rules)} \\
&= \frac{\left(2x+\pi\cos(y)\right)\left(-2\left(\frac{-2y}{2x+\pi\cos(y)}\right)\right)-\left(-2y\right)\left(2-\pi\sin(y)\left(\frac{-2y}{2x+\pi\cos(y)}\right)\right)}{\left(2x+\pi\cos(y)\right)^2} && \text{(substitute for } y'\text{)} \\
&= \frac{8y}{\left(2x+\pi\cos(y)\right)^2}+\frac{4\pi y^2\sin(y)}{\left(2x+\pi\cos(y)\right)^3} && \text{(simplify)}
\end{aligned}$$

The preferred final form for y'' probably depends on the intended uses. ♦

The idea is to differentiate with respect to x, remembering that $y = y(x)$, and using the chain rule. And, whenever y' appears in the computations replace it by its expression in terms of x and y.

Logarithmic Differentiation

If $\ln(f(x))$ is simpler than $f(x)$ there is a technique for finding $f'(x)$ that saves labor. Example 5 illustrates this approach.

EXAMPLE 5. Let $y = \dfrac{\cos(3x)}{\left(\sqrt[3]{x^2+5}\right)^4}$. Find $\dfrac{dy}{dx}$.

Properties of Logarithms
$\ln(AB) = \ln(A) + \ln(B)$
$\ln(A/B) = \ln(A) - \ln(B)$
$\ln(A^B) = B\ln(A)$

SOLUTION The solution to this problem by **logarithmic differentiation** begins by simplifying $\ln(y)$ using properties of logarithms:

$$\begin{aligned}\ln(y) &= \ln(\cos(3x)) - \ln\left(\left(\sqrt[3]{x^2+5}\right)^4\right) && (\ln(A/B) = \ln(A) - \ln(B))\\ &= \ln(\cos(3x)) - \frac{4}{3}\ln\left(x^2+5\right) && (\ln(A^B) = B\ln(A)).\end{aligned}$$

Now y is given implicitly.

By the chain rule, $(\ln(y))' = (1/y)y'$, so that

$$\frac{1}{y}\frac{dy}{dx} = \frac{d}{dx}(\ln(y)) = \frac{d}{dx}\left(\ln(\cos(3x)) - \frac{4}{3}\ln\left(x^2+5\right)\right) = \frac{-3\sin(3x)}{\cos(3x)} - \frac{4}{3}\frac{2x}{x^2+5}.$$

Therefore

$$\frac{dy}{dx} = y\left(-3\tan(3x) - \frac{4}{3}\frac{2x}{x^2+5}\right).$$

Replace y by its formula, getting

$$\frac{dy}{dx} = \frac{\cos(3x)}{\left(\sqrt[3]{x^2+5}\right)^4}\left(-3\tan(3x) - \frac{4}{3}\frac{2x}{x^2+5}\right).$$

To appreciate logarithmic differentiation, find this derivative directly, as requested in Exercise 26. ♦

Summary

If a function is given implicitly, differentiate both sides of the equation it satisfies and solve for the derivative. The derivative is then expressed in terms of the function and its independent variable. To find the second derivative, differentiate the resulting expression, replacing the derivative that appears in it by its formula.

If a function $y(x)$ is a product or a quotient of powers, it may be easier to find $y'(x)$ by differentiating $\ln(y(x))$ implicitly.

EXERCISES for Section 5.2

In Exercises 1 to 4 find $\frac{dy}{dx}$ at (x, y) in two ways: explicitly (solving for y first) and implicitly.

1. $xy = 4$ at $(1, 4)$

2. $x^2 - y^2 = 3$ at $(2, 1)$

3. $x^2y + xy^2 = 12$ at $(3, 1)$

4. $x^2 + y^2 = 100$ at $(6, -8)$

In Exercises 5 to 8 find $\frac{dy}{dx}$ at the specified point by implicit differentiation.

5. $\frac{2xy}{\pi} + \sin(y) = 2$ at $\left(1, \frac{\pi}{2}\right)$

6. $2y^3 + 4xy + x^2 = 7$ at $(1, 1)$

7. $x^5 + y^3x + yx^2 + y^5 = 4$ at $(1, 1)$

8. $x + \tan(xy) = 2$ at $\left(1, \frac{\pi}{4}\right)$

9. Repeat Example 3 by implicit differentiation, but differentiate the objective function and constraint equation with respect to h instead of r.

10. What is the ratio of the radius and height of the cylindrical can of largest volume that can be constructed with a given surface area?

11. Using implicit differentiation, find $D(\arctan(x))$.

12. Using implicit differentiation, find $D(\arcsin(x))$.

In Exercises 13 to 16 find $\frac{dy}{dx}$.

13. $xy^3 + \tan(x + y) = 1$

14. $\sec(x + 2y) + \cos(x - 2y) + y = 2$

15. $-7x^2 + 48xy + 7y^2 = 25$

16. $\sin^3(xy) + \cos(x + y) + x = 1$

In Exercises 17 to 20 implicit differentiation is used to find a second derivative.

17. Assume that $y(x)$ is a differentiable function of x and that $x^3y + y^4 = 2$. Assume that $y(1) = 1$. Find $y''(1)$, following these steps.

(a) Show that $x^3y' + 3x^2y + 4y^3y' = 0$.

(b) Use (a) to find $y'(1)$.

(c) Differentiate the equation in (a) and show that $x^3y'' + 6x^2y' + 6xy + 4y^3y'' + 12y^2(y')^2 = 0$.

(d) Use the equation in (c) to find $y''(1)$. ($y(1)$ and $y'(1)$ are known.)

18. Find $y''(1)$ if $y(1) = 2$ and $x^5 + xy + y^5 = 35$.

19. Find $y'(1)$ and $y''(1)$ if $y(1) = 0$ and $\sin(y) = x - x^3$.

20. Find $y''(2)$ if $y(2) = 1$ and $x^3 + x^2y - xy^3 = 10$.

In Exercises 21 to 24 find y''. Express y'' in terms of x and y (but not y').

21. $y' = (x + y)\sin(x)$

22. $y' = \sin(xy)$

23. $x(y')^3 = (x+3)^2y$

24. $y'\sin(y) + e^{xy} = 0$

---

25. Use implicit differentiation to find the highest and lowest points on the ellipse $x^2 + xy + y^2 = 12$.

26. Differentiate the function in Example 5 directly, without taking logarithms first.

27. Does the tangent line to the curve $x^3 + xy^2 + x^3y^5 = 3$ at $(1,1)$ pass through $(-2,3)$? (Explain.)

In Exercises 28 to 31, find y' two ways:

(a) by using the given explicit formula for y and (b) by simplifying $\ln(y)$ and using implicit differentiation.

28. $y = \sqrt{1+3x}\sqrt[3]{1+2x}$

29. $y = (\cos(3x))^{5/2}(\sin(2x))^{1/3}$

30. $y = \dfrac{\left(1+e^{3x}\right)^4}{\left(1+e^{2x}\right)^3}$

31. $y = \dfrac{(\tan(3x))^4\left(x+x^3\right)^5}{\sqrt{x}}$

Exercises 32 and 33 obtain the formulas for differentiating $x^{1/n}$ and $x^{m/n}$ by implicit differentiation. Here m and n are integers and we assume the functions are differentiable.

32. Let n be a positive integer. Assume that $y = x^{1/n}$ is a differentiable function of x. From $y^n = x$ deduce by implicit differentiation that $y' = \frac{1}{n} x^{1/n-1}$.

33. Let m be a nonzero integer and n a positive integer. Assume that $y = x^{m/n}$ is a differentiable function of x. From $y^n = x^m$ deduce by implicit differentiation that $y' = \frac{m}{n} x^{m/n-1}$.

---

34. For a constant k find $D(x^k)$, $x > 0$, by logarithmic differentiation of $y = x^k$.

35. Let $y = x^x$.
 (a) Find y' by logarithmic differentiation. That is, first take the logarithm of both sides.
 (b) Find y' by first writing the base as $e^{\ln(x)}$. That is, write $y = x^x = \left(e^{\ln(x)}\right)^x = e^{x\ln(x)}$.

36. If $x^3 + y^3 = 1$, find y' and y'' in terms of x and y.

37. Find the first and second derivatives of $y = \sec(x^2)\frac{\sin(x^2)}{x}$.

38. (a) What difficulty arises when you use implicit differentiation to maximize $x^2 + y^2$ subject to $x^2 + 4y^2 = 16$?
 (b) Show that a maximum occurs when $\frac{dy}{dx}$ is not defined. What is the maximum value? Where does it occur?
 (c) The problem can be viewed geometrically as maximizing the square of the distance from the origin for points on the ellipse $x^2 + 4y^2 = 16$. Sketch the ellipse and interpret (b) in terms of it.

5.3 Related Rates

The rate at which one quantity changes may affect the rate at which another quantity connected to it changes. Implicit differentiation is a convenient tool for finding the relation between the two rates, as the next few examples will illustrate.

EXAMPLE 1. An angler has a fish at the end of his line, which is reeled in at 2 feet per second from a bridge 30 feet above the water. At what speed is the fish moving through the water when the amount of line out is 50 feet? 31 feet? Assume the fish is at the surface of the water. (See Figure 5.3.1.)

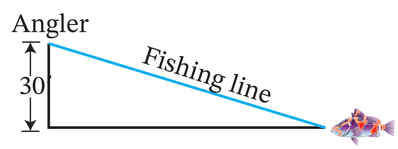


Figure 5.3.1

SOLUTION Our first impression might be that since the line is reeled in at a constant speed, the fish at the end of the line moves through the water at a constant speed. This is not the case.

Let s be the length of the line and x the horizontal distance of the fish from the bridge. (See Figure 5.3.2.)

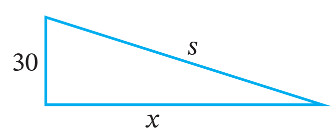


Figure 5.3.2

Since the line is reeled in at the rate of 2 feet per second, s is shrinking, and

$$\frac{ds}{dt} = -2.$$

The rate at which the fish moves through the water is given by the derivative, dx/dt. The problem is to find dx/dt when $s = 50$ and also when $s = 31$.

We need an equation that relates s and x at any time, not just when $s = 50$ or $s = 31$. If we consider only $s = 50$ or $s = 31$, there would be no motion, and no chance to use derivatives.

The quantities x and s are related by the Pythagorean Theorem:

$$x^2 + 30^2 = s^2.$$

Both x and s are functions of time t. Thus both sides of the equation may be differentiated with respect to t, yielding

$$\frac{d(x^2)}{dt} + \frac{d(30^2)}{dt} = \frac{d(s^2)}{dt}$$

or

$$2x\frac{dx}{dt} + 0 = 2s\frac{ds}{dt}.$$

Hence

$$x\frac{dx}{dt} = s\frac{ds}{dt}.$$

This equation provides the relationship among s, x, ds/dt, and dx/dt that is the key to answering the questions posed in this example.

Since $ds/dt = -2$,

$$x\frac{dx}{dt} = (s)(-2).$$

Hence

$$\frac{dx}{dt} = \frac{-2s}{x}.$$

This means the fish moves with velocity $dx/dt = -2s/x$. The negative sign indicates that the fish is being reeled in. The fish's speed is $|dx/dt| = 2s/x$.

When $s = 50$,

$$x^2 + 30^2 = 50^2$$

and we find $x = 40$. Thus when 50 feet of line is out, the speed is

$$\left|\frac{dx}{dt}\right| = \frac{2s}{x} = \frac{2 \cdot 50}{40} = 2.5 \text{ feet per second.}$$

When $s = 31$,

$$x^2 + 30^2 = 31^2.$$

Hence

$$x = \sqrt{31^2 - 30^2} = \sqrt{961 - 900} = \sqrt{61}.$$

Thus when 31 feet of line is out, the fish is moving at the speed of

$$\left|\frac{dx}{dt}\right| = \frac{2s}{x} = \frac{2 \cdot 31}{\sqrt{61}} = \frac{62}{\sqrt{61}} \approx 7.9 \text{ feet per second.}$$

Let us look at the situation from the fish's point of view. When it is x feet from the point in the water directly below the bridge, its speed is $2s/x$ feet per second. Since s is larger than x, its speed is always greater than 2 feet per second. When x is very large, s/x is near 1 and the fish is moving through the water only a little faster than the line is reeled in. However, when the fish is almost at the point under the bridge, x is very small, then $2s/x$ is huge, and the fish finds itself moving at a high speed. ◆

In Example 1 it would be an error to indicate in Figure 5.3.2 that the hypotenuse of the triangle is 50 feet, for if one leg is 30 feet and the hypotenuse is 50 feet, the triangle is completely determined; there is nothing left free to vary with time.

In general, label all the lengths or quantities that can change with letters x, y, s, and so on, even if not all are needed in the solution. Only after you finish differentiating do you determine what the rates are at a specified value of the variable.

The General Procedure

The method used in Example 1 applies to many related rate problems. This is the general procedure, broken into steps:

Algorithm: Procedure for Finding a Related Rate

1. Find an equation that relates the varying quantities.
   (If the quantities are geometric, draw a picture and label the varying quantities with letters.)
2. Differentiate both sides of the equation with respect to time, obtaining an equation that relates the various rates of change.
3. Solve the equation obtained in Step 2 for the unknown rate.
   (Only at this step do you substitute constants for variable.)

REMEMBER: Differentiate before substituting specific values for the variables.
If the order is reversed, there would only be constants to differentiate.

Finding an Acceleration

The method described in Example 1 for determining unknown rates from known ones extends to finding unknown accelerations. Differentiate one more time. Example 2 illustrates this.

EXAMPLE 2. Water flows into a conical tank at the constant rate of 3 cubic meters per second. The radius of the cone is 5 meters and its height is 4 meters. Let $h(t)$ represent the height of the water above the bottom of the cone at time t. Find dh/dt (the rate at which the water is rising in the tank) and d^2h/dt^2 (the rate at which that rate changes) when the tank is filled to a height of 2 meters. (See Figure 5.3.3.)

Figure 5.3.3

SOLUTION Let $V(t)$ be the volume of water in the tank at time t. That water flows into the tank at 3 cubic meters per second is expressed as

$$\frac{dV}{dt} = 3,$$

and, since it is constant,

$$\frac{d^2V}{dt^2} = 0.$$

To find dh/dt and d^2h/dt^2, obtain an equation relating V and h.

When the tank is filled to the height h, the water forms a cone of height h and radius r. (See Figure 5.3.3(b).) By similar triangles,

$$\frac{r}{h} = \frac{5}{4} \qquad \text{or} \qquad r = \frac{5h}{4}.$$

Thus

$$V = \frac{1}{3}\pi r^2 h = \frac{1}{3}\pi\left(\frac{5}{4}h\right)^2 h = \frac{25}{48}\pi h^3.$$

So the equation relating V and h is

$$V = \frac{25\pi}{48}h^3. \tag{5.3.1}$$

From here on, differentiate as often as needed.

Differentiating both sides of (5.3.1) once (using the chain rule) yields

$$\frac{dV}{dt} = \frac{25\pi}{48}\frac{d(h^3)}{dh}\frac{dh}{dt}$$

or

$$\frac{dV}{dt} = \frac{25\pi}{16}h^2\frac{dh}{dt}.$$

Since $dV/dt = 3$ all the time,

$$3 = \frac{25\pi h^2}{16}\frac{dh}{dt},$$

from which it follows that

$$\frac{dh}{dt} = \frac{48}{25\pi h^2} \text{ meters per second.} \tag{5.3.2}$$

As (5.3.2) shows, the larger h is, the slower the water rises. (Why is this to be expected?)

Even though water enters the tank at a constant rate, the height does not rise at a constant rate.

To find dh/dt when $h = 2$ meters, substitute 2 for h in (5.3.2), obtaining

$$\frac{dh}{dt} = \frac{48}{25\pi 2^2} = \frac{12}{25\pi} \approx 0.1528 \text{ meters per second.}$$

Now we turn to the acceleration, d^2h/dt^2. We do not differentiate $dh/dt = 12/(25\pi)$ since it holds only when $h = 2$. We go back to (5.3.2), which holds at any time.

Differentiating (5.3.2) with respect to t yields

$$\frac{d^2h}{dt^2} = \frac{48}{25\pi}\frac{d}{dt}\left(\frac{1}{h^2}\right) = \frac{48}{25\pi}\frac{-2}{h^3}\frac{dh}{dt} = \frac{-96}{25\pi h^3}\frac{dh}{dt}, \tag{5.3.3}$$

which expresses the acceleration in terms of h and dh/dt. Substituting (5.3.2) into (5.3.3) gives

$$\frac{d^2h}{dt^2} = \frac{-96}{25\pi h^3}\frac{48}{25\pi h^2} = \frac{-(96)(48)}{(25\pi)^2h^5} \text{ meters per second per second.} \tag{5.3.4}$$

Equation (5.3.4) tells us that since d^2h/dt^2 is negative, the rate at which the water rises in the tank is decreasing.

The problem also asked for the value of d^2h/dt^2 when $h = 2$. To find it, replace h by 2 in (5.3.4), obtaining

$$\frac{d^2h}{dt^2} = \frac{-(96)(48)}{(25\pi)^2 2^5} = \frac{-144}{625\pi^2} \approx -0.02334 \text{ meters per second per second.}$$

♦

Summary

If two variables are linked by an equation, the rates at which they change are also related. To find an equation involving those two rates, differentiate both sides of the original equation implicitly.

When the variables are position, s, and time, t, the velocity can be found by (implicitly) differentiating the equation relating $s = s(T)$ and t with respect to t. To find the acceleration, simplify differentiate the equation just obtained, involving t, $s = s(t)$, and $v = ds/dt$, with respect to t once again.

EXERCISES for Section 5.3

1. How fast is the fish in Example 1 moving through the water when it is 1 foot horizontally from the bridge?

2. The angler in Example 1 decides to let the line out as the fish swims away. The fish swims away at a constant speed of 5 feet per second relative to the water. How fast is the angler playing out his line when the horizontal distance from the bridge to the fish is (a) 1 foot? (b) 100 feet?

3. A 10-foot ladder is leaning against a wall that is 15 feet high. A person pulls the base of the ladder away from the wall at the rate of 1 foot per second.

- (a) Draw a picture and label the varying lengths by letters and the fixed lengths by numbers.
- (b) Obtain an equation involving the variables in (a).
- (c) Differentiate it with respect to time.
- (d) How fast is the top going down the wall when the base of the ladder is 6 feet from the wall? 8 feet from the wall? 9 feet from the wall?

4. A kite is flying at a height of 300 feet in a horizontal wind.

- (a) Draw a picture and label the varying lengths by letters and the fixed lengths by numbers.
- (b) When 500 feet of string is out, the kite is pulling the string out at a rate of 20 feet per second. What is the kite's velocity? (Assume the string remains straight.)

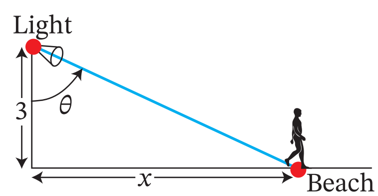


Figure 5.3.4

5. A beachcomber walks 2 miles per hour along the shore as the beam from a rotating light 3 miles offshore follows him. (See Figure 5.3.4.)

- (a) What do you think happens to the rate at which the light rotates as the beachcomber walks farther and farther along the shore away from the lighthouse?
- (b) Let x denote the distance of the beachcomber from the point on the shore nearest the light and θ the angle of the light. Obtain an equation relating θ and x.
- (c) With the aid of (b), show that $\frac{d\theta}{dt} = frac69+x^2$ (radians per hour).
- (d) Does the formula in (c) agree with your guess in (a)?

6. A man 6 feet tall walks at the rate of 5 feet per second away from a street lamp that is 20 feet high. At what rate is his shadow lengthening when he is (a) 10 feet from the lamp? (b) 100 feet from the lamp?

7. A spherical balloon loses air at the rate of 1 cubic inch per second. At what rate is its radius changing when the radius is (a) 10 feet? (b) 20 feet?

> The volume of a sphere with radius r is $V = \frac{4}{3}\pi r^3$.

8. A large spherical balloon is being inflated at the rate of 100 cubic feet per minute. At what rate is the radius increasing when its radius is (a) 2 inches? (b) 1 inch?

9. Bulldozers are moving earth at the rate of 1,000 cubic yards per hour onto a conically shaped hill with half vertex angle $\frac{\pi}{6}$. How fast is the height of the hill increasing when the hill is (a) 20 yards high? (b) 100 yards high?

> The volume of a cone of radius r and height h is $V = \frac{1}{3}\pi r^2 h$.

10. The lengths of the two legs of a right triangle depend on time. One leg, whose length is x, increases at the rate of 5 feet per second, while the other, of length y, decreases at the rate of 6 feet per second. At what rate is the hypotenuse changing when $x = 3$ feet and $y = 4$ feet? Is the hypotenuse increasing or decreasing then?

11. Two sides of a triangle and their included angle are changing with respect to time. The angle increases at the rate of 1 radian per second, one side increases at the rate of 3 feet per second, and the other side decreases at the rate of 2 feet per second. Find the rate at which the area is changing when the angle is $\frac{\pi}{4}$, the first side is 4 feet long, and the second side is 5 feet long. Is the area decreasing or increasing then?

12. The length of a rectangle is increasing at the rate of 7 feet per second, and the width is decreasing at the rate of 3 feet per second. When the length is 12 feet and the width is 5 feet, find the rate of change of
(a) the area , (b) the perimeter, and (c) the length of the diagonal.

NOTATION: **Dot notation**, $\dot{x}$ for $\frac{dx}{dt}$, $\dot{\theta}$ for $\frac{d\theta}{dt}$, $\ddot{x}$ for $\frac{d^2x}{dt^2}$, $\ddot{\theta}$ for $\frac{d^2\theta}{dt^2}$, ..., for derivatives with respect to time was introduced by Newton. It is still widely used, particularly in physics.

Exercises 13 to 17 concern acceleration.

13. What is the acceleration of the fish described in Example 1 when the length of line is (a) 300 feet? (b) 31 feet?

14. A woman on the ground is watching an airplane through a telescope. The airplane is moving in a path that will take it directly over her at a speed of 10 miles per minute at an altitude of 7 miles. At what rate (in radians per minute) is the angle of elevation of the telescope changing when the horizontal distance of the jet from the woman is 24 miles? When the jet is directly above the woman?

15. Find $\ddot{\theta}$ in Exercise 14 when the horizontal distance from the jet is (a) 7 miles and (b) 1 mile.

16. A particle moves on the parabola $y = x^2$ in such a way that $\dot{x} = 3$ throughout the journey. Find the formulas for (a) $\dot{y}$ and (b) $\ddot{y}$.

17. Call one acute angle of a right triangle θ. The adjacent leg has length x and the opposite leg has length y. Assume that x, y, and θ each depend on time t. Obtain an equation
(a) relating x, y, and θ , (b) involving $\dot{x}$, $\dot{y}$, and $\dot{\theta}$, and (c) involving $\ddot{x}$, $\ddot{y}$, and $\ddot{\theta}$.

In (b) and (c), the equations will also involve x, y, and θ.

---

18. At an altitude of x kilometers, the atmospheric pressure decreases at a rate of $128(0.88)^x$ millibars per kilometer. A rocket is rising at the rate of 5 kilometers per second vertically. At what rate is the atmospheric pressure changing (in millibars per second) when the altitude of the rocket is (a) 1 kilometer? (b) 50 kilometers?

19. A woman is walking on a bridge that is 20 feet above a river as a boat passes directly under the center of the bridge (at a right angle to the bridge) at 10 feet per second. At that moment the woman is 50 feet from the center and approaching it at the rate of 5 feet per second.

(a) At what rate is the distance between the boat and woman changing at that moment?
(b) Is the rate increasing, decreasing, or staying the same?

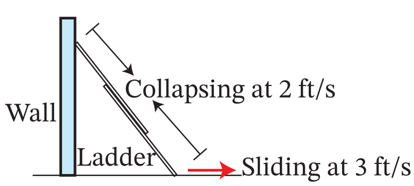


Figure 5.3.5

20. A two-piece extension ladder leaning against a wall is collapsing at the rate of 2 feet per second and the base of the ladder is moving away from the wall at the rate of 3 feet per second. How fast is the top of the ladder moving down the wall when it is 8 feet from the ground and the foot is 6 feet from the wall? (See Figure 5.3.5.)

21. A spherical raindrop evaporates at a rate proportional to its surface area. Show that its radius shrinks at a constant rate.

22. A couple is on a Ferris wheel when the sun is directly overhead. The diameter of the wheel is 50 feet and its speed is 0.01 revolution per second.

(a) What is the speed of their shadows on the ground when they are at a two-o'clock position?
(b) A one-o'clock position?
(c) Show that the shadow is moving fastest when they are at the top or bottom, and slowest when they are at the three-o'clock or nine-o'clock position.

23. Water is flowing into a hemispherical bowl of radius 5 feet at the constant rate of 1 cubic foot per minute.
(a) At what rate is the top surface of the water rising when its height above the bottom of the bowl is 3 feet? 4 feet? 5 feet?
(b) If $h(t)$ is the depth in feet at time t, find $\ddot{h}$ when $h = 3$, 4, and 5 feet.

24. A detective is aiming a flashlight at a door. The axis of the conical beam is perpendicular to the door. Let $x(t)$ be the distance between the detective and the door at time t. Let $A(t)$ be the area of the illuminated disk on the door at time t. The detective is walking towards the door. Explain each answer.
(a) Is $\frac{dx}{dt}$ positive or negative? (b) Is $\frac{dA}{dt}$ positive or negative? (c) Is there a constant k such that $\frac{dA}{dt} = k\frac{dx}{dt}$?

25. The rate at which the variable $B(t)$ changes is proportional to the square of the rate at which the variable $C(t)$ changes. Does it follow that the acceleration of $A(t)$, $A''(t)$, is proportional to the square of $B''(t)$? Explain.

26. Soup is poured into a hemispherical soup bowl of radius a at a constant rate, k cubic centimeters per second. At what rate is the wetted part of the bowl increasing when the soup has depth h?

5.4 Higher Derivatives and the Growth of a Function

The only higher derivative we used so far is the second derivative. In the study of motion, if y denotes position then y'' is acceleration. In the study of graphs, the second derivative determines whether the graph is concave up ($y'' > 0$) or down ($y'' < 0$).

Now we will see how higher derivatives (including the second derivative) influence the growth of a function. The next section uses this information to estimate the error in approximating a function by a polynomial.

Introduction

Imagine that you are in a motionless car parked at the origin of the x-axis. You put your foot to the gas pedal and accelerate. The greater the acceleration, the faster the speed increases and the greater the speed, the farther you travel in a given time. So the acceleration, which is the second derivative of the position function, influences the function. In this section we examine this influence in more detail.

The following lemma is the basis for our analysis. Informally, it says, "If two runners start at the same place at the same time, the slower runner never goes ahead of the faster one."

Lemma 5.4.1

Let $f(x)$ and $g(x)$ be differentiable functions on an interval I. Let a be a number in I where $f(a) = g(a)$. Assume that $f'(x) \le g'(x)$ for x in I. Then $f(x) \le g(x)$ for all x in I to the right of a and $f(x) \ge g(x)$ for all x in I to the left of a.

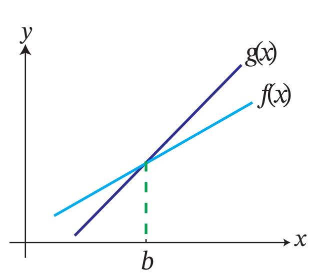


Figure 5.4.1

Figure 5.4.1 makes this plausible when the graphs of f and g are straight lines. To the right of $x = a$ the steeper line lies above the other line. To the left of $x = a$ the steeper line lies below the other line.

Proof of Lemma 5.4.1
Let $h(x) = g(x) - f(x)$. Then $h'(x) = g'(x) - f'(x) \ge 0$. Thus, h is a nondecreasing function on I.

Since $h(a) = 0$, it follows that $h(x) \ge 0$ for $x \ge a$. That is, $g(x) - f(x) \ge 0$, hence $f(x) \le g(x)$ for $x > a$ and in I.

By the same logic, h nondecreasing on I and $h(a) = 0$ imply $h(x) \le 0$ for $x \le a$. Thus, $f(x) \ge g(x)$ for $x < a$ and in I.

Repeated application of Lemma 5.4.1 will enable us to establish a connection between higher derivatives of a function and the function.

The Influence of Higher Derivatives

In the following theorem we name a function $R(x)$ because that will be the notation in the next section when $R(x)$ is the remainder function. The notation $n!$ (read "n **factorial**") for a positive integer n is short for the product of all integers from 1 through n: $n! = n(n-1)\cdots 3\cdot 2\cdot 1$. The symbol 0! is usually defined to be 1.

Theorem 5.4.2: Growth Theorem

Assume that at a the function R and its first n derivatives are zero,

$$R(a) = R'(a) = R''(a) = R^{(3)}(a) = \cdots = R^{(n)}(a) = 0.$$

Assume also that $R(x)$ has continuous derivatives up through the derivative of order $n+1$ in an open interval I containing the numbers a and x. Then, assuming $x > a$, there is a number c in the interval $[a, x]$ such that

$$R(x) = R^{(n+1)}(c)\frac{(x-a)^{n+1}}{(n+1)!}. \tag{5.4.1}$$

Before giving the proof of Theorem 5.4.2, we present some examples illustrating its utility, and a useful extension.

The growth theorem with $n = 1$ and $a = 0$ describes the position of an accelerating car. Let $R(x)$ be the position of the car on the y-axis at time x. One has $R(0) = 0$ (at time 0 the car is at position 0), $R'(0) = 0$ (at time 0 the car is not moving), and R'' describes the acceleration. If the acceleration is constant and equal to k, then as was shown in Section 3.7, the car's position at time x is $kx^2/2$. Whether the car's acceleration is constant or not, (5.4.1) says the car's position at time x is the acceleration at some time multiplied by $x^2/2$.

EXAMPLE 1. Show that $\left|e^x - 1 - x\right| \le \frac{e}{2}x^2$ for x in $(-1, 1)$.

SOLUTION Let $R(x) = e^x - 1 - x$. Since $R'(x) = e^x - 1$ and $R''(x) = e^x$ we see that $R(0) = e^0 - 1 - 0 = 0$ and $R'(0) = e^0 - 1 = 0$. By the growth theorem, with $a = 0$ and $n = 1$, there is a number c in $(-1, 1)$ such that

$$e^x - 1 - x = e^c\frac{(x-0)^2}{2!}.$$

We do not know the value of c, but, since it is less than 1, $e^c < e$. Thus

$$\left|e^x - 1 - x\right| \le \frac{e}{2}x^2.$$

◆

The inequality in Example 1 provides a way to estimate e^x when x is small. For instance, $|e^{0.1} - 1 - 0.1| \le e(0.1)^2/2 = e/200$. The estimate 1.1 for $e^{0.1}$ is off by at most $e/200 \approx 0.013591$.

EXAMPLE 2. Let $R(x) = \cos(x) - 1 + \frac{x^2}{2}$. Show that $|R(x)| \le \frac{\left|x^3\right|}{6}$.

SOLUTION As in Example 1 we use the growth theorem, but now with $a = 0$, $n = 2$, and $x > 0$.

With $n = 2$ we will need the first three derivatives of $R(x) = \cos(x) - 1 + x^2/2$: $R'(x) = -\sin(x) + x$, $R''(x) = -\cos(x) + 1$, and $R^{(3)}(x) = \sin(x)$. Then, $R(0) = 1 - 1 + 0 = 0$, $R'(0) = 0 + 0 = 0$, and $R''(0) = -1 + 1 = 0$.

By the growth theorem, with $a = 0$ and $n = 2$,

$$R(x) = \sin(c)\frac{x^3}{3!} \qquad \text{for some number } c \text{ between 0 and } x.$$

Because $|\sin(x)| \le 1$,

$$|R(x)| = \left|\sin(c)\frac{x^3}{3!}\right| \le \left|(1)\frac{x^3}{6}\right| = \frac{|x|^3}{6}.$$

◆

Example 2 provides an estimate for values of the cosine function for small angles. For instance, if $x = 0.1$ radians, we have

$$\left|\cos(0.1) - 1 + \frac{0.1^2}{2}\right| \le \frac{0.1^3}{6} = 0.00016667 = 1.6667 \times 10^{-4}.$$

Thus, $1 - (0.1^2)/2 = 1 - 0.005 = 0.995$ is an estimate of $\cos(0.1)$ with an error less than $1/6 \times 10^{-3} \approx 1.6667 \times 10^{-4}$. In fact, $\cos(0.1) \approx 0.9950041653$. So, the error is only 0.0000041653. 0.1 radians $= 0.1(180°/\pi) \approx 5.7°$.

An even better bound on the growth of $R(x)$ in Example 2 is possible with no additional work. In addition to $R(0) = R'(0) = R''(0) = 0$, we have $R^{(3)}(0) = \sin(0) = 0$. Then, because $R^{(4)}(x) = \cos(x)$, $R(x) = \cos(c)(x-0)^4/4!$ for some value of c in $[0, x]$, Thus, as $|\cos(c)| \le 1$,

$$|R(x)| \le \left|(1)\frac{x^4}{4!}\right| = \frac{x^4}{24}.$$

This means the difference between the exact value of $\cos(0.1)$ and the estimate $1 - 0.1^2/2 = 0.995$ is no more than $0.1^4/24 = 4.16667 \times 10^{-6}$, showing that the estimate $\cos(0.1) \approx 0.99500$ is accurate to five decimal places. (In fact, $|\cos(0.1) - 0.995| \approx 4.16528 \times 10^{-6}$, so the actual error is just slightly smaller than the error bound produced by the growth theorem.)

In any case, $1 - x^2/2$ is a good estimate of $\cos(x)$ for small values of x. The next section describes how to find polynomials that provide good estimates of functions.

A Refinement of the Growth Theorem

When proving the growth theorem we will establish something stronger:

Theorem 5.4.3: Refined Growth Theorem

In addition to the hypotheses of the Growth Theorem, assume $m \le R^{(n+1)}(t) \le M$ for all t in $[a, x]$. Then

$$R(x) \text{ is between } m\frac{(x-a)^{n+1}}{(n+1)!} \text{ and } M\frac{(x-a)^{n+1}}{(n+1)!}.$$

We remark that the bounds in the refined growth theorem hold even when x is less than a. In these cases $x - a$ is negative and $(x-a)^n$ is negative when n is odd and is positive if n is even.

EXAMPLE 3. Let $R(x) = e^x - \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}\right)$. Show that $\frac{1}{1152} \le R\left(\frac{1}{2}\right) \le \frac{1}{128}$. Use this estimate to obtain approximations, with error bounds, for $\sqrt{e} = e^{1/2}$.

SOLUTION With $R(x) = e^x - (1 + x + x^2/2! + x^3/3!)$ we find $R'(x) = e^x - (1 + x + x^2/2!)$, $R''(x) = e^x - (1 + x)$, $R^{(3)}(x) = e^x - 1$, and $R^{(4)}(x) = e^x$. Thus $R(0) = R'(0) = R''(0) = R^{(3)}(0) = 0$.

For x in $I = (-1, 1)$, $e^{-1} \le e^x = R^{(4)}(x) \le e^1$. Assuming that e is less than 3, we therefore have $1/3 < R^{(4)}(x) < 3$. The refined growth theorem with $a = 0$, $n = 3$, $m = 1/3$, $M = 3$, and $x = 1/2$ gives

$$\frac{1}{3}\frac{(1/2)^4}{4!} \le R(1/2) \le 3\frac{(1/2)^4}{4!}.$$

So

$$\frac{1}{1152} \le \sqrt{e} - \left(1 + \frac{1}{2} + \frac{(1/2)^2}{2!} + \frac{(1/2)^3}{3!}\right) \le \frac{1}{128}$$

or

$$\frac{79}{48} + \frac{1}{1152} \le \sqrt{e} \le \frac{79}{48} + \frac{1}{128}.$$

and finally

$$1.64670 \le \sqrt{e} \le 1.65365.$$

As you can check with your calculator, $\sqrt{e} \approx 1.64872$ to five decimal places. ♦

As Example 3 shows, the refined growth theorem provides both upper and lower bounds on the error made when approximating a function with a polynomial.

Proofs of the Growth Theorems

Repeated application of Lemma 5.4.1 obtains the growth theorem and the refined growth theorem.

Instead of a complete proof of the refined growth theorem, we provide a proof only when $n = 2$. It illustrates the ideas used in the proof for any n.

Proof of the Refined Growth Theorem (Theorem 5.4.3 for $n = 2$

For convenience, we take the case $x > a$. The case with $x < a$ is similar, but is complicated by the fact that $x - a$ is negative and the sign of $(x-a)^n$ depends on whether n is odd or even.

Assume $R(a) = R'(a) = R''(a) = 0$ and $R^{(3)}$ is continuous on $[a, x]$.

Let M be the maximum of $R^{(3)}(t)$ and m the minimum of $R^{(3)}(t)$ on the closed interval $[a, x]$. Thus

$$m \le R^{(3)}(t) \le M \qquad \text{for all } t \text{ in } [a, x].$$

To see what the inequality $R^{(3)}(t) \le M$ implies about $R(x)$ rewrite $R^{(3)}(t)$ as the derivative of $R''(t)$ and rewrite M as the derivative of $M(t-a)$. Thus, we rewrite $R^{(3)}(t) \le M$ as

$$\frac{d}{dt}\left(R''(t)\right) \le \frac{d}{dt}\left(M(t-a)\right).$$

The functions $f(t) = R''(t)$ and $g(t) = M(t-a)$ satisfy $f(a) = g(a) = 0$ and $f'(t) \le g'(t)$ on $[a, x]$. Then, by Lemma 5.4.1,

$$R^{(2)}(t) \le M(t-a) \qquad \text{for all } t \text{ in } [a, x]. \tag{5.4.2}$$

Next, rewrite (5.4.2) as

$$\frac{d}{dt}\left(R'(t)\right) \le \frac{d}{dt}\left(M\frac{(t-a)^2}{2}\right).$$

Let $f(t) = R'(t)$ and $g(t) = M(t-a)^2/2$. Since $f(a) = g(a)$ and $f'(t) \le g'(t)$ on $[a, x]$, the lemma implies that

$$R'(t) \le M\frac{(t-a)^2}{2} \qquad \text{for all } t \text{ in } [a, x]. \tag{5.4.3}$$

Finally, rewrite (5.4.3) as

$$\frac{d}{dt}\left(R(t)\right) \le \frac{d}{dt}\left(M\frac{(t-a)^3}{3\cdot 2}\right).$$

This time, with $f(t) = R(t)$ and $g(t) = M(t-a)^3/3!$, the lemma allows us to conclude that

$$R(t) \le M\frac{(t-a)^3}{3!} \qquad \text{for all } t \text{ in } [a, x]. \tag{5.4.4}$$

Similar reasoning starting with $m \le R^{(3)}(t)$ shows that

$$m\frac{(t-a)^3}{3!} \le R(t) \qquad \text{for all } t \text{ in } [a, x]. \tag{5.4.5}$$

Combining (5.4.4) and (5.4.5) gives

$$m\frac{(t-a)^3}{3!} \le R(t) \le M\frac{(t-a)^3}{3!} \qquad \text{for all } t \text{ in } [a, x].$$

Replacing t by x gives the bounds in the refined growth theorem. ●

The proof of the growth theorem follows immediately from the refined growth theorem. We illustrate the proof with $n = 2$.

Proof of the Growth Theorem (Theorem 5.4.2)

We have just shown that $R(x)$ is between $m(x-a)^3/3!$ and $M(x-a)^3/3!$, and want to show there is a number c in $[a,x]$ such that

$$R(x) = R^{(3)}(c)\frac{(x-a)^3}{3!}.$$

Because $R^{(3)}$ is continuous on $[a,x]$ and m and M are its minimum and maximum values on $[a,x]$, $R^{(3)}(t)$ assumes all values between m and M. Therefore, $R^{(3)}(t)\frac{(x-a)^3}{3!}$, viewed as a function of t with x fixed, assumes all values between $m(x-a)^3/3!$ and $M(x-a)^3/3!$. Since $R(x)$ is between these two values the intermediate value theorem assures us that there is a number c in $[a,x]$ such that

$$R(x) = R^{(3)}(c)\frac{(x-a)^3}{3!}.$$ ●

Summary

If all the derivatives up through the n^{th} derivative of a function R are 0 at a, then

$$R(x) = R^{(n+1)}(c)\frac{(x-a)^{n+1}}{(n+1)!} \qquad \text{for some } c \text{ between } a \text{ and } x.$$

The number c depends on n, as well as on a, x, and the function $R(x)$.

This result, called the growth theorem, is a consequence of the refined growth theorem, which states that if m is the minimum value and M is the maximum value of $R^{(n+1)}(t)$ on $[a,x]$, then

$$R(x) \text{ is between } m\frac{(x-a)^{n+1}}{(n+1)!} \text{ and } M\frac{(x-a)^{n+1}}{(n+1)!}.$$

Both growth theorems are based on Lemma 5.4.1, which, informally speaking, says that when two runners start side-by-side the slower runner never passes the faster one.

EXERCISES for Section 5.4

1. If $f'(x) \geq 3$ for all x in $(-\infty,\infty)$ and $f(0) = 0$, what can be said about $f(2)$? about $f(-2)$?

2. If $f'(x) \geq 2$ for all x in $(-\infty,\infty)$ and $f(1) = 0$, what can be said about $f(3)$? about $f(-3)$?

3. What can be said about $f(2)$ if $f(1) = 0$, $f'(1) = 0$, and $2.5 \leq f''(x) \leq 2.6$ for all x?

4. What can be said about $f(4)$ if $f(1) = 0$, $f'(1) = 0$, and $2.9 \leq f''(x) \leq 3.1$ for all x?

5. A car starts from rest and travels for 4 hours. Its acceleration is always at least 5 miles per hour per hour, but never exceeds 12 miles per hour per hour. What can be said about the distance traveled during the 4 hours?

6. A car starts from rest and travels for 6 hours. Its acceleration is always at least 4.1 miles per hour per hour, but never exceeds 15.5 miles per hour per hour. What can be said about the distance traveled during the 6 hours?

7. State the growth theorem for $x \geq a$ when the first five derivatives of R are continuous and $R(a) = R'(a) = R''(a) = R^{(3)}(a) = R^{(4)}(a) = 0$.

8. State the growth theorem in words, using as little mathematical notation as possible.

9. If $R(1) = R'(1) = R''(1) = 0$, $R^{(3)}(x)$ is continuous on an interval thath includes 1 and 4, and $R^{(3)}(x) \leq 2$ on this interval, what can be said about $R(4)$?

10. If $R(3) = R'(3) = R''(3) = R^{(3)}(3) = R^{(4)}(3) = 0$ and $R^{(5)}(x) \le 6$ on an interval that contains both $x = 3$ and $x = 3.5$, what can be said about $R(3.5)$?

11. Let $R(x) = \sin(x) - \left(x - \frac{x^3}{6}\right)$. Show that

(a) $R(0) = R'(0) = R''(0) = R^{(3)}(0) = 0$.
(b) $R^{(4)}(x) = \sin(x)$.
(c) $|R(x)| \le \frac{x^4}{24}$.
(d) Use $x - \frac{x^3}{6}$ to approximate $\sin(x)$ for $x = \frac{1}{2}$.
(e) Use (c) to estimate the difference between the exact value for $\sin\left(\frac{1}{2}\right)$ and the approximation given in (d).
(f) Explain why $|R(x)| \le \frac{|x|^5}{120}$. How can this be used to obtain a better estimate of the difference between the exact value for $\sin\left(\frac{1}{2}\right)$ and the approximation given in (d)?
(g) By how much does the estimate in (d) differ from $\sin\left(\frac{1}{2}\right)$?

$\frac{1}{2}$ radian $= \frac{1}{2}\frac{180°}{\pi} \approx 29°$

12. Let $R(x) = \cos(x) - \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!}\right)$. Show that

(a) $R(0) = R'(0) = R''(0) = R^{(3)}(0) = R^{(4)}(0) = R^{(5)}(0) = 0$.
(b) $R^{(6)}(x) = -\cos(x)$.
(c) $|R(x)| \le \frac{x^6}{6!}$.
(d) Use $1 - \frac{x^2}{2!} + \frac{x^4}{4!}$ to estimate $\cos(1)$.
(e) By how much does the estimate in (d) differ from $\cos(1)$?

13. Let $R(x) = (1+x)^5 - (1 + 5x + 10x^2)$. Show that

(a) $R(0) = R'(0) = R''(0) = 0$.
(b) $R^{(3)}(x) = 60(1+x)^2$.
(c) $|R(x) \le 40|x|^3$ on $[-1, 1]$.
(d) Use $1 + 5x + 10x^2$ to estimate $(1+x)^5$ for $x = 0.2$.
(e) By how much does the estimate in (d) differ from $(1.2)^5$?

14. If $f(3) = 0$ and $f'(x) \ge 2$ for all x in $(-\infty, \infty)$, what can be said about $f(1)$? Explain.

15. If $f(0) = 3$ and $f'(x) \ge -1$ for all x in $(-\infty, \infty)$, what can be said about $f(2)$ and about $f(-2)$? Explain.

16. Use the polynomial in Example 3 to estimate e, with error bounds.

In Example 2 the polynomial $1 - \frac{x^2}{2}$ was shown to be a good approximation to $\cos(x)$ for x near 0. Exercise 17 shows how this polynomial was chosen.

17. Let $P(x) = a_0 + a_1 x + a_2 x^2$ be a quadratic polynomial. Find the values of a_0, a_1, and a_2 for which

(a) $\cos(0) - P(0) = 0$
(b) $\cos'(0) - P'(0) = 0$
(c) $\cos''(0) - P''(0) = 0$
(d) Let $R(x) = \cos(x) - P(x)$, with the a_i's found in parts (a)-(c). Why are $R(0) = R'(0) = R''(0) = 0$?

---

18. Let $R(x) = \tan(x) - \left(a_0 + a_1 x + a_2 x^2 + a_3 x^3\right)$. Find a_0, a_1, a_2, and a_3 so that $R(0) = R'(0) = R''(0) = R^{(3)}(0) = 0$.

19. Find constants a_0, a_1, a_2, and a_3 so that if $R(x) = \sqrt{1+x} - \left(a_0 + a_1 x + a_2 x^2 + a_3 x^3\right)$ then $R(0) = R'(0) = R''(0) = R^{(3)}(0) = 0$.

Exercises 20 to 24 are related.

20. Because $e > 1$, it follows that $e^x \geq 1$ for every $x \geq 0$.
- (a) Use Lemma 5.4.1 to deduce that $e^x > 1 + x$ for $x > 0$.
- (b) Use (a) and Lemma 5.4.1 to deduce that $e^x > 1 + x + \frac{x^2}{2!}$ for $x > 0$.
- (c) Use (b) and Lemma 5.4.1 to deduce that $e^x > 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}$ for $x > 0$.
- (d) In view of (a), (b), and (c), what general inequality can be proved by this approach?

21. Let k be a positive number. For x in $[0,k]$, $e^x \leq e^k$.
- (a) Deduce that $e^x \leq 1 + e^k x$ for x in $[0,k]$.
- (b) Deduce that $e^x \leq 1 + x + e^k \frac{x^2}{2!}$ for x in $[0,k]$.
- (c) Deduce that $e^x \leq 1 + x + \frac{x^2}{2!} + e^k \frac{x^3}{3!}$ for x in $[0,k]$.
- (d) In view of (a), (b), and (c), what general inequality can be proved by this approach?

22. Combine the results of Exercises 20 and 21 to estimate $e = e^1$ to two decimal places. (Assume $e \leq 3$.)

23. What properties of e^x did you use in Exercises 20 and 21?

24. Let $E(x)$ be a function such that $E(0) = 1$ and $E'(x) = E(x)$ for all x.
- (a) Show that $E(x) \geq 1$ for all $x \geq 0$.
- (b) Use (a) to show that $E(x)$ is an increasing function for all $x \geq 0$. (Show that $E'(x) \geq 1$, for all $x \geq 0$.)
- (c) Show $E(x) \geq 1 + x + \frac{x^2}{2}$ for all $x \geq 0$.

Exercises 25 to 31 show that $\lim_{x\to\infty} \frac{x}{e^x}$, $\lim_{x\to\infty} \frac{\ln(y)}{y}$, $\lim_{x\to 0^+} x\ln(x)$, $\lim_{x\to\infty} \frac{x^k}{b^x}$ $(b > 1)$, and $\lim_{x\to 0^+} x^x$ are closely connected (in the sense that if you know one of these limits you can deduce the other four).

Exercises 25 and 26 use the inequality $e^x > 1 + x + \frac{x^2}{2}$ for all $x > 0$. (See Exercise 20.)

25. Evaluate $\lim_{x\to\infty} \frac{x}{e^x}$.

26. Evaluate $\lim_{y\to\infty} \frac{\ln(y)}{y}$.

Exercises 27 and 28 provide two proofs of the fact that the exponential function grows faster than any power of x.

27. (a) Let n be a positive integer. Write $\frac{x^n}{e^x} = \left(\frac{x}{e^{x/n}}\right)\left(\frac{x}{e^{x/n}}\right)\cdots\left(\frac{x}{e^{x/n}}\right)$. Let $y = x/n$ so that $\frac{x}{e^{x/n}} = \frac{ny}{e^y}$. Use Exercise 25 n times to show that $\lim_{x\to\infty} \frac{x^n}{e^x} = 0$. (b) Deduce that for any fixed number k, $\lim_{x\to\infty} \frac{x^k}{e^x} = 0$.

28. Show that for any positive integer n, $\lim_{x\to\infty} \frac{x^n}{e^x} = 0$, using Exercise 20(d).

29. Evaluate $\lim_{x\to 0^+} x\ln(x)$ as follows: Let $x = \frac{1}{t}$. Then $x\ln(x) = \frac{1}{t}\ln\left(\frac{1}{t}\right) = \frac{-\ln(t)}{t}$, and refer to Exercise 26.

30. Evaluate $\lim_{x\to 0^+} x^x$ as follows: Let $y = x^x$. Then $\ln(y) = x\ln(x)$, a limit that was evaluated in Exercise 29. Explain why $\ln(y) \to 0$ implies $y \to 1$.

31. Evaluate $\lim_{x\to\infty} \frac{x^k}{b^x}$ for any $b > 1$ and k is a positive integer.

---

32. In Example 1 it is shown that $\left|e^x - 1 - x\right| \le \frac{e}{2}x^2$ for all x in $(-1,1)$. Find a bound for

(a) $R(x) = e^x - 1 - x$ on $(-2,1)$
(b) $R(x) = e^x - 1 - x$ on $(-1,2)$
(c) $R(x) = e^x - 1 - x - \frac{x^2}{2}$ on $(-1,1)$
(d) $R(x) = e^x - 1 - x - \frac{x^2}{2}$ on $(-2,1)$
(e) $R(x) = e^x - 1 - x - \frac{x^2}{2}$ on $(-1,2)$
(f) $R(x) = e^x - 1 - x - \frac{x^2}{2} - \frac{x^3}{6}$ on $(-1,1)$
(g) $R(x) = e^x - 1 - x - \frac{x^2}{2} - \frac{x^3}{6}$ on $\left(\frac{-1}{2}, \frac{1}{2}\right)$

33. Explain why $f(a) = g(a)$ and $f'(x) \le g'(x)$ on $[b,a]$ with $a > b$ implies $f(x) \ge g(x)$ for all x in $[b,a]$.

34. Sam was overheard making the following proposal:

> As usual, I can do things more simply than the text. For instance, say $R(a) = R'(a) = R''(a) = 0$ and $R^{(3)}(x) \le M$. I'll show how M affects the size of $R(x)$ for $x > a$.
>
> By the mean value theorem, $R(x) = R(x) - R(a) = R'(c_1)(x-a)$ for some c_1 in $[a,x]$. Then I use the MVT again, this time finding $R'(c_1) = R'(c_1) - R'(a) = R''(c_2)(c_1 - a)$ for some c_2 in $[a,c_1]$. One more application of this idea gives $R''(c_2) = R''(c_2) - R''(a) = R^{(3)}(c_3)(c_2 - a)$.
>
> Since c_1, c_2, and c_3 are in $[a,x]$, I can say that
>
> $$R(x) \le M(x-a)(c_1-a)(c_2-a).$$
>
> Since c_1 and c_2 are in $[a,x]$, I can say that
>
> $$R(x) \le M(x-a)^3.$$
>
> I didn't need that lemma about two runners.

Is Sam correct? Is this a valid substitute for the text's treatment? Explain.

35. Prove the refined growth theorem for $x > a$ and $n = 3$.

36. Prove the refined growth theorem for $x > a$ and $n = 4$.

37. The proof of the refined growth theorem when x is less than a is slightly different than the proof when x is greater than a. Prove it for the case $n = 4$. Here $(x-a)^3$ and $(x-a)$ are negative because $x < a$.

5.5 Taylor Polynomials and Their Errors

We spend years learning how to add, subtract, multiply, and divide. These operations are built into any calculator or computer. Both we and machines can evaluate a polynomial, such as

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n,$$

when x and the coefficients a_0, a_1, a_2, ..., a_n are given. Only multiplication and addition are needed. But how do we evaluate e^x? We resort to our calculators.

Observation 5.5.1: *Why e^x Cannot be a Polynomial*

If e^x were a polynomial in disguise, then it would be easy to evaluate it by finding the polynomial and evaluating it. There are many reasons why e^x cannot be a polynomial. Here are three:

(a) Because e^x equals its own derivative and no polynomial equals its own derivative (other than the polynomial that has constant value 0).

(b) When you differentiate a nonconstant polynomial, you get a polynomial with a lower degree.

(c) Also, $e^x \to 0$ as $x \to -\infty$ and no nonconstant polynomial has this property.

Since we cannot write e^x as a polynomial, we settle for the next best thing. We look for a polynomial that closely approximates e^x. No polynomial can be a good approximation of e^x for all x, since e^x grows too fast as $x \to \infty$. We search instead for a polynomial that is close to e^x for x in some interval.

In this section we develop a method to construct polynomial approximations to functions. Its accuracy can be determined using the growth theorem. As we saw in Section 5.4, higher derivatives play a pivotal role.

Fitting a Polynomial, Near 0

Suppose we want to find a polynomial that closely approximates a function $y = f(x)$ for x near 0. For instance, what polynomial $p(x)$ of the form $a_0 + a_1x + a_2x^2 + a_3x^3$ might produce a good fit?

First we insist that

$$p(0) = f(0) \tag{5.5.1}$$

so the approximation is exact when $x = 0$.

Second, we would like the slope of the graph of $p(x)$ to be the same as that of $f(x)$ when x is 0. Therefore, we require

$$p'(0) = f'(0). \tag{5.5.2}$$

There are many polynomials that satisfy (5.5.1) and (5.5.2). To find the best choices for the four numbers a_0, a_1, a_2, and a_3 we need four equations. One option is to continue the pattern started by (5.5.1) and (5.5.2). We also insist that

$$p''(0) = f''(0) \quad \text{and} \quad p^{(3)}(0) = f^{(3)}(0). \tag{5.5.3}$$

The first equation in (5.5.3) forces the polynomial $p(x)$ to (also) have the same sense of concavity as the function $f(x)$ at $x = 0$. We expect the graphs of $f(x)$ and $p(x)$ to resemble each other, at least for x close to a.

To find the unknowns a_0, a_1, a_2, and a_3 we first compute $p(x)$, $p'(x)$, $p''(x)$, and $p^{(3)}(x)$ at 0. Table 5.5.1 displays the computations that yield formulas for the unknowns in terms of $f(x)$ and its derivatives. For example, we compute $p''(x) = 2a_2 + 3\cdot 2a_3x$ and evaluate it at 0 to obtain $p''(0) = 2a_2 + 3\cdot 2a_3 \cdot 0 = 2a_2$. Then we obtain an equation for a_2 by equating $p''(0)$ and $f''(0)$: $2a_2 = f''(0)$, so $a_2 = f''(0)/2$.

$p(x)$ and its derivatives	Their values at 0	Equation for a_k	Formula for a_k
$p(x) = a_0 + a_1x + a_2x^2 + a_3x^3$	$p(0) = a_0$	$a_0 = f(0)$	$a_0 = f(0)$
$p^{(1)}(x) = a_1 + 2a_2x + 3a_3x^2$	$p^{(1)}(0) = a_1$	$a_1 = f^{(1)}(0)$	$a_1 = f^{(1)}(0)$
$p^{(2)}(x) = 2a_2 + 3\cdot 2a_3x$	$p^{(2)}(0) = 2a_2$	$2a_2 = f^{(2)}(0)$	$a_2 = \frac{1}{2}f^{(2)}(0)$
$p^{(3)}(x) = 3\cdot 2a_3$	$p^{(3)}(0) = 3\cdot 2a_3$	$3\cdot 2a_3 = f^{(3)}(0)$	$a_3 = \frac{1}{3\cdot 2}f^{(3)}(0)$

Table 5.5.1

We can write a general formula for a_k if we let $f^{(0)}(x)$ denote $f(x)$ and recall that $0! = 1$ (by definition), $1! = 1$, $2! = 2\cdot 1 = 2$, and $3! = 3\cdot 2\cdot 1$. According to Table 5.5.1,

$$a_k = \frac{f^{(k)}(0)}{k!}, \qquad k = 0,1,2,3.$$

Therefore

$$p(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3.$$

The coefficient of x^k is completely determined by the k^{th} derivative of f evaluated at 0. It equals the k^{th} derivative of f at 0 divided by $k!$. This suggests the following definition.

Definition: *Taylor Polynomials at* 0

Let n be a nonnegative integer and let f be a function with derivatives at 0 of all orders through n. Then the polynomial

$$f(0) + f^{(1)}(0)x + \frac{f^{(2)}(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n \tag{5.5.4}$$

is called the **n^{th}-order Taylor polynomial of f centered at** 0 and is denoted $P_n(x;0)$.

A Taylor polynomial centered at 0 is also called a **Maclaurin polynomial**.

Whether $P_n(x;0)$ approximates $f(x)$ for x near 0 is not obvious. First we compute some Maclaurin polynomials. Then we use the growth theorem to see how close they are to the function they are designed to approximate.

EXAMPLE 1. Find the Maclaurin polynomial $P_4(x;0)$ associated with $\dfrac{1}{1-x}$.

SOLUTION The first step is to compute the first four derivatives of $1/(1-x)$, then evaluate them (and $1/(1-x)$) at $x = 0$. Dividing these numbers by suitable factorials gives the coefficients of the Maclaurin polynomial. Table 5.5.2 records the computations.

n	$f^{(n)}(a)$	$f^{(n)}(0)$	$\dfrac{f^{(n)}(0)}{n!}$
0	$f(x) = \dfrac{1}{1-x}$	1	1
1	$f'(x) = \dfrac{1}{(1-x)^2}$	1	1
2	$f''(x) = \dfrac{2}{(1-x)^3}$	2	1
3	$f^{(3)}(x) = \dfrac{3\cdot 2}{(1-x)^4}$	$3\cdot 2$	1
4	$f^{(4)}(x) = \dfrac{4\cdot 3\cdot 2}{(1-x)^5}$	$4\cdot 3\cdot 2$	1

Table 5.5.2

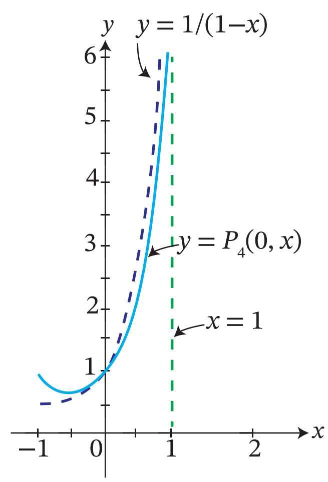


Figure 5.5.1

So the fourth-degree Maclaurin polynomial is

$$P_4(x;0) = 1 + \frac{1}{1!}x + \frac{2}{2!}x^2 + \frac{3\cdot 2}{3!}x^3 + \frac{4\cdot 3\cdot 2}{4!}x^4 = 1 + x + x^2 + x^3 + x^4.$$

Figure 5.5.1 suggests that $P_4(x;0)$ does a good job of approximating $1/(1-x)$ for x near 0. ◆

The calculations in Example 1 motivate the next definitions.

Definition: *Maclaurin Polynomials for* $\frac{1}{1-x}$

The Maclaurin polynomial $P_n(x;0)$ associated with $\frac{1}{1-x}$ is

$$1+x+x^2+x^3+\cdots+x^n.$$

And, because all the derivatives of e^x at 0 are 1,

Definition: *Maclaurin Polynomials for* e^x

The Maclaurin polynomial $P_n(x;0)$ associated with e^x is

$$1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots+\frac{x^n}{n!}.$$

EXAMPLE 2. Find the Maclaurin polynomial $P_5(x;0)$ for $f(x)=\sin(x)$.

SOLUTION Again we make a table for computing the coefficients of the Taylor polynomial centered at 0. (See Table 5.5.3.)

n	$f^{(n)}(x)$	$f^{(n)}(0)\vert$
0	$f^{(0)}(x) = \sin(x)$	$f^{(0)}(0) = \sin(0)=0$
1	$f^{(1)}(x) = \cos(x)$	$f^{(1)}(0) = \cos(0)=1$
2	$f^{(2)}(x) = -\sin(x)$	$f^{(2)}(0) = -\sin(0)=0$
3	$f^{(3)}(x) = -\cos(x)$	$f^{(3)}(0) = -\cos(0)=-1$
4	$f^{(4)}(x) = \sin(x)$	$f^{(4)}(0) = \sin(0)=0$
5	$f^{(5)}(x) = \cos(x)$	$f^{(5)}(0) = \cos(0)=1$

Table 5.5.3

Thus, the Maclaurin polynomial for $f(x)=\sin(x)$ at $x=0$ is

$$\begin{aligned}P_5(x;0)&=f^{(0)}(0)+f^{(1)}(0)x+\frac{f^{(2)}(0)}{2!}x^2+\frac{f^{(3)}(0)}{3!}x^3+\frac{f^{(4)}(0)}{4!}x^4+\frac{f^{(5)}(0)}{5!}x^5\\&=0+(1)x+\frac{0}{2!}x^2+\frac{-1}{3!}x^3+\frac{0}{4!}x^4+\frac{1}{5!}x^5\\&=x-\frac{x^3}{3!}+\frac{x^5}{5!}.\end{aligned}$$

Figure 5.5.2 illustrates the graphs of $P_5(x;0)$ and $\sin(x)$ near 0.

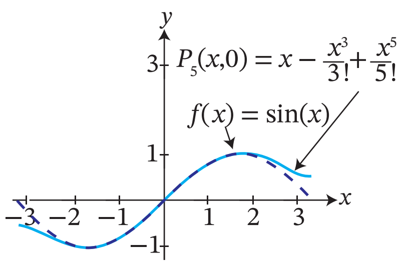


Figure 5.5.2

x	$\sin(x)$	$P_5(x;0)$
0.0	0.000000	0.000000
0.1	0.099833	0.099833
0.5	0.479426	0.479427
1.0	0.841471	0.841667
2.0	0.909297	0.933333
π	0.000000	0.524044
2π	0.000000	46.546732

Table 5.5.4

Having found the fifth-order Maclaurin polynomial for $\sin(x)$, how good of an approximation is it? Table 5.5.4 compares values computed to six-decimal-place accuracy for inputs both near 0 and far from 0. The closer x is to 0, the better the approximation is. When x is large, $\sin(x)$ stays between -1 and 1 but $P_5(x;0)$ gets large. ♦

Definition: *Maclaurin Polynomials for* $\sin(x)$

The Maclaurin polynomials associated with $\sin(x)$ have only odd powers and their terms alternate in sign:

$$P_m(x;0) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \pm \frac{x^m}{m!}.$$

The $\pm$ in front of $x^m/m!$ indicates the coefficient is either positive or negative. For the terms involving x, x^5, x^9, ..., the coefficient is $+1$. For x^3, x^7, x^{11}, ... it is -1. If m is odd, $m = 2n+1$ for some integer n. If n is even, the coefficient of x^{2n+1} is $+1$. If n is odd, the coefficient of x^{2n+1} is -1. The compact notation to write the typical summand is

$$(-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

So we may write

$$P_{2n+1}(x;0) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

Taylor Polynomials Centered at a

We may be interested in estimating a function $f(x)$ near a number a other than 0. Then we express the approximating polynomial in terms of powers of $x - a$ instead of powers of $x = x - 0$ and make the derivatives of the approximating polynomial, evaluated at a, coincide with the derivatives of the function at a. Calculations similar to those that gave us the Maclaurin polynomial (5.5.4) produce the polynomial that will be called a Taylor polynomial centered at a. (If a is not 0, it is not called a Maclaurin polynomial.)

Definition: *Taylor Polynomials of* n^{th} *order,* $P_n(x;a)$.

If the function f has derivatives through order n at a, then the **n^{th}-order Taylor polynomial of f centered at a** is defined as

$$f(a) + f^{(1)}(a)(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

and is denoted $P_n(x;a)$.

The Taylor polynomial $P_n(x,a)$ has the property that $P_n^{(i)}(x,a) = f^{(i)}(x)$, $i = 0, 1, \dots, n$, when both functions are evaluated at $x = a$.

EXAMPLE 3. Find the n^{th}-order Taylor polynomial centered at a for $f(x) = e^x$.

SOLUTION All the derivatives of e^x evaluated at a are e^a. Thus

$$P_n(x;a) = e^a + e^a(x-a) + \frac{e^a}{2!}(x-a)^2 + \frac{e^a}{3!}(x-a)^3 + \cdots + \frac{e^a}{n!}(x-a)^n.$$

◆

The Error in Using A Taylor Polynomial

There is no point using $P_n(x;a)$ to estimate $f(x)$ if we have no idea how large the difference between $f(x)$ and $P_n(x;a)$ may be. So let us look at the difference between these two numbers.

Define the **remainder** to be the difference between the function $f(x)$ and the Taylor polynomial $P_n(x;a)$. Denote the remainder as $R_n(x;a)$. Then

$$f(x) = P_n(x;a) + R_n(x;a).$$

We will be interested in the absolute value of the remainder and will call $|R_n(x;a)|$ the **error** in using $P_n(x;a)$ to approximate $f(x)$. While the sign of the remainder indicates if the approximation is too large or too small, the error reflects only the distance to the exact value.

Theorem 5.5.2: The Lagrange Form of the Remainder

Assume that $f(x)$ has continuous derivatives of orders through $n+1$ in an interval that includes the numbers a and x. Let $P_n(x;a)$ be the n^{th}-order Taylor polynomial associated with $f(x)$ in powers of $x-a$. Then there is a number c between a and x such that

$$R_n(x;a) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

Proof of the Lagrange Form of the Remainder (Theorem 5.5.2)

For simplicity, denote the remainder $R_n(x;a) = f(x) - P_n(x;a)$ by $R(x)$.

Since $P_n(a;a) = f(a)$,

$$R(a) = f(a) - P_n(a;a) = f(a) - f(a) = 0.$$

Similarly, repeated differentiation of $R(x)$ leads to

$$R^{(k)}(x) = f^{(k)}(x) - P_n^{(k)}(x;a),$$

for each integer k, $1 \le k \le n$. From the definition of $P_n(x;a)$,

$$R^{(k)}(a) = f^{(k)}(a) - P_n^{(k)}(a;a) = 0.$$

Since $P_n(x;a)$ is a polynomial of degree at most n, its $(n+1)^{\text{st}}$ derivative is 0. As a result, the $(n+1)^{\text{st}}$ derivative of $R(x)$ is the same as the $(n+1)^{\text{st}}$ derivative of $f(x)$. Thus, $R(x)$ satisfies all the assumptions of the growth theorem. The conclusion is (5.4.1) in Section 5.4 with $R^{(n+1)}(c)$ replaced by $f^{(n+1)}(c)$. ●

EXAMPLE 4. Discuss the error in using $x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!}$ to estimate $\sin(x)$ for $x > 0$.

SOLUTION Example 2 showed that $x - x^3/3! + x^5/5!$ is the fifth-order Maclaurin polynomial, $P_5(x;0)$, associated with $\sin(x)$. In this case $f(x) = \sin(x)$ and each derivative of $f(x)$ is either $\pm\sin(x)$ or $\pm\cos(x)$. Therefore, $\left|f^{n+1}(c)\right|$ is at most 1, and we have

$$\frac{\left|f^{5+1}(c)\right|}{6!}x^6 \le \frac{x^6}{6!}.$$

Then

$$\left|\sin(x) - \left(x - \frac{x^3}{6} + \frac{x^5}{120}\right)\right| \le \frac{|x|^6}{6!} = \frac{x^6}{720}.$$

For instance, with $x = \frac{1}{2}$,

$$\left|\sin\left(\frac{1}{2}\right) - \left(\left(\frac{1}{2}\right) - \frac{1}{6}\left(\frac{1}{2}\right)^3 + \frac{1}{120}\left(\frac{1}{2}\right)^5\right)\right| \le \frac{1}{720}\left(\frac{1}{2}\right)^6 = \frac{1}{(64)(720)} = \frac{1}{46,080} \approx 0.0000217 = 2.17\times10^{-5}.$$

This means the approximation

$$P_5\left(\frac{1}{2};0\right) = \frac{1}{2} - \frac{1}{3!}\left(\frac{1}{2}\right)^3 + \frac{1}{5!}\left(\frac{1}{2}\right)^5 = \frac{1}{2} - \frac{1}{48} + \frac{1}{3840} = \frac{1841}{3840} \approx 0.4794271$$

differs from $\sin\left(\frac{1}{2}\right)$ by less than 2.17×10^{-5}. Therefore at least the first four decimal places are correct. The exact value of $\sin(1/2)$ to eight decimal places is 0.47942553 and our estimate is correct to five decimal places. By comparison, a calculator gives $\sin(1/2) \approx 0.479426$, which is also correct to five decimal places. ♦

Observation 5.5.3: ***A Second Error Estimate***

Because every even-order derivatives of $\sin(x)$, evaluated at $x = 0$, is zero, $P_{2n}(x;a) = P_{2n-1}(x;a)$ for every positive integer n. In particular, $P_6(x;a) = P_5(x;a)$. This realization provides a new error estimate for this approximation:

$$\left|\sin(x) - \left(x - \frac{x^3}{6} + \frac{x^5}{120}\right)\right| = |\sin(x) - P_6(x;0)| \le |R_6(x;0)| \le \left|\frac{|x|^7}{7!}\right| = \frac{x^7}{5040}.$$

In particular,

$$R_6(\frac{1}{2};0) \le \frac{1}{5040}\left(\frac{1}{2}\right)^7 \le \frac{1}{645,120} \approx 1.55\times10^{-6}.$$

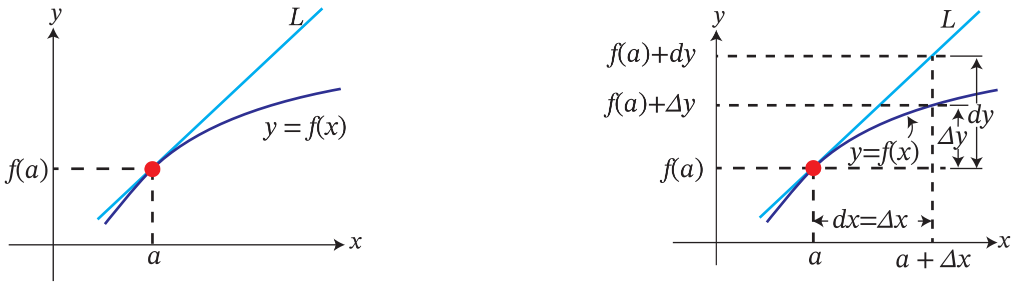
Thus, the fact that this approximation is accurate to five decimal places is not as surprising as it might have seemed when first discovered in Example 4.

The Linear Approximation $P_1(x;a)$

The graph of the Taylor polynomial $P_1(x;a) = f(a) + f'(a)(x-a)$ is a line that passes through the point $(a, f(a))$ and has the same slope as f does at a. That means that the graph of $P_1(x;a)$ is the tangent line to the graph of f at $(a, f(a))$. It is customary to call $P_1(x;a) = f(a) + f'(a)(x-a)$ the **linear approximation** to $f(x)$ for x near a. It is often denoted $L(x)$. Figure 5.5.3(a) shows the graphs of f and L near the point $(a, f(a))$.

We mention this to make a connection with the definition of the derivative: $f'(a) = \lim_{\Delta x\to0}\Delta y/\Delta x$. Let x be a number close to a and define $\Delta x = x - a$ and $\Delta y = f(a+\Delta x) - f(a)$. Often Δx is denoted by dx and $f'(a)dx$ is defined to be dy, as shown in Figure 5.5.3(b). Then dy is an approximation to Δy and $f(a)+dy$ is an approximation to $f(a+\Delta x) = f(a)+\Delta y$.

The expressions dx and dy are called **differentials**. In the seventeenth century, dx and dy referred to **infinitesimals**, infinitely small numbers. Leibniz viewed the derivative as the quotient dy/dx, and his notation for the derivative persists more than three centuries later.



(a) (b)

Figure 5.5.3

Warning: $\frac{dy}{dx}$ is Not a Fraction

The derivative is not a quotient. It is the limit of quotients. This distinction is very important.

The next example uses the linear approximation to estimate $\sqrt{x}$ near $x = 1$.

EXAMPLE 5. Use $P_1(x;1)$ to estimate $\sqrt{x}$ for x near 1. Then discuss the error.

SOLUTION With $f(x) = \sqrt{x}$, then $f'(x) = 1/(2\sqrt{x})$, and $f'(1) = 1/2$. The linear approximation of $f(x)$ near $a = 1$ is

$$P_1(x;1) = f(1) + f'(1)(x-1) = 1 + \frac{1}{2}(x-1)$$

and the remainder is

$$R_1(x;1) = \sqrt{x} - \left(1 + \frac{1}{2}(x-1)\right).$$

Table 5.5.5 shows how rapidly $R_1(x;1)$ approaches 0 as $x \to 1$ and compares it with $(x-1)^2$.

x	$R_1(x;1)$	$(x-1)^2$	$\frac{R_1(x;1)}{(x-1)^2}$
2.0	$\sqrt{2} - \left(1+\frac{1}{2}(2-1)\right) \approx -0.08578643$	1	-0.08579
1.5	$\sqrt{1.5} - \left(1+\frac{1}{2}(1.5-1)\right) \approx -0.02525512$	0.25	-0.10102
1.1	$\sqrt{1.1} - \left(1+\frac{1}{2}(1.1-1)\right) \approx -0.00119115$	0.01	-0.11912
1.01	$\sqrt{1.01} - \left(1+\frac{1}{2}(1.01-1)\right) \approx -0.00001243$	0.0001	-0.12438

Table 5.5.5

The final column in Table 5.5.5 shows that $R_1(x;1)/(x-1)^2$ is nearly constant. Because $(x-1)^2 \to 0$ as $x \to 0$, this means $R_1(x;1)$ approaches 0 at the same rate as the square of $(x-1)$.

Since the Lagrange form for $R_1(x;1)$ is approximately $(f''(1)/2)(x-1)^2$ when x is near 1, $R_1(x;1)/(x-1)^2$ should be near $f''(1)/2$ when x is near 1. As a check, we compute $f''(1)/2$. We have $f''(x) = -x^{-3/2}/4$. Thus $f''(1)/2 = (1/2)(-1/4) = -1/8 = -0.125$. This is consistent with the final column of Table 5.5.5. ♦

Summary

Assume f is a function with n derivatives on an interval that contains the number a. The n^{th}-order Taylor polynomial at a is

$$P_n(x;a) = f(a) + f^{(1)}(a)(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

The "zeroth derivative" of a function is defined to be the function itself. This allows us to say simply that the derivatives $P_n^{(k)}(x;a)$ coincide with $f^{(k)}(a)$ for $k = 0, 1, \ldots, n$.

In addition to $P_n(x;a)$ and f having the same value at a, their first n derivatives at a also agree.

If $a = 0$, $P_n(x;0)$ is call a Maclaurin polynomial.

The Maclaurin polynomial associated with e^x, $\sin(x)$, $\cos(x)$, and $1/(1-x)$ are given in Table 5.5.6.

$f(x)$	$P_n(x;0)$
e^x	$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$
$\sin(x)$	$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^n \frac{x^{2n+1}}{(n+1)!}$
$\cos(x)$	$1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^n \frac{x^{2n}}{(2n)!}$
$\frac{1}{1-x}$	$1 + x + x^2 + x^3 + \cdots + x^n$

Table 5.5.6

The remainder in using the Taylor polynomial of order n to estimate a function involves the $(n+1)^{\text{st}}$ derivative:

$$R_n(x;a) = f(x) - P_n(x;a) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$$

where c is a number between a and x . The error is the absolute value of the remainder, $|R_n(x;a)|$.

The linear approximation to a function near a is $L(x) = P_1(x;a) = f(a) + f'(a)(x-a)$. The differentials are $dx = x - a$ and $dy = f'(a)dx$. While dx and Δx are identical, dy is only an approximation of $\Delta y = f(x+\Delta x) - f(x)$.

EXERCISES for Section 5.5

Use a graphing calculator or computer algebra system to assist with the computations and graphing involved with many of these exercises.

1. Give at least three reasons $\sin(x)$ cannot be a polynomial.

In Exercises 2 to 13 compute the Taylor polynomials. Graph $f(x)$ and $P_n(x;a)$ on the same axes on a domain centered at a. The graph of $P_1(x;a)$ is the tangent line at the point $(a, f(a))$.

2. $f(x) = \frac{1}{1+x}$, $P_1(x;0)$ and $P_2(x;0)$

3. $f(x) = \frac{1}{1+x}$, $P_1(x;1)$ and $P_2(x;1)$

4. $f(x) = \ln(1+x)$, $P_1(x;0)$, $P_2(x;0)$, and $P_3(x;0)$

5. $f(x) = \ln(1+x)$, $P_1(x;1)$, $P_2(x;1)$, and $P_3(x;1)$

6. $f(x) = e^x$, $P_1(x;0)$, $P_2(x;0)$, $P_3(x;0)$, and $P_4(x;0)$

7. $f(x) = e^x$, $P_1(x;2)$, $P_2(x;2)$, $P_3(x;2)$, and $P_4(x;2)$

8. $f(x) = \arctan(x)$, $P_1(x;0)$, $P_2(x;0)$, and $P_3(x;0)$

9. $f(x) = \arctan(x)$, $P_1(x;-1)$, $P_2(x;-1)$, and $P_3(x;-1)$

10. $f(x) = \cos(x)$, $P_2(x;0)$ and $P_4(x;0)$

11. $f(x) = \sin(x)$, $P_7(x;0)$

12. $f(x) = \cos(x)$, $P_6\left(x;\frac{\pi}{4}\right)$

13. $f(x) = \sin(x)$, $P_7\left(x;\frac{\pi}{4}\right)$

---

14. Can there be a polynomial $p(x)$ such that $\sin(x) = p(x)$ for all x in the interval $[1, 1.0001]$? Explain.

15. Can there be a polynomial $p(x)$ such that $\ln(x) = p(x)$ for all x in the interval $[1, 1.0001]$? Explain.

16. State the Lagrange formula for the error in using a Taylor polynomial as an estimate of the value of a function. Use as little mathematical notation as you can.

In Exercises 17 to 22 obtain the Maclaurin polynomial of order n associated with the function.

17. $\dfrac{1}{1-x}$

18. e^x

19. e^{-x}

20. $\sin(x)$

21. $\cos(x)$

22. $\dfrac{1}{1+x}$

Exercises 23 to 26 are related.

23. Let $f(x) = (1+x)^3$.
(a) Find $P_3(x;0)$ and $R_3(x;0)$. (b) Check that your answer to (a) is correct by multiplying out $(1+x)^3$.

24. Let $f(x) = (1+x)^4$.
(a) Find $P_4(x;0)$ and $R_4(x;0)$. (b) Check that your answer to (a) is correct by multiplying out $(1+x)^4$.

25. Let $f(x) = (1+x)^5$. Using $P_5(x;0)$, show that $(1+x)^5 = 1 + 5x + \frac{5\cdot4}{1\cdot2}x^2 + \frac{5\cdot4\cdot3}{1\cdot2\cdot3}x^3 + \frac{5\cdot4\cdot3\cdot2}{1\cdot2\cdot3\cdot4}x^4 + \frac{5\cdot4\cdot3\cdot2\cdot1}{1\cdot2\cdot3\cdot4\cdot5}x^5$.

Definition: *Binomial Coefficient*

For a positive integer n and a nonnegative integer k, with $k \le n$, the **binomial coefficient** is

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{1\cdot2\cdot3\cdots k} = \frac{n!}{k!(n-k)!}.$$

Thus, the formula for $(1+x)^5$ found in Exercise 25 can also be written as

$$(1+x)^5 = \binom{5}{0} + \binom{5}{1}x + \binom{5}{2}x^2 + \binom{5}{3}x^3 + \binom{5}{4}x^4 + \binom{5}{5}x^5.$$

Using $P_n(x;0)$ one can show that for any positive integer n

$$(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n-1}x^{n-1} + \binom{n}{n}x^n = \sum_{k=0}^{n}\binom{n}{k}x^k.$$

Theorem 5.5.4: Binomial Theorem

For any real numbers a and b and any positive integer n: $(a+b)^n = \sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k}$.

26. (a) Using algebra (no calculus) derive the binomial theorem for $(a+b)^3$ from the binomial theorem for $(1+x)^3$.

(b) Verify the binomial theorem for $(a+b)^{12}$ from the special case $(1+x)^{12} = \sum_{k=0}^{12} \binom{12}{k} x^k$.

---

27. Let $f(x) = \sqrt{x}$.

(a) What is the linear approximation $P_1(x;4)$ to $\sqrt{x}$?

(b) Fill in the following table.

x	$R_1(x;4) = f(x) - P_1(x;4)$	$(x-4)^2$	$\dfrac{R_1(x;4)}{(x-4)^2}$
5.0			
4.1			
4.01			
3.99			

(c) Compute $\frac{1}{2}f''(4)$. Explain its relationship with the entries in the fourth column of the table in (b).

28. Repeat Exercise 27 for the linear approximation to $\sqrt{x}$ at $a = 3$. Use $x = 4$, 3.1, 3.01, and 2.99.

29. Assume $f(x)$ has continuous first and second derivatives and that $4 \le f''(x) \le 5$ for all x.

(a) What can be said about the error in using $f(2) + f'(2)(x-2)$ to approximate $f(x)$?

(b) How small should $x - 2$ be to be sure that the error — the absolute value of the remainder — is less than or equal to 0.005? (This ensures the approximate value is correct to 2 decimal places.)

30. Let $f(x) = 2 + 3x + 4x^2$. Find (a) $P_2(x;0)$, (b) $P_3(x;0)$, (c) $P_2(x;5)$, and (d) $P_3(x;5)$.

31. (a) What can be said about the degree of the polynomial $P_n(x;0)$?

(b) When is the degree of $P_n(x;0)$ less than n?

(c) When is the degree of $P_n(x;a)$ less than n?

32. For $f(x) = \dfrac{1}{1-x}$ the error $R_n(x;0)$ in using a Maclaurin polynomial $P_n(x;0)$ to estimate the function can be calculated exactly. Show that it equals $\left|\dfrac{x^{n+1}}{1-x}\right|$.

33. (a) Find $P_5(x;0)$ for $f(x) = \ln(1+x)$.

(b) Find $P_n(x;0)$?

(c) Use $P_5(x;0)$ to estimate $\ln(1.05)$ and bound the error.

In Exercises 34 and 35, use a calculator or computer to help evaluate the Taylor polynomials

34. Let $f(x) = e^x$.

(a) Find $P_{10}(x;0)$.

(b) Compute $f(x)$ and $P_{10}(x;0)$ at $x = 1$, $x = 2$, and $x = 4$.

35. Let $f(x) = \ln(x)$.

(a) Find $P_{10}(x;1)$.

(b) Compute $f(x)$ and $P_{10}(x;1)$ at $x = 1$, $x = 2$, and $x = 4$.

Exercises 36 to 39 involve even and odd functions.

A function is even if $f(-x) = f(x)$ and is odd if $f(-x) = -f(x)$.

36. Show that if f is an odd function, f' is an even function.

37. Show that if f is an even function, f' is an odd function.

38. (a) Which polynomials are even functions?
(b) If f is an even function, are its associated Maclaurin polynomials necessarily even functions? Explain.

39. (a) Which polynomials are odd functions?
(b) If f is an odd function, are its associated Maclaurin polynomials necessarily odd functions? Explain.

---

40. Let $f(x) = e^{-1/x^2}$ if $x \neq 0$ and $f(0) = 0$. This exercise constructs Maclaurin polynomials that do not approximate the associated function. (a) Find $f'(0)$. (b) Find $f''(0)$. (c) Find $P_2(x;0)$. (d) What is $P_{100}(x;0)$?

41. Suppose $f^{(4)}(x) > 0$ on an interval, show that $f(x) \geq f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \frac{1}{6}f'''(a)(x-a)^3$ for each x in that interval.

42. We defined the Taylor polynomials but did not show that at a their derivatives equal those of the function. Show that if $P(x) = P_4(x;a)$, the Taylor polynomial of order 4 associated with $f(x)$, then $P(a) = f(a)$, $P'(a) = f'(a)$, $P''(a) = f''(a)$, $P^{(3)}(a) = f^{(3)}(a)$, and $P^{(4)}(a) = f^{(4)}(a)$.

43. The quantity $\sqrt{1 - \frac{v^2}{c^2}}$ occurs in the theory of relativity. Here v is an object's velocity and c is the velocity of light. Justify the following approximations that physicists use:

REALITY CHECK: Even for a rocket v/c is very small.

(a) $\sqrt{1 - \frac{v^2}{c^2}} \approx 1 - \frac{1}{2}\frac{v^2}{c^2}$

(b) $\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \approx 1 + \frac{1}{2}\frac{v^2}{c^2}$

44. If $P_n(x;0)$ is the Maclaurin polynomial associated with $f(x)$, is $P_n(-x;0)$ the Maclaurin polynomial associated with $f(-x)$? Explain.

45. Let $P(x)$ be the Maclaurin polynomial of the second order associated with $f(x)$. Let $Q(x)$ be the Maclaurin polynomial of the second order associated with $g(x)$. What part, if any, of $P(x)Q(x)$ is a Maclaurin polynomial associated with $f(x)g(x)$? Explain.

5.6 L'Hôpital's Rule for Evaluating Certain Limits

There are two types of limits in calculus: those that can be evaluated at a glance, and those that require some work to evaluate. In Section 2.4 we called limits that can be evaluated easily **determinate** and those that require some work **indeterminate**.

For instance $\lim_{x\to\pi/2} \sin(x)/x$ is clearly $1/(\pi/2) = 2/\pi$. That's easy. But the value of $\lim_{x\to 0} \sin(x)/x$ is not obvious. In Section 2.2 we used a diagram of circles, sectors, and triangles to show that this limit is 1.

In this section we describe a technique for evaluating several common types of indeterminate limits, for instance

$$\lim_{x\to a} f(x)/g(x)$$

when both $f(x)$ and $g(x)$ approach 0 as x approaches a. The numerator is trying to drag $f(x)/g(x)$ toward 0 at the same time as the denominator is trying to make the quotient large. L'Hôpital's rule helps determine which term wins or whether there is a compromise.

L'Hôpital is pronounced lope-ee-tall.

Indeterminate Limits

The following limits are called **indeterminate** because they cannot be determined without knowing more about the functions f and g.

$$\lim_{x\to a}\frac{f(x)}{g(x)}, \text{ where } \lim_{x\to a} f(x) = 0 \text{ and } \lim_{x\to a} g(x) = 0$$

$$\lim_{x\to a}\frac{f(x)}{g(x)}, \text{ where } \lim_{x\to a} f(x) = \infty \text{ and } \lim_{x\to a} g(x) = \infty$$

L'Hôpital's rule provides a way for dealing with these limits and limits that can be transformed to those forms. In short, l'Hôpital's rule is not a magic way to evaluate all limits; it applies only in certain special situations.

Case 1: Zero-over-Zero Limits

Theorem 5.6.1: L'Hôpital's Rule (zero-over-zero case)

Let f and g be differentiable over some open interval that contains a. Assume also that $g'(x)$ is not 0 for any x in that interval except perhaps at a.

If

$$\lim_{x\to a} f(x) = 0, \quad \lim_{x\to a} g(x) = 0, \quad \textit{and} \quad \lim_{x\to a}\frac{f'(x)}{g'(x)} = L,$$

then

$$\lim_{x\to a}\frac{f(x)}{g(x)} = L.$$

Observation 5.6.2: *Concise Version of l'Hôpital's Rule*

To evaluate the limit of a quotient that is indeterminate with form $0/0$, evaluate the limit of the quotient of their derivatives (not the derivative of the quotient).

The proof of Theorem 5.6.1 will be discussed after some examples.

EXAMPLE 1. Find $\displaystyle\lim_{x\to 1}\frac{x^5-1}{x^3-1}$.

SOLUTION Here

$$a = 1,\ f(x) = x^5 - 1, \text{ and } g(x) = x^3 - 1.$$

The assumptions of l'Hôpital's rule are satisfied because

$$\lim_{x\to 1}(x^5-1) = 0 \text{ and } \lim_{x\to 1}(x^3-1) = 0.$$

According to l'Hôpital's rule,

$$\lim_{x\to 1}\frac{x^5-1}{x^3-1} \overset{\text{l'H}}{=} \lim_{x\to 1}\frac{(x^5-1)'}{(x^3-1)'} \text{ provided the second limit exists.}$$

NOTATION: Each application of l'Hôpital's rule is indicated by $\overset{\text{l'H}}{=}$. The two limits are not actually equal until, and unless, the one involving the quotient of the derivatives is known to exist.

Returning our attention to this example, to evaluate the second limit, we see

$$\begin{aligned}
\lim_{x\to 1}\frac{(x^5-1)'}{(x^3-1)'} &= \lim_{x\to 1}\frac{5x^4}{3x^2} && \text{(differentiation of numerator and denominator)}\\
&= \lim_{x\to 1}\frac{5}{3}x^2 && \text{(algebra)}\\
&= \frac{5}{3}.
\end{aligned}$$

Thus, because this limit exists, l'Hôpital's rule says the original limit also exists, and has the same value:

$$\lim_{x\to 1}\frac{x^5-1}{x^3-1} = \frac{5}{3}.$$

◆

Sometimes it may be necessary to apply l'Hôpital's rule more than once, as in the next example.

EXAMPLE 2. Find $\lim_{x\to 0}\frac{\sin(x)-x}{x^3}$.

SOLUTION As $x \to 0$, both numerator and denominator approach 0. By l'Hôpital's rule,

$$\begin{aligned}
\lim_{x\to 0}\frac{\sin(x)-x}{x^3} &\overset{\text{l'H}}{=} \lim_{x\to 0}\frac{(\sin(x)-x)'}{(x^3)'}\\
&= \lim_{x\to 0}\frac{\cos(x)-1}{3x^2}.
\end{aligned}$$

As $x \to 0$, both $\cos(x)-1 \to 0$ and $3x^2 \to 0$. So use l'Hôpital's rule again:

$$\begin{aligned}
\lim_{x\to 0}\frac{\cos(x)-1}{3x^2} &\overset{\text{l'H}}{=} \lim_{x\to 0}\frac{(\cos(x)-1)'}{(3x^2)'}\\
&= \lim_{x\to 0}\frac{-\sin(x)}{6x}.
\end{aligned}$$

Both $\sin(x)$ and $6x$ approach 0 as $x \to 0$. Use l'Hôpital's rule yet another time:

$$\begin{aligned}
\lim_{x\to 0}\frac{-\sin(x)}{6x} &\overset{\text{l'H}}{=} \lim_{x\to 0}\frac{(-\sin(x))'}{(6x)'}\\
&= \lim_{x\to 0}\frac{-\cos(x)}{6}\\
&= \frac{-1}{6}.
\end{aligned}$$

After three applications of l'Hôpital's rule we find that

$$\lim_{x\to 0}\frac{\sin(x)-x}{x^3} = -\frac{1}{6}.$$

◆

The last application of l'Hôpital's rule is unnecessary as $\lim_{x\to 0}\sin(x)/x = 1$ was developed in Section 2.2.

Sometimes a limit may be simplified before l'Hôpital's rule is applied. For instance,

$$\lim_{x\to 0}\frac{(\sin(x)-x)\cos^5(x)}{x^3}.$$

Since $\lim_{x\to 0}\cos^5(x) = 1$, we have

$$\lim_{x\to 0}\frac{(\sin(x)-x)\cos^5(x)}{x^3} = \left(\lim_{x\to 0}\frac{\sin(x)-x}{x^3}\right)\cdot 1,$$

which, by Example 2, is $-1/6$. This saves a lot of work, as may be checked by finding the limit using l'Hôpital's rule without separating $\cos^5(x)$, but the product rule and algebra get progressively more complicated each time l'Hôpital's rule is applied.

Theorem 5.6.1 concerns limits as $x \to a$. L'Hôpital's rule also applies if $x \to \infty$, $x \to -\infty$, $x \to a^+$, or $x \to a^-$. In the first case, we would assume that $f(x)$ and $g(x)$ are differentiable in some interval (c, ∞) and $g'(x)$ is not zero there. In the case of $x \to a^+$, assume that $f(x)$ and $g(x)$ are differentiable in some open interval (a, b) and $g'(x)$ is not 0 there.

Case 2: Infinity-over-Infinity Limits

Theorem 5.6.1 concerns the limit of $f(x)/g(x)$ when both $f(x)$ and $g(x)$ approach 0. A similar problem arises when both $f(x)$ and $g(x)$ get arbitrarily large as $x \to a$ or as $x \to \infty$. The behavior of the quotient $f(x)/g(x)$ will be influenced by how rapidly $f(x)$ and $g(x)$ grow.

If $\lim_{x\to a} f(x) = \infty$ and $\lim_{x\to a} g(x) = \infty$, then $\lim_{x\to a}(f(x)/g(x))$ is an indeterminate form. The next theorem presents a l'Hôpital rule for this case.

Theorem 5.6.3: L'Hôpital's Rule (infinity-over-infinity case)

Let f and g be defined and differentiable for all x larger than some number c. Then, if $g'(x)$ is not zero for all $x > c$,

$$\lim_{x\to\infty} f(x) = \infty, \quad \lim_{x\to\infty} g(x) = \infty, \quad \textit{and} \quad \lim_{x\to\infty} \frac{f'(x)}{g'(x)} = L,$$

it follows that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = L.$$

Just as was noted for Theorem 5.6.1, a similar result holds for limits with $x \to a$, $x \to a^-$, $x \to a^+$, or $x \to -\infty$. Moreover, Theorem 5.6.3 holds when $\lim_{x\to\infty} f(x)$ and $\lim_{x\to\infty} g(x)$ could both be $-\infty$, or one could be ∞ and the other $-\infty$.

EXAMPLE 3. Find $\displaystyle\lim_{x\to\infty} \frac{\ln(x)}{x^2}$.

SOLUTION Since $\ln(x) \to \infty$ and $x^2 \to \infty$ as $x \to \infty$, we may use l'Hôpital's rule in the infinity-over-infinity form. We have

$$\lim_{x\to\infty} \frac{\ln(x)}{x^2} \overset{\text{l'H}}{=} \lim_{x\to\infty} \frac{(\ln(x))'}{(x^2)'} = \lim_{x\to\infty} \frac{1/x}{2x} = \lim_{x\to\infty} \frac{1}{2x^2} = 0.$$

Hence $\lim_{x\to\infty}\left(\ln(x)/x^2\right) = 0$. ◆

Example 3 says that while $\ln(x)$ and x^2 both grow without bound, $\ln(x)$ grows much more slowly than x^2.

EXAMPLE 4. Find $\displaystyle\lim_{x\to\infty} \frac{x - \cos(x)}{x}$.

SOLUTION Both numerator and denominator approach ∞ and $x \to \infty$. Trying l'Hôpital's rule, we obtain

$$\lim_{x\to\infty} \frac{x - \cos(x)}{x} \overset{\text{l'H}}{=} \lim_{x\to\infty} \frac{(x - \cos(x))'}{x'} = \lim_{x\to\infty} \frac{1 + \sin(x)}{1}.$$

L'Hôpital's rule may fail to provide an answer.

But $\lim_{x\to\infty}(1 + \sin(x))$ does not exist, since $\sin(x)$ oscillates back and forth from -1 to 1 as $x \to \infty$.
What can we conclude about the limit in Example 4?

L'Hôpital's rule says that if $\lim_{x\to\infty} f'(x)/g'(x)$ exists, then $\lim_{x\to\infty} f(x)/g(x)$ exists and has the same value. It says nothing when $\lim_{x\to\infty} f'(x)/g'(x)$ does not exist. In fact, in this case, it is not difficult to evaluate the limit in Example 4 directly:

$$\begin{aligned}\lim_{x\to\infty} \frac{x-\cos(x)}{x} &= \lim_{x\to\infty}\left(1-\frac{\cos(x)}{x}\right) && (\text{ algebra })\\ &= 1-0=1. && (\text{ since } |\cos(x)|\le 1\text{ }).\end{aligned}$$

♦

We now turn our attention to understanding why Theorem 5.6.3 is true. Imagine that $f(t)$ and $g(t)$ describe the locations on the x-axis of two cars at time t. Call the cars the f-car and the g-car. Assume the cars are on endless journeys, that is, $\lim_{t\to\infty} f(t)=\infty$ and $\lim_{t\to\infty} g(t)=\infty$. Their velocities are $f'(t)$ and $g'(t)$. See Figure 5.6.1. Assume that as time $t\to\infty$ the f-car tends to travel at a speed closer and closer to L times the speed of the g-car. That is, assume that

$$\lim_{t\to\infty}\frac{f'(t)}{g'(t)}=L.$$

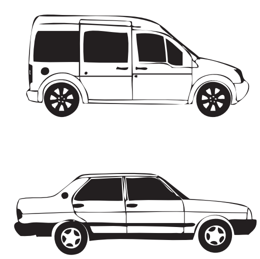


Figure 5.6.1

No matter how the two cars move in the short run, in the long run the f-car will tend to travel about L times as far as the g-car; that is,

$$\lim_{t\to\infty}\frac{f(t)}{g(t)}=L.$$

Transforming Limits So l'Hôpital's Rule is Useful

Many limits can be transformed to limits to which l'Hôpital's rule applies. For instance, the problem of finding

$$\lim_{x\to0^+} x\ln(x)$$

does not fit into l'Hôpital's rule, since it does not involve the quotient of two functions. As $x\to0^+$, one factor, x, approaches 0 and the other factor $\ln(x)$, approaches $-\infty$. This is another type of indeterminate limit, involving a small number times a large number (zero-times-infinity).

Observation 5.6.4: *Limits with Form Zero-Times-Infinity*

It is not obvious how a product involving a small number times a large number behaves in the limit. The limit might not exist and, when it does exist, the value could be positive, zero, or negative. In fact, the limit can turn out to have any value.

A little algebra transforms the zero-times-infinity case into a form to which l'Hôpital's rule applies, as the next example illustrates.

EXAMPLE 5. Find $\lim_{x\to0^+} x\ln(x)$.

SOLUTION Rewrite $x\ln(x)$ as a quotient, $\ln(x)/(1/x)$. Then

$$\lim_{x\to0^+}\ln(x)=-\infty \text{ and } \lim_{x\to0^+}\frac{1}{x}=\infty,$$

so by l'Hôpital's rule,

$$\begin{aligned}\lim_{x\to0^+}\frac{\ln(x)}{1/x} &\overset{\text{l'H}}{=} \lim_{x\to0^+}\frac{1/x}{-1/x^2} && (\text{l'Hôpital's rule: } \infty/\infty)\\ &= \lim_{x\to0^+}(-x) && (\text{simplification } (x\ne0)).\end{aligned}$$

Thus, because this limit exists, Theorem 5.6.3 tells us that

$$\lim_{x\to 0^+} \frac{\ln(x)}{1/x} = \lim_{x\to 0^+} (-x) = 0,$$

from which it follows that $\lim_{x\to 0^+} x\ln(x) = 0$. ♦

From $\lim_{x\to 0^+} x\ln(x) = 0$ we conclude that the factor x, which approaches 0, dominates the factor $\ln(x)$ which approaches $-\infty$ at a slower rate.

The final example illustrates another type of limit that can be found by first relating it to limits to which l'Hôpital's rule applies.

EXAMPLE 6. Evaluate $\lim\limits_{x\to 0^+} x^x$, or determine that it does not exist.

CHALLENGE: Before reading the solution of Example 6, spend a few minutes investigating the behavior of $y = x^x$ for positive values of x near 0. Create graphs. Compute values. What do you think this limit will be?

SOLUTION Since the limit involves an exponential it does not fit directly into l'Hôpital's rule. Maybe some algebraic manipulations can change it to a form covered by l'Hôpital's rule. Let

$$y = x^x.$$

Then

$$\ln(y) = \ln\left(x^x\right) = x\ln(x)$$

By Example 5,

$$\lim_{x\to 0^+} x\ln(x) = 0.$$

Therefore, $\lim_{x\to 0^+} \ln(y) = 0$. By the definition of $\ln(y)$ and the continuity of the exponential function,

$$\lim_{x\to 0^+} y = \lim_{x\to 0^+} \exp(\ln(y)) = \exp\left(\lim_{x\to 0^+} \ln(y)\right) = e^0 = 1.$$

Hence $x^x \to 1$ as $x \to 0^+$. How does this compare with your findings before reading this solution? ♦

Concerning the Proof of the $\frac{0}{0}$-Case of l'Hôpital's Rule (Theorem 5.6.1)

A complete proof of Theorem 5.6.1 is found in Exercises 69 to 72. The following argument is intended to make the l'Hôpital's rule plausible. To do so, consider the special case where f, f', g, and g' are continuous throughout an open interval containing a — so they are defined at a. Assume that $g'(x) \neq 0$ throughout the interval. Since we have $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$, it follows by continuity that $f(a) = 0$ and $g(a) = 0$.

Assume that $\lim_{x\to a} \frac{f'(x)}{g'(x)} = L$. Then

$$\begin{aligned}
\lim_{x\to a} \frac{f(x)}{g(x)} &= \lim_{x\to a} \frac{f(x)-f(a)}{g(x)-g(a)} && (\text{ since } f(a)=0 \text{ and } g(a)=0\,) \\
&= \lim_{x\to a} \frac{\frac{f(x)-f(a)}{x-a}}{\frac{g(x)-g(a)}{x-a}} && (\text{ algebra }) \\
&= \frac{\lim_{x\to a} \frac{f(x)-f(a)}{x-a}}{\lim_{x\to a} \frac{g(x)-g(a)}{x-a}} && (\text{ limit of quotient = quotient of limits }) \\
&= \frac{f'(a)}{g'(a)} && (\text{ definitions of } f'(a) \text{ and } g'(a)\,) \\
&= \frac{\lim_{x\to a} f'(x)}{\lim_{x\to a} g'(x)} && (\, f',\, g' \text{ continuous, by assumption }) \\
&= \lim_{x\to a} \frac{f'(x)}{g'(x)} && (\text{ quotient of limits = limit of quotients }) \\
&= L && (\text{ by assumption }).
\end{aligned}$$

Consequently, when all of the assumptions are met, $\lim_{x\to a} \frac{f(x)}{g(x)} = L$.

Summary

We described l'Hôpital's rule, which is a technique for dealing with limits of the indeterminate forms zero-over-zero and infinity-over-infinity. In both cases

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)},$$

if the second limit exists. The second limit involves the quotient of two derivatives, $f'(x)/g'(x)$, not the derivative of the quotient. Table 5.6.1 shows how some limits of other indeterminate forms can be converted into these forms.

Indeterminate Forms	Name	Conversion Method	New Form
$f(x)g(x)$; $f(x)\to 0,\ g(x)\to\infty$	Zero-times-infinity $(0\cdot\infty)$	Write as $\frac{f(x)}{1/g(x)}$ or as $\frac{g(x)}{1/f(x)}$	$\frac{0}{0}$ or $\frac{\infty}{\infty}$
$f(x)^{g(x)}$; $f(x)\to 1,\ g(x)\to\infty$	One-to-infinity (1^{∞})	Let $y=f(x)^{g(x)}$; take $\ln(y)$, find the limit of $\ln(y)$, and then find the limit of $y=e^{\ln(y)}$	$\ln(y)$ has form $\infty\cdot 0$
$f(x)^{g(x)}$; $f(x)\to 0,\ g(x)\to 0$	Zero-to-zero (0^0)	Same as for 1^{∞}	$\ln(y)$ has form $0\cdot\infty$.

Table 5.6.1

EXERCISES for Section 5.6

In Exercises 1 to 16 find each limit. Before using l'Hôpital's rule, check that it applies. Identify all uses of l'Hôpital's rule, including the type of indeterminate form.

1. $\lim_{x\to 2} \frac{x^3-8}{x^2-4}$

2. $\lim_{x\to 1} \frac{x^7-1}{x^3-1}$

3. $\lim_{x\to 0} \frac{\sin(3x)}{\sin(2x)}$

4. $\lim_{x\to 0} \frac{\sin(x^2)}{(\sin(x))^2}$

5. $\lim_{x\to 0} \frac{\sin(5x)\cos(3x)}{x}$

6. $\lim_{x\to 0} \frac{\sin(5x)\cos(3x)}{x-\frac{\pi}{2}}$

7. $\lim_{x\to \pi/2} \frac{\sin(5x)\cos(3x)}{x}$

8. $\lim_{x\to \pi/2} \frac{\sin(5x)\cos(3x)}{x-\frac{\pi}{2}}$

9. $\lim_{x\to\infty} \frac{x^3}{e^x}$

10. $\lim_{x\to\infty} \frac{x^5}{3^x}$

11. $\lim_{x\to 0} \frac{1-\cos(x)}{x^2}$

12. $\lim_{x\to 0} \frac{\sin(x)-x}{(\sin(x))^3}$

13. $\lim_{x\to 0} \frac{\tan(3x)}{\ln(1+x)}$

14. $\lim_{x\to 1} \frac{\cos(\pi x/2)}{\ln(x)}$

15. $\lim_{x\to 2} \frac{(\ln(x))^2}{x}$

16. $\lim_{x\to 0} \frac{\arcsin(x)}{e^{2x}-1}$

In Exercises 17 to 22 transform the problem into one to which l'Hôpital's rule applies and find the limit. Identify all uses of l'Hôpital's rule, including the type of indeterminate form.

17. $\lim_{x\to 0}(1-2x)^{1/x}$

18. $\lim_{x\to 0}(1+\sin(2x))^{\csc(x)}$

19. $\lim_{x\to 0^+}(\sin(x))^{(e^x-1)}$

20. $\lim_{x\to 0^+} x^2\ln(x)$

21. $\lim_{x\to 0^+}(\tan(x))^{\tan(2x)}$

22. $\lim_{x\to 0^+}(e^x-1)\ln(x)$

In Exercises 23 to 51 find the limits, or explain why the limit does not exist. Use l'Hôpital's rule only if it applies. Identify all uses of l'Hôpital's rule, including the type of indeterminate form.

Warning: Do Not Overuse l'Hôpital's Rule.

Remember that l'Hôpital's rule, carelessly applied, may give a wrong answer or no answer.

23. $\lim_{x\to\infty} \frac{2^x}{3^x}$

24. $\lim_{x\to\infty} \frac{2^x+x}{3^x}$

25. $\lim_{x\to\infty} \frac{\log_2(x)}{\log_3(x)}$

26. $\lim_{x\to 1} \frac{\log_2(x)}{\log_3(x)}$

27. $\lim_{x\to\infty} \left(\frac{1}{x}-\frac{1}{\sin(x)}\right)$

28. $\lim_{x\to\infty} \frac{x^2+3\cos(5x)}{x^2-2\sin(4x)}$

29. $\lim_{x\to\infty} \frac{e^x-1/x}{e^x+1/x}$

30. $\lim_{x\to 0} \frac{3x^3+x^2-x}{5x^3+x^2+x}$

31. $\lim_{x\to\infty} \frac{3x^3+x^2-x}{5x^3+x^2+x}$

32. $\lim_{x\to\infty} \frac{\sin(x)}{4+\sin(x)}$

33. $\lim_{x\to\infty} x\sin(3x)$

34. $\lim_{x\to\infty} \left(\sqrt{x^2+3}-\sqrt{x^2+4x}\right)$

35. $\lim_{x\to 0}(\cos(x))^{1/x}$

36. $\lim_{x\to 0^+} x^{1/x}$

37. $\lim_{x\to 0}(1+x)^{1/x}$

38. $\lim_{x\to 0}\left(1+x^2\right)^x$

39. $\lim_{x\to \pi/2} \frac{\tan(x)}{2x-\pi}$

40. $\lim_{x\to 1} \frac{x^2-1}{x^3-1}$

41. $\lim_{x\to 0} \frac{xe^x(1+x)^3}{e^x-1}$

42. $\lim_{x\to 0} \frac{xe^x\cos^2(6x)}{e^{2x}-1}$

43. $\lim_{x\to 1^+}(x-1)\ln(x-1)$

44. $\lim_{x\to 0}(\csc(x)-\cot(x))$

45. $\lim_{x\to 0} \frac{\csc(x)-\cot(x)}{\sin(x)}$

46. $\lim_{x\to 0} \frac{5^x-3^x}{\sin(x)}$

47. $\lim_{x\to 0} \frac{(\tan(x))^5-(\tan(x))^3}{1-\cos(x)}$

48. $\lim_{x\to 2} \frac{x^3+8}{x^2+5}$

49. $\lim_{x\to \pi/4} \frac{\sin(5x)}{\sin(3x)}$

50. $\lim_{x\to 0}\left(\frac{1}{1-\cos(x)}-\frac{2}{x^2}\right)$

51. $\lim_{x\to 0} \frac{\arcsin(x)}{\arctan(2x)}$

---

52. In Figure 5.6.2(a) the unit circle is centered at O, BQ is a vertical tangent line, and the length of BQ is the same as the arc length $\widehat{BP}$ What happens to the point E as $Q \to B$?

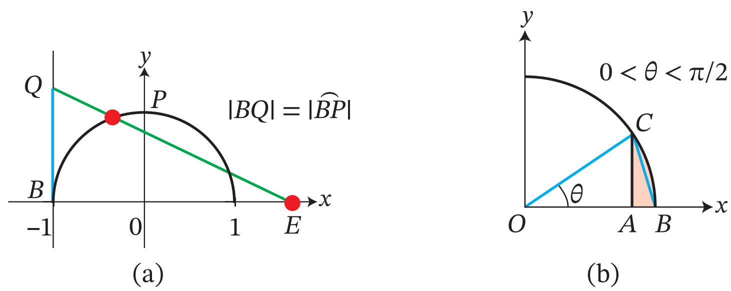


Figure 5.6.2

53. Exercise 42 of Section 2.2 asked you to guess a certain limit. Now that limit will be computed.

In Figure 5.6.2(b), which shows a circle, let $f(\theta)$ be the area of triangle ABC and $g(\theta)$ be the area of the shaded region formed by deleting triangle OAC from sector OBC.

(a) Why is $f(\theta)$ smaller than $g(\theta)$? (b) What would you guess is the value of $\lim_{\theta\to 0} \frac{f(\theta)}{g(\theta)}$? (c) Find $\lim_{\theta\to 0} \frac{f(\theta)}{g(\theta)}$.

54. In a thermodynamic analysis of surface tension the following limit was encountered: $\lim_{a\to 1} \frac{1-a}{1-a(1-C)-a^2C}$, where C is a constant. CONTRIBUTED BY: Anthony Wexler

(a) Find the limit with the use of l'Hôpital's rule. (b) Find the limit without using l'Hôpital's rule.

55. The following appears in an economics text: "Consider the production function $y = k\left(\alpha x_1^{-\rho} + (1-\alpha)x_2^{-\rho}\right)^{-1/\rho}$, where k, α, x_1, and x_2 are positive constants and $\alpha < 1$. Taking the limit as $\rho \to 0^+$, we find that $\lim_{\rho\to 0^+} y = kx_1^{\alpha}x_2^{1-\alpha}$, which is the Cobb-Douglas function, as expected."

Fill in the details.

56. Sam proposes the following proof for Theorem 5.6.1: "Since $\lim_{x\to a^+} f(x) = 0$ and $\lim_{x\to a^+} g(x) = 0$, I will define $f(a) = 0$ and $g(a) = 0$. Next I consider $x > a$ but near a. I now have continuous functions f and g defined on the closed interval $[a, x]$ and differentiable on the open interval (a, x). So, using the mean value theorem, I conclude that there is a number c, $a < c < x$, such that

$$\frac{f(x)-f(a)}{x-a} = f'(c) \quad \text{and} \quad \frac{g(x)-g(a)}{x-a} = g'(c).$$

Since $f(a) = 0$ and $g(a) = 0$, these equations tell me that $f(x) = (x-a)f'(c)$ and $g(x) = (x-a)g'(c)$. Thus $\frac{f(x)}{g(x)} = \frac{f'(c)}{g'(c)}$. And, finally, $\lim_{x\to a^+} \frac{f(x)}{g(x)} = \lim_{x\to a^+} \frac{f'(c)}{g'(c)}$."

Sam made one error. What is it?

57. Find $\lim_{x\to 0^+} \left(\frac{1+2^x}{x}\right)^{1/x}$.

58. R. P. Feynman, in *Lectures in Physics*, wrote: "Here is the quantitative answer of what is right instead of kT. This expression $\frac{\hbar\omega}{e^{\hbar\omega/kT}-1}$ should, of course, approach kT as $\omega \to 0$. .... See if you can prove that it does — learn how to do the mathematics."

All symbols, except ω, denote constants. Do the mathematics to prove Feynman is correct.

59. Graph $y = x^x$ for $0 < x \le 1$, showing its minimum point.

In Exercises 60 to 62 graph the function, showing (a) where it is increasing and decreasing, (b) where the function has any asymptotes, and (c) how the function behaves for x near 0.

60. $f(x) = (1+x)^{1/x}$ for $x > -1$, $x \neq 0$

61. $y = x\ln(x)$

62. $y = x^2\ln(x)$

---

63. In each part below, is it possible to determine $\lim_{x\to a} f(x)^{g(x)}$ without further information about the functions? If so, find the limit. If not, explain why.

(a) $\lim_{x\to a} f(x) = 0$, $\lim_{x\to a} g(x) = 7$

(b) $\lim_{x\to a} f(x) = 2$, $\lim_{x\to a} g(x) = 0$

(c) $\lim_{x\to a} f(x) = 0$, $\lim_{x\to a} g(x) = 0$

(d) $\lim_{x\to a} f(x) = 0$, $\lim_{x\to a} g(x) = \infty$

(e) $\lim_{x\to a} f(x) = \infty$, $\lim_{x\to a} g(x) = 0$

(f) $\lim_{x\to a} f(x) = \infty$, $\lim_{x\to a} g(x) = -\infty$

64. Sam is angry. "Now I know why calculus books are so long. This one spent a whole page (page 51) showing that $\lim_{x\to 0} \frac{\sin(x)}{x}$ is 1. They could have saved space and me a lot of trouble if they had just used l'Hôpital's rule."

Is Sam right, for once?

65. Jane says, "I can get $\lim_{x\to 0} \frac{e^x - 1}{x}$ easily. It's just the derivative of e^x evaluated at 0. I don't need l'Hôpital's rule." Is Jane right, or has Sam's influence affected her ability to reason?

66. Give an example of functions f and g such that $\lim_{x\to 0} f(x) = 1$, $\lim_{x\to 0} g(x) = \infty$, and $\lim_{x\to 0} f(x)^{g(x)} = 2$.

67. Obtain l'Hôpital's rule for $\lim_{x\to\infty} \frac{f(x)}{g(x)}$ from the case $\lim_{t\to 0^+} \frac{f(t)}{g(t)}$. (Let $t = 1/x$.)

68. Find the limit of $(1^x + 2^x + 3^x)^{1/x}$ as (a) $x \to 0$, (b) $x \to \infty$, and (c) $x \to -\infty$.

Exercise 72 outlines the proof of Theorem 5.6.1. It depends on the following result, which you will prove in Exercise 70.

Theorem 5.6.5: Generalized Mean Value Theorem

Let f and g be two functions that are continuous on $[a,b]$ and differentiable on (a,b). Assume that $g'(x)$ is never 0 for x in (a,b). Then there is a number c in (a,b) such that

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}.$$

69. During a time interval one car travels twice as far as another car. Use the generalized mean value theorem to show that there is at least one instant when the first car is traveling exactly twice as fast as the second car.

70. To prove the generalized mean value theorem, introduce the function

$$h(x) = f(x) - f(a) - \frac{f(b) - f(a)}{g(b) - g(a)}(g(x) - g(a)). \tag{5.6.1}$$

Show that $h(b) = 0$ and $h(a) = 0$. Then apply Rolle's Theorem to h on (a,b). (Rolle's Theorem is Theorem 4.1.4.)

71. The function h in (5.6.1) is similar to the function h used in the proof of the mean value theorem (Theorem 4.1.6 in Section 4.1). Check that $h(x)$ is the vertical distance between the point $(g(x), f(x))$ and the line through $(g(a), f(a))$ and $(g(b), f(b))$.

72. To prove Theorem 5.6.1, assume its hypotheses are satisfied. Define $f(a) = 0$ and $g(a) = 0$, so that f and g are continuous at a and

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)}.$$

Show that the hypotheses of the generalized mean value theorem (Theorem 5.6.5) are satisfied. Show that the generalized mean value theorem's conclusion is the conclusion of Theorem 5.6.1, l'Hôpital's rule in the zero-over-zero case.

Exercises 73 and 74 are related. Do not assume f and g are differentiable.

73. If $\lim_{t\to\infty} f(t) = \infty$, $\lim_{t\to\infty} g(t) = \infty$, and $\lim_{t\to\infty} \frac{f(t)}{g(t)} = 3$, what can be said about $\lim_{t\to\infty} \frac{\ln(f(t))}{\ln(g(t))}$?

74. If $\lim_{t\to\infty} f(t) = \infty$, $\lim_{t\to\infty} g(t) = \infty$, and $\lim_{t\to\infty} \frac{\ln(f(t))}{\ln(g(t))} = 1$, must $\lim_{t\to\infty} \frac{f(t)}{g(t)} = 1$? Explain.

---

75. Assume that f, f', and f'' are defined in $[-1, 1]$ and are continuous. Also, $f(0) = 0$, $f'(0) = 0$, and $f''(0) > 0$.

(a) Sketch what the graph of f may look like for x in $[0, a]$, where a is a small positive number.

(b) Interpret the quotient $Q(a) = \dfrac{\int_0^a f(x)\,dx}{af(a) - \int_0^a f(x)\,dx}$ in terms of the graph in (a).

(c) What do you think happens to $Q(a)$ as $a \to 0$?

(d) Find $\lim_{a\to 0} Q(a)$.

(Nothing is known about f''', which may not exist.)

76.
SAM: I bet I can find $\lim_{x\to 0} \frac{e^x - 1 - x - \frac{x^2}{2}}{x^3}$ by using the Maclaurin polynomial $P_2(x;0)$ for e^x and paying attention to the error.

Is Sam right?

Exercises 77 and 78 will be used in Exercises 79 and 80.

77. Find $\lim_{x\to 0} \frac{\tan(x) - x}{2x - \sin(2x)}$.

78. Find $\lim_{x\to 0} \frac{\tan(x) - x}{x - \sin(x)}$.

79. Let $P(n)$ be the perimeter of the regular polygon with n sides that circumscribes a circle of radius 1. Similarly, let $p(n)$ be the perimeter of the inscribed regular polygon of n sides. When n is large, which is the better estimate of the perimeter of the circle? To decide, examine the limit of $\frac{P(n) - 2\pi}{2\pi - p(n)}$. Form an opinion before answering.

80. Let $A(n)$ be the area of a regular polygon with n sides that circumscribes a circle of radius 1. Similarly, let $a(n)$ be the area of an inscribed regular polygon of n sides. When n is large, which is the better estimate of the area of the circle? To decide, examine the limit of $\frac{A(n) - \pi}{\pi - a(n)}$. Form an opinion before answering.

Exercises 81 to 85 build upon the estimates of the circumference of a unit circle by inscribed and circumscribed regular n-gons found in Exercise 79. The perimeter of the inscribed n-gon is $p(n) = 2n\sin(\pi/n)$ and the perimeter of the circumscribed n-gon is $P(n) = 2n\tan(\pi/n)$. These exercises provide more information about the errors in these estimates; $e(n) = 2\pi - p(n)$ and $E(n) = P(n) - 2\pi$.

81. Find $\lim_{n\to\infty} \dfrac{E(n)}{1/n}$, as follows:

(a) Introduce $f(x) = \dfrac{2x\tan(\pi/x) - 2\pi}{1/x}$ and try to find $\lim_{x\to\infty} f(x)$.

(b) If finding the limit in (a) is too difficult, try replacing $1/x$ in (a) by t, rewriting $f(x)$ as

$$g(t) = \frac{(2/t)\tan(\pi t) - 2\pi}{t} = \frac{2\tan(\pi t) - 2\pi t}{t^2}.$$

Then find $\lim_{t\to\infty} g(t)$, which is the same as the limit in (a), $\lim_{x\to\infty} f(x)$.

(c) Which is larger for large n, $E(n)$ or $\dfrac{1}{n}$?

82. (a) Find k such that $\lim_{n\to\infty} \dfrac{E(n)}{1/n^k}$ is neither 0 nor ∞. That value of k indicates how rapidly the error $E(n)$ approaches 0.

(b) Show that for that value of k, $\dfrac{E(n)}{1/n^k}$ approaches $\dfrac{2}{3}\pi^3$ as n approaches ∞.

83. Now we turn to the error $e(n) = 2\pi - 2n\sin(\pi/n)$. Show that for large n, $e(n)$ is much smaller than $1/n$.

84. Find k such that $\lim_{n\to\infty} \dfrac{e(n)}{1/n^k}$ is neither 0 nor ∞. For that k the limit is $\pi^3/3$.

85. Comparing Exercises 82(b) and 84 shows that $E(n)$ tends to be about twice $e(n)$.

(a) Why might $Q(n) = \dfrac{1}{3}(P(n) + 2p(n))$ be a better estimate than either $P(n)$ or $p(n)$?

(b) Find k such that $\lim_{n\to\infty} \dfrac{Q(n)}{1/n^k}$ is neither 0 nor ∞.

(c) Is $Q(n)$ a better estimate than either $P(n)$ or $p(n)$?

---

86. One point made in Exercise 1 in CIE 3 is that, everything else being the same, accounts compounded more often have more in them at the end of a full year. This problem shows that $(1+\frac{1}{n})^n$ indeed increases as n increases by showing that $(1+x)^{1/x}$ decreases for $x > 0$.

(a) Show that $f(x) = (1+x)^{1/x}$, for $x > 0$, is a decreasing function.

(b) Use (a) to show that $g(n) = \left(1 + \dfrac{1}{n}\right)^n$, for $n > 0$, is an increasing function

5.7 Natural Growth and Decay

In 2023 the population of the United States was about 340 million and growing at a rate of about 0.5% (roughly 1.7 million people) a year. The world population was about 7.869 billion and growing at a rate of about 0.9% (roughly 73 million people) a year. In each case the rate of change of the population is proportional to its current size. That is, both are examples of natural growth.

Natural Growth

Let $P(t)$ be the size of a population at time t. If its rate of growth is proportional to its size, there is a positive constant k such that

If $P(t)$ denotes the US population with time t in years, then $k \approx 0.005$.

$$\frac{dP(t)}{dt} = kP(t). \tag{5.7.1}$$

To find an explicit formula for $P(t)$ as a function of t, rewrite (5.7.1) as

$$\frac{1}{P(t)}\frac{dP(t)}{dt} = k. \tag{5.7.2}$$

Because $P(t)$ is positive, the left-hand side of (5.7.2) can be rewritten as the derivative of $\ln(P(t))$ and so

$$\frac{d(\ln(P(t)))}{dt} = \frac{d(kt)}{dt}.$$

Therefore, recalling Observation 3.6.2, there is a constant C such that

$$\ln(P(t)) = kt + C. \tag{5.7.3}$$

From (5.7.3) it follows, by the definition of a logarithm, that

$$P(t) = e^{kt+C} = e^C e^{kt}.$$

Since C is a constant, so is e^C, which we give a simpler name, A. We have obtained the following compact formula for the population, $P(t)$:

Definition: *Natural Growth*

The equation for **natural growth** is

$$P(t) = Ae^{kt}$$

where k is a positive constant. Because $P(0) = Ae^{k(0)} = A$, the coefficient A is called the **initial population** and the constant k is called the **(natural) growth rate** of the population.

Because of the presence of the exponential e^{kt}, natural growth is also called **exponential growth**.

EXAMPLE 1. The size of the world population at the beginning of 2011 was approximately 6.916 billion. At the beginning of 2012 it was 6.992 billion. Assume that the growth rate remains constant.

(a) What is the growth constant k?

(b) What would the population be in 2032?

(c) When will the population double in size?

SOLUTION Let $P(t)$ be the population, in billions, at time t. For convenience, measure time starting in the year 2011 so $t = 0$ corresponds to 2011 and $t = 1$ to 2012. Then $P(0) = 6.916$ and $P(1) = 6.992$. The natural growth equation describing the population in billions at time is

$$P(t) = 6.916e^{kt}.$$

(a) To find k, use $P(1) = 6.916e^{k\cdot 1}$ to obtain

$$\begin{aligned} 6.916e^k &= 6.992 \\ e^k &= \frac{6.992}{6.916} && \text{(divide both sides by 6.916)} \\ k &= \ln\left(\frac{6.992}{6.916}\right) \approx 0.0109. && \text{(take natural logarithm of both sides)} \end{aligned}$$

The growth rate is approximately 1.09%. We now have a complete definition of the growth equation for the world's population: $P(t) = 6.916e^{0.0109t}$.

(b) The year 2032 corresponds to $t = 21$, so in the year 2032 the population, in billions, would be

$$P(21) = 6.916e^{0.0109\cdot 21} = 6.916e^{0.229} \approx 6.916(1.257) \approx 8.693.$$

The Bureau of the Census estimates that the world population in 2032 will be about 8.450 billion. It assumes that the growth rate will go down.

(c) The population will double when it reaches $2P(0) = 2(6.916) = 13.832$ billion. We solve for t in the equation $P(t) = 13.832$ as follows:

$$\begin{aligned} 6.916e^{kt} &= 13.832 \\ e^{kt} &= 2 && \text{(divide both sides by 6.916)} \\ kt &= \ln(2) && \text{(take natural logarithm of both sides)} \\ t &= \frac{\ln(2)}{k} \approx \frac{0.693}{0.0109} \approx 63.578. && \text{(divide both sides by } k = 0.0109\text{)} \end{aligned}$$

The world population will double approximately 63 years after 2011, during the summer of 2074. ◆

Observation 5.7.1: *Doubling Time*

The time it takes for a population to double is called the **doubling time** and is denoted t_2. Exponential growth is often described by its doubling time t_2 rather than by its growth constant k. If you know either t_2 or k you can figure out the other, as they are related by the equation

$$t_2 = \frac{\ln(2)}{k}$$

that appeared during part (c) of the solution to Example 1.

EXAMPLE 2. Find the doubling time if the growth rate is 2 percent per year.

SOLUTION The growth rate is 2 percent, so we set $k = 0.02$. Then $t_2 = \ln(2)/k \approx 0.693/0.02 = 34.65$ years. ◆

Exponential growth may also be described in terms of an annual percentage increase, such as "The population is growing 6% annually." That is, each year the population is multiplied by the factor 1.06: $P(t+1) = P(t)(1.06)$.

From the exponential growth function, we see that

$$P(t+1) = P(0)e^{k(t+1)} = P(0)e^{kt}e^k = P(t)e^k.$$

That is, during each unit of time the population is magnified by a factor of e^k. When k is small, $e^k \approx 1 + k$. Consequently we can approximate 6 percent annual growth by letting $k = 0.06$. The approximation is valid whenever the growth rate is only a few percent. Since population figures are themselves only approximations, setting the growth constant k equal to the annual percentage rate is a reasonable tactic in many cases.

Note that in Example 2 the initial population was not known, and was not needed to determine the doubling time for this population. The doubling time depends only on the growth rate. This is one of the defining properties of natural growth.

Natural Decay

Historical Note: The Scientist, The Senator, and Half-Life

In 1963, during the hearings before the Senate Foreign Relations Committee on the nuclear test ban treaty, this exchange took place between Glen Solborg, winner of the Nobel prize for chemistry in 1951, and Senator James W. Fulbright.

SEABORG: Tritium is used in a weapon, and it decays with a half-life of about 12 years. But the plutonium and uranium have such long half-lives that there is no detectable change in a human lifetime.

FULBRIGHT: I am sure this seems to be a very naive question, but why do you refer to half-life rather than whole life? Why do you measure by half-lives?

SEABORG: Here is something that I could go into a very long discussion on.

FULBRIGHT: I probably wouldn't benefit adequately from a long discussion. It seems rather odd that you should call it a half-life rather than its whole life.

SEABORG: Well, I will try. If we have, let us say, one million atoms of a material like tritium, in 12 years half of those will be transformed into a decay product and you will have 500,000 atoms. Then, in another 12 years, half of what remains transforms, so you have 250,000 atoms left. And so forth.
On that basis it never all decays, because half is always left, but of course you finally get down to where your last atom is gone.

As Glen Seaborg observes in the above exchange with Senator Fulbright, radioactive elements decay at a rate proportional to the amount present. The time it takes for half the initial amount to decay is denoted $t_{1/2}$ and is called the element's **half-life**.

Similarly, in medicine one speaks of the half-life of a drug administered to a patient, meaning the time required for half the drug to be removed from the body. This half-life depends on the drug and the patient, and can be from 20 minutes for penicillin to 2 weeks for quinacrine, an antimalarial drug. This half-life is critical to determining how frequently a drug can be administered. Some elderly patients have died from overdoses before it was realized that the half-life of some drugs is longer in the elderly than in the young.

Letting $P(t)$ represent the amount present at time t, we have

$$P'(t) = kP(t), \qquad k < 0,$$

where k is the **decay constant**. This is the same equation as (5.7.1), so

$$P(t) = P(0)e^{kt},$$

as before, except now k is a negative number. Since k is negative, the factor e^{kt} is a decreasing function of t.

Just as the doubling time is related to (positive) k by the equation $t_2 = (\ln(2))/k$, the half-life is related to (negative) k by the equation $t_{1/2} = (\ln(1/2))/k$, which can be rewritten as $t_{1/2} = -(\ln(2))/k$.

EXAMPLE 3. The Chernobyl nuclear reactor accident in April 1986 released radioactive cesium 137 into the air. The half-life of ^{137}Cs is 27.9 years.

(a) Find the decay constant k of ^{137}Cs.
(b) When will only one-fourth of an initial amount remain?
(c) When will only 20 percent of an initial amount remain?

SOLUTION

(a) The formula for the half-life can be solved for k to give $k = -\ln(2)/t_{1/2} \approx -0.693/27.9 \approx -0.0248$.
(b) The time needed to determine when only 1/4 of the original amount remains can be done without the aid of any formulas. Since $1/4 = 1/2 \cdot 1/2$, in two half-lives only one-quarter of an initial amount remains. After $2(27.9) = 55.8$ years one-quarter of the original amount remains.

(c) We want to find t such that only 20 percent remains. While we know the answer is greater than 55.8 years (since 20% is less than 25%), finding the exact time requires using the formula for $P(t)$.
We want to find the time t such that $P(t) = 0.20P(0)$. That is, we want to solve $P(0)e^{kt} = 0.20P(0)$. Canceling $P(0)$ from both sides of the equation leads to $e^{kt} = 0.20$. Taking logarithms, $kt = \ln(0.20)$. And, finally, we find that $t = \ln(0.20)/k$. Since $k \approx -0.0248$, this gives $t \approx -1.609/-0.0248 \approx 64.9$ years. After 64.9 years (that is, in the spring of 2051) 20% of the original amount of radioactive cesium 137 will still be present at Chernobyl. ♦

Summary

We developed the mathematics of growth or decay that is proportional to the amount present. This required solving the differential equation

$$\frac{dP}{dt} = kP$$

where k is a constant, positive for growth and negative for decay. The solution is

$$P(t) = Ae^{kt}$$

where A is $P(0)$, the amount present when $t = 0$.

In the case of growth, the time for the quantity to double (the "doubling time") is denoted t_2. In the case of decay, the time when only half the original amount survives is denoted $t_{1/2}$, the half-life. We have

$$t_2 = \frac{\ln(2)}{k} \qquad \text{and} \qquad t_{1/2} = \frac{\ln(1/2)}{k} = -\frac{\ln(2)}{k}.$$

EXERCISES for Section 5.7

1. Using no mathematical symbols, describe the basic assumption in natural growth and decay.

2. (a) Show that exponential growth can be expressed as $P(t) = Ab^t$ for constants A and b.
(b) What can be said about b?

3. (a) Show that exponential decay can be expressed as $P(t) = Ab^t$ for constants A and b.
(b) What can be said about b?

4. If $P(t) = 30e^{0.2t}$ what are the initial size and the doubling time?

5. If $P(t) = 30e^{-0.2t}$ what are the initial size and the half-life?

6. What is the doubling time for a population always growing at 1% a year?

7. What is the half-life for a population always shrinking at 1% a year?

8. A quantity is increasing according to the law of natural growth. The amount present at time $t = 0$ is A. It will double when $t = 10$. Express the amount at time t in the form (a) Ae^{kt} and (b) Ab^t.

9. The mass of a bacterial culture after t hours is $10 \cdot 3^t$ grams. Find the
(a) initial amount, (b) growth constant k, and (c) percent increase in any period of 1 hour.

10. Let $f(t) = 3 \cdot 2^t$. (a) Solve the equation $f(t) = 12$. (b) Solve the equation $f(t) = 5$. (c) Find k such that $f(t) = 3e^{kt}$.

11. A population growing exponentially has a doubling time of 5 years. How long will it take to quadruple in size?

12. The population of Latin America has a doubling time of 27 years. Estimate the percent it grows per year.

13. At 1:00 P.M. a bacterial culture weighed 100 grams. At 4:30 P.M. it weighed 250 grams. Assuming that it grows at a rate proportional to the amount present, find (a) the time when it is 400 grams and (b) its growth constant.

14. A bacterial culture grows from 100 to 400 grams in 10 hours according to the law of natural growth.
(a) How much was present after 3 hours? (b) How long will it take the mass to double? quadruple? triple?

15. A radioactive substance disintegrates at the rate of 0.05 grams per day when its mass is 10 grams.
(a) How much of the substance will remain after t days if the initial amount is A? (b) What is its half-life?

16. In July 2021 the population of Mexico was 130 million and of the United States 335 million and the annual growth rates were 0.7% for the United States and 1.04% for Mexico. At these rates, when would the two nations have the same size population?

17. The size of the population in India was 699 million in 1980 and 1,183 million in 2007. What is its doubling time t_2? How does this compare with the fact that India's population in 2022 was 1,407 million? What does this say about the actual growth rate in India between 2007 and 2022?

18. The newspaper article shown in Figure 5.7.1(a) illustrates the rapidity of exponential growth.

(a) Is the figure of \$14 billion correct? Assume that the interest is compounded annually.

(b) What interest rate would be required to produce an account of \$14 billion if interest were compounded once a year?

(c) Answer (b) for continuous compounding, which is another term for natural growth (a bank account increasing at a rate proportional to the amount in the account at any instant).


U.S. HIT FOR 200-YR. DEBT

LAS VEGAS—An autograph dealer is demanding the U.S. government pay off a 200-year-old note. At seven percent interest, the debt amounts to $14 billion.

It was issued to Haym Salomon on March 27, 1782, by finance chief Robert Morse in return for a $30,000 loan.

The note is payable to the bearer—but the statute of limitations ran out on it some 150 years ago.

Two Japanese Banks Lower An Interest Rate to 0.001%

Only 69,315 Years to Double Your Money

(a) (b)

Figure 5.7.1

19. The headline shown in Figure 5.7.1(b) appeared in 2002. Is the number 69,315 correct? Explain.

20. Carbon 14 (chemical symbol ^{14}C), an isotope of carbon, is radioactive and has a half-life of approximately 5,730 years. If the ^{14}C concentration in a piece of wood of unknown age is half of the concentration in a present-day live specimen, then it is about 5,730 years old, assuming that ^{14}C concentrations in living objects remain about the same. This gives a way of estimating the age of an undated specimen. Show that if A_C is the concentration of ^{14}C in a live (contemporary) specimen and A_u is the concentration of ^{14}C in a specimen of unknown age, then the age of the undated material is about $8,300\ln(A_C/A_u)$ years. This method, called **radiocarbon dating**, is reliable up to about 70,000 years.

21. From a letter to an editor in a newspaper:

> I've been hearing bankers and investment advisers talk about something called the "rule of 72." Could you explain what it means?

The "rule of 72" tells you the number of months needed for an account to double in value. To find the number of years before your account's value doubles, divide the yield into 72.

For instance, if your bank account pays interest at 4% per year, and the interest is reinvested in the account, the account will double in about $72/4 = 18$ years. If, instead of reinvestsing the interest in the account, it was put under the mattress, it would take $100/4 = 25$ years to double in value.

(a) Explain the rule of 72. What number should be used instead of 72? (b) Why do you think 72 is used?

22. Benjamin Franklin conjectured that the population of the United States would double every 20 years, beginning in 1751, when the population was 1.3 million.

(a) If Franklin's conjecture was right, what would the population of the United States be in 2012?

(b) In 2012 the population was 313 million. Assuming natural growth, what is the doubling time?

23. (Doomsday equation) A differential equation of the form $dP/dt = kP^{1.01}$ is called a **doomsday equation**. The rate of growth is just slightly higher than that for natural growth. Solve the differential equation to find $P(t)$. How does $P(t)$ behave as t increases? Does $P(t)$ increase forever?

24. The following are all mathematically the same:

(i) A drug is administered in a dose of A grams to a patient and gradually leaves the system through excretion.

(ii) Initially there is an amount A of smoke in a room. The air conditioner is turned on and gradually the smoke is removed.

(iii) Initially there is an amount A of a pollutant in a lake, when further dumping of toxic materials is prohibited. The rate at which water enters the lake equals the rate at which it leaves. (Assume the pollution is thoroughly mixed.)

In each case, let $P(t)$ be the amount present at time t (whether drug, smoke, or pollutant); answer these questions.

(a) Why is it reasonable to assume that there is a constant k such that for small intervals of time, Δt, $\Delta P \approx kP(t)\Delta t$?

(b) From (a) deduce that $P(t) = Ae^{kt}$.

(c) Is k positive or negative? (How do you know?)

25. **Newton's law of cooling** assumes that an object cools at a rate proportional to the difference between its temperature and the room temperature. Denote the room temperature as A. The differential equation for Newton's law of cooling is $dy/dt = k(y - A)$ where k and A are constants.

(a) Explain why k is negative.

(b) Draw the slope field for the differential equation when $k = -1/2$.

(c) Use (b) to conjecture the behavior of $y(t)$ as $t \to \infty$.

(d) Solve for y as a function of t.

(e) Draw the graph of $y(t)$ on the slope field produced in (b).

(f) Find $\lim_{t\to\infty} y(t)$.

26. The sun emits electromagnetic waves with wavelengths ranging from 2 nm (X-rays) to 10 m (radio waves); (green) visible light has a wavelength of about 500 nm. Let $I(x)$ be the intensity of visible light at a depth of x m in the ocean. As x increases, $I(x)$ decreases.

(a) Why is it reasonable to assume that there is a constant k (negative) such that $\Delta I \approx kI(x)\Delta x$ for small Δx?

(b) Deduce that $I(x) = I(0)e^{kx}$, where $I(0)$ is the intensity of visible light at the surface.

(c) Use the fact that the intensity of visible light at a depth of 1 m is only one-fourth as intense at the surface to estimate k.

27. A particle moving through a liquid meets a drag force proportional to the velocity; so its acceleration is proportional to its velocity. Let x denote its position and v its velocity at time t. Assume $v > 0$.

(a) Show that there is a positive constant k such that $\frac{dv}{dt} = -kv$.

(b) Show that there is a constant A such that $v = Ae^{-kt}$.

(c) Show that there is a constant B such that $x = -\frac{A}{k}e^{-kt} + B$.

(d) How far does the particle travel as t goes from 0 to ∞? (Is it a finite or infinite distance?)

28. (a) Show that the natural growth function $P(t) = Ae^{kt}$ can be written in terms of A and t_2 as $P(t) = A \cdot 2^{t/t_2}$.

(b) Check that the function found in (a) is correct when $t = 0$ and $t = t_2$.

29. (a) Express the natural decay function $P(t) = Ae^{kt}$ in terms of A and $t_{1/2}$.

(b) Check that the function found in (a) is correct when $t = 0$ and $t = t_{1/2}$.

30. A population is growing exponentially. At time 0 it is P_0. At time u it is P_u.

(a) Show that at time t it is $P_0\left(\frac{P_u}{P_0}\right)^{t/u}$.

(b) Check that the formula in (a) gives the correct population when $t = 0$ and $t = u$.

31. Let $P(t) = Ae^{kt}$. Then $\frac{P(t+1)-P(t)}{P(t)} = e^k - 1$. Show that when k is small, $e^k - 1 \approx k$. That means the relative change in one unit of time is approximately k.

32. A fish population increases at a rate proportional to the size of the population. It is also being harvested at a constant rate. Let $P(t)$ be the size of the population at time t.

(a) Show that there are positive constants h and k such that for small Δt, $\Delta P \approx kP\Delta t - h\Delta t$. What are the units on h and k?

(b) What differential equation is obtained in the limit when $\Delta t \to 0^+$?

(c) Verify that $P(t) = \left(P_0 - \frac{h}{k}\right)e^{kt} + \frac{h}{k}$ satisfies both $P(0) = P_0$ and the differential equation found in (b).

(d) Describe the behavior of $P(t)$ in the cases $h = kP(0)$, $h > kP(0)$, and $h < kP(0)$.

33. The half-life of a drug administered to a patient is 8 hours. It is given in a 1-gram dose every 8 hours.

(a) How much is there in the patient just after the second dose is administered?

(b) How much is there in the patient just after the third dose? The fourth dose?

(c) Let $P(t)$ be the amount in the patient at t hours after the first dose. Graph $P(t)$ for a period of 48 hours. ($P(t)$ has meaning for all values of t, not just for integers.)

(d) Does the amount in the patient get arbitrarily large as time goes on?

34. The half-life of the drug in Exercise 33 is 16 hours when administered to a different patient. Answer, for this patient, the questions in Exercise 33.

35. The half-life of a drug in a patient is $t_{1/2}$ hours. It is administered every h hours. Can the concentration of the drug get arbitrarily high? Explain your answer.

Exercises 36 to 38 introduce and analyze the **inhibited** or **logistic growth** model. It will be encountered in the CIE about petroleum at the end of Chapter 12.

36. In many cases of growth there is a finite upper bound M which the population cannot exceed. Why is it plausible to assume that there is a constant k such that

$$\frac{dP}{dt} = kP(t)\left(1 - \frac{P(t)}{M}\right) \qquad 0 < P(t) < M? \tag{5.7.4}$$

37. (a) Find the constant a that makes $P(t) = \frac{M}{1+ae^{-Mkt}}$ a solution of the differential equation (5.7.4).

(b) Find $\lim_{t\to\infty} P(t)$. Is this value reasonable?

(c) Assume the initial population is $P(0) = P_0$. Express a in terms of the problem's parameters: k, M, and P_0.

38. By considering (5.7.4) in Exercise 36 directly (not the explicit formula in Exercise 37), show that

(a) P is an increasing function (of t).

(b) The maximum rate of change of P occurs when $P(t) = \frac{M}{2}$.

(c) The graph of $P(t)$ has an inflection point.

---

39. The populations of two countries are growing exponentially but at different rates. One is described by $A_1e^{k_1t}$, the other by $A_2e^{k_2t}$, and k_1 is not equal to k_2. Is their total population growing exponentially? That is, are there constants A and k such that the formula for their total population has the form Ae^{kt}? Explain your answer.

40. Assume c_1, c_2, and c_3 are distinct constants. Can there be constants A_1, A_2, and A_3, not all 0, such that $A_1 e^{c_1 x} + A_2 e^{c_2 x} + A_3 e^{c_3 x} = 0$ for all x?

41. If two functions describe natural growth does their (a) product? (b) quotient? (c) sum?

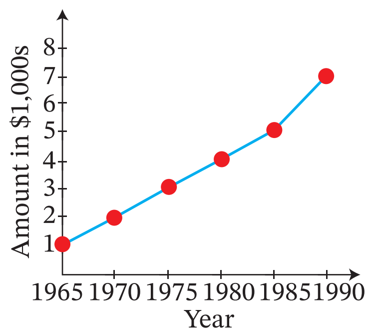


Figure 5.7.2

42. A financial advisor, trying to persuade a client to invest in Standard Coagulated Mutual Fund, shows him the graph in Figure 5.7.2, that records the value of a similar investment made in the fund in 1965. "Look! In the first 5 years the investment increased $1,000," the salesman observed, "but in the past 5 years it increased by $2,000. It's really improving. Look at the part of the graph from 1985 to 1990."

The investor replied, "Hogwash. Though your graph is steeper from 1985 to 1990, in fact, the rate of return is less than from 1965 to 1970. In fact, that was your best period."

(a) If the percentage return on the accumulated investment remains the same over each 5-year period as the first 5-year period, sketch the graph.

(b) Explain the investor's reasoning.

5.8 Hyperbolic Functions and Their Inverses

Certain combinations of the exponential functions e^t and e^{-t} occur often enough — for instance, in the study of the shape of electrical transmission or suspension cables — to be given names. They are called **hyperbolic functions**. This section defines the hyperbolic functions and presents their basic properties.

The Hyperbolic Functions

Definition: *The Hyperbolic Cosine*

Let t be a real number. The **hyperbolic cosine** of t, denoted $\cosh(t)$, is $\cosh(t) = \dfrac{e^t + e^{-t}}{2}$.

One of the first properties of the hyperbolic cosine function is that

$$\cosh(-t) = \frac{e^{-t} + e^{-(-t)}}{2} = \frac{e^t + e^{-t}}{2} = \cosh(t).$$

Thus, the cosh function is even and its graph is symmetric with respect to the vertical axis. Furthermore, writing

$$\cosh(t) = \frac{e^t}{2} + \frac{e^{-t}}{2},$$

observe that the second term, $e^{-t}/2$, is positive and, as $t \to \infty$, approaches 0. Thus, for $t > 0$ and large, the graph of $\cosh(t)$ is just a little above the graph of $e^t/2$. This, together with the fact that $\cosh(0) = (e^0 + e^{-0})/2 = 1$, is the basis for Figure 5.8.1(a). It shows that the graph of $y = \cosh(t)$ is asymptotic to the graph of $y = e^t/2$ as $t \to \infty$ and to the graph of $y = e^{-t}/2$ as $t \to -\infty$. The curve $y = \cosh(t)$ in Figure 5.8.1(a) is called a **catenary**, from the Latin *catena* meaning chain. It describes the shape of a free-hanging chain.

Definition: *The Hyperbolic Sine*

Let t be a real number. The **hyperbolic sine** of t, denoted $\sinh(t)$, is $\sinh(t) = \dfrac{e^t - e^{-t}}{2}$.

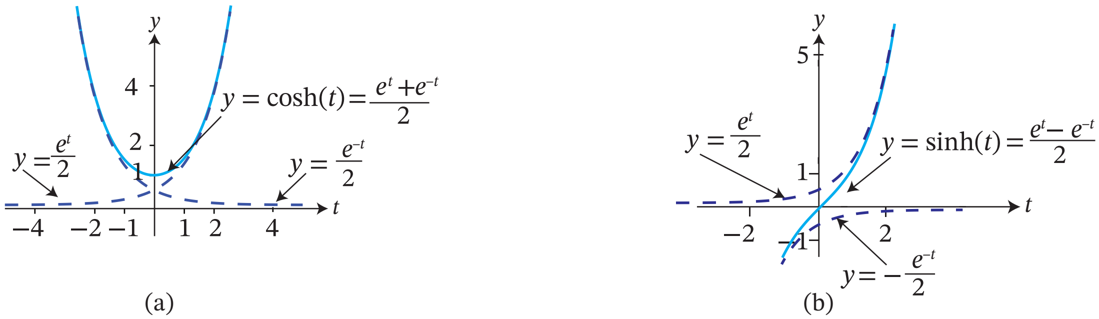


Figure 5.8.1

Because $\sinh(-t) = -\sinh(t)$, the hyperbolic sine function is an odd function. Thus, the graph of $y = \sinh(t)$ is symmetric with respect to the origin. This graph lies below the graph of $e^t/2$ and, since $e^{-t}/2 \to 0$ as $t \to \infty$, the graphs of $y = \sinh(t)$ and $y = e^t/2$ approach each other. Thus, by symmetry, the graph of $y = \sinh(t)$ is asymptotic to the graph of $y = -e^t/2$ as $t \to -\infty$. Figure 5.8.1 (b) shows the graph of $\sinh(t)$.

"cosh" is pronounced as written; it rhymes with "gosh"; "sinh" is pronounced "sinch", rhyming with "pinch".

The graphs of $\sinh(t)$ and $\sin(t)$ exhibit very different behaviors. As $|t|$ becomes large, the hyperbolic sine becomes large, $\lim_{t\to\infty} \sinh(t) = \infty$ and $\lim_{t\to-\infty} \sinh(t) = -\infty$, while $-1 \le \sin(t) \le 1$. There is a similar contrast between $\cosh(t)$ and $\cos(t)$.

Observation 5.8.1: *Initial Observations about Hyperbolic Functions*

We have demonstrated that $\cosh(x)$ and $\cos(x)$ are both even functions. And, $\sinh(x)$ and $\sin(x)$ are both odd functions.

While the trigonometric functions are periodic, the hyperbolic functions are not; the hyperbolic functions are unbounded. In fact, the graphs of $y = \cosh(x)$ and $y = \sinh(x)$ are both asymptotic to an exponential function.

Example 1 shows why the functions $(e^t + e^{-t})/2$ and $(e^t - e^{-t})/2$ are called **hyperbolic**.

EXAMPLE 1. Show that for any real number t the point with coordinates $(x, y) = (\cosh(t), \sinh(t))$ lies on the hyperbola $x^2 - y^2 = 1$.

SOLUTION Compute $x^2 - y^2 = \cosh^2(t) - \sinh^2(t)$ and see whether it simplifies to 1. We have

$$\begin{aligned}
\cosh^2(t) - \sinh^2(t) &= \left(\frac{e^t + e^{-t}}{2}\right)^2 - \left(\frac{e^t - e^{-t}}{2}\right)^2 && \text{(definitions)}\\
&= \frac{e^{2t} + 2e^t e^{-t} + e^{-2t}}{4} - \frac{e^{2t} - 2e^t e^{-t} + e^{-2t}}{4} && \text{(expand powers)}\\
&= \frac{2+2}{4} && \text{(simplification)}\\
&= 1.
\end{aligned}$$

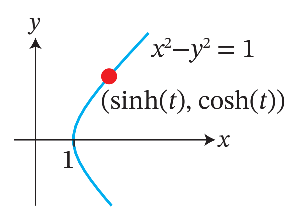


Figure 5.8.2

Because $\cosh(t) \ge 1$, the point $(\cosh(t), \sinh(t))$ is on the right half of the hyperbola $x^2 - y^2 = 1$, as shown in Figure 5.8.2. ♦

Observation 5.8.2: ***More Comparisons Between Circular and Hyperbolic Functions***

The trigonometric functions are also known as **circular functions** because the points $(x, y) = (\cos(\theta), \sin(\theta))$ form the circle $x^2 + y^2 = 1$.

Example 1 provides justification for calling $\cosh(t)$ and $\sinh(t)$ hyperbolic functions: the points $(x, y) = (\cosh(t), \sinh(t))$ form the hyperbola $x^2 - y^2 - 1$.

Other comparisons between the circular and hyperbolic functions will be noted as we learn about their derivatives (and antiderivatives).

But, first, we introduce four more hyperbolic functions: hyperbolic tangent, hyperbolic secant, hyperbolic cotangent, and hyperbolic cosecant. They are defined as:

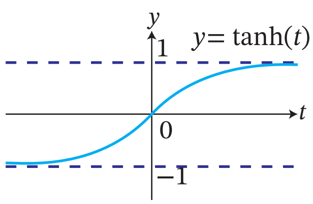


Figure 5.8.3

$$\tanh(t) = \frac{\sinh(t)}{\cosh(t)}, \qquad \operatorname{sech}(t) = \frac{1}{\cosh(t)},$$
$$\coth(t) = \frac{\cosh(t)}{\sinh(t)}, \qquad \operatorname{csch}(t) = \frac{1}{\sinh(t)}.$$

Each hyperbolic function can be expressed in terms of exponentials. For instance,

$$\tanh(t) = \frac{(e^t - e^{-t})/2}{(e^t + e^{-t})/2} = \frac{e^t - e^{-t}}{e^t + e^{-t}}.$$

As $t \to \infty$, $e^t \to \infty$ and $e^{-t} \to 0$. Thus $\lim_{t\to\infty} \tanh(t) = 1$. Similarly, $\lim_{t\to-\infty} \tanh(t) = -1$. Figure 5.8.3 shows a graph of $y = \tanh(t)$.

The Derivatives of the Hyperbolic Functions

The derivatives of the hyperbolic sine and cosine functions can be computed directly from the definitions in terms of exponential functions. For instance,

$$\begin{aligned}
(\cosh(t))' &= \left(\frac{e^t + e^{-t}}{2}\right)' && \text{(definition of } \cosh(t) \text{)}\\
&= \frac{e^t - e^{-t}}{2} && \text{(derivative of exponentials)}\\
&= \sinh(t) && \text{(definition of } \sinh(t) \text{).}
\end{aligned}$$

The derivatives of the other four hyperbolic functions can be from their definitions in terms of $\sinh(t)$ and $\cosh(t)$ and appropriate differentiation formulas:

$$\begin{aligned}
(\tanh(t))' &= \left(\frac{\sinh(t)}{\cosh(t)}\right)' && \text{(definition of } \tanh(t) \text{)}\\
&= \frac{\cosh(t)(\sinh(t))' - \sinh(t)(\cosh(t))'}{(\cosh(t))^2} && \text{(Product Rule)}\\
&= \frac{\cosh^2(t) - \sinh^2(t)}{\cosh^2(t)} && \text{(derivative of sinh and cosh)}\\
&= \left(\frac{1}{\cosh(t)}\right)^2 && \text{(see Example 1)}\\
&= \operatorname{sech}^2(t) && \text{(definition of } \operatorname{sech}(t) \text{).}
\end{aligned}$$

Table 5.8.1 summarizes the full list of six hyperbolic functions and their derivatives. The formulas are reminiscent of those for the derivatives of trigonometric functions, except for some of the signs.

Function	Derivative
$\cosh(t)$	$\sinh(t)$
$\sinh(t)$	$\cosh(t)$
$\tanh(t)$	$\operatorname{sech}^2(t)$
$\coth(t)$	$-\operatorname{csch}^2(t)$
$\operatorname{sech}(t)$	$-\operatorname{sech}(t)\tanh(t)$
$\operatorname{csch}(t)$	$-\operatorname{csch}(t)\coth(t)$

Table 5.8.1

The Inverses of the Hyperbolic Functions

Inverse hyperbolic functions appear on some calculators and in tables of mathematical functions. As the hyperbolic functions are expressed in terms of exponential functions, each inverse hyperbolic function can be expressed in terms of logarithms. They provide useful antiderivatives as well as solutions to some differential equations.

Consider the inverse of $\sinh(t)$ first. Since $\sinh(t)$ is increasing, it is one-to-one; there is no need to restrict its domain. To find its inverse, it is necessary to solve the equation

$$x = \sinh(t)$$

for t as a function of x. We have

$$\begin{aligned} x &= \frac{e^t - e^{-t}}{2} && (\text{ definition of } \sinh(t)\ ) \\ 2x &= e^t - \frac{1}{e^t} && (\ e^{-t} = 1/e^t\ ) \\ 2xe^t &= (e^t)^2 - 1 && (\text{ multiply by } e^t\ ) \end{aligned}$$

or

$$(e^t)^2 - 2xe^t - 1 = 0. \tag{5.8.1}$$

Since (5.8.1) is quadratic in e^t, the quadratic formula yields

$$e^t = \frac{2x \pm \sqrt{(-2x)^2 + 4}}{2} = x \pm \sqrt{x^2+1}.$$

Observe that $e^t > 0$ and $\sqrt{x^2+1} > x$. As a result, the solution with the plus sign is kept and the one with the minus sign is rejected. Thus

$$e^t = x + \sqrt{x^2+1} \qquad \text{and} \qquad t = \ln\left(x + \sqrt{x^2+1}\right).$$

In conclusion, a definition of the inverse hyperbolic sine is

$$\operatorname{arcsinh}(x) = \sinh^{-1}(x) = \ln\left(x + \sqrt{x^2+1}\right).$$

Computing $\operatorname{arctanh}(x)$ is a little different. Since the derivative of $\tanh(t)$ is $\operatorname{sech}^2(t)$, the function $\tanh(t)$ is increasing and has an inverse. However, $|\tanh(t)| < 1$, so the inverse function will be defined only for $|x| < 1$. Computations similar to those for $\operatorname{arcsinh}(x)$ show that

$$\operatorname{arctanh}(x) = \tanh^{-1}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right) \qquad |x| < 1.$$

Function	Formula	Derivative	Domain
$\operatorname{arccosh}(x)$	$\ln(x+\sqrt{x^2-1})$	$\dfrac{1}{\sqrt{x^2-1}}$	$x \geq 1$
$\operatorname{arcsinh}(x)$	$\ln(x+\sqrt{x^2+1})$	$\dfrac{1}{\sqrt{x^2+1}}$	all x
$\operatorname{arctanh}(x)$	$\dfrac{1}{2}\ln\left(\dfrac{1+x}{1-x}\right)$	$\dfrac{1}{1-x^2}$	$\lvert x\rvert < 1$
$\operatorname{arccoth}(x)$	$\dfrac{1}{2}\ln\left(\dfrac{x+1}{x-1}\right)$	$\dfrac{1}{1-x^2}$	$\lvert x\rvert > 1$
$\operatorname{arcsech}(x)$	$\ln\left(\dfrac{1+\sqrt{1-x^2}}{x}\right)$	$\dfrac{-1}{x\sqrt{1-x^2}}$	$0 < x \leq 1$
$\operatorname{arccsch}(x)$	$\ln\left(\dfrac{1+\sqrt{1+x^2}}{x}\right)$	$\dfrac{-1}{\lvert x\rvert\sqrt{1+x^2}}$	$x \neq 0$

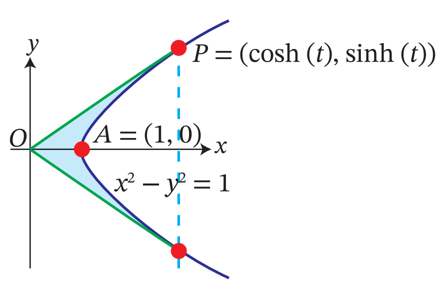
Table 5.8.2

Inverses of the other hyperbolic functions are ted similarly. The functions $\operatorname{arccosh}(x)$ and $\operatorname{arcsech}(x)$ are chosen to be positive. Their formulas are in Table 5.8.2.

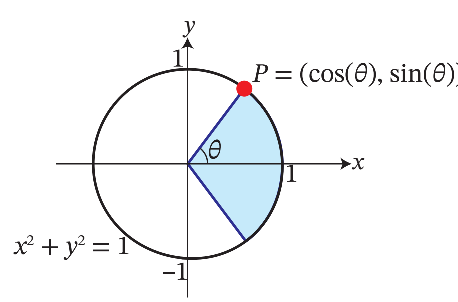
The derivative of each inverse hyperbolic function is found by differentiating the formulas in the second column, and simplifying.

Some Geometry of Hyperbolic Functions

The point $(\cosh(t), \sinh(t))$ lies on the graph of the hyperbola $x^2 - y^2 = 1$. (See Example 1.) The parameter t, which can be any number, has a geometric interpretation: it is the area of the shaded region in Figure 5.8.4(a). This corresponds to the fact that a sector of the unit circle with angle 2θ has area θ, as shown in Figure 5.8.4(b). (See Exercise 64 in the Section 6.S.)



(a)



(b)

Figure 5.8.4

Summary

We introduced the six hyperbolic functions and their inverses, including sinh, cosh, tanh, and their inverses arcsinh, arccosh, and arctanh. Because they are expressible in terms of exponentials, square roots, and logarithms, they do not add to the collection of elementary functions. However, some of them are especially convenient.

EXERCISES for Section 5.8

1. (a) Compute $\cosh(t)$ and $\frac{1}{2}e^t$ for $t = 0, 1, 2, 3$, and 4.

(b) Using (a) and the fact that $\cosh(t)$ is an even function, graph $y = \cosh(t)$ and $y = \frac{1}{2}e^t$ on the same axes.

2. (a) Compute $\tanh(t)$ for $t = 0, 1, 2$, and 3.

(b) Show that $\tanh(t)$ is an odd function.

(c) Use the information in (a) and (b) to graph $y = \tanh(t)$.

In Exercises 3 to 5 obtain the derivatives of the functions and express them in terms of hyperbolic functions.

3. $\sinh(x)$ **4.** $\operatorname{sech}(x)$ **5.** $\coth(x)$

---

6. (a) Compute $\sinh(t)$ and $\cosh(t)$ for $t = -3, -2, -1, 0, 1, 2$, and 3.

(b) Plot the seven points $(x, y) = (\cosh(t), \sinh(t))$ found in (a).

(c) Explain why the points plotted in (b) lie on the hyperbola $x^2 - y^2 = 1$.

7. (a) Show that $\operatorname{sech}^2(x) + \tanh^2(x) = 1$. (b) What equation links $\sec(\theta)$ and $\tan(\theta)$?

In Exercises 8 to 16 use the definitions of the hyperbolic functions to verify the identities.

NOTE: While there are many similarities with the corresponding identities for trigonometric functions, the differences are equally important.

8. $\cosh(x + y) = \cosh(x)\cosh(y) + \sinh(x)\sinh(y)$

9. $\sinh(x + y) = \sinh(x)\cosh(y) + \cosh(x)\sinh(y)$

10. $\cosh(x - y) = \cosh(x)\cosh(y) - \sinh(x)\sinh(y)$

11. $\sinh(x - y) = \sinh(x)\cosh(y) - \cosh(x)\sinh(y)$

12. $\cosh(2x) = \cosh^2(x) + \sinh^2(x)$

13. $\sinh(2x) = 2\sinh(x)\cosh(x)$

14. $2\sinh^2\left(\frac{x}{2}\right) = \cosh(x) - 1$

15. $2\cosh^2\left(\frac{x}{2}\right) = \cosh(x) + 1$

16. $\tanh(x + y) = \dfrac{\tanh(x) + \tanh(y)}{1 + \tanh(x)\tanh(y)}$

In Exercises 17 to 19 obtain a formula for the function in terms of logarithms.

17. $\operatorname{arctanh}(x)$ **18.** $\operatorname{arcsech}(x)$ **19.** $\operatorname{arccosh}(x)$

In Exercises 20 to 23 show that the derivative of the first expression is the second expression.

20. $\operatorname{arccosh}(x)$; $\dfrac{1}{\sqrt{x^2 - 1}}$ **21.** $\operatorname{arcsinh}(x)$; $\dfrac{1}{\sqrt{x^2 + 1}}$ **22.** $\operatorname{arcsech}(x)$; $\dfrac{-1}{x\sqrt{1 - x^2}}$ **23.** $\operatorname{arccsch}(x)$; $\dfrac{1}{x\sqrt{1 + x^2}}$

---

24. Find any inflection points on the curve $y = \tanh(x)$.

25. Graph $y = \sinh(x)$ and $y = \operatorname{arcsinh}(x)$ on the same axes. Show any inflection points.

26. Hyperbolic functions are encountered in the study of motion in which the resistance of the medium is proportional to the square of the velocity. In that application, a body starts from rest and falls x meters in t seconds. Let g (a constant) be the acceleration due to gravity ($g = 9.8\ \text{m/s}^2$). It can be shown that there is a constant $V > 0$ such that $x = \frac{V^2}{g}\ln\left(\cosh\left(\frac{gt}{V}\right)\right)$.

(a) Find the velocity $v(t) = \frac{dx}{dt}$ (as a function of t).

(b) Show that $\lim_{t\to\infty} v(t) = V$.

(c) Compute the acceleration $a(t) = \frac{dv}{dt}$.

(d) Show that $\frac{a(t)}{g} + \left(\frac{v(t)}{V}\right)^2 = 1$.

(e) What is the limit of the acceleration as $t \to \infty$?

(f) What is $\lim_{t\to\infty} x(t)$?

27. This exercise concerns antiderivatives of $f(x) = \frac{1}{\sqrt{ax+b}\sqrt{cx+d}}$.

(a) Show that $\frac{2}{\sqrt{-ac}}\arctan\left(\sqrt{\frac{-c(ax+b)}{a(cx+d)}}\right)$ is an antiderivative of $f(x)$ when $a > 0$ and $c < 0$.

(b) Show that $\frac{2}{\sqrt{ac}}\operatorname{arctanh}\left(\sqrt{\frac{c(ax+b)}{a(cx+d)}}\right)$ is an antiderivative of $f(x)$ when $a > 0$ and $c > 0$.

5.S Chapter Summary

This chapter showed the derivative at work: applying it to practical problems, estimating errors, and evaluating some limits.

To determine the extrema of a quantity we need to find a function that tells how it depends on other quantities. Then, finding the extrema is like finding the highest or lowest points on the graph of the function.

When two varying quantities are related by an equation, differentiate both sides of that equation to find a relationship between their rates of change.

The next two sections formed a unit that presents one of the main uses of higher derivatives: to estimate errors when approximating a function by a polynomial and later, in Section 6.5, to estimate errors in approximating area under a curve by trapezoids and parabolas.

The key to the growth theorem is that if R is a function such that

$$0 = R(a) = R'(a) = R''(a) = \cdots = R^{(n)}(a)$$

and in some interval around a we know $R^{(n+1)}(x)$ is continuous, then there is a number c between a and x, which depends on x, such that

$$R(x) = R^{(n+1)}(c)\frac{(x-a)^{n+1}}{(n+1)!} \qquad \text{for all } x \text{ in } [a,x].$$

This gives us information on how rapidly $R(x)$ can grow for x near a. It was used to control the error when using a polynomial to approximate a function.

A likely candidate for the polynomial of degree n that closely resembles a given function f near $x = a$ is the one whose derivatives at a, up through order n, agree with those of f at a. That polynomial is

$$P(x) = P_n(x;a) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

Because the polynomial was chosen so that $P^{(k)}(a) = f^{(k)}(a)$ for all k up through n, the remainder function function $R(x) = f(x) - P(x)$ has all its derivatives up through order n at a equal to 0. Moreover, since the $(n+1)^{\text{st}}$

derivative of any polynomial of degree at most n is identically 0, $R^{(n+1)}(x) = f^{(n+1)}(x)$. Using these facts we obtained Lagrange's formula for the error:

$$\frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \qquad \text{for some } c \text{ between } a \text{ and } x.$$

Section 5.6 concerned l'Hôpital's rule, a tool for computing limits, such as the limit of a quotient whose numerator and denominator both approach zero (0/0) or both approach ∞ (∞/∞). Other indeterminate forms, including $0\cdot\infty$, $\infty-\infty$, 0^0, 1^∞, must be converted to one of these two forms before using l'Hôpital's rule.

The chapter concludes with sections on natural growth and decay and the hyperbolic functions.

EXERCISES for Section 5.S

1. Arrange the following numbers in order of increasing size as $x \to \infty$.
(a) $1000x$ (b) $\log_2(x)$ (c) $\sqrt{x}$ (d) $(1.0001)^x$ (e) $\log_{1000}(x)$ (f) $0.01x^3$

In Exercises 2 to 28 find the limits, if they exist. Indicate any uses of l'Hôpital's rule, including the type of indeterminate form.

2. $\displaystyle\lim_{u\to\infty}\left(\frac{u+1}{u}\right)^{u+1}\frac{1}{\sqrt{u}}$

3. $\displaystyle\lim_{x\to\infty}\left(\frac{x+2}{x+1}\right)^{x+3}$

4. $\displaystyle\lim_{x\to\infty}\left(\frac{x}{x+1}\right)^{x+1}$

5. $\displaystyle\lim_{x\to3}\frac{x-3}{\cos(\pi x)}$

6. $\displaystyle\lim_{x\to3}\frac{x-3}{\sin(\pi x)}$

7. $\displaystyle\lim_{x\to\infty}\frac{\sqrt{1+x^2}}{x}$

8. $\displaystyle\lim_{x\to\infty}\frac{\sqrt{1+x^2}}{\sqrt{2+x^2}}$

9. $\displaystyle\lim_{x\to\infty}\frac{(1+x^2)^{1/2}}{(2+x^2)^{1/3}}$

10. $\displaystyle\lim_{x\to\infty}\frac{1+x+x^2}{2+3x+4x^2}$

11. $\displaystyle\lim_{x\to1}\frac{\ln(x)\tan\left(\frac{\pi x}{4}\right)}{\cos\left(\frac{\pi x}{2}\right)}$

12. $\displaystyle\lim_{x\to0}\frac{f(3+x)-f(3)}{x}$ where $f(x)=(x^2+5)\sin^2(3x)$

13. $\displaystyle\lim_{x\to0}\frac{(x+2)(\cos(5x)-1)}{(x+3)(\cos(7x)-1)}$

14. $\displaystyle\lim_{x\to\infty}\frac{\ln(6x)-\ln(5x)}{\ln(7x)-\ln(6x)}$

15. $\displaystyle\lim_{x\to\infty}\frac{\ln(6x)-\ln(5x)}{x\ln(7x)-x\ln(6x)}$

16. $\displaystyle\lim_{x\to\pi}\frac{e^{-x^2}\sin(x)}{x^2-\pi^2}$

17. $\displaystyle\lim_{x\to\pi}\frac{\ln(x^3-\sin(x))-3\ln(\pi)}{x-\pi}$

18. $\displaystyle\lim_{x\to\infty}\left(\frac{x+2}{x+1}\right)^{2x}$

19. $\displaystyle\lim_{x\to\pi}\frac{\sin^4(x)}{(\pi^4-x^4)^2}$

20. $\displaystyle\lim_{x\to\infty}\frac{\sec^4(x)\tan(3x)}{\sin(2x)}$

21. $\displaystyle\lim_{x\to1}\frac{e^{3x}(x^2-1)}{\cos(\sqrt{2}x)\tan(3x-3)}$

22. $\displaystyle\lim_{t\to0}\frac{e^{3(x+t)}-e^{3x}}{5t}$

23. $\displaystyle\lim_{t\to\infty}\frac{e^{3(x+t)}-e^{3x}}{5t}$

24. $\displaystyle\lim_{x\to0}\left(\frac{1+2^x}{2}\right)^{1/x}$

25. $\displaystyle\lim_{x\to0}\left(\frac{1+2^x}{1+3^x}\right)^{1/x}$

26. $\displaystyle\lim_{x\to0}(1+0.005x)^{20x}$

27. $\displaystyle\lim_{x\to\infty}(1+0.003x)^{20/x}$

28. $\displaystyle\lim_{x\to0}(1+0.003x)^{20/x}$

In Exercises 29 to 36 find the derivative of the given function.

29. $(\cos(x))^{1/x^2}$

30. $(\cos^2(3x))^{\cos^2(2x)}$

31. $\ln\left(\sec^2(3x)\sqrt{1+x^2}\right)$

32. $\displaystyle\ln\left(\sqrt{e^{x^3}}\right)\frac{\tan^2(2x)}{(1+\cos(2x))^4}$

33. $\displaystyle\frac{5+3x+7x^2}{58-4x+x^2}\;\frac{\tan^2(2x)}{(1+\cos(2x))^4}$

34. $\displaystyle\frac{\tan^2(2x)}{(1+\cos(2x))^4}$

35. $f(x)=\begin{cases} x^2\sin\left(\frac{\pi}{x}\right) & \text{if } x\neq0 \\ 0 & \text{if } x=0\end{cases}$

36. $f(x)=\begin{cases} \dfrac{\sin(\pi x)}{x} & \text{if } x\neq0 \\ \pi & \text{if } x=0\end{cases}$

37. (a) Find $P_1(x;64)$ for $f(x)=\sqrt{x}$. (b) Use $P_1(x;64)$ to estimate $\sqrt{67}$. (c) Bound the error in the estimate in (b).

38. (a) Show that when x is small $\sqrt[3]{1+x}$ is approximately $1+x/3$. (b) Use (a) to estimate $\sqrt[3]{0.94}$ and $\sqrt[3]{1.06}$.

39. (a) Show that when x is small $\dfrac{1}{\sqrt[3]{1+x}}$ is approximately $1-\dfrac{x}{3}$. (b) Use (a) to estimate $\dfrac{1}{\sqrt[3]{0.94}}$ and $\dfrac{1}{\sqrt[3]{1.06}}$.

40. (a) Find the Maclaurin polynomial of degree 6 associated with $\cos(x)$.

(b) Use (a) to estimate $\cos\left(\dfrac{\pi}{4}\right)$.

(c) What is the error between the estimate found in (b) and the exact value, $\dfrac{\sqrt{2}}{2}$?

(d) What is the Lagrange bound for the error?

In Exercises 41 to 52 determine whether the limit exists, and, if it does exist, find its value.

41. $\lim\limits_{x\to 1}\dfrac{1-e^x}{1-e^{2x}}$

42. $\lim\limits_{x\to 0}\dfrac{x}{\sqrt{1+x^2}}$

43. $\lim\limits_{x\to 0}\dfrac{1-e^x}{1-e^{2x}}$

44. $\lim\limits_{x\to\infty}\dfrac{x^2}{(1+x^3)^{2/3}}$

45. $\lim\limits_{x\to\infty} x^2\sin(x)$

46. $\lim\limits_{x\to 8}\dfrac{2^x-2^8}{x-8}$

47. $\lim\limits_{x\to 1}\dfrac{e^{x^2}-e^x}{x-1}$

48. $\lim\limits_{x\to 4}\dfrac{2^x+2^4}{x+4}$

49. $\lim\limits_{x\to 0}\dfrac{\sin(x)-e^{2x}}{x}$

50. $\lim\limits_{x\to 0}\dfrac{e^{3x}\sin(2x)}{\tan(3x)}$

51. $\lim\limits_{x\to 0}\dfrac{\sqrt{1+x^2}-1}{\sqrt[3]{1+x^2}-1}$

52. $\lim\limits_{x\to\pi/2}\dfrac{\sin(x)\cos(x)}{x-\frac{\pi}{2}}$

53. If $\lim\limits_{x\to\infty} f'(x)=3$ and $\lim\limits_{x\to\infty} g'(x)=3$, what, if anything, can be said about

(a) $\lim\limits_{x\to\infty}\dfrac{f(x)}{3x}$ (b) $\lim\limits_{x\to\infty}(g(x)-f(x))$ (c) $\lim\limits_{x\to\infty}\dfrac{f(x)}{g(x)}$ (d) $\lim\limits_{x\to\infty}(f(x)-3x)$ (e) $\lim\limits_{x\to\infty}\dfrac{(f(x))^3}{(g(x))^3}$

54. Let $f(x)=(5x^3+x+2)^{20}$. Find (a) $f^{(60)}(4)$ and (b) $f^{(61)}(2)$.

55. Figure 5.S.1(a) shows a rectangle whose base lies on the x-axis and which is inscribed in the parabola $y=1-x^2$. (a) Find the rectangle of largest perimeter. (b) Find the rectangle of largest area.

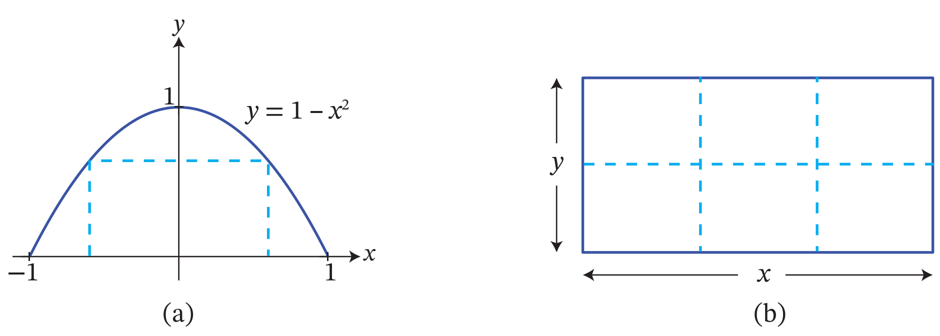


Figure 5.S.1

56. A farmer has 200 feet of fence that he wants to use to enclose a rectangle divided into six congruent rectangles, as shown in Figure 5.S.1(b). He wishes to enclose a maximum area. (a) If x is near 0, what is the area, approximately? (b) How large can x be? (c) In the case that produces the maximum area, which do you think will be larger, x or y? Why? (d) Find the dimensions x and y that maximize the area.

57. A rectangle of perimeter 12 inches is spun around one of its edges to produce a circular cylinder.

(a) For which rectangle is the area of the curved surface of the cylinder a maximum?

(b) For which rectangle is the volume of the cylinder a maximum?

Exercises 58 and 59 are related.

58. Consider isosceles triangles whose equal sides have length a and whose angle where the equal sides meet is θ. For which θ is the area of the triangle a maximum? Solve this problem (a) using calculus and (b) without using calculus.

59. Let c and d be positive numbers. Figure 5.S.2 shows a line segment through $P = (c, d)$ whose ends are on the positive x- and y-axes. Let θ be the acute angle between the line and the x-axis. Show that the angle that produces the shortest line segment through P is $\theta = \alpha$ where $\tan^3(\alpha) = d/c$.

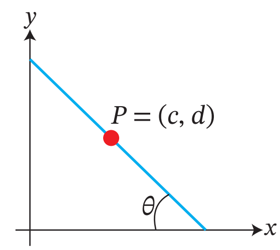


Figure 5.S.2

---

60. Define $g(a)$ to be the area of the triangle formed by the tangent to $y = \frac{1}{x}$ at $\left(a, \frac{1}{a}\right)$, the x-axis, and the y-axis.

(a) Before finding a formula for $g(a)$, how do you think this area changes as a increases?

(b) Find $\lim_{a\to\infty} g(a)$.

(c) Does the limit in (b) agree with your opinion in (a)?

(d) Now that you know the answer to (b), find this limit using elementary geometry.

(e) What if the line through $\left(a, \frac{1}{a}\right)$ is not tangent, but has a constant slope $k < 0$?

Exercises 61 to 64 are all related.

61. Let (a, b) be a point in the first quadrant. Each line through (a, b) with a negative slope, together with the x- and y-axes, forms a triangle. Find the slope of the line that minimizes the area of the triangle.

62. (a) Use the result from Exercise 61 to draw a slope field with the slope at (a, b) being the slope of the line that minimizes the area of the triangle formed by the line and the x- and y-axes.

(b) Sketch three curves that follow the slope field created in (a).

(c) If the point (a, b) is on C, one of the curves drawn in (b), why is the tangent to C at (a, b) the line through (a, b) that minimizes the area?

63. (a) Show that if $y = f(x)$ is a curve suggested by the slope field in Exercise 62, then $\frac{1}{y}\frac{dy}{dx} = -\frac{1}{x}$.

(b) Deduce that $f(x) = \frac{k}{x}$, where k is a constant (different for each curve).

64. Exercises 61 to 63 show, in particular, that the tangent to the graph of $y = \frac{1}{x}$ at a point P is the line through P that minimizes the area of the triangles formed by a line through P and the positive x- and y-axes. Establish this directly.

---

65. A semicircle of radius $r \le 1$ rests upon a semicircle of radius 1, as shown in Figure 5.S.3(a). The length of PQ, the segment from the origin of the lower circle to the top of the upper circle, is a function of r, $f(r)$.

(a) Find $f(0)$ and $f(1)$. (b) Find $f(r)$. (c) Maximize $f(r)$, testing the maximum by the second derivative.

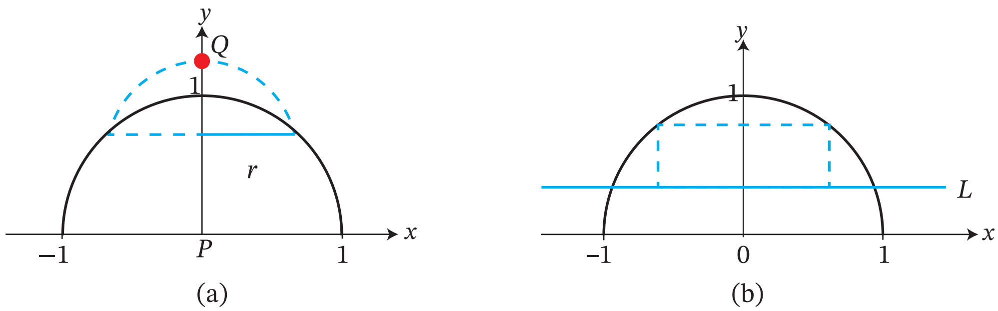


Figure 5.S.3

Exercises 66 to 68 are independent, but related. They contain a surprise.

66. Figure 5.S.3(b) shows the top half of the unit circle $x^2 + y^2 = 1$, the line L whose equation is $y = \frac{1}{3}$, and a rectangle with base on L, inscribed in the circle. Find the rectangle with base on L that has (a) minimum perimeter and (b) maximum perimeter.

67. As Exercise 66 but this time the line L has the equation $y = \frac{1}{2}$.

---

68. The analyses in Exercises 66 and 67 are different. Let the line L have the equation $y = c$, $0 < c < 1$. For which values of c is the analysis like that for (a) Exercise 66? (b) Exercise 67?

69. A. Bellemans, in "Power demand in walking and pace optimization," *Amer. J. Physics* 49(1981), pp. 25–27, develops a model for the work spent while walking. At one point he writes "$H = L(1 - \cos(\gamma))$ or, to a sufficient approximation for the present purpose, $H = \frac{L\gamma^2}{2}$." Justify this approximation when γ is small.

70. Two houses, A and B, are a distance p apart. They are on the same side of the road at distances q and r, respectively, from it. Find the length of the shortest path that goes from A to the road, and then to B.
(a) Use calculus. (b) Use only elementary geometry.

71. For a constant k, determine $\lim_{x\to\infty} x\left(e^{-k} - \left(1 - \frac{k}{x}\right)^x\right)$.

72. For a constant k, determine $\lim_{x\to\infty} x\left(e^{k} - \left(1 + \frac{k}{x}\right)^x\right)$.

In each of Exercises 73–74 there are three functions and their Maclaurin polynomials. While the notation is slightly nonstandard, it will be convenient to use $P_n(f(x))$ to denote the Maclaurin polynomial of order n associated with a function $f(x)$. For example, $P_2(e^x) = 1 + x + \frac{x^2}{2}$.

73. Because $e^x e^{-x} = 1$, we might expect that the product of Maclaurin polynomials $P_n(e^x)\,P_n(e^{-x}) = 1$. But, this cannot be true because the degree of the product is $2n$. However, there is a relation, which can be examined with the aid of the following steps.

(a) Check that $P_1(e^x)\,P_1(e^{-x}) = 1 - x^2$, which agrees with $P_1(1) = 1 + 0x$ through the linear term.

(b) Check that $P_2(e^x)\,P_2(e^{-x}) = 1 + \frac{x^4}{4}$, which agrees with $P_3(1) = 1 + 0x + 0x^2 + 0x^3$ through the x^3 term.

(c) Check that $P_3(e^x)\,P_3(e^{-x}) = 1 - \frac{x^4}{12} - \frac{x^6}{36}$, which agrees with $P_3(1) = 1 + 0x + 0x^2 + 0x^3$ through the x^3 term.

(d)

n	1	2	3	4	5	6	7	8	9	10
agreement	1	3	3							

(e) Use the results found in (d) to conjecture how the table continues for larger values of n.

74. Repeat Exercise 73, comparing the number of terms that agree in the product of Maclaurin polynomials $P_n(e^x)\,P_n(e^x)$ and the Maclaurin polynomial of the product $P_n\left(e^{2x}\right)$.

75. Repeat Exercise 73, comparing the number of terms that agree in the product of Maclaurin polynomials $P_n\left(\frac{1}{1+x}\right)P_n\left(\frac{1}{1-x}\right)$ and the Maclaurin polynomial of the product $P_n\left(\frac{1}{1-x^2}\right)$.

---

76. What can be said about $f(10)$ if $f(1)=5$, $f'(1)=3$, and $f''(x)<4$ for x in $(-10,20)$?

77. The demand for a product is influenced by its price. In one example an economics text links the amount sold (x) to the price (P) by the equation $x=b-aP$, where b and a are positive constants. As the price increases the sales decrease. The cost of producing x items is an increasing function $C(x)=c+kx$, where c and k are positive constants.

(a) Express P in terms of x.
(b) Express the total revenue $R(x)$ in terms of x.
(c) What is the economic significance of c?
(d) What is the economic significance of k?
(e) Let $E(x)$ be the profit, that is, the revenue minus the cost. Express $E(x)$ as a function of x.
(f) Which value of x produces the maximum profit?
(g) The marginal revenue is defined as $\frac{dR}{dx}$ and the marginal cost as $\frac{dC}{dx}$. Show that for the value of x that produces the maximum profit, $\frac{dR}{dx}=\frac{dC}{dx}$.
(h) What is the economic significance of $\frac{dR}{dx}=\frac{dC}{dx}$ in (g)?

78. This exercise concerns a function used to describe the consumption of a finite resource, such as petroleum. Denote by Q the total amount available. For a positive constant a and a negative constant b, the amount consumed by time t is $y(t)=\frac{Q}{1+ae^{bt}}$.

(a) Show that $\lim_{t\to\infty} y(t)=Q$ and $\lim_{t\to-\infty} y(t)=0$. Why are these realistic?
(b) Show that $y(t)$ has an inflection point when $t=-\frac{\ln(a)}{b}$.
(c) Show that at the inflection point, $y(t)=\frac{Q}{2}$, that is, half the resource has been used up.
(d) Sketch the graph of $y(t)$.
(e) Where is $y'(t)$, the rate of using the resource, greatest?

The same function describes limited growth that is bounded by Q, so called **logistic growth**.

79. One hundred cubic yards of debris are added to a landfill every day. The operator decides to pile the debris up in the form of a cone whose base angle is $\frac{\pi}{4}$.

(a) At what rate is the height of the cone increasing when the height is (i) 10 yards? (ii) 20 yards? (iii) 100 yards?
(b) How long will it take to make a cone (i) 100 yards high? (ii) 300 yards high?

The volume of a circular cone is one-third the product of its height and the area of its circular base.

80. A wine dealer has a case of wine that he could sell today for \$100. Or, he could decide to store it, letting it age, and sell later for a higher price. Assume he could sell in t years for \$ $100e^{\sqrt{t}}$. To decide which option to choose he computes the **present value** of the sale. When should he sell the wine?

If the interest rate is r, the present value of one dollar t years hence is e^{-rt}.

Exercises 81 to 83 are related.

81. A differentiable function is defined throughout $(-\infty, \infty)$. Its derivative is 0 at exactly two inputs.

(a) Can there be exactly one relative extremum?
(b) Could it have two relative maxima?
(c) What is the greatest number of relative extrema possible?
(d) What is the least number?

82. A differentiable function is defined throughout $(-\infty, \infty)$. Its derivative is 0 at exactly three inputs and the function approaches 0 as x approaches ∞.

(a) Can there be exactly two relative extremum?
(b) Could it have three relative maxima?
(c) What is the greatest number of relative extrema possible?
(d) What is the least number?

83. A differentiable function is defined throughout $(-\infty, \infty)$. Its derivative is 0 at exactly two inputs and the function approaches the same finite limit as x approaches ∞ and $-\infty$.

(a) Can there be exactly one relative extremum?
(b) Could it have two relative maxima?
(c) What is the greatest number of relative extrema possible?
(d) What is the least number?

---

84. In the paper cited in Exercise 69, Bellemans writes "The total mechanical power required for walking is $P(v,a) = \dfrac{\alpha M v^3}{a} + \dfrac{\beta M g v}{L} a$. Enlarging the pace, a, at a constant speed v, lowers the first term and increases the second one so that the formula predicts an optimal pace $a^*(v)$, minimizing $P(v,a)$."

In the formula, α, M, β, g, and L are constants. Verify that (a) $a^*(v) = \left(\dfrac{\alpha}{\beta}\right)^{1/2} \left(\dfrac{L}{g}\right)^{1/2} v$ and (b) the corresponding minimum power is $P(v, a^*(v)) = 2(\alpha\beta)^{1/2} \left(\dfrac{g}{L}\right)^{1/2} M v^2$.

"One would therefore expect that, when walking naturally on the flat at a fixed velocity, a subject will adjust its pace automatically to the optimum value corresponding to the minimum work expenditure. This has indeed been verified experimentally."

85. Figure 5.S.4(a) shows two points A and B a mile apart and both at a distance a from the river CD. Sam is at A. He will walk in a straight line to the river at 4 mph, fill a pail, then continue on to B at 3 mph. He wishes to do this in the shortest time.

(a) For the fastest route which angle in Figure 5.S.4(a) do you expect to be larger, α or β?
(b) Show that for the fastest route $\sin(\alpha)/\sin(\beta)$ equals $4/3$.

86. A fence b feet high is a feet from a tall building, whose wall contains BC, as shown in Figure 5.S.4(b). Find the angle θ that minimizes the length of AB.

87. (a) Show that if $p_n(x)$ is the n^{th}-order Maclaurin polynomial associated with $f(x)$, then $p_n'(x)$ is the $(n-1)^{\text{st}}$-order Maclaurin polynomial associated with $f'(x)$.

(b) Use (a) to find the 6^{th}-order Maclaurin polynomial for $\dfrac{1}{(1-x)^2}$.

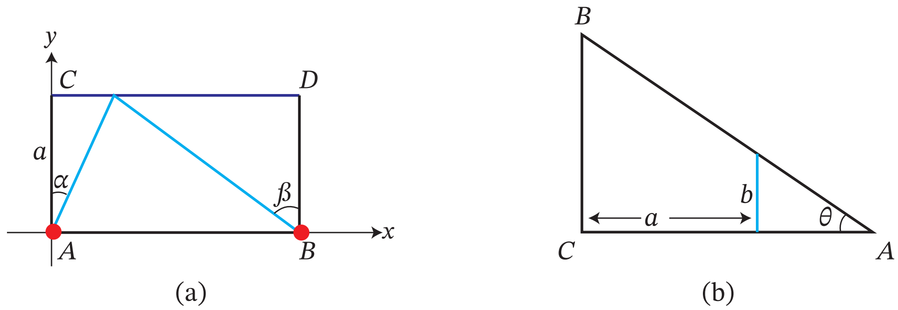


Figure 5.S.4

88. Let $P_n(x)$ be the n^{th}-order Maclaurin polynomial associated with e^x. For how large an x can one be sure that (a) $|e^x - P_1(x)| < 0.01$? (b) $|e^x - P_2(x)| < 0.01$? (c) $|e^x - P_3(x)| < 0.01$?

89. A number b is **algebraic** if there is a nonzero polynomial $\sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$, with all a_i rational, such that $\sum_{i=0}^{n} a_i b^i = 0$. In other words, b is algebraic if there is a function f that satisfies $f(b) = 0$, all derivatives of f at 0 are rational, but not all zero, and there is a positive integer m such that $D^m(f) = 0$. (Recall that D is the differentiation operator.)

We call a number b **almost algebraic** if b is not algebraic and there is a function f with $f(b) = 0$, all derivatives of f at 0 are rational, but not all zero, and there is a nonzero polynomial $p(D)$ such that $p(D)(f) = 0$. For example, if $p(x) = x^2 + 1$ then $p(D)(f) = D^2(f) + f = f'' + f$.

Show that π is almost algebraic.

Kepler, the astrologer and astronomer, to celebrate his wedding in 1613, ordered some wine, which was available in cylindrical barrels of various shapes and sizes. He was surprised by the way the volume of a barrel was measured.

Here is the measurement method Kepler observed: A ruler was pushed through the opening in the side of the barrel (used to fill the barrel) until it came to a stop at the edge of a circular base. The length of the part of the ruler inside the barrel was used to determine the volume of the barrel. Figure 5.S.5 shows the configuration of the barrel and ruler.

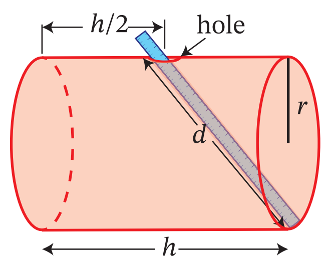


Figure 5.S.5

Exercise 90 explores this process in more detail. Part (d) is what Kepler discovered.

CHALLENGE: Answer (d) two different ways, one without implicit differentiation and the other with it.

90. The barrel in Figure 5.S.5 has radius r, height h, and volume V. The length of the ruler inside the barrel is d.
(a) Using common sense, show that d does not determine V.
(b) How small can V be for a given value of d?
(c) Using calculus, show that the maximum volume for a given d occurs when $h = \dfrac{2d}{\sqrt{3}}$ and $r = \dfrac{d}{\sqrt{6}}$.
(d) Show that to maximize the volume the height must be $\sqrt{2}$ times the diameter.

---

91. Let m and n be positive integers. Let $f(x) = \sin^m(x)\cos^n(x)$ for x in $\left[0, \dfrac{\pi}{2}\right]$.
(a) For which x is $f(x)$ a minimum? (b) For which x is $f(x)$ a maximum? (c) What is the maximum value of $f(x)$?

92. (a) Let $P(x)$ be a polynomial such that $D^2(x^2P(x)) = 0$. Show that $P(x) = 0$.
(b) Does the same conclusion follow if instead we assume $D^2(xP(x)) = 0$?

93. Translate the following news item into the language of calculus: "The one positive sign during the quarter was a slowing in the rate of increase in home foreclosures."

94. Assume that $f(x)$ is defined on $[0,\infty)$, has a continuous positive second derivative, and $\lim_{x\to\infty} f(x) = 0$.
(a) Can $f(x)$ ever be negative?
(b) Can $f'(x)$ ever be positive?
(c) What are the possible shapes for the graph of $y = f(x)$?
(d) Give an explicit formula for one function with these properties.

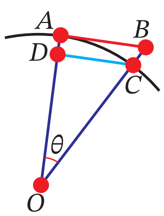


Figure 5.S.6

Exercise 95 is a challenge to your intuition.

95. In Figure 5.S.6 AB is tangent to an arc of a circle, OA is a radius and DC is parallel to AB.
(a) Observe that regions ADC and ABC share a common side, namely arc $\overparen{AC}$. What do you think happens to the ratio of the area of ABC to the area of ADC as $\theta \to 0$?
(b) Using calculus, find the limit of the ratio as $\theta \to 0$.

96. Evaluate the following limits. (a) $\lim_{x\to 0}\left(\frac{1+2^x}{2}\right)^{1/x}$ and (b) $\lim_{x\to 0}\left(\frac{1+2^x}{1+3^x}\right)^{1/x}$.

97. Evaluate the following limits (a) $\lim_{x\to\infty}\frac{x(x+1)^x}{x^{x+1}}$ and (b) $\lim_{x\to\infty}\frac{(x+1)^x}{x^{x+1}}$.

98. **Jane:** I wonder which is bigger, 2001^{2000} or 2000^{2001}?
Sam: Obviously the one with the bigger base.
Jane: But its exponent is smaller than the exponent of the other.
Sam: I think the base has more influence.
Jane: And I think the exponent has more impact.

Settle the dispute by examining the ratio $\frac{2001^{2000}}{2000^{2001}}$.

99. **Sam:** I can use Taylor polynomials to get l'Hôpital's theorem.
Jane: How so?
Sam: For instance, I write $f(x) = f(0) + f'(0)x + f''(c)\frac{x^2}{2}$ and $g(x) = g(0) + g'(0)x + g''(d)\frac{x^2}{2}$.
Jane: O.K.
Sam: Since $\lim_{x\to 0} f(x)$ and $\lim_{x\to 0} g(x)$ are both zero I have $f(0) = g(0) = 0$. I can write, after canceling some x's

$$\frac{f(x)}{g(x)} = \frac{f'(0) + f''(c)\frac{x}{2}}{g'(0) + g''(d)\frac{x}{2}}.$$

Jane: But you don't know the second derivatives.
Sam: It doesn't matter. I just take limits and get

$$\lim_{x\to 0}\frac{f(x)}{g(x)} = \lim_{x\to 0}\frac{f'(0) + f''(c)\frac{x}{2}}{g'(0) + g''(d)\frac{x}{2}}.$$

So

$$\lim_{x\to 0}\frac{f(x)}{g(x)} = \frac{f'(0)}{g'(0)} = \lim_{x\to 0}\frac{f'(x)}{g'(x)}.$$

There you have it.

Jane: Let me check your steps.

Check Sam's steps and comment on this (proposed) method.

100. Graph $y = x^2 \ln(x)$, showing extrema and inflection points. What is $\lim_{x\to 0^+} x^2\ln(x)$?

101. If $P(x)$ is a Maclaurin polynomial associated with $f(x)$, what is the Maclaurin polynomial of the same order associated with $f(2x)$?

102. Find the Maclaurin polynomial of order 6 associated with $\frac{1}{e^x}$.

103. Find the Maclaurin polynomial of order 6 associated with $\sin(x)\cos(x)$.

104. The center $(x, 0)$ of a circle C_1 of radius 1 is at a distance x, with $1 < x < 3$, from the center $(0,0)$ of a circle C_2 of radius 2. AB is the chord joining the two points the circles have in common. Let A_1 be the area within C_1 to the left of the chord and A_2 the area within C_2 to the right of the chord.

(a) Which is larger, A_1 or A_2?

(b) If $\lim_{x\to 3^-}\frac{A_2}{A_1}$ exists, what do you think it is?

(c) Determine whether the limit in (b) exists. If it does, find it.

105. In Exercise 104, let O_1 be the center of C_1 and O_2 the center of C_2. What happens to the ratio of the area common to the two circles and the area of the quadrilateral AO_1BO_2 as $x \to 3^-$?

106. Let $g(x) = f(x^2)$.

(a) Express the Maclaurin polynomial for $g(x)$ up through the x^4 term in terms of f and its derivatives.

(b) How is the answer in (a) related to a Maclaurin polynomial associated with f?

107. Find $\lim_{x\to\pi/2^-}(\sec(x) - \tan(x))$ (a) using l'Hôpital's rule and (b) without using l'Hôpital's rule.

108. Every six hours a patient takes an amount A of a medicine. Once in the patient, the amount decays exponentially. In six hours it declines from A to kA, where k is less than 1 (and positive), and then increases to $A + kA$ when the second dose of medicine is taken. At the end of the second six-hour period, $t = 12$, the amount in the system is $k(A + kA)$. At this instant the third dose is taken, increasing the amount of medicine in the system to $A + kA + k^2A$.

(a) Graph the general shape of the amount of medicine in the patient as a function of time.

(b) When a pill is taken at the end of n six-hour periods how much is in the system?

(c) Does the amount in the system become arbitrarily large? (If so, this could be dangerous.)

The constant k depends on many factors, such as the age of the patient. For this reason, a dosage tested on a 20-year old may be lethal on a 70-year. See also Exercise 32 in Section 11.1.

The remaining exercises offer an opportunity to practice differentiation. In each case show that the derivative of the first expression is the second one.

109. $\arctan\left(\frac{x}{a}\right)$, $\frac{a}{x^2+a^2}$.

110. $\frac{2(3ax-2b)}{15a^2}\sqrt{(ax+b)^3}$, $x\sqrt{ax+b}$.

111. $\sin(ax) - \frac{1}{3}\sin^3(ax)$, $a\cos^3(ax)$.

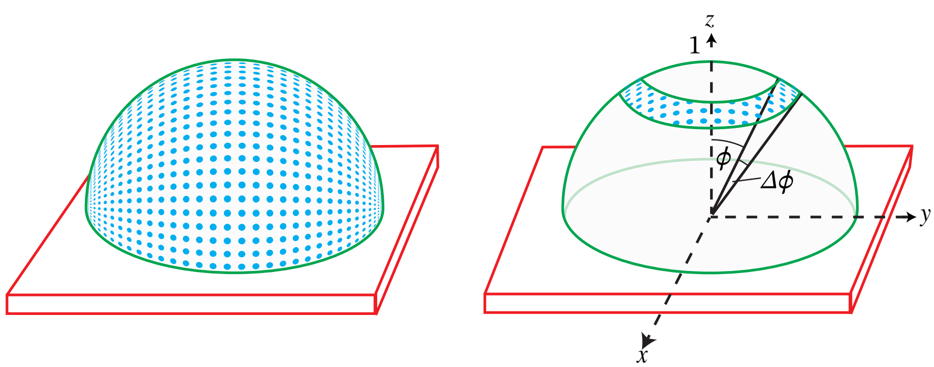
112. $e^{ax}(a\cos(bx) + b\sin(bx))$, $(a^2+b^2)e^{ax}\cos(bx)$.

Calculus is Everywhere # 7

The Uniform Sprinkler

One day one of the authors (SS) realized that his sprinkler did not water his lawn evenly. Placing empty cans throughout the lawn, he discovered that some places received as much as nine times as much water as others. That meant some parts of the lawn were getting too much water and others not enough.

The sprinkler, which had no moving parts, consisted of a hemisphere, with holes distributed uniformly on its surface, as in Figure C.7.1(a). Though the holes were uniformly spaced, water was not supplied uniformly to the lawn. Why not?



(a) (b)

Figure C.7.1

A little calculus answered that question and showed how the holes should be placed to have an equitable distribution.

Assume that the radius of the spherical head is 1, that the speed of the water as it left the head was the same at any hole, and disregard air resistance.

Our objective is to find how much water is contributed to the lawn by the uniformly spaced holes in a narrow band of width $\Delta\phi$ near the angle ϕ, as shown in Figure C.7.1(a). To be sure the jet was not blocked by the grass, the angle ϕ is assumed to be no more than $\pi/4$. Water from the band wets the ring shown in Figure C.7.2.

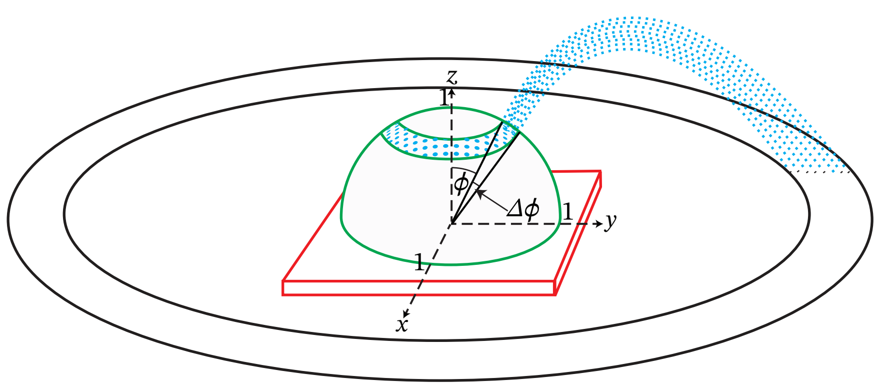


Figure C.7.2

The area of the band on the sprinkler is then roughly $2\pi\sin(\phi)\,\Delta\phi$. As shown in Section 9.3 (Exercises 24 and 25), water from the band lands at a distance from the sprinkler of about

$$x = kv^2\sin(2\phi).$$

Here k is a constant and v is the speed of the water as it leaves the sprinkler. The width of the ring on the lawn is roughly

$$\Delta x = 2kv^2\cos(2\phi)\Delta\phi.$$

Since its radius is approximately $kv^2\sin(2\phi)$, its area is approximately

$$2\pi x\,\Delta x \approx 2\pi\left(kv^2\sin(2\phi)\right)\left(2kv^2\cos(2\phi)\,\Delta\phi\right),$$

which is proportional to $\sin(2\phi)\cos(2\phi)$, hence to $\sin(4\phi) = 2\sin(2\phi)\cos(2\phi)$.

Thus the water supplied by the band is proportional to $\sin(\phi)$ but the area watered by it was proportional to $\sin(4\phi)$. The ratio

$$\frac{\sin(4\phi)}{\sin(\phi)} \quad \text{is proportional to} \quad \frac{\text{Area watered on lawn}}{\text{Volume of water from band}} \tag{C.7.1}$$

is the key to understanding why the distribution was not uniform and to finding out how the holes should be placed to water the lawn uniformly.

By l'Hôpital's rule, the fraction approaches 4 as ϕ approaches zero:

$$\lim_{\phi\to 0}\frac{\sin(4\phi)}{\sin(\phi)} = 4. \tag{C.7.2}$$

For angles ϕ near 0 the ratio is near 4. When ϕ is $\pi/4$, that ratio is $\sin(\pi)/\sin(\pi/4) = 0$, and water was supplied much more heavily far from the sprinkler than near it. To compensate for this the number of holes in the band should be proportional to $\sin(4\phi)/\sin(\phi)$. Then the amount of water is proportional to the area watered, and watering is therefore uniform.

Professor Anthony Wexler of the Mechanical Engineering Department of UC-Davis calculated where to drill the holes and made a prototype, which produced a beautiful fountain and a more even supply of water. Moreover, if some of the holes were removed, it would water a rectangular lawn.

We offered the idea to the firm that made the biased sprinkler. After keeping the prototype for half a year, it turned it down because "it would compete with the product we have."

As water becomes more expensive our uniform sprinkler may be used to water many a lawn — if lawns survive the current drought. In fact, the State of California has introduced criteria on how uniformly a sprinkler must distribute water. Recently, we mentioned our sprinkler to another manufacturer, who showed great interest.

EXERCISES for CIE C.7

1. Show that the limit (C.7.2) is 4 (a) using only trigonometric identities and (b) using l'Hôpital's rule.

2. An oscillating sprinkler goes back and forth at a fixed angular speed.
(a) Does it water a lawn uniformly? (b) If not, how would you modify it to provide more uniform coverage?

Chapter 6

The Definite Integral

Up to this point we have been concerned with the derivative, which provides local information, such as the slope at a particular point on a curve or the velocity at a particular time. Now we introduce the second major concept of calculus, the definite integral. The definite integral provides global information, such as the area under a curve.

Section 6.1 motivates the definite integral through three of its applications. Section 6.2 defines the definite integral and Section 6.3 explores the basic properties of definite integrals. Section 6.4 develops the connection between the derivative and the definite integral, which culminates in the Fundamental Theorems of Calculus. The derivative turns out to be essential for evaluating many definite integrals. We conclude, in Section 6.5, with an introduction to ways to estimate definite integrals.

Chapters 2 to 6 form the core of calculus. Later chapters are mostly variations or applications of the key ideas found in these five chapters.

6.1 Three Problems That Are One Problem

We introduce the definite integral with three problems. At first glance they may seem unrelated, but by the end of the section it will be clear that they represent one basic problem in different guises. They lead up to the concept of the definite integral, defined in the next section.

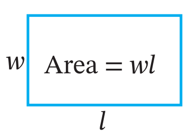


Figure 6.1.1

Estimating an Area

To find the area of a rectangle, multiply its length by its width (see Figure 6.1.1). But how can we find the area of the region in Figure 6.1.1? In this section we will show how to make accurate estimates of that area. The technique we use will lead to the definition of the definite integral of a function.

> **PROBLEM 1** Estimate the area of the region bounded by the curve $y = x^2$, the x-axis, and the vertical line $x = 3$, as shown in Figure 6.1.2(a).

Since we know how to find the area of a rectangle, we use rectangles to approximate the region. Figure 6.1.2(b) shows an approximation by six rectangles whose total area is more than the area under the parabola. Figure 6.1.2(c) shows a similar approximation whose area is less than the area under the parabola.

In each case we break the interval $[0,3]$ into six intervals of width $\frac{1}{2}$. To find the areas of the overestimate and of the underestimate, we find the heights of the rectangles. They are determined by the curve $y = x^2$. Let us examine only the overestimate, leaving the underestimate for the Exercises.

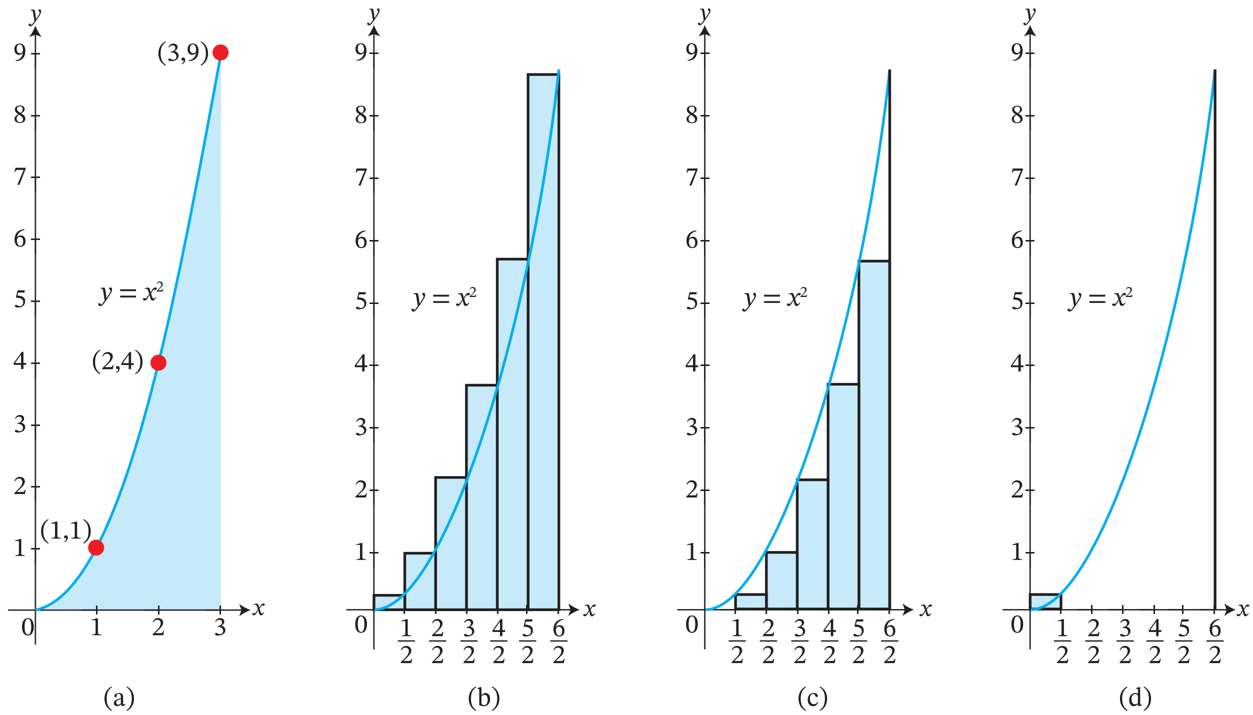


Figure 6.1.2

There are six rectangles in the overestimate shown in Figure 6.1.2(b). The smallest rectangle is shown in Figure 6.1.2(d). Its height is equal to the value of x^2 when $x = 1/2$, or $(1/2)^2$. The area is $(1/2)^2(1/2)$, the product of its height and its width. The areas of the other rectangles can be found similarly. In each case evaluate x^2 at the right end of the rectangle's base to find the height. The total area of the rectangles is

$$\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{2}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{3}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{4}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{5}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{6}{2}\right)^2\left(\frac{1}{2}\right).$$

This equals

$$\frac{1}{8}\left(1^2+2^2+3^2+4^2+5^2+6^2\right)=\frac{91}{8}=11.375.$$

The area under the parabola is therefore less than 11.375.

To get a closer estimate we should use more rectangles. Figure 6.1.3 shows an overestimate in which there are 12 rectangles. Each rectangle has width $3/12 = 1/4$. The total area of the overestimate is

$$\left(\frac{1}{4}\right)^2\left(\frac{1}{4}\right)+\left(\frac{2}{4}\right)^2\left(\frac{1}{4}\right)+\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)+\cdots+\left(\frac{12}{4}\right)^2\left(\frac{1}{4}\right).$$

This equals

$$\frac{1}{4^3}\left(1^2+2^2+3^2+\cdots+12^2\right)=\frac{650}{64}=10.15625.$$

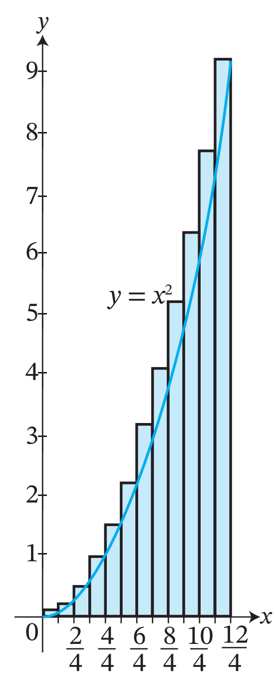


Figure 6.1.3

Now we know the area under the parabola is less than 10.15625.

To get closer estimates, cut the interval $[0,3]$ into more sections, maybe 100 or 10,000 or more, and calculate the total area of the corresponding rectangles. (This is an easy computation on a computer.)

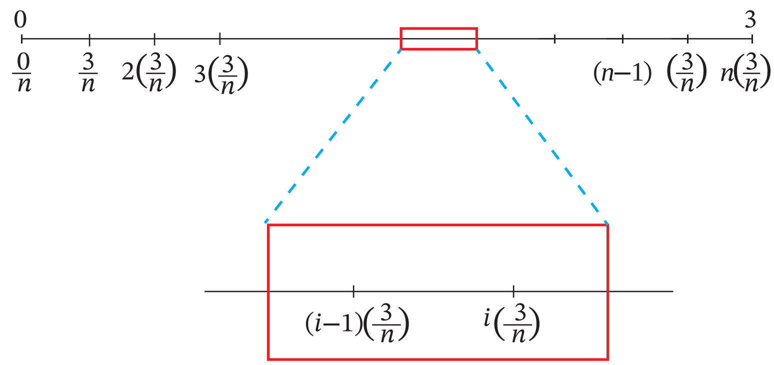


Figure 6.1.4

In general, divide $[0,3]$ into n sections of equal length. The length of each section is then $3/n$. Their endpoints are shown in Figure 6.1.4.

For each integer $i = 1,\ 2,\ \ldots, n$, the i^{th} section from the left has endpoints $(i-1)(3/n)$ and $i(3/n)$, as shown in the inset in Figure 6.1.4.

Because x^2 is increasing for $x > 0$, evaluating x^2 at the right-hand endpoint of each interval produces an overestimate of the area under the curve. Multiply the height by the width of the interval, getting

$$\left(i\left(\frac{3}{n}\right)\right)^2 \frac{3}{n} = 3^3\,\frac{i^2}{n^3}.$$

Sum the overestimates for all n intervals:

$$3^3\,\frac{1^2}{n^3} + 3^3\,\frac{2^2}{n^3} + 3^3\,\frac{3^2}{n^3} + \cdots + 3^3\,\frac{(n-1)^2}{n^3} + 3^3\,\frac{n^2}{n^3}$$

which simplifies to

$$3^3\left(\frac{1^2+2^2+3^2+\cdots+(n-1)^2+n^2}{n^3}\right). \tag{6.1.1}$$

In summation notation this equals

$$\frac{3^3}{n^3}\sum_{i=1}^{n} i^2.$$

We have seen that the overestimates become more accurate as the number of intervals increases. We would like to know what happens to the values of the overestimates as n gets larger. Does

$$\lim_{n\to\infty} \frac{1^2+2^2+3^2+\cdots+(n-1)^2+n^2}{n^3} = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} i^2}{n^3} \tag{6.1.2}$$

exist? If it does, call it L. Then the area would be $3^3 L$.

The numerator becomes large, tending to make the fraction large. But the denominator also becomes large, which tends to make the fraction small. Once again we encounter one of the limit battles that occur in the foundation of calculus.

To estimate L, use, say, $n = 6$. Then we have

$$\frac{1}{6^3}\left(1^2+2^2+3^2+4^2+5^2+6^2\right) = \frac{91}{216} \approx 0.42130.$$

Try a larger value of n to get a closer estimate of L.

If we knew L we would know the area under the parabola and above the interval $[0,3]$, for the area is $3^3 L$. Since we do not know L, we don't know the area. We will find L indirectly in this section. You may want to compute the quotient in (6.1.2) for some n and guess what L is. With $n = 12$, the estimate is $650/12^3 = 650/1728 \approx 0.37616$.

Historical Note: Archimedes and $\sum_{i=1}^{n} i^2$

Archimedes, some 2200 years ago, found a short formula for $\sum_{i=1}^{n} i^2$ that enabled him to evaluate the limit.

REFERENCE: S. Stein, *Archimedes: What did he do besides cry Eureka?*, Mathematical Association of America, 1999.

Estimating a Distance Traveled

The units are consistent: $\frac{\text{mi}}{\text{hr}} \times \text{hr} = \text{mi}$.

If you drive at a constant speed of v miles per hour for a period of t hours, you travel vt miles:

$$\text{Distance} = \text{Speed} \times \text{Time} = vt \text{ miles}.$$

How would you compute the total distance traveled if your speed was not constant? (Imagine that the odometer, which records distance traveled, was broken. However, the speedometer is still working fine, so you know the speed at any instant.) The next problem illustrates how you could make accurate estimates of the total distance traveled.

PROBLEM 2 A snail is crawling and knows that she is traveling at the rate of t^2 feet per minute at time t minutes. For instance, after half a minute, she is slowly moving at the rate of $(1/2)^2$ feet per minute. After three minutes she is moving along at 3^2 feet per minute. Estimate how far she has traveled during the three minutes.

The speed during the three-minute trip increases from 0 to 9 feet per minute. During shorter time intervals, such a wide fluctuation does not occur. As in Problem 1, cut the three minutes of the trip into six equal intervals, each 1/2 minute long, and use them to estimate the total distance covered. Represent time by a line segment cut into six parts of equal length, as in Figure 6.1.5.

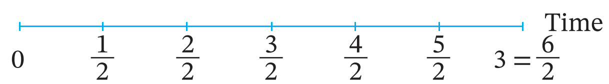


Figure 6.1.5

Speed increases as t increases.

Consider the distance the snail travels during one of the six half-minute intervals, say during the interval $[3/2, 4/2]$. At the beginning of the interval her speed was $(3/2)^2$ feet per minute and at the end she was going $(4/2)^2$ feet per minute. The highest speed during this half-minute was $(4/2)^2$ feet per minute. Therefore, she traveled at most $(4/2)^2(1/2)$ feet during the time interval $[3/2, 4/2]$. Similar reasoning applies to the other half-minute periods. Adding the upper estimates for the distance traveled, we find that the total distance traveled is less than

$$\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{2}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{3}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{4}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{5}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{6}{2}\right)^2\left(\frac{1}{2}\right)=\frac{91}{8}.$$

If we divide the time interval into n equal sections of duration $\frac{3}{n}$, the right endpoint of the i^{th} interval is $i\left(\frac{3}{n}\right)$. At that time the speed is $(3i/n)^2$ feet per minute. So the distance covered during the i^{th} interval of time is less than

$$\underbrace{\left(\frac{3i}{n}\right)^2}_{\text{max speed}}\underbrace{\frac{3}{n}}_{\text{time}}=\frac{3^3 i^2}{n^3}.$$

The total overestimate is then

$$3^3\frac{1^2}{n^3}+3^3\frac{2^2}{n^3}+3^3\frac{3^2}{n^3}+\cdots+3^3\frac{(n-1)^2}{n^3}+3^3\frac{n^2}{n^3}$$

or

$$3^3\left(\frac{1^2+2^2+3^2+\cdots+(n-1)^2+n^2}{n^3}\right). \tag{6.1.3}$$

Observation 6.1.1: *Problems 1 and 2 are the Same Problem*

The final mathematical expressions in the area problem (6.1.1) and the distance problem (6.1.3) are the same. Thus, the area and distance have the same upper estimates. Their lower estimates are also the same, as can be checked. In each case the limit of the expression is 3^3L. The two applications are really the same mathematical problem.

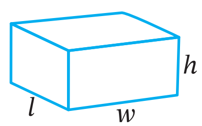


Figure 6.1.6

Estimating a Volume

The volume of a rectangular box is the product of its length, width, and height. See Figure 6.1.6. Finding the volume of a pyramid or ball requires more work. The next example illustrates how we can estimate the volume inside a tent.

PROBLEM 3 Estimate the volume inside a tent with a square floor of side 3 feet, whose vertical pole, 3 feet long, is located above one corner of the floor. The tent is shown in Figure 6.1.7(a).

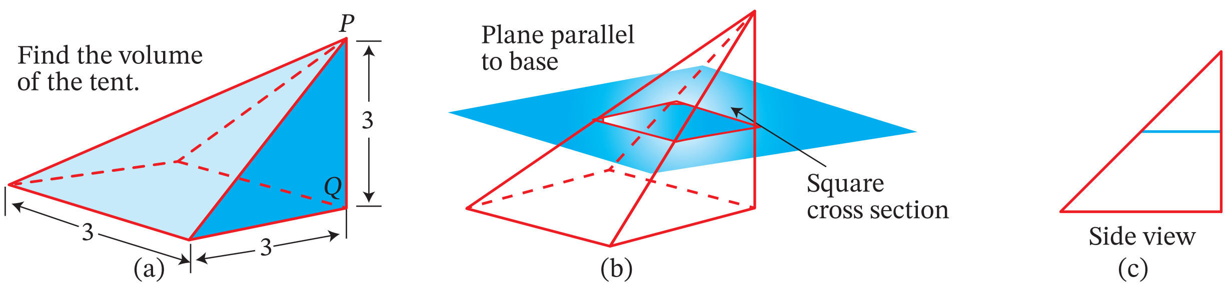


Figure 6.1.7

The cross section of the tent made by a plane parallel to the base is a square, as shown in Figure 6.1.7(b). The width of the square equals its distance from the top of the pole, as shown in Figure 6.1.7(c). Using this, we can approximate the volume inside the tent with rectangular boxes with square cross sections.

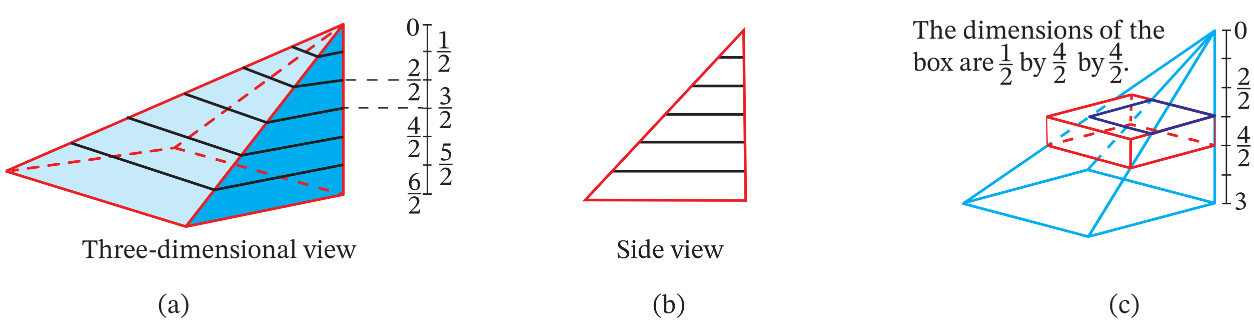


Figure 6.1.8

Cut a vertical line, representing the pole, into six sections of equal length, each 1/2 foot long. Draw the corresponding square cross section of the tent, as in Figure 6.1.8(a).

Use them to form rectangular boxes. The part of the tent corresponding to the interval $[3/2, 4/2]$ on the pole has a square base with sides $4/2$ feet. The box with this square as a base and height $1/2$ foot encloses completely the part of the tent corresponding to $[3/2, 4/2]$. (See Figure 6.1.8(c).) The volume of the box is $(4/2)^2(1/2)$ cubic feet. Figures 6.1.9(a) and (b) show six boxes, whose total volume is greater than the volume of the tent.

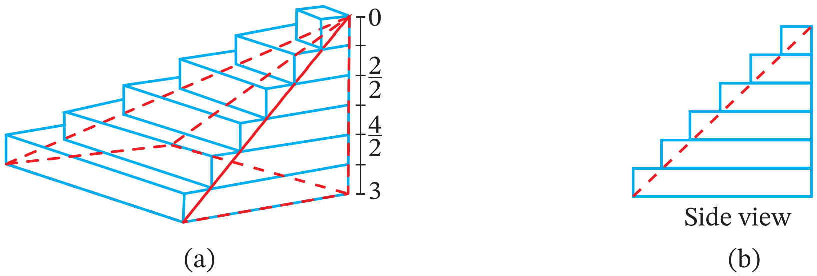


Figure 6.1.9

Since the volume of each box is the area of its base times its height, the total volume of the six boxes is

$$\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{2}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{3}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{4}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{5}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{6}{2}\right)^2\left(\frac{1}{2}\right)=\frac{91}{8}\text{ cubic feet.}$$

This sum, which we have encountered twice before, equals 11.375. It is an overestimate of the volume of the tent. Better estimates can be obtained by cutting the pole into shorter pieces. The arithmetic for the tent volume is the same as for the area and distance problems.

Observation 6.1.2: *And, Problem 3 is Also the Same as Problems 1 and 2*

We now know that the number describing the volume of the tent is the same as the number describing the area under the parabola and also the length of the snail's journey, $3^3 L$. The arithmetic of the estimates is the same in each case.

A Neat Bit of Geometry

If we knew the limit L in (6.1.1), we could then answer all three problems. But we have not found L. Luckily, there is a way to find the volume of the tent without knowing L.

The key is that three identical copies of the tent fill up a cube of side 3 feet. To see why, imagine a flashlight at one corner of the cube, aimed into the cube, as in Figure 6.1.10(a).

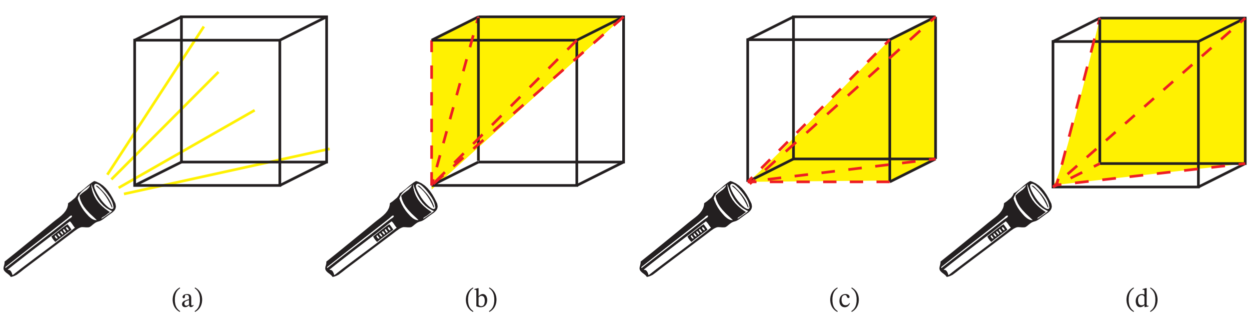


Figure 6.1.10

The flashlight illuminates the three square faces not meeting the corner at the flashlight. The rays from the flashlight to the top, side, and back, as shown in Figure 6.1.10(b), (c), and (d), respectively, fill out a copy of the tent.

This is like finding the area of a right triangle by arranging two copies of it to form a rectangle.

Since three copies of the tent fill a cube of volume $3^3 = 27$ cubic feet, the tent has volume 9 cubic feet. From this, we see that the area under the parabola above $[0,3]$ is 9 and the snail travels 9 feet. The limit L must be $1/3$, since the area under the parabola is both 9 and $3^3 L$. That is,

$$\lim_{n\to\infty} \frac{1^2+2^2+3^2+\cdots+(n-1)^2+n^2}{n^3} = \lim_{n\to\infty} \frac{1}{n^3}\sum_{i=1}^{n} i^2 = \frac{1}{3}.$$

Summary

Using upper estimates, we showed that problems concerning area, distance traveled, and volume were the same problem. We were studying a problem concerning a particular function, x^2, over a particular interval $[0,3]$. We resolved all three problems by cutting a cube into three congruent pieces. This chapter will develop general techniques that will make such a special device unnecessary.

EXERCISES for Section 6.1

Exercises 1 to 21 concern estimates of areas under curves.

1. In Problem 1 we broke the interval $[0,3]$ into six sections. Instead, break $[0,3]$ into four sections of equal lengths and estimate the area under $y = x^2$ and above $[0,3]$ as follows.

(a) Draw the four rectangles whose total area is larger than the area under the curve. The value of x^2 at the right-hand endpoint of each section determines the height of each rectangle. (Use graph paper. If you do not have some in you , look online.)

(b) On the diagram in (a), show the height and width of each rectangle.

(c) Find the total area of the four rectangles.

2. Like Exercise 1, but this time obtain an underestimate of the area by using the value of x^2 at the left-hand endpoint of each section to determine the height of the rectangle.

3. Estimate the area under $y = x^2$ and above the interval $[1,2]$ using the five rectangles with equal widths shown in Figure 6.1.11(a).

4. Repeat Exercise 3 with the five rectangles in Figure 6.1.11(b).

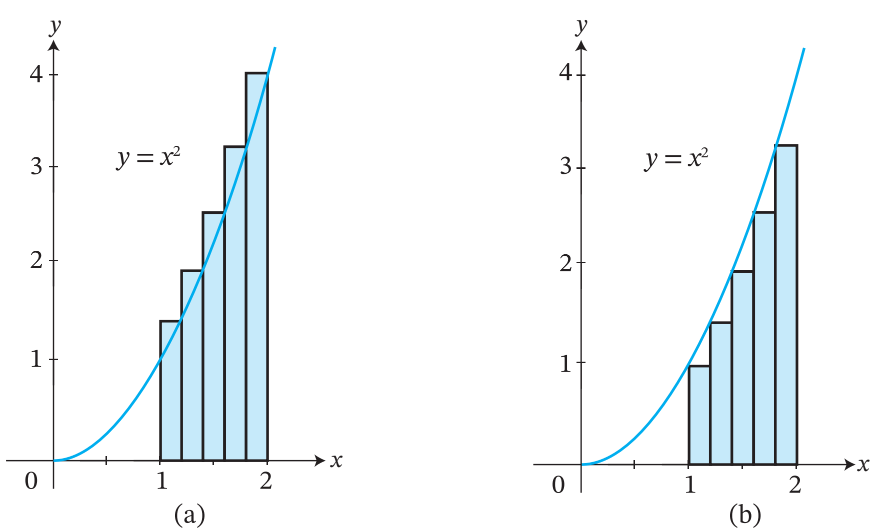


Figure 6.1.11

5. Evaluate (a) $\sum_{i=1}^{4} i^2$, (b) $\sum_{i=1}^{4} 2^i$, and (c) $\sum_{n=3}^{4} (n-3)$.

6. Evaluate (a) $\sum_{i=1}^{4} i^3$, (b) $\sum_{i=2}^{5} 2^i$, and (c) $\sum_{k=1}^{4} (k^3-k^2)$.

7. Figure 6.1.12(a) shows the curve $y=\frac{1}{x}$ above the interval $[1,2]$ and an approximation to the area under the curve by five rectangles of equal width.

(a) Make a large copy of Figure 6.1.12(a).
(b) On your diagram show the height and width of each rectangle.
(c) Find the total area of the five rectangles.
(d) Find the total area of the five rectangles in Figure 6.1.12(b).
(e) On the basis of (c) and (d), what can you say about the area under the curve $y=1/x$ and above $[1,2]$?

Exercise 8 and 9 develop underestimates for each of the problems considered in this section.

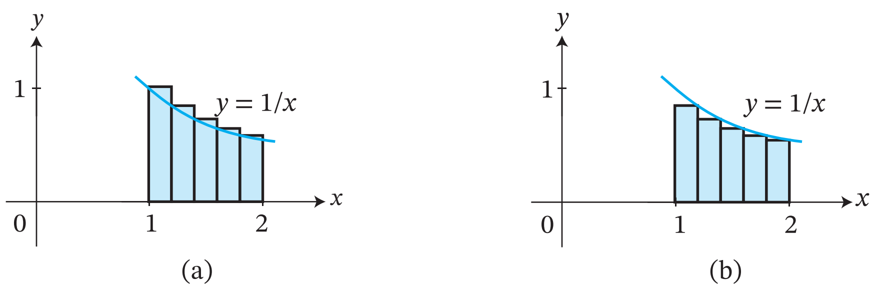


Figure 6.1.12

8. In Problem 1 we found overestimates for the area under the parabola x^2 over the interval $[0,3]$. Here we obtain underestimates for this area as follows.

(a) Break $[0,3]$ into six sections of equal lengths and draw the six rectangles whose total area is smaller than the area under the curve.
(b) Because x^2 is increasing on $[0,3]$, the left endpoint of each section determines the height of each rectangle. Show the height and width of each rectangle you drew in (a).
(c) Find the total area of the six rectangles.

9. Repeat Exercise 8 with twelve sections of equal lengths.

---

10. Use the information found in Exercises 3 and 4 to fill in the blanks:

The area in Problem 1 is certainly less than ________ but larger than ________.

11. Consider the area under $y=2^x$ and above $[-1,1]$.

(a) Graph the curve and estimate the area by eye.
(b) Make an overestimate of the area, using four sections of equal width.
(c) Make an underestimate of the area, using four sections of equal width.

12. Estimate the area in Problem 1, using the division of $[0,3]$ into four sections with endpoints 0, 1, $\frac{5}{3}$, $\frac{11}{4}$, and 3 (see Figure 6.1.13).

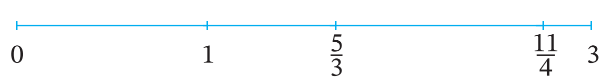


Figure 6.1.13

(a) Estimate the area when the right-hand endpoints of each section are used to find the heights of the rectangles.

(b) Repeat (a), using the left-hand endpoints of each section to find the heights of the rectangles.

(c) Repeat (a) computing the heights of the rectangles at the points $\frac{1}{2}$, $\frac{3}{2}$,

---

In each of Exercises 13 to 18 complete the following four steps.

(a) Draw the region.

(b) Draw six rectangles of equal widths whose total area overestimates the area of the region.

(c) On your diagram indicate the height and width of each rectangle.

(d) Find the total area of the six rectangles accurate to two decimal places.

13. Under $y = x^2$, above $[2,3]$.

14. Under $y = \frac{1}{x}$, above $[2,3]$.

15. Under $y = x^3$, above $[0,1]$.

16. Under $y = \sqrt{x}$, above $[1,4]$.

17. Under $y = \sin(x)$, above $\left[[0, \frac{\pi}{2}\right]$.

18. Under $y = \ln(x)$, above $[1,e]$.

---

19. Estimate the area under $y = x^2$ and above $[-1,2]$ by dividing the interval into six sections of equal lengths.

(a) Draw the six rectangles that form an overestimate for the area under the curve. You cannot do this using only left endpoints or only right endpoints.

(b) Find the total area of all six rectangles.

(c) Repeat (a) and (b) to find an underestimate for this area.

20. Estimate the area between the curve $y = x^3$, the x-axis, and the vertical line $x = 6$ using a division into

(a) six sections of equal lengths using heights at the left endpoints

(b) six sections of equal lengths using heights at the right endpoints

(c) three sections of equal lengths using heights at the midpoints

(d) six sections of equal lengths using heights at the midpoints

21. Estimate the area below the curve $y = \frac{1}{x^2}$ and above $[1,7]$ following the directions in Exercise 20.

22. To estimate the area in Problem 1 the interval $[0,3]$ is divided into n sections of equal lengths. Using the right-hand endpoint of each of the n sections provides an overestimate. Using the left-hand endpoint provides an underestimate.

(a) Show that the estimates differ by $\frac{27}{n}$.

(b) How large should n be chosen in order to be sure the difference between the upper estimate and the area under the parabola is less than 0.01?

23. An electron is being accelerated in such a way that its velocity is t^3 kilometers per second after t seconds. Estimate how far it travels in the first 4 seconds, as follows:

(a) Draw the interval $[0,4]$ as the time axis and cut it into eight sections of equal length.

(b) Using the sections in (a), make an estimate that is too large.

(c) Using the sections in (a), make an estimate that is too small.

24. A business that now shows no profit seeks to increase its profit flow gradually in the next 3 years until it reaches a rate of 9 million dollars per year. At the end of the first half year the rate is to be $1/4$ million dollars per year and at the end of 2 years, 4 million dollars per year. In general, at the end of t years, where t is between 0 and 3, the rate of profit is to be t^2 million dollars per year. Estimate the total profit during its first 3 years if the plan is successful using (a) six intervals and left endpoints, (b) six intervals and right endpoints, and (c) six intervals and midpoints.

25. Oil is leaking out of a tank at the rate of 2^{-t} gallons per minute after t minutes. Describe how you would estimate how much oil leaks out during the first 10 minutes. Illustrate your procedure by computing one estimate.

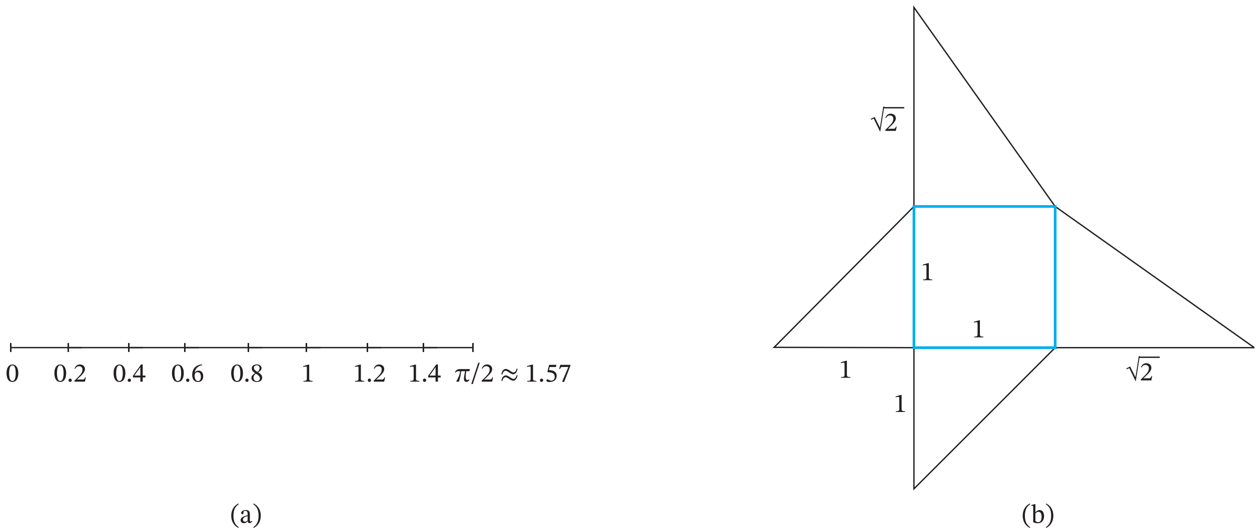


Figure 6.1.14

26. Estimate the area of the region under the curve $y = \sin(x)$ and above the interval $\left[0, \frac{\pi}{2}\right]$, cutting the interval as shown in Figure 6.1.14(a) and using (a) left endpoints, (b) right endpoints, and (c) midpoints.

NOTE: All but the last section are of the same length.

27. Make three copies of the tent in Problem 3 (on page 295) by folding a pattern as shown in Figure 6.1.14(b). Check that they fill up a cube.

28. A right circular cone has a height of 3 feet and a radius of 3 feet. (See Figure 6.1.15(a).) Estimate its volume by the sum of the volumes of six cylindrical slabs, as we estimated the volume of the tent using six rectangular slabs.

(a) Make a large and neat diagram that shows the six cylinders used in making an overestimate.
(b) Compute the total volume of the six cylinders in (a).
(c) Make a separate diagram showing a corresponding underestimate.
(d) Compute the total volume of the six cylinders in (c). (One of the cylinders has radius 0.)

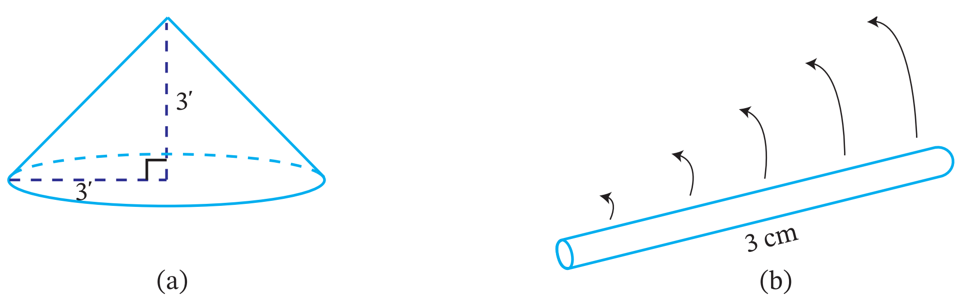


Figure 6.1.15

29. The kinetic energy of an object of mass m grams and speed v centimeters per second is defined as $\frac{1}{2}mv^2$ ergs. In a machine a uniform rod 3 centimeters long and weighing 32 grams rotates once per second around one of its ends as shown in Figure 6.1.15(b). Estimate the kinetic energy of this rod by cutting it into six sections, each 1/2 centimeter long, and taking as the speed of a section the speed of its midpoint.

Exercises 30 and 31 are related.

30. Archimedes showed that $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$.

(a) Check that the formula is correct for $n = 1$.
(b) Show that if the formula is correct for the integer n, it is also correct for the next integer, $n+1$.
(c) Why do (a) and (b) together show that Archimedes' formula holds for all positive integers n?

NOTE: This type of proof is known as **mathematical induction**.

31. (a) Explain why the area of the region under the curve $y = x^2$ and above the interval $[0, b]$ is $\lim_{n\to\infty} \sum_{i=1}^{n} \left(\frac{bi}{n}\right)^2 \frac{b}{n}$.

(b) Use Exercise 30 to evaluate the limit.
(c) Give an explicit formula for the area of the region under $y = x^2$ and above $[0, b]$.
(d) For $0 < a < b$, what is the area under the curve $y = x^2$ and above $[a, b]$?

---

32. The function $f(x)$ is increasing for x in the interval $[a, b]$ and is positive. To estimate the area under the graph of $y = f(x)$ and above $[a, b]$ divide the interval $[a, b]$ into n sections of equal lengths. Then form an overestimate B (for "big") using right-hand endpoints of the sections and an underestimate S (for "small") using left-hand endpoints. Express the difference between the two estimates, $B - S$, as simply as possible.

33. Express the sum $\sum_{i=1}^{n} \ln\left(\frac{i+1}{i}\right)$ as simply as possible. (So that you could compute the sum in the fewest steps.)

In Exercises 34 to 39 differentiate the expression (with respect to x).

34. $(1+x^2)^{4/3}$

35. $\frac{1}{6}\cos^3(2x) - \frac{1}{2}\cos(2x)$

36. $x^3\sqrt{x^2-1}\tan(5x)$

37. $\frac{(1+x^3)\sin(3x)}{\sqrt[3]{5x}}$

38. $\frac{3}{8(2x+3)^2} - \frac{1}{4(2x+3)}$

39. $\frac{3x}{8} + \frac{3}{32}\sin(4x) + \frac{1}{8}\cos^3(2x)\sin(2x)$

In Exercises 40 to 50 give an antiderivative of the expression.

40. $(x+2)^3$

41. $(x^2+1)^2$

42. $x\sin\left(x^2\right)$

43. e^{3x}

44. 2^x

45. $\frac{3}{x}$

46. $x^3 + \frac{1}{x^3}$

47. $\frac{1}{1+x^2}$

48. $\frac{1}{x^2}$

49. $\frac{1}{\sqrt{x}}$

50. $\frac{4}{\sqrt{1-x^2}}$

6.2 The Definite Integral

We now introduce the other main concept in calculus, the definite integral of a function over an interval.

The preceding section was not really about area under a parabola, distance a snail traveled, or volume of a tent. Common to all three was the procedure we carried out with the function x^2 and the interval $[0,3]$: Cut the interval into small pieces, evaluate the function somewhere in each section, form sums, and see how those sums behave as we choose the sections smaller and smaller.

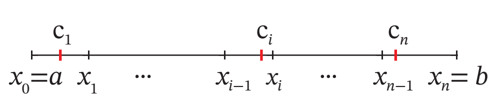


Figure 6.2.1

Here is the general procedure. We have a function f defined on an interval $[a, b]$. We cut, or partition, the interval into n sections by the numbers $x_0 = a$, x_1, x_2, ..., x_{n-1}, $x_n = b$, as in Figure 6.2.1. The sections $[a, x_1]$, $[x_1, x_2]$, ..., $[x_{n-1}, b]$ form a **partition** of $[a, b]$.

The sections need not all be of the same length, though, for convenience, they usually will be.

We pick a **sampling number** in each interval, c_1 in $[x_0, x_1]$, c_2 in $[x_1, x_2]$, ..., c_i in $[x_{i-1}, x_i]$, ..., c_n in $[x_{n-1}, x_n]$ as in Figure 6.2.1. In Section 6.1, the c_i's were mostly either right-hand or left-hand endpoints or midpoints. However, they can be anywhere in a section.

Next we bring in the function f. (In Section 6.1 the function was x^2.) We evaluate that function at each c_i and form the sum

$$f(c_1)(x_1 - x_0) + f(c_2)(x_2 - x_1) + \cdots + f(c_i)(x_i - x_{i-1}) + \cdots + f(c_{n-1})(x_{n-1} - x_{n-2}) + f(c_n)(x_n - x_{n-1}). \tag{6.2.1}$$

Rather than continue to write out such a long expression, we choose to take advantage of the fact that each term in (6.2.1) is the function value at the sampling number multiplied by the length of the section. Expressed in Σ-notation the sum in (6.2.1) is:

$$\sum_{i=1}^{n} f(c_i)(x_i - x_{i-1}). \tag{6.2.2}$$

If the length of section i is written as $\Delta x_i = x_i - x_{i-1}$, the expression for the sum becomes even more compact:

$$\sum_{i=1}^{n} f(c_i)\Delta x_i. \tag{6.2.3}$$

If all the sections have the same length, each Δx_i equals $(b-a)/n$ since the length of $[a, b]$ is $b-a$. Let Δx denote $(b-a)/n$. We can write (6.2.2) and (6.2.3) as

$$\sum_{i=1}^{n} f(c_i)\left(\frac{b-a}{n}\right) \qquad \text{or as} \qquad \Delta x \sum_{i=1}^{n} f(c_i). \tag{6.2.4}$$

The final step is to investigate what happens to the sums of the form (6.2.3) (or (6.2.4)) as the lengths of the sections approach 0. That is, we try to find

$$\lim_{\text{all } \Delta x_i \text{ approach } 0} \sum_{i=1}^{n} f(c_i)\Delta x_i. \tag{6.2.5}$$

The sums in (6.2.1)–(6.2.4) are called **Riemann sums** in honor of the nineteenth century German mathematician, Bernhard Riemann (1826–1866).

In advanced mathematics it is proved that if f is continuous on $[a, b]$ then the limit in (6.2.5) exists. This brings us to the definition of the definite integral.

The Definite Integral

Definition: *Definite Integral of a function f over an interval* $[a, b]$.

Assume that f is a continuous function defined at least on the interval $[a, b]$. The limit of sums of the form $\sum_{i=1}^{n} f(c_i)\Delta x_i$, for partitions of $[a, b]$ where every Δx_i approaches 0, exists (no matter how the sampling numbers c_i are chosen). The limiting value is called the **definite integral of f over the interval** $[a, b]$ and is denoted

$$\int_a^b f(x)\, dx.$$

The symbol $\int$ comes from "S," for "sum." The "dx" reminds us that we are working with very short sections. Both symbols were introduced by the seventeenth century German philosopher, mathematician, scientist, and diplomat, Gottfried Leibniz (1646–1716).

The limit in the definition is a little unusual. It requires the length of every segment within the partition to approach 0. It is not sufficient to consider partitions of $[a, b]$ with more and more segments as there could be segments with lengths that do not approach 0. Another way of stating the requirement is that the length of the largest segment in the partition must approach zero.

EXAMPLE 1. Express the area under $y = x^2$ and above $[0,3]$ as a definite integral.

SOLUTION Here the function is $f(x) = x^2$ and the interval is $[0,3]$. As we saw in the previous section, the area equals the limit of Riemann sums

$$\lim_{\text{largest } \Delta x_i \to 0} \sum_{i=1}^{n} c_i^2 \Delta x_i = \int_0^3 x^2 \, dx. \tag{6.2.6}$$

◆

The dx suggests the length of a small section of the x-axis and denotes the **variable of integration** (usually x, as in this case). The function $f(x)$ is called the **integrand**, while the numbers a and b are called the **limits of integration**; a is the **lower limit of integration** and b is the **upper limit of integration**.

The symbol $\int_a^b x^2 \, dx$ is read as "the integral from a to b of x^2." Not referring to x, we could say, "the integral from a to b of the squaring function". There is nothing special about x. We could just as well have used the letter t, or any other letter. (We typically pick a letter near the end of the alphabet, since letters near the beginning customarily denote constants.) The notations

$$\int_a^b x^2 \, dx, \qquad \int_a^b t^2 \, dt, \qquad \int_a^b z^2 \, dz, \qquad \int_a^b u^2 \, du, \qquad \int_a^b \theta^2 \, d\theta$$

all denote the same number, the definite integral of the squaring function from a to b. We could express (6.2.6) as

$$\int_a^b ()^2 \, d(),$$

but we find it more convenient to use some letter to name the independent variable. Since the letter chosen to represent the variable has no significance of its own, it is called a **dummy variable**. Later the interval of integration will be $[a, x]$, where x is a variable, instead of $[a, b]$, where both a and b are constants. Were we to write $\int_a^x x^2 \, dx$, it would be easy to think there is some relation between the x in x^2 and the x in the upper limit of integration. To avoid confusion, we use a different dummy variable and write, for example, $\int_a^x t^2 \, dt$ in such cases.

It is important to realize that area, distance traveled, and volume are applications of the definite integral. It is a mistake to link the definite integral too closely with one of its applications, such as the area under a curve. (If we think of it as area, then we may be surprised by its appearance in the study of distance, volume, work, etc.) The definite integral $\int_a^b f(x) \, dx$ is also called the **Riemann integral**.

Observation 6.2.1: *Derivatives and Definite Integrals are Both Limits*

Slope and velocity are interpretations or applications of the derivative, which is a purely mathematical concept defined as a limit:

$$\text{Derivative of } f \text{ at } x = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

Similarly, area, total distance, and volume are interpretations of the definite integral, which is also defined as a limit:

$$\text{Definite Integral of } f \text{ over } [a, b] = \lim_{\text{largest } \Delta x_i \to 0} \sum_{i=1}^{n} f(c_i) \Delta x_i.$$

The Definite Integral of a Constant Function

Let us use the definition to evaluate the definite integral of a constant function.

EXAMPLE 2. Assume that f is the function whose value at any number x is 4; that is, f is the constant function given by $f(x) = 4$. Use only the definition of the definite integral to compute $\int_1^3 f(x)\,dx$.

SOLUTION Every partition of the interval $[1,3]$ has $x_0 = 1$ and $x_n = 3$. See Figure 6.2.2.

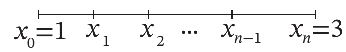


Figure 6.2.2

No matter how the sampling numbers c_i are chosen, $f(c_i) = 4$, and so the approximating sum equals

$$\sum_{i=1}^{n} f(c_i)\Delta x_i = \sum_{i=1}^{n} f(c_i)(x_i - x_{i-1}) = \sum_{i=1}^{n} 4(x_i - x_{i-1})$$

Now

$$\sum_{i=1}^{n} 4(x_i - x_{i-1}) = 4\sum_{i=1}^{n}(x_i - x_{i-1}) = 4\cdot(x_n - x_0) = 4\cdot 2 = 8$$

because the sum of the widths of the sections is the width of the interval $[1,3]$, namely 2. All approximating sums have the same value, namely 8. For every partition,

$$\sum_{i=1}^{n} f(c_i)\Delta x_i = \sum_{i=1}^{n} f(c_i)(x_i - x_{i-1}) = 8.$$

Thus, as sections are chosen smaller and smaller, the values of the Riemann sums are always 8. So the limit is 8 and: $\int_1^3 4\,dx = 8$. ◆

We could have guessed the value of $\int_1^3 4\,dx$ by interpreting the definite integral as area. To do so, draw a rectangle of height 4 and base $[1,3]$. (See Figure 6.2.3.) Since the area of a rectangle is its base times its height, it follows again that $\int_1^3 4\,dx = 8$.

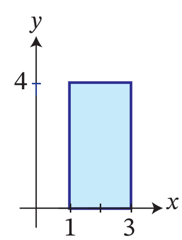


Figure 6.2.3

Similar reasoning shows that for a constant function that has the value c,

$$\int_a^b c\,dx = c(b-a) \qquad (c \text{ is a constant function}).$$

The Definite Integral of x

EXAMPLE 3. Use the area interpretation of the definite integral to predict the value of $\int_a^b x\,dx$.

SOLUTION When $0 < a < b$ the integrand is positive and the area in question then lies above the x-axis, as shown in Figure 6.2.4(a). Two copies of this region form a rectangle of width $b-a$ and height $a+b$, as shown in Figure 6.2.4(b).

Thus, the area shown in Figure 6.2.4(a) is half of $(b-a)(b+a) = b^2 - a^2$: $\int_a^b x\,dx = (b^2 - a^2)/2$. ◆

In Exercise 34 at the end of this section, we find $\int_a^b x\,dx$ directly from the definition in terms of Riemann sums. Then, in Exercise 44 in Section 6.4, we show that this result holds for any nonzero values of a and b with $b > a$. It is a trivial matter to extend these results to the cases where $a = b$ and where one or both of a and b is zero. In the end,

$$\int_a^b x\,dx = \frac{1}{2}(b^2 - a^2) \qquad \text{for all values of } a \text{ and } b.$$

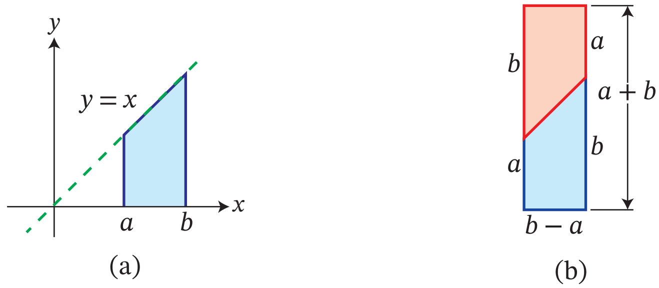


Figure 6.2.4

The Definite Integral of x^2

EXAMPLE 4. Find $\int_0^b x^2\,dx$ by examining the approximating sums when all the sections have the same length, as they did in Section 6.1.

SOLUTION Pick a positive integer n and cut the interval $[0,b]$ into n sections of length $\Delta x = b/n$ as in Figure 6.2.5. Then the points of subdivision are $0, \Delta x, 2\Delta x, \dots, (n-1)\Delta x$, and $n\Delta x = b$.

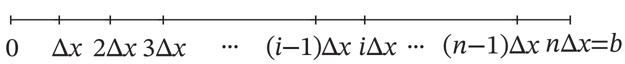


Figure 6.2.5

In the section $[(i-1)\Delta x, i\Delta x]$ we pick the right-hand endpoint as the sampling number. The approximating sum is

$$\sum_{i=1}^{n} (i\Delta x)^2\,(\Delta x) = (\Delta x)^3 \sum_{i=1}^{n} i^2.$$

Since $\Delta x = b/n$, these overestimates can be written as

$$\frac{b^3}{n^3}\sum_{i=1}^{n} i^2. \tag{6.2.7}$$

In Section 6.1 (see Exercise 31) we used geometry to find that

$$\lim_{n\to\infty} \frac{1}{n^3}\sum_{i=1}^{n} i^2 = \frac{1}{3}.$$

Thus, (6.2.7) approaches $b^3/3$ as n increases, and we conclude that

$$\int_0^b x^2\,dx = \frac{b^3}{3}.$$

When $b=3$, we have $b^3/3 = 9$, agreeing with the three problems in Section 6.1. ♦

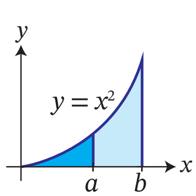


Figure 6.2.6

Geometry also suggests the value of $\int_a^b x^2\,dx$, for $0 \le a < b$. Interpret $\int_a^b x^2\,dx$ as the area under $y = x^2$ and above $[a,b]$. It is equal to the area under $y=x^2$ and above $[0,b]$ minus the area under $y=x^2$ and above $[0,a]$, as shown in Figure 6.2.6. Thus

$$\int_a^b x^2\,dx = \frac{b^3}{3} - \frac{a^3}{3}.$$

The Definite Integral of 2^x

EXAMPLE 5. Use the definition of the definite integral to evaluate $\int_0^b 2^x \, dx$. (Assume $b > 0$.)

SOLUTION Divide the interval $[0, b]$ into n sections of equal length, $d = b/n$. We will evaluate the integrand at the left-hand endpoint of each section. Call it c_i, $c_i = (i-1)d$. The approximating sum has one term for each section. The contribution from the i^{th} section is

$$2^{c_i} d = 2^{(i-1)d} d.$$

The total estimate is the sum

$$2^0 d + 2^d d + 2^{2d} d + \cdots + 2^{(i-1)d} d + \cdots + 2^{(n-1)d} d$$

or

$$d\left(1 + 2^d + (2^d)^2 + \cdots + (2^d)^i + \cdots + (2^d)^{n-1}\right). \tag{6.2.8}$$

The terms inside the large parentheses in (6.2.8) form a geometric progression with n terms, whose first term is 1 and whose ratio is 2^d. RECALL: The sum of a geometric progression: $a + ar + ar^2 + \cdots + ar^{n-1} = a(1-r^n)/(1-r)$. Thus the sum in (6.2.8) simplifies to

$$\frac{1-(2^d)^n}{1-2^d}.$$

Therefore this typical underestimate is

$$\frac{d(1-(2^d)^n)}{1-2^d} = \frac{d(1-2^{dn})}{1-2^d} = \frac{d(1-2^b)}{1-2^d}. \tag{6.2.9}$$

In the last step we used the fact that $dn = b$ to rewrite (6.2.9) as

$$\frac{d}{2^d - 1}\left(2^b - 1\right). \tag{6.2.10}$$

It remains to take the limit as n increases without bound. To find what happens to (6.2.10) as $n \to \infty$, we must investigate how $d/(2^d - 1)$ behaves as d approaches 0 from the right. Though we have not met this quotient before, we have seen its reciprocal, $(2^d - 1)/d$. It occurs in the definition of the derivative of 2^x at $x = 0$:

$$\lim_{x\to 0} \frac{2^x - 2^0}{x} = \lim_{x\to 0} \frac{2^x - 1}{x}.$$

As we saw in Example 4 in Section 3.4, the derivative of 2^x is $2^x \ln(2)$. Thus $D(2^x)$ at $x = 0$ is $\ln(2)$. Hence

$$\lim_{d\to 0} \frac{d}{2^d - 1}(2^b - 1) = \left(2^b - 1\right) \lim_{d\to 0} \frac{1}{\left(\frac{2^d - 1}{d}\right)} = \frac{2^b - 1}{\ln(2)}.$$

We conclude that

$$\int_0^b 2^x \, dx = \frac{1}{\ln(2)}\left(2^b - 1\right).$$

◆

To evaluate $\int_a^b 2^x \, dx$ with $b > a \geq 0$, we reason as we did when we generalized $\int_0^b x^2 \, dx$ to $\int_a^b x^2 \, dx$:

$$\int_a^b 2^x \, dx = \int_0^b 2^x \, dx - \int_0^a 2^x \, dx = \frac{2^b - 1}{\ln(2)} - \frac{2^a - 1}{\ln(2)} = \frac{2^b}{\ln(2)} - \frac{2^a}{\ln(2)}.$$

Summary

We defined the definite integral of a function $f(x)$ over an interval $[a,b]$ to be the limit of sums of the form $\sum_{i=1}^{n} f(c_i)\Delta x_i$ created from partitions of $[a,b]$. It is a purely mathematical idea. We could estimate $\int_a^b f(x)\,dx$ with a calculator without any application in mind. However, the definite integral has many applications. Three of them are area under a curve, distance traveled, and volume.

Table 6.2.1 contains a lot of information. Compare the first three lines with the fourth, which describes the fundamental definition of integral calculus. In the table all the functions, whether cross-sectional length, velocity, or cross-sectional area, are denoted by the same symbol $f(x)$. Underlying the applications is one mathematical concept, the definite integral,

$$\int_a^b f(x)\,dx = \lim_{\text{all } \Delta x_i \text{ approach } 0} \sum_{i=1}^{n} f(c_i)\Delta x_i.$$

The concepts summarized inTable 6.2.1 will be used often. It is essential to keep the definition of the number $\int_a^b f(x)\,dx$ clear.

IMPORTANT FACT: Every definite integral is a limit of Riemann sums.

$f(x)$	$\sum_{i=1}^{n} f(c_i)(x_i - x_{i-1})$	$\int_a^b f(x)\,dx$
Length of a set's cross section	Approximate area of the set	The area of the set
Velocity	Approximation to the distance traveled	The distance traveled
Cross sectional area of a solid	Approximate volume of the solid	The volume of a solid
A function value	A Riemann sum	The limit of the sums as the largest Δx_i approaches 0

Table 6.2.1

EXERCISES for Section 6.2

1. When $f(x)$ is decreasing for x in $[a,b]$, is an approximating sum for $\int_a^b f(x)\,dx$ with left-hand endpoints as sampling points too large or too small? Explain in complete sentences.

2.
(a) Define "the definite integral of $f(x)$ from a to b, $\int_a^b f(x)\,dx$."
(b) Define the definite integral, using as few mathematical symbols as you can.
(c) Give three applications of the definite integral.

3. Use the formula for $\int_a^b x^2\,dx$ to find the area under the curve $y = x^2$ and (a) above the interval $[0,5]$, (b) above the interval $[0,4]$, and (c) above the interval $[4,5]$.

In Exercises 4 to 7 evaluate the given summation.

4. (a) $\sum_{i=1}^{3} i$, (b) $\sum_{i=3}^{7}(2i+3)$, and (c) $\sum_{d=1}^{3} d^2$.

5. (a) $\sum_{i=2}^{4} i^2$, (b) $\sum_{j=2}^{4} j^2$, and (c) $\sum_{i=1}^{3}(i^2+i)$.

6. (a) $\sum_{i=1}^{4} 1^i$, (b) $\sum_{k=2}^{6}(-1)^k$, and (c) $\sum_{j=1}^{150} 3$.

7. (a) $\sum_{i=3}^{5} \frac{1}{i}$, (b) $\sum_{i=0}^{4} \cos(2\pi i)$, and (c) $\sum_{i=1}^{3} 2^{-i}$.

In Exercises 8 to 11 write the sums in Σ-notation. Do not evaluate them.

8. (a) $1+2+2^2+2^3+\cdots+2^{100}$, (b) $x^3+x^4+x^5+x^6+x^7$, and (c) $\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\cdots+\frac{1}{102}+\frac{1}{103}$.

9. (a) $\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{100}$, (b) $\frac{1}{4}+\frac{1}{6}+\frac{1}{8}+\frac{1}{10}+\frac{1}{12}+\frac{1}{14}$, and (c) $\frac{1}{1^2}+\frac{1}{3^2}+\frac{1}{5^2}+\cdots+\frac{1}{101^2}$.

10. (a) $x_0^2(x_1-x_0)+x_1^2(x_2-x_1)+x_2^2(x_3-x_2)$ and (b) $x_1^2(x_1-x_0)+x_2^2(x_2-x_1)+x_3^2(x_3-x_2)$.

11. (a) $8t_0^2(t_1-t_0)+8t_1^2(t_2-t_1)+\cdots+8t_{99}^2(t_{100}-t_{99})$ and (b) $8t_1^2(t_1-t_0)+8t_2^2(t_2-t_1)+\cdots+8t_n^2(t_n-t_{n-1})$.

12. Figure 6.2.7 shows the curve $y=x^2$. What is the ratio between the shaded area and the area of rectangle $ABCD$?

13.

(a) Use the definition of definite integral to evaluate $\int_0^b e^x\,dx$. (See Example 5.)

(b) From (a), deduce that for $0 \le a \le b$, $\int_a^b e^x\,dx = e^b - e^a$.

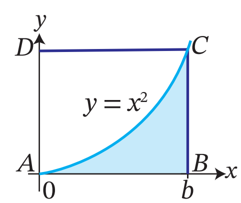


Figure 6.2.7

14.

(a) Use the definition of definite integral to evaluate $\int_0^b 3^x\,dx$.

(b) From (a), deduce that for $0 \le a < b$, $\int_a^b 3^x\,dx = (3^b - 3^a)/\ln(3)$.

15. Knowing that $\int_a^b f(x)\,dx = \lim_{n\to\infty}\sum_{i=1}^{n} f(c_i)\Delta x$ gives us a way to evaluate some limits of sums that would otherwise be difficult to evaluate. Write the following limits as definite integrals. Do not evaluate them.

(a) $\lim_{n\to\infty}\sum_{i=1}^{n} e^{i/n}\frac{1}{n}$ (b) $\lim_{n\to\infty}\sum_{i=1}^{n} \frac{1}{1+\left(1+\frac{2i}{n}\right)^2}\frac{2}{n}$ (c) $\lim_{n\to\infty}\sum_{i=1}^{n} \sin\left(\frac{i\pi}{n}\right)\frac{\pi}{n}$ (d) $\lim_{n\to\infty}\sum_{i=1}^{n} \left(2+\frac{3i}{n}\right)^4\frac{3}{n}$

In Exercises 16 to 18 evaluate $\sum_{i=1}^{n} f(c_i)(x_i - x_{i-1})$ for the given function ($f(x)$), interval ($[a,b]$), partition (x_i), and sampling numbers (c_i).

16. $f(x)=\sqrt{x}$; $[1,5]$; $x_0=1$, $x_1=3$, $x_2=5$; $c_1=1$, $c_2=4$ ($n=2$)

17. $f(x)=\sqrt[3]{x}$; $[0,10]$; $x_0=0$, $x_1=1$, $x_2=4$, $x_3=10$; $c_1=0$, $c_2=1$, $c_3=8$ ($n=3$)

18. $f(x)=1/x$; $[1,2]$; $x_0=1$, $x_1=1.25$, $x_2=1.5$, $x_3=1.75$, $x_4=2$; $c_1=1$, $c_2=1.25$, $c_3=1.6$, $c_4=2$ ($n=4$)

In Exercises 19 to 22 partition the interval into four sections of equal lengths. Estimate the definite integral using sampling numbers chosen to be (a) the left-hand endpoints and (b) the right-hand endpoints.

19. $\int_1^2 \frac{dx}{x^2}$ **20.** $\int_1^5 \ln(x)\, dx$ **21.** $\int_1^5 \frac{2^x}{x}\, dx$ **22.** $\int_0^1 \sqrt{1+x^3}\, dx$

---

23. Write the following expression using summation notation. $c^{n-1} + c^{n-2}d + c^{n-3}d^2 + \cdots + cd^{n-2} + d^{n-1}$.

24. Assume that $f(x) \le -3$ for all x in $[1,5]$. What can be said about the value of $\int_1^5 f(x)\, dx$? Explain, in detail, using the definition of the definite integral.

25. A rocket's varying speed is $f(t)$ miles per second at time t seconds. Let $t_0, \ldots, t_n$ be a partition of $[a,b]$, and let $T_1, \ldots, T_n$ be sampling numbers c_i. What is the physical interpretation of each of the following quantities?

(a) $t_i - t_{i-1}$ (b) $f(T_i)$ (c) $f(T_i)(t_i - t_{i-1})$ (d) $\sum_{i=1}^{n} f(T_i)(t_i - t_{i-1})$ (e) $\int_a^b f(t)\, dt$

Exercises 26 and 27 are related.

26. (a) Sketch $y = \cos(x)$ for x in $\left[0, \frac{\pi}{2}\right]$.

(b) Estimate by eye the area under the curve and above $\left[0, \frac{\pi}{2}\right]$.

(c) Partition $\left[0, \frac{\pi}{2}\right]$ into three equal sections and use them to overestimate the area under the curve.

(d) Use the same partition to provide an underestimate of the area under the curve.

27. Repeat Exercise 26 for the area under the curve $y = e^{-x}$ above $[0,3]$.

---

28. Show that the volume of a right circular cone of radius a and height h is $\frac{\pi a^2 h}{3}$.

29. The speed of an automobile at time t is $s(t)$ feet per second. The graph of s for t in $[0,20]$ is shown in Figure 6.2.8(a). Explain, in complete sentences, why the shaded area under the curve is the distance traveled during the time interval $[10,20]$.

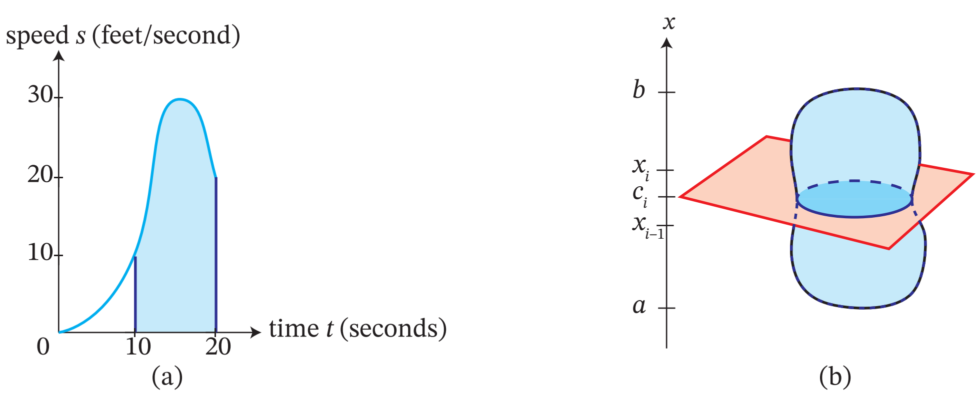


Figure 6.2.8

30. For x in $[a,b]$, define $A(x)$ to be the area of the cross section of a solid perpendicular to the x-axis at x (think of slicing a potato). Denote by $x_0, x_1, \ldots, x_n$ a partition of $[a,b]$ and by $c_1, \ldots, c_n$ the corresponding sampling numbers. What is the geometric interpretation of the following quantities? (See Figure 6.2.8(b).)

(a) $x_i - x_{i-1}$ (b) $A(c_i)$ (c) $A(c_i)(x_i - x_{i-1})$ (d) $\sum_{i=1}^{n} A(c_i)(x_i - x_{i-1})$ (e) $\int_a^b A(x)\,dx$

31.
(a) Set up a definite integral $\int_a^b f(x)\,dx$ that equals the volume of the headlight in Figure 6.2.9(a) whose cross section by a plane perpendicular to the x-axis at x is a disk whose radius is $\sqrt{\frac{x}{\pi}}$.
(b) Evaluate the definite integral found in (a).

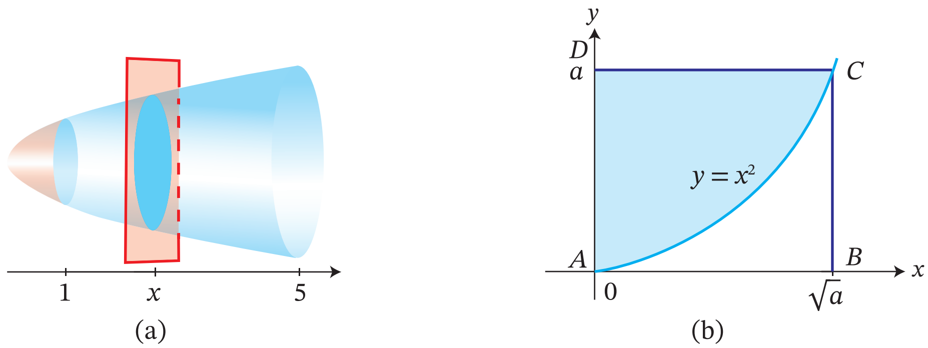


Figure 6.2.9

32.
(a) By considering the area of region ACD in Figure 6.2.9(b), show that $\int_0^a \sqrt{x}\,dx = \frac{2}{3}a^{3/2}$.
(b) Use (a) to evaluate $\int_a^b \sqrt{x}\,dx$ when $0 < a < b$.

Let f be a function defined for the positive integers. A sum of the form $\sum_{i=1}^{n}(f(i+1) - f(i))$ is called a **telescoping sum**. To show why, write it out:

$$(f(2) - f(1)) + (f(3) - f(2)) + (f(4) - f(3)) + \cdots + (f(n) - f(n-1)) + (f(n+1) - f(n)).$$

In general, the first number in each term cancels with the second number in the next term, leaving only two uncanceled numbers: $-f(1)$ and $f(n+1)$. The sum shrinks like a collapsible telescope and has value $f(n+1) - f(1)$. Exercises 33 to 36 involve telescoping sums.

33. (a) Show that $\sum_{i=1}^{n}\left((i+1)^2 - i^2\right) = (n+1)^2 - 1$.
(b) From (a), show that $\sum_{i=1}^{n}(2i+1) = (n+1)^2 - 1$.
(c) From (b), show that $n + 2\sum_{i=1}^{n} i = (n+1)^2 - 1$.
(d) From (c), show that $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

34. Exercise 33 showed that $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$. Use this to find $\int_0^b x\,dx$ directly from the definition of the definite integral, not by interpreting it as an area. No picture is needed.

35. (a) Starting with the telescoping sum $\sum_{i=1}^{n}\left((i+1)^3 - i^3\right)$ show that $n + 3\sum_{i=1}^{n} i^2 + 3\sum_{i=1}^{n} i = (n+1)^3 - 1$.

(b) Use (a) to show that $\sum_{i=1}^{n} i^2 = \frac{1}{6}n(n+1)(2n+1)$.

(c) Use (b) to show that $\int_0^b x^2\,dx = \frac{b^3}{3}$. (See Exercise 34.)

36. (a) Using the techniques of Exercises 33 to 35, find a short formula for the sum $\sum_{i=1}^{n} i^3$.

(b) Use it to show that $\int_0^b x^3\,dx = \frac{b^4}{4}$.

---

37. The function $f(x) = \frac{1}{x}$ has a remarkable property, namely, that for a and b greater than 1, $\int_1^a \frac{1}{x}\,dx = \int_b^{ab} \frac{1}{x}\,dx$. In words, magnifying the interval $[1,a]$ by a positive number b does not change the value of the definite integral. The following steps show why this is so.

(a) Let $x_0 = 1, x_1, x_2, \ldots, x_n = a$ divide the interval $[1,a]$ into n sections. Using left-hand endpoints write an approximating sum for $\int_1^a \frac{1}{x}\,dx$.

(b) Let $bx_0 = b, bx_1, bx_2, \ldots, bx_n = ab$ divide the interval $[b,ab]$ into n sections. Using left-hand endpoints write an approximating sum for $\int_b^{ab} \frac{1}{x}\,dx$.

(c) Explain why $\int_1^a \frac{1}{x}\,dx = \int_b^{ab} \frac{1}{x}\,dx$.

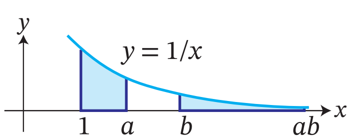


Figure 6.2.10

38. Let $L(t) = \int_1^t \frac{1}{x}\,dx$, $t > 1$. (See Figure 6.2.10.)

(a) Show that $L(a) = L(ab) - L(b)$.

(b) By (a), conclude that $L(ab) = L(a) + L(b)$.

(c) What familiar function has the property listed in (b)?

Historical Note: The Natural Logarithm as an Area

Gregory St. Vincent noticed property (a) in 1647, and his friend A.A. de Sarasa saw that (b) followed. Euler, in the 18th century, recognized that $L(x)$ is the logarithm of x to the base e. Thus the area under the hyperbola $y = 1/x$ and above $[1,a]$, $a > 1$, is $\ln(a)$. It can be shown that for a in $(0,1)$, the negative of the area below the curve and above $[a,1]$ is $\ln(a)$. (See C. H. Edwards Jr., *The Historical Development of the Calculus*, Springer-Verlag, 1994, 154–158.)

39. In Exercise 13 it was shown that for $0 \le a \le b$, $\int\limits_a^b e^x\,dx = e^b - e^a$.

(a) Use this and a diagram to show that $\int\limits_{e^a}^{e^b} \ln(x)\,dx = e^b(b-1) - e^a(a-1)$.

(b) From (a), deduce that for $1 \le c \le d$, $\int\limits_c^d \ln(x)\,dx = (d\ln(d) - d) - (c\ln(c) - c)$.

(c) By differentiating $x\ln(x) - x$, show that it is an antiderivative of $\ln(x)$.

40.

(a) Estimate $\int\limits_1^2 \frac{1}{x}\,dx$ by dividing $[1,2]$ into n equal-length sections and using right-hand endpoints as the sampling points.

(b) Deduce from (a) that $\lim\limits_{n\to\infty} \sum\limits_{i=n+1}^{2n} \frac{1}{i} = \lim\limits_{n\to\infty}\left(\frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n}\right)$ is the area under $y = \frac{1}{x}$ and above $[1,2]$.

(c) Let $g(n) = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n}$. Show that $\frac{1}{2} \le g(n) < 1$ and $g(n+1) > g(n)$.

Exercises 41 to 45 are related. Exercises 41 and 42 describe a method devised by Pierre Fermat (1601–1665) to find the area under $y = x^k$ and above $[1,b]$ ($b > 1$ and k a number) by using approximating sums. **Fermat's method** is based on an unusual partition of $[1,b]$.

41. For $b > 1$ and n a positive integer define $r(n)$ by $r(n) = b^{1/n}$.

(a) For $b = 5$, find $r(n)$ for $n = 1, 2, 3$, and 10.

(b) The calculations in (a) suggest that $\lim\limits_{n\to\infty} r(n) = 1$. Show that this is correct.

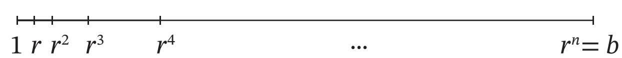


Figure 6.2.11

42. Fermat's method begins by introducing, for a positive number n, the number r such that $r^n = b$. Exercise 41 shows r approaches 1 as n increases. Then he partitioned $[1,b]$ using the numbers r, r^2, r^3, ..., r^{n-1}, as shown in Figure 6.2.11. The n sections are $[1,r]$, $[r,r^2]$, ..., $[r^{n-1},r^n] = [r^{n-1},b]$.

(a) Show that the width of the i^{th} section, $[r^{i-1},r^i]$, is $r^{i-1}(r-1)$.

(b) Using the left-hand endpoint of each section as the sampling number, obtain an underestimate of $\int\limits_1^b x^2\,dx$.

(c) Show that the estimate in (b) is $\frac{b^3-1}{1+r+r^2}$.

(d) Find $\lim\limits_{n\to\infty} \frac{b^3-1}{1+r+r^2}$. Remember that r depends on n.

(e) What does (d) imply about $\int\limits_1^b x^2\,dx$?

43. Use Fermat's method (see Exercise 42) to find $\int_1^b x^3\, dx$.

44. Use Fermat's method (see Exercise 42) to find $\int_1^b x^4\, dx$.

45. Obtain an overestimate for $\int_1^b x^2\, dx$ by repeating Exercise 42(b) using the right-hand endpoint as the sampling number for each section. What is the limit as $n \to \infty$?

46.
(a) Obtain an underestimate and an overestimate of $\int_0^{\pi/2} \cos(x)\, dx$ that differ by at most 0.1.

Remember that angles are measured in radians.

(b) Average the two estimates in (a).

(c) If $\int_0^{\pi/2} \cos(x)\, dx$ is a famous number, what do you think it is?

47. By considering the approximating sums in the definition of a definite integral, show that $\int_3^4 \frac{dx}{(x+5)^3} = \int_2^3 \frac{dx}{(x+6)^3}$.

48. For a continuous function f defined for all x are $\int_a^b f(x+1)\, dx$ and $\int_{a+1}^{b+1} f(x)\, dx$ equal?

49. For continuous functions f and g defined for all x does $\int_a^b \big(f(x)g(x)\big)\, dx = \left(\int_a^b f(x)\, dx\right)\left(\int_a^b g(x)\, dx\right)$?

If your answer is "yes", explain why (an example does not suffice); if your answer is "no", explain why not (a counterexample or two will suffice).

50. If f is an increasing function such that $f(1) = 3$ and $f(6) = 7$, what can be said about $\int_2^4 f(x)\, dx$? Explain.

51. (a) Using formulas already developed, evaluate $G(x) = \int_1^x t^2\, dt.$, (b) Find $G'(x)$., (c) Repeat (a) and (b) for $G(x) = \int_1^x 2^t\, dt.$, and (d) In view of (b) and (c), what might the derivative of $\int_a^x f(t)\, dt$ be for any continuous function?.

In Exercises 52 to 59 give two antiderivatives for the following expressions.

52. x^2 **53.** e^{-4x} **54.** 2^x **55.** $\sin(3x)$

56. $\frac{1}{x^3}$ **57.** $\frac{1}{2x+1}$ **58.** $\frac{3}{1+9x^2}$ **59.** $\frac{4}{\sqrt{1-x^2}}$

6.3 Properties of Antiderivatives and Definite Integrals

In 3.4 we defined an antiderivative of a function $f(x)$ as a function $F(x)$ whose derivative is $f(x)$. For instance, x^3 is an antiderivative of $3x^2$. So is x^3+2024. Keep in mind that an antiderivative is a function.

In this section we discuss properties of antiderivatives and definite integrals. They will be needed in Section 6.4 where we obtain a relation between antiderivatives and definite integrals that will be a great time-saver in evaluating many (but not all) definite integrals.

Notation for an Antiderivative

We have not yet introduced a symbol for an antiderivative of a function. The following notation is standard:

NOTATION: Any antiderivative of f is denoted $\int f(x)\,dx$.

For instance, $x^3 = \int 3x^2\,dx$. This equation is read "x^3 is an antiderivative of $3x^2$". That means that the derivative of x^3 is $3x^2$. It is true that $x^3+2024 = \int 3x^2\,dx$, since x^3+2024 is also an antiderivative of $3x^2$. That does not mean that the functions x^3 and x^3+2024 are equal. All it means is that these two functions both have the same derivative, $3x^2$. The symbol $\int 3x^2\,dx$ refers to any function whose derivative is $3x^2$.

If $F'(x) = f(x)$ we write $F(x) = \int f(x)\,dx$. The function $f(x)$ is called the **integrand**. The function $F(x)$ is called an antiderivative of $f(x)$. The symbol for an antiderivative, $\int f(x)\,dx$, is similar to the symbol for a definite integral, $\int_a^b f(x)\,dx$, but they denote different concepts. An antiderivative is often called an integral or indefinite integral, but it should not be confused with a definite integral. The symbol $\int f(x)\,dx$ denotes a function, any function whose derivative is $f(x)$. The symbol $\int_a^b f(x)\,dx$ denotes a number — one that is defined by a limit of certain sums. The value of the definite integral may vary as the interval $[a,b]$ changes.

It is important to remember that $\int f(x)\,dx$ is a function and $\int_a^b f(x)\,dx$ is a number.

We apologize for the use of such similar notations, $\int f(x)\,dx$ and $\int_a^b f(x)\,dx$, for such distinct concepts. However, it is not for us to undo over three centuries of custom. Rather, the symbols $\int f(x)\,dx$ and $\int_a^b f(x)\,dx$ must be read carefully. You can distinguish between $\int f(x)\,dx$ and $\int_a^b f(x)\,dx$ just as you can distinguish between such similar-looking words as "density" and "destiny" or "nuclear" and "unclear."

Properties of Antiderivatives

The book's Table of Integrals (in Appendix A) lists many antiderivatives. One example is $\int \sin(x)\,dx = -\cos(x)$. Of course, $-\cos(x)+17$ also is an antiderivative of $\sin(x)$. In Section 4.1 it was shown that if F and G have the same derivative on an interval, they differ by a constant. The most common choice for this constant is C. So $F(x)-G(x) = C$ or $F(x) = G(x)+C$. For emphasis, we state this as a theorem.

This result was anticipated back in Section 3.6.

The theorem asserts that if $F(x)$ is an antiderivative for $f(x)$, then any other antiderivative of $f(x)$ is of the form $F(x)+C$ for some constant C.

Theorem 6.3.1: Antiderivatives Agree up to a Constant

If F and G are both antiderivatives of f on some interval, then there is a constant C such that $F(x) = G(x)+C$.

When using an antiderivative, it is best to include the constant C. For example,

$$\int 5\,dx = 5x+C, \qquad \int e^x\,dx = e^x + C, \quad \text{and} \quad \int \sin(2x)\,dx = \frac{-1}{2}\cos(2x)+C.$$

Many tables of integrals, including the one in this book, omit the $+C$.

We also know the following two derivatives:

$$\frac{d}{dx}\left(\int x^3\, dx\right) = x^3 \qquad \text{and} \qquad \frac{d}{dx}\left(\int \sin(2x)\, dx\right) = \sin(2x).$$

Are the equations profound or trivial? Read them aloud and decide.

The first says, "The derivative of an antiderivative of x^3 is x^3." It is true simply because that is how we defined the antiderivative. We know that

$$\frac{d}{dx}\left(\int \frac{\ln(1+x^2)}{\sin^2(x)}\, dx\right) = \frac{\ln(1+x^2)}{\sin^2(x)}$$

even though we cannot write out a formula for an antiderivative of $\ln(1+x^2)/\sin^2(x)$.

This sort of an inverse relationship is not new to us. For example, we know that the square of the square root of 7 is 7 and that $e^{\ln(3)} = 3$. Both of these facts are a properties of inverse functions.

In the same way, the fact that antidifferentiation and differentiation are inverse operations yields the following general result.

Observation 6.3.2: ***Inverse Relationship between Derivatives and Antiderivatives***

$$\frac{d}{dx}\left(\int f(x)\, dx\right) = f(x).$$

Every property of derivatives gives us a corresponding property of antiderivatives. Three of the most important properties of antiderivatives are in the next theorem.

Theorem 6.3.3: Properties of Antiderivatives

Assume that f and g are functions with antiderivatives $\int f(x)\, dx$ and $\int g(x)\, dx$. Then

(a) $\int c f(x)\, dx = c \int f(x)\, dx$ *for any constant c.*

(b) $\int (f(x) + g(x))\, dx = \int f(x)\, dx + \int g(x)\, dx.$

(c) $\int (f(x) - g(x))\, dx = \int f(x)\, dx - \int g(x)\, dx.$

Proof of Three Properties of Antiderivatives (Theorem 6.3.3)

(a) Before we prove that $\int cf(x)\, dx = c\int f(x)\, dx$, we stop to see what it means. This equation says that "c times an antiderivative of $f(x)$ is an antiderivative of c times $f(x)$". If $F(x)$ is an antiderivative of $f(x)$, then the equation says "c times $F(x)$ is an antiderivative of c times $f(x)$". To determine if this statement is true we must differentiate $cF(x)$ and check that we get $cf(x)$. So, we compute $(cF(x))'$:

$$\begin{aligned}(cF(x))' &= cF'(x) && (\ c \text{ is a constant}\ )\\ &= cf(x) && (\ F \text{ is antiderivative of } f\ ).\end{aligned}$$

Thus $cF(x)$ is indeed an antiderivative of $cf(x)$. Therefore, we may write

$$cF(x) = \int cf(x)\, dx.$$

Since $F(x) = \int f(x)\, dx$, we conclude that

$$c\int f(x)\, dx = \int cf(x)\, dx.$$

(b) The proof is similar. We show that $\int f(x)\,dx + \int g(x)\,dx$ is an antiderivative of $f(x)+g(x)$. To do this we compute the derivative of $\int f(x)\,dx + \int g(x)\,dx$:

$$\begin{aligned}\frac{d}{dx}\left(\int f(x)\,dx + \int g(x)\,dx\right) &= \frac{d}{dx}\left(\int f(x)\,dx\right) + \frac{d}{dx}\left(\int g(x)\,dx\right) && (\text{ derivative of a sum })\\ &= f(x)+g(x). && (\text{ definition of antiderivatives })\end{aligned}$$

(c) The proof is similar to the one for (b).

EXAMPLE 1. Find (a) $\int 6\cos(x)\,dx$, (b) $\int (6\cos(x)+3x^2)\,dx$, and (c) $\int \left(6\cos(x) - \frac{5}{1+x^2}\right)\,dx$.

SOLUTION

(a) Part (a) of the theorem is used to move the 6 (a constant) past the integral sign. We then have:

$$\begin{aligned}\int 6\cos(x)\,dx &= 6\int \cos(x)\,dx && (\text{ part (a) of the theorem })\\ &= 6\sin(x)+C && (\text{ } (\sin(x))' = \cos(x) \text{ })\end{aligned}$$

The $+C$ is added as the last step in finding an antiderivative.

(b) This antiderivative with an integrand that is the sum of two functions is found as follows:

$$\begin{aligned}\int (6\cos(x)+3x^2)\,dx &= \int 6\cos(x)\,dx + \int 3x^2\,dx && (\text{ part (b) of the theorem })\\ &= 6\sin(x)+x^3+C. && (\text{ part (a) of the Example and } (x^3)' = 3x^2 \text{ })\end{aligned}$$

Separate constants are not needed for each antiderivative, only one $+C$ for the overall antiderivative.

(c) Antiderivatives when the integrand is a difference of two functions is found in a similar manner:

$$\begin{aligned}\int \left(6\cos(x) - \frac{5}{1+x^2}\right)\,dx &= \int 6\cos(x)\,dx - \int \frac{5}{1+x^2}\,dx && (\text{ part (c) of the theorem })\\ &= 6\int \cos(x)\,dx - 5\int \frac{1}{1+x^2}\,dx && (\text{ part (a) of the theorem, twice })\\ &= 6\sin(x) - 5\arctan(x) + C && (\text{ } (\sin(x))' = \cos(x),\ (\arctan(x))' = 1/(1+x^2) \text{ }).\end{aligned}$$

The last two parts of Theorem 6.3.3 extend to any finite number of functions. For instance,

$$\int (f(x)-g(x)+h(x))\,dx = \int f(x)\,dx - \int g(x)\,dx + \int h(x)\,dx.$$

Theorem 6.3.4: Antiderivative of x^a

For $a \neq -1$, $\int x^a\,dx = \frac{x^{a+1}}{a+1} + C.$

Proof of Antiderivative of x^a (Theorem 6.3.4)

As we have done for all other antiderivative results, this result is proven by verifying that the derivative of the antiderivative is the integrand: Assume $a \neq -1$, then

$$\begin{aligned}\left(\frac{x^{a+1}}{a+1}\right)' &= \frac{(a+1)x^{(a+1)-1}}{a+1}\\ &= x^a.\end{aligned}$$

EXAMPLE 2. Find $\int \left(\frac{3}{\sqrt{1-x^2}} - \frac{2}{x} + \frac{1}{x^3} \right) dx$, $0 < x < 1$.

SOLUTION Recall the $(\arcsin(x))' = 1/\sqrt{1-x^2}$ (for $|x| < 1$) and $(\ln(x))' = 1/x$ (for $x > 0$). Then, for $0 < x < 1$,

$$\begin{aligned}\int \left(\frac{3}{\sqrt{1-x^2}} - \frac{2}{x} + \frac{1}{x^3} \right) dx &= 3\int \frac{1}{\sqrt{1-x^2}}\, dx - 2\int \frac{1}{x}\, dx + \int x^{-3}\, dx \\ &= 3\arcsin(x) - 2\ln(x) + \frac{x^{-2}}{-2} + C \\ &= 3\arcsin(x) - 2\ln(x) - \frac{1}{2x^2} + C.\end{aligned}$$

♦

Properties of Definite Integrals

So far, in the definite integral, $\int_a^b f(x)\, dx$, b is larger than a. It will be useful to be able to speak about the definite integral from a to b even if b is less than or equal to a. The following definitions allow us to do that. We will use them in the proofs of the two fundamental theorems of calculus in the next section.

Definition: *Definite Integral from a to b, $\int_a^b f(x)\, dx$, where $b < a$*

If b is less than a, then $\int_a^b f(x)\, dx = -\int_b^a f(x)\, dx$.

EXAMPLE 3. Compute $\int_3^0 x^2\, dx$, the integral from 3 to 0 of x^2.

SOLUTION The symbol $\int_3^0 x^2\, dx$ is defined as $-\int_0^3 x^2\, dx$. As was shown in Section 6.2, $\int_0^3 x^2\, dx = 9$. Thus

$$\int_3^0 x^2\, dx = -9.$$

♦

The definite integral is defined using partitions of an interval. Because partitions do not have sections of length 0, we need to define $\int_a^a f(x)\, dx$.

Definition: *Definite Integral from a to a, $\int_a^a f(x)\, dx$*

$$\int_a^a f(x)\, dx = 0$$

With these definitions the symbol $\int_a^b f(x)\, dx$ is defined for any numbers a and b and any continuous function f, assuming $f(x)$ is defined for x between a and b. It is no longer necessary that a be less than b.

The definite integral has several properties, some of which we will be using in this section and some in later chapters. Justifications of them are provided immediately after the following theorem.

Theorem 6.3.5: Properties of the Definite Integral

If f and g are continuous functions, and c is a constant, then

1. ***Moving a Constant Past $\int_a^b$***

$$\int_a^b c f(x)\,dx = c\int_a^b f(x)\,dx$$

2. ***Definite Integral of a Sum***

$$\int_a^b (f(x)+g(x))\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$$

3. ***Definite Integral of a Difference***

$$\int_a^b (f(x)-g(x))\,dx = \int_a^b f(x)\,dx - \int_a^b g(x)\,dx$$

4. ***Definite Integral of a Non-Negative Function***
If $f(x) \geq 0$ for all x in $[a,b]$, $a<b$, then

$$\int_a^b f(x)\,dx \geq 0.$$

5. ***Definite Integrals Preserve Order***
If $f(x) \geq g(x)$ for all x in $[a,b]$, $a<b$, then

$$\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx.$$

6. ***Sum of Definite Integrals Over Adjoining Intervals***
If a, b, and c are numbers, then

$$\int_a^c f(x)\,dx + \int_c^b f(x)\,dx = \int_a^b f(x)\,dx.$$

7. ***Bounds on Definite Integrals***
If m and M are numbers such that $m \leq f(x) \leq M$ for all x between a and b, then

$$m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a) \qquad \text{if } a<b$$

and

$$m(b-a) \geq \int_a^b f(x)\,dx \geq M(b-a) \qquad \text{if } a>b$$

Proof of Property 1 (Theorem 6.3.5)

Take the case $a<b$. The equation $\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx$ resembles part (a) of Theorem 6.3.3 about antideriva-

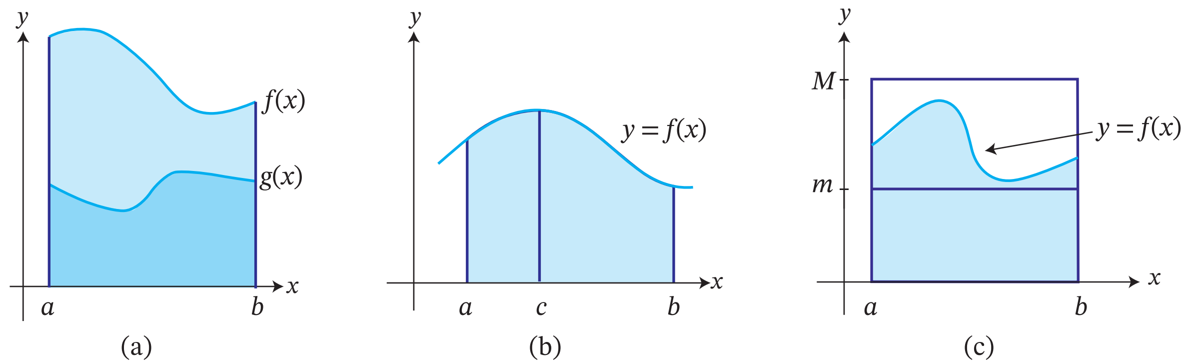


Figure 6.3.1

tives: $\int cf(x)\,dx = c\int f(x)\,dx$. However, its proof is different, since $\int_a^b cf(x)\,dx$ is defined as a limit of sums.

We have

$$\begin{aligned}
\int_a^b cf(x)\,dx &= \lim_{\text{all }\Delta x_i \to 0^+} \sum_{i=1}^n cf(c_i)\Delta x_i && (\text{ definition of definite integral })\\
&= \lim_{\text{all }\Delta x_i \to 0^+} c\sum_{i=1}^n f(c_i)\Delta x_i && (\text{ distributive law })\\
&= c\lim_{\text{all }\Delta x_i \to 0^+} \sum_{i=1}^n f(c_i)\Delta x_i && (\text{ property of limits })\\
&= c\int_a^b f(x)\,dx. && (\text{ definition of definite integral }).
\end{aligned}$$

Similar arguments justify Properties 2–4 of Theorem 6.3.5. We will make plausible arguments for Properties 5–7 by interpreting them in terms of areas for positive functions.

Plausibility Argument for Property 5 of Theorem 6.3.5 when $f(x) \ge g(x) \ge 0$

This amounts to the assertion that when the graph of $y = f(x)$ is always at least as high as the graph of $y = g(x)$, then the area of a region under the curve $y = f(x)$ is greater than or equal to the area under the curve $y = g(x)$ above a given interval. (See Figure 6.3.1(a).)

Plausibility Argument for Property 6 of Theorem 6.3.5

When $a < c < b$ and $f(x)$ assumes only positive values, this property asserts that the area of the region below the graph of $y = f(x)$ and above the interval $[a,b]$ is the sum of the areas of the regions below the graph and above the smaller intervals $[a,c]$ and $[c,b]$. Figure 6.3.1(b) shows that this is plausible.

Plausibility Argument for Property 7 of Theorem 6.3.5

The inequalities in this property compare the area under the graph of $y = f(x)$ with the areas of two rectangles, one of height M and one of height m. (See Figure 6.3.1(c).) For $a < b$, the area of the larger rectangle is $M(b-a)$ and the area of the smaller rectangle is $m(b-a)$.

The Mean Value Theorem for Definite Integrals

The mean value theorem for derivatives says that under suitable conditions $f(b) - f(a) = f'(c)(b-a)$ for some number c in $[a,b]$. The mean value theorem for definite integrals is similar. First we state it geometrically.

If $f(x)$ is positive and $a < b$, then $\int_a^b f(x)\,dx$ can be interpreted as the area of the shaded region in Figure 6.3.2(a).

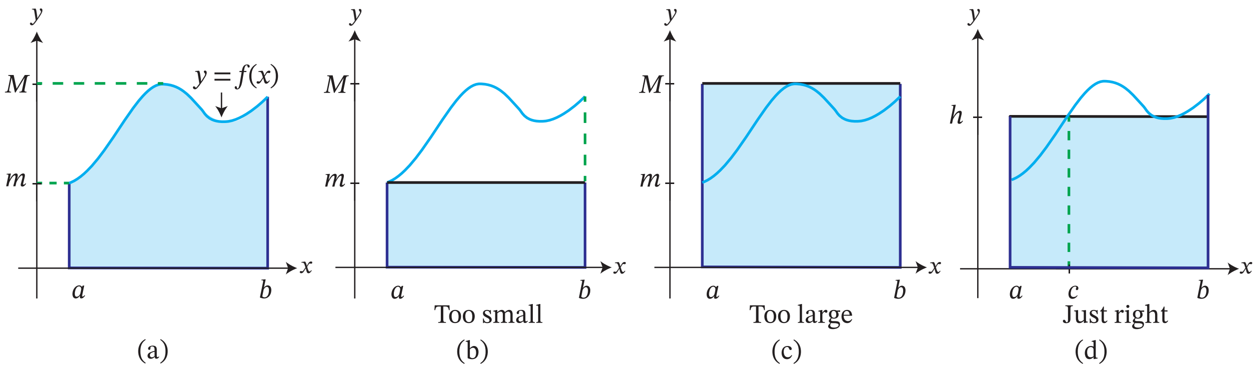


Figure 6.3.2

Let m be the minimum and M the maximum value of $f(x)$ for x in $[a,b]$. We assume that $m < M$. (What would change if $m = M$? Why is $m > M$ not possible?) The area of the rectangle of height M is larger than the shaded area; the area of the rectangle of height m is smaller than the shaded area. (See Figures 6.3.2(b) and (c).) Therefore, there is a rectangle, with height h somewhere between m and M, whose area is the same as the area under the curve $y = f(x)$. (See Figure 6.3.2(d).) Hence $\int_a^b f(x)\,dx = h(b-a)$.

Because h is a number between m and M, by the intermediate value property for continuous functions there is a number c in $[a,b]$ such that $f(c) = h$. (See Figure 6.3.2(d).) Hence,

$$\text{Area of shaded region under curve} = f(c)(b-a).$$

This suggests the mean value theorem for definite integrals.

Theorem 6.3.6: Mean Value Theorem for Definite Integrals

If a and b are numbers, and if f is a continuous function defined between a and b, then there is a number c between a and b such that

$$\int_a^b f(x)\,dx = f(c)(b-a).$$

Proof of the Mean Value Theorem for Definite Integrals (Theorem 6.3.6)

This proof uses only properties of the definite integral.

Suppose $a < b$. Let M be the maximum and m the minimum value of $f(x)$ on $[a,b]$. Property 7, combined with division by $b-a$, gives

$$m \le \frac{1}{b-a}\int_a^b f(x)\,dx \le M,$$

Because f is continuous on $[a,b]$, by the intermediate value property of Section 2.5 there is a number c in $[a,b]$ such that

$$f(c) = \frac{1}{b-a}\int_a^b f(x)\,dx$$

and the theorem is proved without depending on a picture. ●

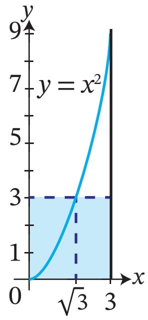


Figure 6.3.3

EXAMPLE 4. Verify the mean value theorem for definite integrals when $f(x) = x^2$ and $[a,b] = [0,3]$.

SOLUTION In Section 6.2 it was shown that $\int_0^3 x^2\,dx = 9$. Since $f(x) = x^2$, we are looking for c in $[0,3]$ such that

$$\int_0^3 x^2\,dx = 9 = c^2(3-0).$$

That is, $9 = 3c^2$, so $c^2 = 3$, and $c = \sqrt{3}$. (See Figure 6.3.3.) The rectangle with height $f(\sqrt{3}) = 3$ and base $[0,3]$ has the same area as the region under the curve $y = x^2$ and above $[0,3]$. ♦

The Average Value of a Function

When $f(x)$ is a continuous function defined on $[a,b]$, what shall we mean by "the average value of $f(x)$ over $[a,b]$"? We cannot add the values of $f(x)$ for all x's in $[a,b]$ and divide by the number of x's, since there are an infinite number of them. However, we can work with the average (or mean) of n numbers $a_1, a_2, \dots, a_n$, which is their sum divided by n:

$$\frac{1}{n}\sum_{i=1}^{n} a_i.$$

This suggests how to define "the average value of $f(x)$ over $[a,b]$". Choose a large integer n and partition $[a,b]$ into n sections of equal length, $\Delta x = (b-a)/n$. Let the c_i be sampling points chosen from each section. Then an estimate of the average would be

$$\frac{1}{n}\left(f(c_1) + f(c_2) + \cdots + f(c_n)\right). \tag{6.3.1}$$

Since $\Delta x = (b-a)/n$, it follows that $1/n = \Delta x/(b-a)$. Therefore, (6.3.1) can be rewritten as

$$\frac{1}{b-a}\sum_{i=1}^{n} f(c_i)\Delta x.$$

But $\sum_{i=1}^{n} f(c_i)\Delta x$ is an estimate of $\int_a^b f(x)\,dx$. It follows that, as $n \to \infty$, this average of the n function values approaches $\left(\int_a^b f(x)\,dx\right)/(b-a)$. This motivates the definition of average value of a function over an interval.

Definition: ***Average Value of a Function over an Interval.***

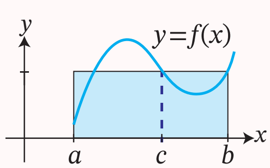


Figure 6.3.4

Let $f(x)$ be defined on the interval $[a,b]$. Assume that $\int_a^b f(x)\,dx$ exists. The **average value** or **mean value** of f on $[a,b]$ is defined to be

$$\frac{1}{b-a}\int_a^b f(x)\,dx.$$

Geometrically speaking (if $f(x)$ is positive), the average value of $f(x)$ on an interval $[a,b]$ is the height of a rectangle that has base $[a,b]$ and the same area as the area of the region under the curve $y = f(x)$ and above $[a,b]$. (See Figure 6.3.4.) The average value of $f(x)$ over $[a,b]$ is between its maximum and minimum values but it is not necessarily their average. Nor is it necessarily the average of $f(a)$ and $f(b)$.

EXAMPLE 5. Find the average value of 2^x over the interval $[1,3]$.

SOLUTION The average value of 2^x over $[1,3]$ by definition equals $\frac{1}{3-1}\int_1^3 2^x\,dx$. By Example 5 in Section 6.2,

$$\int_1^3 2^x\,dx = \frac{1}{\ln(2)}(2^3-2^1) = \frac{6}{\ln(2)}.$$

Hence,

$$\text{Average value of } 2^x \text{ over } [1,3] = \frac{1}{3-1}\frac{6}{\ln(2)} = \frac{3}{\ln(2)} \approx 4.2381.$$

◆

Observation 6.3.7:

Observe that the average value of $f(x) = 2^x$ on $[1,3]$ is not equal to the average of the function values at the endpoints, namely $\left(f(1)+f(3)\right)/2 = (2^1+2^3)/2 = 5$, nor is it equal to the function evaluated at the midpoint of the interval $[1,3]$, namely $f((1+3)/2) = f(2) = 2^2 = 4$.

The Zero-Integral Principle

Let f be a continuous function on the interval $[a,b]$. Suppose that for every subinterval $[c,d]$ of $[a,b]$, $\int_c^d f(x)\,dx$ is zero. The constant function $f(x)=0$ has this property. We now show that it is the only function with this property.

Let $f(x)$ be a continuous function on $[a,b]$ that is not the constant function 0. Then there is a number q in $[a,b]$ such that $f(q)=p$ is not zero. We consider the case when p is positive. (The case when p is negative can be treated the same way. See Exercise 48.)

By the permanence property (Theorem 2.5.5 in Section 2.5), there is a subinterval $[c,d]$ of $[a,b]$ where the function values remain larger than $p/2$. Then $\int_c^d f(x)\,dx$ is at least $p/2$ times the length of the interval $[c,d]$ and hence not 0. This contradicts the assumption that $\int_c^d f(x)\,dx = 0$ for all subintervals $[c,d]$ of the domain of f. As a result, the hypothesis must be false and so f is zero on $[a,b]$.

Theorem 6.3.8: Zero-Integral Principle

If f is a continuous function on $[a,b]$, with the property that $\int_c^d f(x)\,dx = 0$ for every subinterval $[c,d]$ of $[a,b]$, then $f(x)=0$ for all x on $[a,b]$.

Summary

We introduced the notation $\int f(x)\,dx$ for an **antiderivative** of $f(x)$ and stated several of its properties. We also defined the definite integral $\int_a^b f(x)\,dx$ for all numbers a and b, and stated several of its properties. This discussion culminated with the presentation, proof, and application of three ideas involving definite integrals: the mean value theorem for definite integrals, the average value of a function on an interval, and the zero-integral property.

The mean value theorem for definite integrals asserts that for a continuous function $f(x)$, $\int_a^b f(x)\,dx$ equals $f(c)$ times $(b-a)$ for at least one value of c in $[a,b]$.

The quantity $\left(\int_a^b f(x)\,dx\right)/(b-a)$ is called the average value (or mean value) of $f(x)$ over $[a,b]$. For a positive function it can be thought of as the height of a rectangle whose area is the same as the area of the region under the curve $y=f(x)$.

The zero-integral principle, which says that if the integral of a continuous function is 0 *over each interval within a given interval*, then, within the given interval, it must be the constant function with value 0.

EXERCISES for Section 6.3

In Exercises 1 to 14 evaluate the antiderivative. Remember to add a constant. Check the answer by differentiation.

1. $\int 5x^2\, dx$

2. $\int (2x - x^3 + x^5)\, dx$

3. $\int \frac{7}{x^2}\, dx$

4. $\int \left(6x^2 + 2x^{-1} + \frac{1}{\sqrt{x}}\right) dx$

5. $\int (\sin(2x) - 3\cos(x))\, dx$

6. $\int (2\sin(x) + \cos(3x))\, dx$

7. (a) $\int e^x\, dx$, (b) $\int e^{x/3}\, dx$

8. (a) $\int \cos(x)\, dx$, (b) $\int \cos(2x)\, dx$

9. (a) $\int \sin(x)\, dx$, (b) $\int \sin(3x)\, dx$

10. $\int \sec(x)\tan(x)\, dx$

11. $\int (\sec(x))^2\, dx$

12. $\int (\csc(x))^2\, dx$

13. $\int \frac{1}{1+x^2}\, dx$

14. $\int \frac{1}{\sqrt{1-x^2}}\, dx$

---

15. State the mean value theorem for definite integrals in words, using no mathematical symbols.

16. Define the average value of a function over an interval, using no mathematical symbols.

In Exercises 17 to 20 evaluate each definite integral or antiderivative.

17. (a) $\int_2^5 x^2\, dx$, (b) $\int_5^2 x^2\, dx$, and (c) $\int_5^5 x^2\, dx$.

18. (a) $\int_1^2 x\, dx$, (b) $\int_2^1 x\, dx$, and (c) $\int_3^3 x\, dx$.

19. (a) $\int x\, dx$ and (b) $\int_3^4 x\, dx$.

20. (a) $\int 3x^2\, dx$ and (b) $\int_1^4 3x^2\, dx$.

---

21. If $2 \le f(x) \le 3$, what can be said about $\int_1^6 f(x)\, dx$?

22. If $-1 \le f(x) \le 4$, what can be said about $\int_{-2}^7 f(x)\, dx$?

23. Write a sentence or two in your own words that tells what the symbols $\int f(x)\, dx$ and $\int_a^b f(x)\, dx$ mean. Include examples. Use as few mathematical symbols as possible.

24. Let $f(x)$ be a differentiable function. In this exercise you will determine if $f(x) = \int \frac{df}{dx}\, dx$ is true or false.

(a) Pick several functions and test if the equation is true in general.

(b) Determine if the equation is true and justify your conclusion.

The mean value theorem for definite integrals asserts that if $f(x)$ is continuous throughout the interval with endpoints a and b, then $\int_a^b f(x)\,dx = f(c)(b-a)$ for some number c in $[a,b]$. In Exercises 25 to 28 find $f(c)$ and at least one value of c in $[a,b]$.

25. $f(x) = 2x$, $[a,b] = [1,5]$

26. $f(x) = 5x+2$, $[a,b] = [1,2]$

27. $f(x) = x^2$, $[a,b] = [0,4]$

28. $f(x) = x^2 + x$, $[a,b] = [1,4]$

---

29. If $\int_1^2 f(x)\,dx = 3$ and $\int_1^5 f(x)\,dx = 7$, find

(a) $\int_2^1 f(x)\,dx$ and (b) $\int_2^5 f(x)\,dx$.

30. If $\int_1^3 f(x)\,dx = 4$ and $\int_1^3 g(x)\,dx = 5$, find

(a) $\int_1^3 \big(2f(x)+6g(x)\big)\,dx$ and (b) $\int_3^1 \big(f(x)-g(x)\big)\,dx$.

---

31. If the maximum value of $f(x)$ on $[a,b]$ is 7 and its minimum value on $[a,b]$ is 4, what can be said about

(a) $\int_a^b f(x)\,dx$? (b) the mean value of $f(x)$ on $[a,b]$?.

32. Let $f(x) = c$ (constant) for all x in $[a,b]$. Find the average value of $f(x)$ on $[a,b]$.

In Exercises 33 to 36, find the minimum, maximum, and average value of the given function on the given interval.

33. $f(x) = x^2$, $[2,3]$

34. $f(x) = x^2$, $[0,5]$

35. $f(x) = 2^x$, $[0,4]$

36. $f(x) = 2^x$, $[2,4]$

---

37. (a) Let a, b, and c be constants. Assume that the integral of $\left(ax^2+bx+c\right)^2$ over every interval is zero. Find a, b, and c.

(b) Assume that the integral of ax^2+bx+c over every interval is zero. Find a, b, and c.

38. Let a and b be constants. Assume that the integral of $ae^{x^3} + b\cos^{10}(x)$ over every interval is zero. Find a and b.

39. Prove the mean value theorem for definite integrals when $b < a$. (Use the definition of $\int_a^b f(x)\,dx$ when $b < a$.)

40. Is $\int f(x)g(x)\,dx$ always equal to $\int f(x)\,dx \int g(x)\,dx$? Are they ever equal? Explain.

41. (a) Show that $\frac{1}{3}\sin^3(x)$ is not an antiderivative of $\sin^2(x)$.

(b) Use the identity $\sin^2(x) = \frac{1}{2}(1-\cos(2x))$ to find an antiderivative of $\sin^2(x)$.

(c) Verify your answer in (b) by differentiation.

42. Define $f(x) = \begin{cases} -x & 0 < x \le 1 \\ -1 & 1 < x \le 2 \\ 1 & 2 < x \le 3 \\ 4-x & 3 < x \le 4. \end{cases}$

(a) Sketch the graphs of $y = f(x)$ and $y = \left(f(x)\right)^2$ on the interval $[0,4]$.
(b) Find the average value of f on the interval $[0,4]$.
(c) The **root mean square** (RMS) of a function g on $[a,b]$ is defined as $\sqrt{\dfrac{1}{b-a}\displaystyle\int_a^b \left(g(x)\right)^2\, dx}$.

Find the root mean square value of f on the interval $[0,4]$.
(d) Why is it not surprising that your answer in (b) is zero and your answer in (c) is positive?

INTERESTING FACT: The voltage of the alternating current is defined using RMS.

43. Consider the following short exchange between Sam and Jane.

SAM: The text makes the average value of a function on $[a,b]$ too hard.
JANE: How so?
SAM: It's easy. Just average $f(a)$ and $f(b)$, that is, $\dfrac{1}{b-a}\displaystyle\int_a^b f(x)\, dx = \dfrac{1}{2}\left(f(b)+f(a)\right)$.
JANE: That sure is easier.

(a) Show that Sam is correct when $f(x)$ is any polynomial of degree 0 or 1.
(b) Is Sam's suggestion always correct? Is it ever correct? Explain.

Exercise 44 describes the famous **Buffon needle** problem, now over 200 years old. Exercise 45 is related, but is not nearly as famous.

44. On the floor there are parallel lines a distance d from each other, such as the edges of slats. You throw a straight wire of length d on the floor at random. Sometimes when it comes to rest it crosses a line, sometimes it does not.

(a) Find a floor with parallel lines, perform the experiment at least 20 times, and calculate the percentage of times the wire crosses a line.
(b) If the wire makes an angle θ with a line perpendicular to the lines, show that the probability that it crosses a line is $\cos(\theta)$.
(c) Find the average value of that probability. That is the probability that the wire crosses a line.
(d) How close is the experimental value in (a) to the theoretical value in (c)?

45. An infinite floor is composed of congruent square tiles arranged as in a checkerboard. You have a straight wire whose length is the same as the length of a side of a square. The edges of the squares form lines in perpendicular directions. What is the probability that when you throw the wire at random it crosses two lines, one in each of the two perpendicular directions? This is related to Exercise 44, the classic Buffon needle problem. (One way to check if your answer is reasonable is to collect data by carrying out the experiment.) $1/\pi$

---

46. The average value of a function $f(x)$ on $[1,3]$ is 4. On $[3,6]$ its average value is 5. What is its average value on $[1,6]$? Explain your answer.

47. Assume that f and g are continuous functions and that $\displaystyle\int_a^b f(x)\, dx$ equals $\displaystyle\int_a^b g(x)\, dx$ for every interval $[a,b]$. Show that $f(x)$ equals $g(x)$ for all x.

48. The justification of the zero-integral principle in this section showed that if the integrand is positive at a point leads to a contradiction of the hypothesis of the zero-integral property (that the definite integral of $f(x)$ over every subinterval is zero). Show that the assumption that the integrand is negative at a point also leads to a contradiction of the zero-integral property. Then, with these two possibilities excluded, the only remaining possibility is that the integrand is zero at every point in the interval.

49. This exercise evaluates two definite integrals that appear often in applications, but for which are not easily evaluated using the FTC.

(a) Draw the graphs of $y = \cos^2(x)$ and $y = \sin^2(x)$. From the picture, decide how $\int_0^{\pi/2} \cos^2(x)\,dx$ and $\int_0^{\pi/2} \sin^2(x)\,dx$ compare.

(b) Using (a) and a trigonometric identity, show that $\int_0^{\pi/2} \cos^2(x)\,dx = \frac{\pi}{4} = \int_0^{\pi/2} \sin^2(x)\,dx$.

(c) Evaluate $\int_0^{\pi} \cos^2(x)\,dx$.

In Exercises 50 and 51 verify the equations. Differentiate the right-hand side and see that the result is the integrand on the left-hand side. (The number a is a constant.)

50. $\int x^2 \sin(ax)\,dx = \frac{2x}{a^2}\sin(ax) + \frac{2}{a^3}\cos(ax) - \frac{x^2}{a}\cos(ax) + C$

51. $\int x \sin^2(ax)\,dx = \frac{x^2}{4} - \frac{x}{4a}\sin(2ax) - \frac{1}{8a^2}\cos(2ax) + C$

6.4 The Fundamental Theorem of Calculus

ALERT #1: Many people say this is the most important section in this, or any, calculus book.

In this section we obtain two closely related theorems. They are called the Fundamental Theorems of Calculus I and II, or simply the Fundamental Theorem of Calculus (FTC). The first part of the FTC provides a way to evaluate a definite integral if you know an antiderivative of the integrand. That means that the derivative, developed in Chapter 3, has yet another application.

The second Fundamental Theorem of Calculus (FTC II) tells how rapidly the value of a definite integral changes as the interval of integration changes.

ALERT #2: Our numbering of the two parts of the FTC might seem a bit curious: FTC I is the result that is used most often. But, FTC II is used to prove FTC I. (Some books reverse this numbering. Stay alert!)

Observation 6.4.1: *Overview of the Fundamental Theorem of Calculus*

- FTC I gives a way to evaluate $\int_a^b f(x)\,dx$.
- FTC II gives a way to evaluate $\frac{d}{dx}\left(\int_a^x f(t)\,dt\right)$.

TERMINOLOGY: The FTC is "fundamental" because of the way it connects the separate ideas of antiderivative (inverse of differentiation) and definite integral (area under a curve).

Motivation for the Fundamental Theorem of Calculus I

In Section 6.2 we found that $\int_a^b c\ dx = cb - ca$ and $\int_a^b x\ dx = b^2/2 - a^2/2$. In the same section we found that $\int_a^b x^2\ dx = b^3/3 - a^3/3$ by knowing that congruent lopsided tents filled a cube. Using the formula for the sum of a geometric series, we showed that $\int_a^b 2^x\ dx = 2^b/\ln(2) - 2^a/\ln(2)$.

All four definite integrals follow a similar pattern:

$$\int_a^b c\ dx = cb - ca \qquad \int_a^b x\ dx = \frac{b^2}{2} - \frac{a^2}{2} \qquad \int_a^b x^2\ dx = \frac{b^3}{3} - \frac{a^3}{3} \qquad \int_a^b 2^x\ dx = \frac{2^b}{\ln(2)} - \frac{2^a}{\ln(2)}.$$

Next, compute an antiderivative (indefinite integral) of each of the integrands:

$$\int c\ dx = cx \qquad \int x\ dx = \frac{x^2}{2} \qquad \int x^2\ dx = \frac{x^3}{3} \qquad \int 2^x\ dx = \frac{2^x}{\ln(2)}.$$

Observation 6.4.2: *Evaluation of Definite Integrals*

In each case the definite integral equals the difference between the values of an antiderivative of the integrand evaluated at b and at a, the two endpoints of the interval.

This suggests that maybe if $F(x)$ is an antiderivative of the integrand $f(x)$, then

$$\int_a^b f(x)\ dx = F(b) - F(a). \tag{6.4.1}$$

If this is correct, then, instead of evaluating a definite integral by a special device, such as cutting up a cube or summing a geometric series, we can do it if we know an antiderivative of the integrand.

We may reason using velocity and distance to provide further evidence for (6.4.1). Picture a particle moving upwards on the y-axis. At time t it is at position $F(t)$. The velocity at time t is $F'(t)$.

Observation 6.4.3: *Physical Justification of FTC I*

We saw that the definite integral of the velocity from time a to time b tells the change in position over the same time interval, that is,

the definite integral of the velocity = the final position−the initial position = change in position.

In symbols,

$$\int_a^b F'(t)\ dt = F(b) - F(a). \tag{6.4.2}$$

If we give $F'(t)$ the name $f(t)$, then we can restate (6.4.2) as:

$$\text{If } f(t) = F'(t) \text{, then } \int_a^b f(t)\ dt = F(b) - F(a).$$

In other words,

$$\text{If } F \text{ is an antiderivative of } f \text{, then } \int_a^b f(t)\ dt = F(b) - F(a).$$

Definition: *Notation for Evaluating Definite Integrals*

The difference $F(b) - F(a)$ is denoted $F(t)\Big|_a^b$.

The formulas we found for the integrands c, x, x^2, and 2^x and reasoning about motion are all consistent with the first part of the **Fundamental Theorem of Calculus**.

Theorem 6.4.4: Fundamental Theorem of Calculus I

If f is continuous on $[a,b]$ and if F is an antiderivative of f, then

$$\int_a^b f(x)\,dx = F(x)\Big|_a^b = F(b) - F(a).$$

This theorem says that to evaluate the definite integral of f from a to b look for an antiderivative of f, then evaluate the antiderivative at b and subtract the value of the antiderivative at a. The difference is the value of the definite integral. To apply the theorem it is necessary to find an antiderivative of the integrand f. For many functions it is easy to do this. For some it is possible, but for others an antiderivative expressible in terms of the common functions cannot be found. Some techniques for finding antiderivatives are discussed in Chapter 7.

Example 1 shows the power of FTC I.

EXAMPLE 1. Use the fundamental theorem of calculus to evaluate $\int_0^{\pi/2} \cos(x)\,dx$.

SOLUTION Since $(\sin(x))' = \cos(x)$, $\sin(x)$ is an antiderivative of $\cos(x)$. By FTC I,

$$\int_0^{\pi/2} \cos(x)\,dx = \sin(x)\Big|_0^{\pi/2} = \sin\left(\frac{\pi}{2}\right) - \sin(0) = 1 - 0 = 1.$$

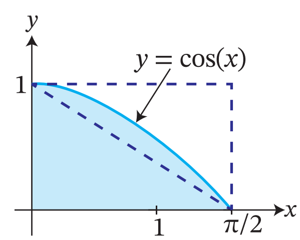


Figure 6.4.1

This tells us that the area under the curve $y = \cos(x)$ and above $[0, \pi/2]$, shown in Figure 6.4.1, is 1.

The result is reasonable since the region lies inside a rectangle of area $1 \times \pi/2 = \pi/2 \approx 1.57$ and it contains a triangle of area $(1 \times \pi/2)/2 = \pi/4 \approx 0.79$. ♦

Observation 6.4.5: *Example 1 Revisited*

Recall that antiderivatives are not unique. How would the evaluation in Example 1 change if we used $\sin(x) + 5$ as the antiderivative of $\cos(x)$? Does this give a different answer? What about other antiderivatives of $\cos(x)$?

Motivation for the Fundamental Theorem of Calculus II

Assume that f is a continuous function such that $f(x)$ is positive for x in $[a,b]$. For x in $[a,b]$, denote by $G(x)$ the area of the region under the graph of f and above the interval $[a,x]$, as shown in Figure 6.4.2(a). Note that $G(a) = 0$ since the interval $[a,a]$ has length 0.

We will compute the derivative of $G(x)$, that is,

$$G'(x) = \lim_{\Delta x \to 0} \frac{\Delta G}{\Delta x} = \lim_{\Delta x \to 0} \frac{G(x + \Delta x) - G(x)}{\Delta x}.$$

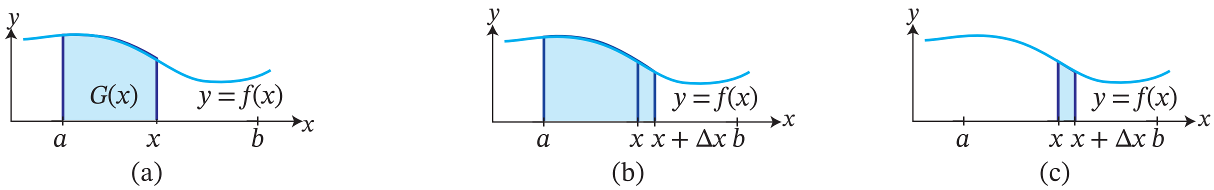


Figure 6.4.2

This is one of several occasions when we go back to the definition of the derivative as a limit. For simplicity keep Δx positive. Then $G(x+\Delta x)$ is the area under the curve $y = f(x)$ above the interval $[a, x+\Delta x]$. If Δx is small, $G(x+\Delta x)$ is only slightly larger than $G(x)$, as shown in Figure 6.4.2(b). Then $\Delta G = G(x+\Delta x) - G(x)$ is the area of the shaded strip in Figure 6.4.2(c).

When Δx is small, the narrow shaded strip above $[x, x+\Delta x]$ resembles a rectangle of base Δx and height $f(x)$ with area $f(x)\Delta x$. Therefore, it seems reasonable that when Δx is small,

$$\frac{\Delta G}{\Delta x} \approx \frac{f(x)\Delta x}{\Delta x} = f(x).$$

Thus it seems plausible that

$$\lim_{\Delta x \to 0} \frac{\Delta G}{\Delta x} = f(x)$$

or, recognizing the limit as the derivative of G,

$$G'(x) = f(x).$$

In words, the derivative of the area of the region under the graph of f and above $[a, x]$ with respect to x is the value of f at x.

In terms of definite integrals, if f is a continuous function, and $G(x) = \int_a^x f(t)\, dt$, we expect

$$\frac{d}{dx}\left(\int_a^x f(t)\, dt\right) = f(x).$$

NOTATION: The integrand is expressed using t as the variable to avoid using x to denote both an endpoint of the interval and part of the integrand.

This equation says that the derivative of the definite integral of f with respect to the right end of the interval is f evaluated at that end. This is the substance of the second part of the fundamental theorem of calculus. It tells how rapidly the definite integral changes as we change the upper limit of integration.

Theorem 6.4.6: Fundamental Theorem of Calculus II

Assume that f is continuous on the interval $[a, b]$. Define

$$G(x) = \int_a^x f(t)\, dt \qquad \textit{for all } a \le x \le b.$$

Then G is differentiable on $[a, b]$ and its derivative is f:

$$G'(x) = \frac{d}{dx}\int_a^x f(t)\, dt = f(x).$$

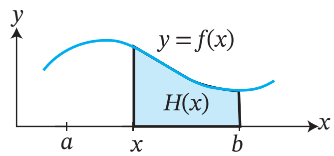


Figure 6.4.3

As a consequence of FTC II, every continuous function is the derivative of some function.

There is a similar theorem for definite integrals with a lower limit that is a variable: $H(x) = \int_x^b f(t)\,dt$. For such a function, its derivative is $H'(x) = -f(x)$. Figure 6.4.3 shows why there is a minus sign: the area shrinks as x increases. (See Exercise 69 for a complete proof of this fact.)

EXAMPLE 2. Give an example of an antiderivative of $\dfrac{\sin(x)}{x}$.

SOLUTION There are many antiderivatives of $\sin(x)/x$. Any two differ by a constant. The general shape of these antiderivatives can be seen in the slope field for $y' = \sin(x)/x$ shown in Figure 6.4.4(a). But, there is no simple expression for these antiderivatives.

RECALL: Slope fields were introduced in Section 3.6.

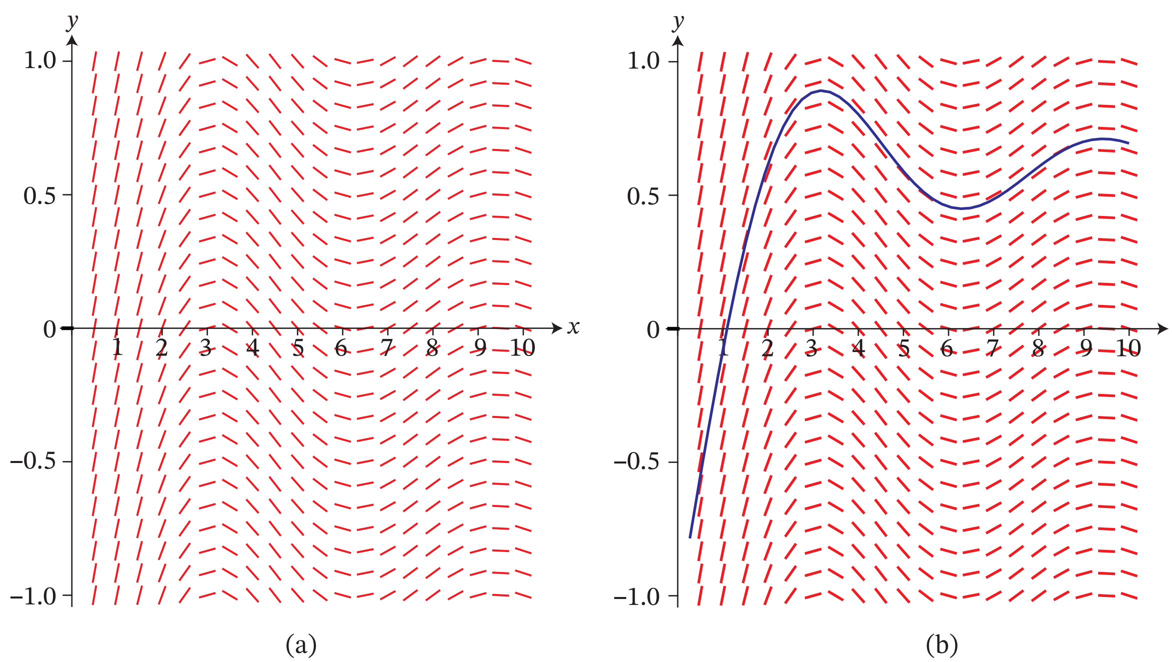


Figure 6.4.4

Let $G(x) = \int_1^x \sin(t)/t\,dt$. By FTC II, $G'(x) = \sin(x)/x$, so $G(x)$ is an antiderivative of $\sin(x)/x$. In fact, it is the antiderivative of $\sin(x)/x$ that passes through the point $(1,0)$. The graph of $y = G(x)$ is shown in Figure 6.4.4(b). ◆

One might expect the answer in Example 2 to be an explicit formula expressed in terms of the functions discussed in Chapters 2 and 3. We repeat the observation first made in Section 3.6 that the derivative of every elementary function is an elementary function and that in the nineteenth century Joseph Liouville proved that there are elementary functions that do not have elementary antiderivatives. Nobody will ever find an explicit formula in terms of elementary functions for an antiderivative of $\frac{\sin(x)}{x}$. (The proof of Liouville's result is reserved for a graduate course.)

EXAMPLE 3. Give an example of an antiderivative of $\dfrac{\sin(\sqrt{x})}{\sqrt{x}}$.

SOLUTION This integrand appears more complicated than $\sin(x)/x$, yet, unlike $\sin(x)/x$, it has an elementary antiderivative, namely $-2\cos(\sqrt{x})$. To check, we differentiate $y = -2\cos(\sqrt{x})$. To setup to use the chain rule, we write $y = -2\cos(u)$ where $u = \sqrt{x}$. Therefore,

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = -2(-\sin(u))\frac{1}{2\sqrt{x}} = \frac{\sin(\sqrt{x})}{\sqrt{x}}.$$

◆

Because the antiderivatives of $\sin(\sqrt{x})/\sqrt{x}$ are elementary functions, $\int_1^2 \sin(\sqrt{x})/\sqrt{x}\,dx$ would be easy to calculate.

Any antiderivative of e^x is of the form $e^x + C$, an elementary function. However, no antiderivative of e^{-x^2} is elementary. Statisticians define the **error function** to be $\operatorname{erf}(x) = (2/\sqrt{\pi})\int_0^x e^{-t^2}\,dt$. Except that $\operatorname{erf}(0) = 0$, there is no easy way to evaluate $\operatorname{erf}(x)$. Approximate values of special functions such as the error function can be obtained from mathematical software and even a few calculators.

Net Area

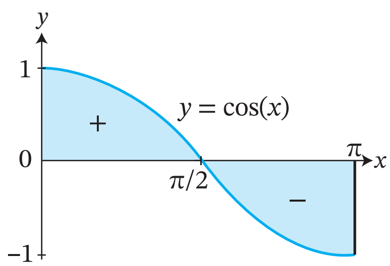


Figure 6.4.5

When we evaluate $\int_0^\pi \cos(x)\,dx$, we obtain $\sin(\pi) - \sin(0) = 0 - 0 = 0$. What does this say about areas? Inspection of Figure 6.4.5 shows what is happening.

For x in $[\pi/2, \pi]$, $\cos(x)$ is negative and the curve $y = \cos(x)$ lies below the x-axis. If we interpret the corresponding area as negative, then we see that it cancels the area from 0 to $\pi/2$. When we say that $\int_a^b f(x)\,dx$ represents the area under the curve $y = f(x)$, we mean that it represents the area between the curve and the x-axis, with area below the x-axis taken as negative. $\int_a^b f(x)\,dx$ is the **net area** between $y = f(x)$ and the x-axis for x in $[a, b]$. It can be positive, zero, or negative.

EXAMPLE 4. Evaluate $\displaystyle\int_1^2 \frac{1}{x^2}\,dx$ by the Fundamental Theorem of Calculus I.

SOLUTION In order to apply FTC I we need an antiderivative of $1/x^2$. In Section 6.3 we saw that

$$\int x^a\,dx = \frac{1}{a+1}x^{a+1} + C \qquad (a \neq -1).$$

With $a = -2$ we have

$$\int \frac{1}{x^2}\,dx = \int x^{-2}\,dx = \frac{1}{(-2)+1}x^{(-2)+1} + C = \frac{1}{-1}x^{-1} + C = \frac{-1}{x} + C.$$

Then, by FTC I

$$\int_1^2 \frac{1}{x^2}\,dx = \left(\frac{-1}{x} + C\right)\Bigg|_1^2 = \left(\frac{-1}{2} + C\right) - \left(\frac{-1}{1} + C\right) = \frac{-1}{2} - (-1) = \frac{1}{2}.$$

The C's cancel! ◆

Observation 6.4.7: ***The*** $+C$ ***is Not Needed When Applying FTC I***

When using FTC I, if the antiderivative includes $+C$, it cancels when computing $F(b) - F(a)$. For this reason the constant of integration is generally omitted when finding an antiderivative to evaluate a definite integral.

FTC I asserts that

$$\underbrace{\int_1^2 \frac{1}{x^2}\,dx}_{\text{The definite integral is a limit of sums}} = \underbrace{\left(\int \frac{1}{x^2}\,dx\right)\Bigg|_1^2}_{\text{The difference when an antiderivative is evaluated at 2 and at 1}}$$

The symbols on the right and left of the equal sign are so similar that it is tempting to think that the equation is obvious or says nothing whatsoever, but that is not so. It is an instance of FTC I.

Some Terms and Notation

NOTATION: Integrals can be written in many equivalent ways. For instance, we write $\int (1/x^2)\,dx$ as $\int dx/x^2$, merging the 1 with the dx.

The related processes of computing $\int_a^b f(x)\,dx$ and of finding an antiderivative $\int f(x)\,dx$ are both called **integrating** $f(x)$. The term refers to two separate processes: computing a number $\int_a^b f(x)\,dx$ or finding a function $\int f(x)\,dx$.

In practice, both FTC I and FTC II are called the Fundamental Theorem of Calculus. The context makes it clear which one is meant.

Proofs of the Two Parts of the FTC

We now prove both parts of the fundamental theorem of calculus without referring to motion, area, or examples. The proofs use only the mathematics of functions and limits. As previously noted, the proof of FTC I makes use of FTC II. For this reason we prove FTC II first.

The second fundamental theorem of calculus asserts that the derivative of $G(x) = \int_a^x f(t)\,dt$ is $f(x)$. We previously presented an argument using areas of regions. But definite integrals are defined in terms of approximating sums, not areas, so it was not a proof. We now give a proof that uses only properties of definite integrals.

Proof of FTC II (Theorem 6.4.6)

We wish to show that $G'(x) = f(x)$. To do this we make use of the definition of the derivative of a function.

We have the following chain of equalities:

$$\begin{aligned}
G'(x) &= \lim_{\Delta x\to 0} \frac{G(x+\Delta x)-G(x)}{\Delta x} && (\text{ definition of derivative })\\
&= \lim_{\Delta x\to 0} \frac{1}{\Delta x}\left(\int_a^{x+\Delta x} f(t)\,dt - \int_a^x f(t)\,dt\right) && (\text{ definition of } G\text{ })\\
&= \lim_{\Delta x\to 0} \frac{1}{\Delta x}\left(\int_a^{x} f(t)\,dt + \int_x^{x+\Delta x} f(t)\,dt - \int_a^x f(t)\,dt\right) && (\text{ Property 6 in Section 6.3 })\\
&= \lim_{\Delta x\to 0} \frac{1}{\Delta x}\int_x^{x+\Delta x} f(t)\,dt && (\text{ canceling like terms in numerator })\\
&= \lim_{\Delta x\to 0} \frac{f(c)\Delta x}{\Delta x} && (\text{ MVT for definite integrals: } c \text{ is between } x \text{ and } x+\Delta x\text{ })\\
&= \lim_{\Delta x\to 0} f(c) && (\text{ canceling } \Delta x\text{, which is not zero })\\
&= f(x) && (\text{ continuity of } f; c\to x \text{ as } \Delta x\to 0\text{ }).
\end{aligned}$$

Hence, we discover that $G'(x) = f(x)$, which is the conclusion of FTC II. ●

A similar argument, with the additional step that $\int_x^b f(t)\, dt = -\int_b^x f(t)\, dt$, confirms that

$$\frac{d}{dx}\int_x^b f(t)\, dt = -f(x).$$

The first fundamental theorem of calculus asserts that if $F' = f$, then $\int_a^b f(x)\, dx = F(b) - F(a)$. Recall that we saw that this is true in a few specific cases: when $f(x) = c$, $f(x) = x$, $f(x) = x^2$, and $f(x) = 2^x$. We now prove the theorem in the general case when the integrand is any continuous function on the interval $[a, b]$.

Proof of FTC I (Theorem 6.4.4)

We assume that f is a continuous function on $[a, b]$ and that $F' = f$. We wish to show that $F(b) - F(a) = \int_a^b f(x)\, dx$.

Define $G(x)$ to be $\int_a^x f(t)\, dt$. By FTC II, G is an antiderivative of f. Since F and G are both antiderivatives of f, they differ by a constant, say C. That is, $F(x) = G(x) + C$. Then, starting with the difference $F(b) - F(a)$, we find

$$\begin{aligned}
F(b) - F(a) &= (G(b) + C) - (G(a) + C) \\
&= G(b) - G(a) && (\text{ } C\text{'s cancel }) \\
&= \int_a^b f(t)\, dt - \int_a^a f(t)\, dt && (\text{ definition of } G\text{ }). \\
&= \int_a^b f(t)\, dt && (\text{ } \textstyle\int_a^a f(t)\, dt = 0\text{ }).
\end{aligned}$$

This confirms that FTC I is true for any continuous function on an interval $[a, b]$, the proof of FTC I is complete. ●

Summary

This section linked the two basic ideas of calculus, the derivative (more precisely, the antiderivative) and the definite integral.

FTC I says that if you can find a formula for an antiderivative F of f, then you can evaluate the definite integral $\int_a^b f(x)\, dx$:

$$\int_a^b f(x)\, dx = F(b) - F(a).$$

FTC II says that if f is continuous then it has an antiderivative, namely $G(x) = \int_a^x f(t)\, dt$. The function G might not be an elementary function, but $G'(x) = f(x)$. Regardless of whether you can find an explicit representation for the antiderivatives of a function, their general shape can be seen in the slope field for $dy/dx = f(x)$.

EXERCISES for Section 6.4

1. State, using mathematical notation, FTC I.

2. State, using mathematical notation, FTC II.

3. Using only words, no mathematical symbols, state FTC I.

4. Using only words, no mathematical symbols, state FTC I.

5. Evaluate (a) $x^3\big|_1^2$, (b) $x^2\big|_{-1}^2$, and (c) $\cos(x)\big|_0^\pi$.

6. Evaluate (a) $(x + \sec(x))\big|_0^{\pi/4}$, (b) $\left.\frac{1}{x}\right|_2^3$, and (c) $\sqrt{x-1}\,\big|_5^{10}$.

In Exercises 7 to 20 use FTC I to evaluate the definite integrals.

7. $\int_1^2 5x^3\,dx$

8. $\int_{-1}^3 2x^4\,dx$

9. $\int_1^4 (x+5x^2)\,dx$

10. $\int_1^2 (6x-3x^2)\,dx$

11. $\int_{\pi/6}^{\pi/3} 5\cos(x)\,dx$

12. $\int_{\pi/4}^{3\pi/4} 3\sin(x)\,dx$

13. $\int_0^{\pi/2} \sin(2x)\,dx$

14. $\int_0^{\pi/6} \cos(3x)\,dx$

15. $\int_4^9 5\sqrt{x}\,dx$

16. $\int_1^9 \frac{1}{\sqrt{x}}\,dx$

17. $\int_1^8 \sqrt[3]{x^2}\,dx$

18. $\int_2^4 \frac{4}{x^3}\,dx$

19. $\int_0^1 \frac{dx}{1+x^2}$

20. $\int_{1/4}^{1/2} \frac{dx}{\sqrt{1-x^2}}$

In Exercises 21 to 26 find the average value of the given function over the given interval.

21. x^2; $[3,5]$

22. x^4, $[1,2]$

23. $\sin(x)$, $[0,\pi]$

24. $\cos(x)$, $\left[0,\frac{\pi}{2}\right]$

25. $\sec^2(x)$, $\left[\frac{\pi}{6},\frac{\pi}{4}\right]$

26. $\sec(2x)\tan(2x)$, $\left[\frac{\pi}{8},\frac{\pi}{6}\right]$

In Exercises 27 to 34 find the indicated quantity.

27. The area of the region under the curve $y=3x^2$ and above $[1,4]$.

28. The area of the region under the curve $y=\frac{1}{x^2}$ and above $[2,3]$.

29. The area of the region under the curve $y=6x^4$ and above $[-1,1]$.

30. The area of the region under the curve $y=\sqrt{x}$ and above $[25,36]$.

31. The distance an object travels from time $t=1$ s to time $t=2$ s, if its velocity at time t s is t^5 feet per second.

32. The distance an object travels from time $t=1$ s to time $t=8$ s, if its velocity at time t s is $7\sqrt[3]{t}$ feet per second.

33. The volume of a solid located between the planes at $x=1$ and at $x=5$, if the cross-sectional area of the intersection of the solid with the plane perpendicular to the x-axis through the point $(x,0,0)$ is $6x^3$ square centimeters.

34. The volume of a solid located between the planes at $x=1$ and at $x=5$, if the cross-sectional area of the intersection of the solid with the plane perpendicular to the x-axis through the point $(x,0,0)$ is $\frac{1}{x^3}$ square centimeters.

---

35. Let f be a continuous function. Estimate $f(7)$, if $\int_5^7 f(x)dx=20.4$ and $\int_5^{7.05} f(x)dx=20.53$.

36. Determine if each expression is a function or a number: (a) $\int x^2\,dx$, (b) $\left.\int x^2\,dx\right|_1^3$, and (c) $\int_1^3 x^2\,dx$.

37. (a) Determine which of the following numbers is defined as a limit of sums. $\left.\int x^2\,dx\right|_1^2$ and $\int_1^2 x^2\,dx$. (b) How is the other number defined? (c) Why are they equal?

38. There is no elementary antiderivative of $\sin(x^2)$. Does $\sin(x^2)$ have an antiderivative? Explain.

39. Determine whether each statement is true or false. Explain each answer. (a) Every elementary function has an elementary derivative. (b) Every elementary function has an elementary antiderivative.

40. (a) Draw the slope field for $\dfrac{dy}{dx} = \dfrac{e^{-x}}{x}$ for $x > 0$.

(b) Use (a) to sketch the graph of an antiderivative of $\dfrac{e^{-x}}{x}$.

(c) On the slope field drawn in (a), sketch the graph of $f(x) = \displaystyle\int_1^x \frac{e^{-t}}{t}\,dt$.

Exercises 41 and 42 illustrate why FTC I can be applied using any antiderivative of the integrand.

41. Evaluate the definite integral $\displaystyle\int_a^b x\,dx$ using the following antiderivatives of $f(x) = x$.

(a) $F(x) = \dfrac{1}{2}x^2 + 1$ (b) $F(x) = \dfrac{1}{2}x^2 - 3$ (c) $F(x) = \dfrac{1}{2}x^2 + C$

42. Evaluate the definite integral $\displaystyle\int_a^b 2^x\,dx$ using the following antiderivatives of $f(x) = 2^x$.

(a) $F(x) = \dfrac{1}{\ln(2)}2^x + 11$ (b) $F(x) = \dfrac{1}{\ln(2)}2^x - 7$ (c) $F(x) = \dfrac{1}{\ln(2)}2^x + C$

---

43. Let $F(x) = \displaystyle\int_0^x e^{t^2}\,dt$ for all x.

(a) Does the graph of $F(x)$ have inflection points? If so, find them. (b) Sketch the graph of $F(x)$.

44. Area was used in Section 6.2 to develop $\displaystyle\int_a^b x\,dx = \frac{b^2}{2} - \frac{a^2}{2}$ when $0 < a < b$. To see that this result is true for all nonzero values of a and b (with $b > a$) we consider the following two additional cases:

(a) If $a < b < 0$, interpret the definite integral in terms of signed area.

(b) If $a < 0 < b$, divide the interval $[a, b]$ into two pieces and work with signed areas.

45. Find $\dfrac{dy}{dx}$, (a) if $y = \displaystyle\int \sin(x^2)\,dx$, (b) if $y = 3x + \displaystyle\int_{-2}^{3} \sin(x^2)\,dx$, and (c) if $y = \displaystyle\int_{-2}^{x} \sin(t^2)\,dt$.

In Exercises 46 to 49 differentiate the given functions.

46. (a) $\displaystyle\int_1^x t^4\,dt$ and (b) $\displaystyle\int_x^1 t^4\,dt$.

47. (a) $\displaystyle\int_1^x \sqrt[3]{1+\sin(t)}\,dt$ and (b) $\displaystyle\int_1^{x^2} \sqrt[3]{1+\sin(t)}\,dt$.

48. (a) $\displaystyle\int_{-1}^x 3^{-t}\,dt$ and (b) $\displaystyle\int_{-1}^{x^3} 3^{-t}\,dt$.

49. $\displaystyle\int_{2x}^{3x} t\tan(t)\,dt$ for $\dfrac{-\pi}{6} < x < \dfrac{\pi}{6}$

---

50. Figure 6.4.6(a) shows the graph of a function $f(x)$ for x in $[1,3]$. Let $G(x) = \displaystyle\int_1^x f(t)\,dt$. Graph $y = G(x)$ for x in $[1,3]$ as well as you can. Explain your reasoning.

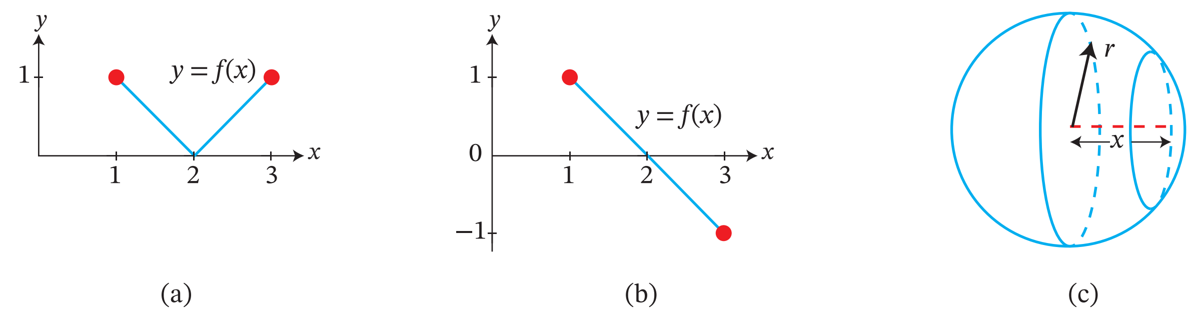


Figure 6.4.6

51. Figure 6.4.6(b) shows the graph of a function $f(x)$ for x in $[1,3]$. Let $G(x) = \int_1^x f(t)\,dt$. Graph $y = G(x)$ for x in $[1,3]$ as well as you can. Explain your reasoning.

52. A plane at a distance x from the center of the ball of radius r, $0 \le x \le r$, meets the ball in a disk. (See Figure 6.4.6(c).)

(a) Show that the radius of the disk is $\sqrt{r^2 - x^2}$.
(b) Show that the area of the disk is $\pi r^2 - \pi x^2$.
(c) Using the FTC, find the volume of the ball.

53. Assume that $v(t)$ is the velocity at time t of an object moving on a straight line. It may be positive or negative.

(a) What is the physical meaning of $\int_a^b v(t)\,dt$? Explain.
(b) What is the physical meaning of the slope of the graph of $y = v(t)$? Explain.
(c) What is the physical meaning of $\int_a^b |v(t)|\,dt$? Explain.

54. Let f be a continuous function. Show that $\frac{d}{dx}\int_x^b f(t)\,dt = -f(x)$

(a) by using the definition of derivative as a limit
(b) by using properties of the definite integral and FTC II.

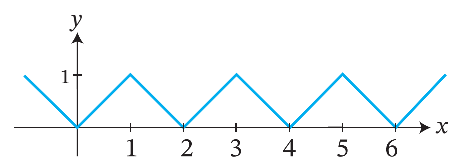


Figure 6.4.7

55. Find the derivative (with respect to x) of $\cos(t^2)\Big|_{2x}^{3x}$.

56. Figure 6.4.7 shows the graph of a function f. Let $A(x)$ be the area under the graph of f and above the interval $[1,x]$. (a) Find $A(1)$, $A(2)$, and $A(3)$. (b) Find $A'(1)$, $A'(2)$, and $A'(3)$.

57. Let R be a function with continuous second derivative R''. Assume $R(1) = 2$, $R'(1) = 6$, $R(3) = 5$, and $R'(3) = 8$. Evaluate $\int_1^3 R''(x)\,dx$.

58. If $f(x) = \int_{-1}^x \sin^3\left(e^{t^2}\right)dt$, find $f'(1)$.

59. If $\int_1^x f(t)dt = \sin^3(5x)$, find $f'(3)$.

60. Differentiate $\int_{x^2}^{x^3} e^{t^2}\,dt$.

61. Differentiate $\int_{x^2}^{5} \sin^{10}(3t)dt$.

62. (a) If $\int_x^{x+4} g(t)\,dt = 5$ for all x, what is known about the graph of the function g? (b) Can you construct $g(x)$?

63. Jane and Sam are, once again, talking about what they read in this section.

JANE: $\int_a^b f(x)\,dx$ is a number.
SAM: But if I treat b as a variable, then it is a function.
JANE: How can it be both a number and a function?
SAM: It depends on what "it" means.
JANE: You can't get out of this so easily.

Which student is correct? Explain your answer.

64. Show that if we knew that every continuous function has an antiderivative, then FTC I would imply FTC II.

65. (a) Show that for a constant function, $f(x) = c$, the average value of f over $[a,b]$ is the same as the value of the function at the midpoint of the interval $[a,b]$.

(b) Give an example of a nonconstant function f such that for any interval $[a,b]$, $\frac{1}{b-a}\int_a^b f(t)\,dt = f\left(\frac{a+b}{2}\right)$.

(c) Show that if a continuous function f on $(-\infty,\infty)$ satisfies the equation in (b), it is differentiable.

(d) Find all continuous functions that satisfy the equation in (b).

66. Find all continuous functions, f, such that their average over $[0,t]$ equals $f(t)$ for all t.

67. Give a geometric explanation of the following properties of definite integrals:

(a) If f is an even function, then $\int_{-a}^{a} f(t)\,dt = 2\int_0^a f(t)\,dt$. (b) If f is an odd function, then $\int_{-a}^{a} f(t)\,dt = 0$.

(c) If f is a periodic function with period p, then, for positive integers m and n, $\int_{mp}^{np} f(t)\,dt = (n-m)\int_0^p f(t)\,dt$.

68. How often should a machine be overhauled? This depends on the rate $f(t)$ at which it depreciates ($f(t) > 0$) and the cost A of an overhaul. Denote the time between overhauls by T.

(a) Explain why you would like to minimize $g(T) = \frac{1}{T}\left(A + \int_0^T f(t)\,dt\right)$.

(b) Find $\frac{dg}{dT}$.

(c) Show that if $\frac{dg}{dT} = 0$, then $f(T) = g(T)$.

(d) Is this reasonable? Explain.

69. Let $f(x)$ be a continuous function with only positive values. Define $H(x) = \int_x^b f(t)\,dt$ for all $a \le x \le b$.

(a) Interpreting the definite integral as a region's area, draw the regions whose areas are $H(x)$ and $H(x+\Delta x)$.

(b) Is $H(x+\Delta x) - H(x)$ positive or negative? (Assume Δx is positive.)

(c) Draw the region whose area is related to $H(x+\Delta x) - H(x)$.

(d) When Δx is small, estimate $H(x+\Delta x) - H(x)$ in terms of f.

(e) Use (d) to evaluate the derivative $H'(x)$: $\frac{dH}{dx} = \lim_{\Delta x \to 0} \frac{H(x+\Delta x) - H(x)}{\Delta x}$.

70. Give an example of a function f such that $f(4) = 0$ and $f'(x) = \sqrt[3]{1+x^2}$.

71. The function $\dfrac{e^x}{x}$ does not have an elementary antiderivative. Show that its reciprocal, $\dfrac{x}{e^x}$, does have an elementary antiderivative.

72. Use FTC II to explain why, if u and v are differentiable functions,

(a) $\displaystyle \frac{d}{dx}\int_a^{v(x)} f(t)\,dt = f(v(x))v'(x)$

(b) $\displaystyle \frac{d}{dx}\int_{u(x)}^{b} f(t)\,dt = -f(u(x))u'(x)$

(c) $\displaystyle \frac{d}{dx}\int_{u(x)}^{v(x)} f(t)\,dt = f(v(x))v'(x) - f(u(x))u'(x)$

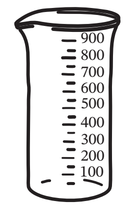


Figure 6.4.8

73. Say that you want to find the area of a plane cross section of a rock. One way to find it is by sawing the rock in two and measuring the area directly. Suppose you do not want to ruin the rock. Suppose you have a measuring glass, as shown in Figure 6.4.8, that gives excellent volume measurements in cubic centimeters. How could you use it to get a good estimate of the cross-sectional area?

74. Here is another conversation between Sam and Jane. Find Sam's error(s).

SAM: It's obvious that if a function $f(x)$ is increasing, then so is its average value, $g(b) = \dfrac{1}{b}\displaystyle\int_0^b f(x)\,dx$. Just look at the graph of $f(x)$.

JANE: I'm very leery of any claim that starts with "it's obvious".

SAM: In that case, I'll use calculus. The derivative of $g(b)$ is $g'(b) = \dfrac{1}{b^2}\left(bf(b) - \displaystyle\int_0^b f(x)\,dx\right)$. Since both factors are positive, $g'(b)$ is positive. By the way, if the average value increases, so must $f(x)$. After all, the factor in parenthese, $bf(b) - \displaystyle\int_0^b f(x)\,dx$, is not negative. So $bf(b) \geq \displaystyle\int_0^b f(x)\,dx$. Differentiate both sides of this to get $f(b) + bf'(b) \geq f(b)$. The $f(b)$'s cancel, showing that $bf'(b) \geq 0$. Since b is positive, $f'(b)$ is nonnegative. So $f(x)$ is not decreasing.

JANE: But my friend, Dean Hickerson, who edited the exercises, constructed a function that is not increasing even though its average value does increase.

SAM: So what is this impossible function?

JANE: It's just a polynomial, $f(x) = 2x^3 - 9x^2 + 12x$, whose derivative is $f'(x) = 6x^2 - 18x + 12 = 6(x-1)(x-2)$. Thus $f(x)$ decreases for x in $[1, 2]$. However, the derivative of the average value function is $g'(b) = \dfrac{3}{2}(b-2)^2$, which is never negative.

SAM: Did you check that?

JANE: Yes, but you better check it yourself.

SAM: I'm sure I'm right.

JANE: Everyone should recite this little reminder every morning. "I could be wrong, I could be wrong, I could be wrong."

75. Find all continuous functions, f, whose average value on the interval $[0, b]$ is a nondecreasing function of b? (Assume $b > 0$.)

6.5 Estimating a Definite Integral

It is easy to evaluate $\int_0^1 x^2\sqrt{1+x^3}\,dx$ by the fundamental theorem of calculus. The integrand has an elementary antiderivative, as can be checked by differentiating $\frac{2}{9}(1+x^3)^{3/2}$. However, an antiderivative of $\sqrt{1+x^3}$ is not elementary, so $\int_0^1 \sqrt{1+x^3}\,dx$ cannot be evaluated as easily. Instead we estimate its value. This section describes four ways to do this.

Approximation by Rectangles

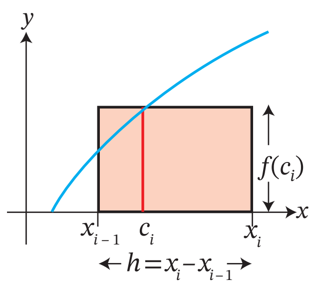


Figure 6.5.1

The definite integral $\int_a^b f(x)\,dx$ is, by definition, a limit of sums of the form

$$\sum_{i=1}^{n} f(c_i)(x_i - x_{i-1}).$$

Any such sum is an estimate of $\int_a^b f(x)\,dx$.

The area of a rectangle gives a local estimate of the area under the graph of $y = f(x)$ above the interval $[x_{i-1}, x_i]$. See Figure 6.5.1. The sum of the areas of the rectangles is an estimate of the area under the curve.

To use rectangles to estimate $\int_a^b f(x)\,dx$, divide the interval $[a, b]$ into n sections of equal length by the $n+1$ numbers $a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$.

The width of each section is $h = (b-a)/n$. Then choose a sampling number c_i in the i^{th} section, $i = 1, 2, \dots, n$ and form the Riemann sum $\sum_{i=1}^{n} f(c_i)h$. By the definition of the definite integral, this sum is an estimate of the definite integral.

Choosing the sections to have the same length is not necessary, but simplifies the arithmetic.

Denoting $f(x_i)$ by y_i, and using the left endpoint x_{i-1} of each interval $[x_{i-1}, x_i]$ as the sampling number, yields the **left endpoint rectangular estimate**.

Formula 6.5.1: *Left Endpoint Rectangular Estimate for* $\int_a^b f(x)\,dx$

The left endpoint rectangular estimate for $\int_a^b f(x)\,dx$ with n equal width rectangles is

$$\int_a^b f(x)\,dx \approx h(y_0 + y_1 + y_2 + \cdots + y_{n-2} + y_{n-1}) \qquad \text{where } h = \frac{b-a}{n}$$

If the right endpoints are used instead of the left endpoints, we have the **right endpoint rectangular estimate**:

Formula 6.5.2: *Right Endpoint Rectangular Estimate for* $\int_a^b f(x)\,dx$

The right endpoint rectangular estimate for $\int_a^b f(x)\,dx$ with n equal width rectangles is

$$\int_a^b f(x)\,dx \approx h(y_1 + y_2 + \cdots + y_{n-1} + y_n) \qquad \text{where } h = \frac{b-a}{n}$$

We will illustrate this and other ways to estimate a definite integral by estimating $\int_0^1 \frac{dx}{1+x^2}$. We chose it because it can be computed by the FTC:

$$\int_0^1 \frac{dx}{1+x^2} = \arctan(x)|_0^1 = \arctan(1) - \arctan(0) = \frac{\pi}{4} \approx 0.785398.$$

That enables us to judge the accuracy of each method.

EXAMPLE 1. Use the left endpoint rectangular estimate with four equal width rectangles to estimate $\int_0^1 \frac{dx}{1+x^2}$. Repeat using the right endpoint rectangular estimate.

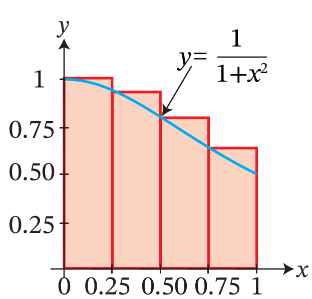
SOLUTION Since the length of $[0,1]$ is 1, each of the four sections of equal length has length $1/4 = 0.25$. Using the height at the left endpoint of each segment, the sum of the areas of the rectangles is

$$\begin{aligned}\frac{1}{1+0^2}\cdot\frac{1}{4} + \frac{1}{1+\left(\frac{1}{4}\right)^2}\cdot\frac{1}{4} + \frac{1}{1+\left(\frac{2}{4}\right)^2}\cdot\frac{1}{4} + \frac{1}{1+\left(\frac{3}{4}\right)^2}\cdot\frac{1}{4} &= \frac{1}{4}\left(1+\frac{16}{17}+\frac{16}{20}+\frac{16}{25}\right)\\ &\approx \frac{1}{4}(1.00000+0.94118+0.80000+0.64000)\\ &= \frac{1}{4}(3.38118) \approx 0.84529.\end{aligned}$$

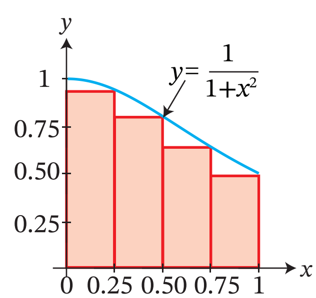
The corresponding estimate using the height at the right endpoint of each segment is

$$\begin{aligned}\frac{1}{1+\left(\frac{1}{4}\right)^2}\cdot\frac{1}{4} + \frac{1}{1+\left(\frac{2}{4}\right)^2}\cdot\frac{1}{4} + \frac{1}{1+\left(\frac{3}{4}\right)^2}\cdot\frac{1}{4} + \frac{1}{1+\left(\frac{4}{4}\right)^2}\cdot\frac{1}{4} &= \frac{1}{4}\left(\frac{16}{17}+\frac{16}{20}+\frac{16}{25}+\frac{1}{2}\right)\\ &\approx \frac{1}{4}(0.94118+0.80000+0.64000+0.50000)\\ &= \frac{1}{4}(2.88118) \approx 0.72029.\end{aligned}$$

◆



(a)



(b)

Figure 6.5.2

As Figure 6.5.2 shows, the left endpoint rectangle estimate is an overestimate and the right endpoint rectangle estimate is an underestimate. The exact value is $\arctan(1) = \pi/4 \approx 0.785398$. The error in each estimate is about 0.06 with only four rectangles.

Approximation by Trapezoids

Trapezoids can also be used to find local estimates of the area under the graph of $y = f(x)$ above the interval $[x_{i-1}, x_i]$. The idea is shown in Figure 6.5.3(a).

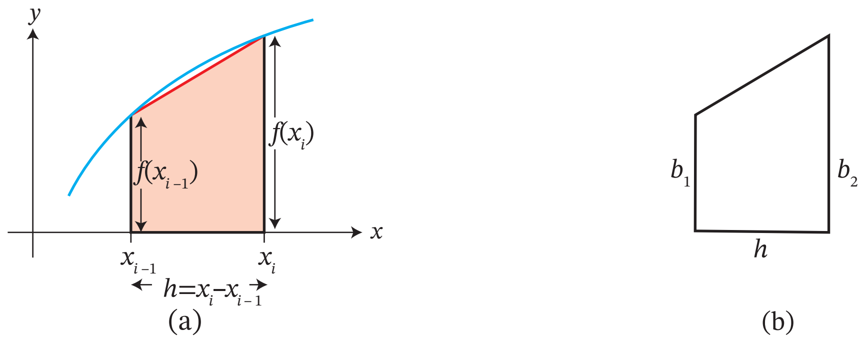


Figure 6.5.3

The area, A, of a trapezoid with base width h and side lengths b_1 and b_2 is the product of the base width and the average of the two side lengths: $A = (b_1 + b_2)h/2$. (See Figure 6.5.3(b).)

The formula for the trapezoidal estimate of $\int_a^b f(x)\,dx$ follows from an argument like the ones for the rectangular estimates.

Let n be a positive integer. Divide the interval $[a, b]$ into n sections of equal length $h = (b-a)/n$ with

$$x_0 = a,\ x_1 = a+h,\ x_2 = a+2h,\ \ldots,\ x_n = a+nh = b.$$

Denote $f(x_i)$ by y_i. The local estimate of the area under $y = f(x)$ and above $[x_{i-1}, x_i]$ is

$$\frac{1}{2}(y_{i-1} + y_i)h.$$

Summing the n local estimates of area gives the formula for the trapezoidal estimate of $\int_a^b f(x)\,dx$:

$$\frac{y_0+y_1}{2}\cdot h + \frac{y_1+y_2}{2}\cdot h + \cdots + \frac{y_{n-1}+y_n}{2}\cdot h$$

Factoring $h/2$ and collecting like terms gives us the **trapezoidal method**.

Formula 6.5.3: *Trapezoidal Estimate for* $\int\limits_a^b f(x)\,dx$

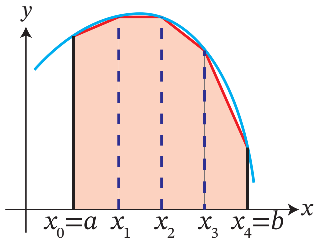
The trapezoidal estimate for $\int_a^b f(x)\,dx$ with n equal width trapezoids is

$$\int\limits_a^b f(x)\,dx \approx \frac{h}{2}\left(y_0 + 2y_1 + 2y_2 + \cdots + 2y_{n-1} + y_n\right) \qquad \text{where } h = \frac{b-a}{n} \tag{6.5.1}$$

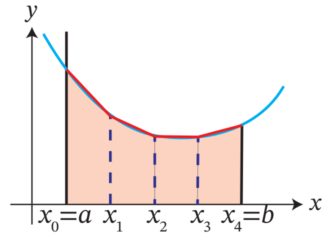
There are n sections of width $h = (b-a)/n$, each corresponding to one trapezoid. The function is evaluated at $n+1$ points, including both ends of the interval $[a, b]$.

The values y_0 and y_n have coefficient 1 while the other y_i's have coefficient 2. This is due to the double counting of the edges common to two trapezoids.

If $f(x)$ is a polynomial of the form $A + Bx$, its graph is a straight line. The top edge of each approximating trapezoid coincides with the graph. The approximation (6.5.1) in this case gives the exact value of $\int_a^b f(x)\,dx$. There is no error.



(a)



(b)

Figure 6.5.4

Figures 6.5.4(a) and (b) illustrate the trapezoidal estimate for the case $n = 4$ for concave down and concave up functions.

In Figure 6.5.4(a) the function is concave down and the trapezoidal estimate underestimates $\int_a^b f(x)\,dx$. When the curve is concave up the trapezoids overestimate, as shown in Figure 6.5.4(b). In both cases the trapezoids appear to give a better approximation of $\int_a^b f(x)\,dx$ than the same number of rectangles. We expect the trapezoidal method to provide a better estimate of a definite integral than those we obtain using rectangles.

EXAMPLE 2. Use the trapezoidal method with $n = 4$ to estimate $\displaystyle\int_0^1 \frac{dx}{1+x^2}$.

SOLUTION In this case $a = 0$, $b = 1$, and $n = 4$, so $h = (1-0)/4 = 1/4$. The four trapezoids are shown in Figure 6.5.5. The trapezoidal estimate is

$$\frac{h}{2}\left(f(0) + 2f\left(\frac{1}{4}\right) + 2f\left(\frac{2}{4}\right) + 2f\left(\frac{3}{4}\right) + f(1)\right).$$

Now, $h/2 = (1/4)/2 = 1/8$. To compute the sum of the five terms involving values of $f(x) = 1/(1+x^2)$, make a list as shown in Table 6.5.1.

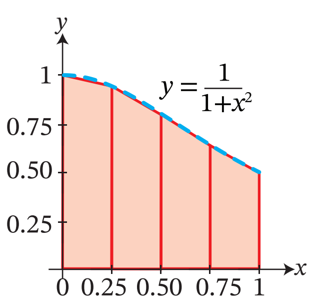


Figure 6.5.5

x_i	$f(x_i)$	coefficient	summand	decimal form
0	$\frac{1}{1+0^2}$	1	$1\cdot\frac{1}{1+0}$	1.00000
$\frac{1}{4}$	$\frac{1}{1+\left(\frac{1}{4}\right)^2}$	2	$2\cdot\frac{1}{1+\frac{1}{16}}$	1.88235
$\frac{2}{4}$	$\frac{1}{1+\left(\frac{2}{4}\right)^2}$	2	$2\cdot\frac{1}{1+\frac{4}{16}}$	1.60000
$\frac{3}{4}$	$\frac{1}{1+\left(\frac{3}{4}\right)^2}$	2	$2\cdot\frac{1}{1+\frac{9}{16}}$	1.28000
$\frac{4}{4}$	$\frac{1}{1+\left(\frac{4}{4}\right)^2}$	1	$1\cdot\frac{1}{1+\frac{16}{16}}$	0.50000

Table 6.5.1

The resulting trapezoidal sum is

$$\frac{1}{8}(1.00000 + 1.88235 + 1.60000 + 1.28000 + 0.50000) = \frac{1}{8}(6.26236) = 0.78279.$$

Thus

$$\int_0^1 \frac{dx}{1+x^2} \approx 0.7828.$$

◆

The trapezoid estimate differs from the definite integral by about 0.0026, which is much smaller than 0.06, the error in the rectangular methods.

Comparison of Rectangular and Trapezoidal Estimates

If we divide out the 2 in the trapezoidal estimate, it takes the form

$$h\left(\frac{y_0}{2} + y_1 + y_2 + \cdots + y_{n-1} + \frac{y_n}{2}\right). \tag{6.5.2}$$

which looks much like the rectangular estimate. It has $n+1$ summands, while the rectangular estimate has only n summands. However, if $f(a)$ happens to equal $f(b)$, that is, if $y_0 = y_n$, then (6.5.2) can be written either as $h(y_0 + y_1 + y_2 + \cdots + y_{n-1})$ (the left endpoint rectangular estimate) or as $h(y_1 + y_2 + \cdots + y_{n-1} + y_n)$ (the right endpoint rectangular estimate). In the special case when $f(a) = f(b)$ the three estimates for $\int_a^b f(x)\,dx$ coincide.

Observation 6.5.1: *Trapezoid = (Left + Right)/2*

The trapezoid estimate with n subintervals is the average of the left endpoint rectangular estimate and the right endpoint rectangular estimate:

$$h\left(\frac{y_0}{2} + y_1 + y_2 + \cdots + y_{n-1} + \frac{y_n}{2}\right) = \frac{1}{2}\left(h(y_0 + y_1 + y_2 + \cdots + y_{n-1}) + h(y_1 + y_2 + \cdots + y_{n-1} + y_n)\right)$$

Simpson's Estimate: Approximation by Parabolas

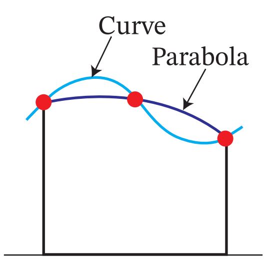


Figure 6.5.6

In the trapezoidal estimate a curve is approximated by straight lines. Simpson's estimate for $\int_a^b f(x)\,dx$ approximates a curve by parabolas. Given three points on a curve, there is a unique parabola of the form $y = Ax^2 + Bx + C$ that passes through them, as shown in Figure 6.5.6. (See Exercise 28.) The area under that parabola is used to approximate the area under the curve.

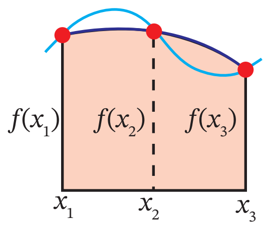
The computations leading to the formula for the area under the parabola are more involved than those for the area of a trapezoid. (They are outlined in Exercises 28 and 29.) However, the final formula is fairly simple. Let the three points be $(x_1, f(x_1))$, $(x_2, f(x_2))$, $(x_3, f(x_3))$, with $x_1 < x_2 < x_3$, $x_2 - x_1 = h$, and $x_3 - x_2 = h$, as shown in Figure 6.5.7(a). The shaded area under the parabola turns out to be

$$\frac{h}{3}\left(f(x_1) + 4f(x_2) + f(x_3)\right).$$

To estimate $\int_a^b f(x)\,dx$, we pick an even number n and use $n/2$ parabolic arcs, each of width $2h$. The case $n = 6$, which uses three parabolas, is illustrated in Figure 6.5.7(b). As in the trapezoidal method, we start with a partition of $[a, b]$ into n sections of equal width, h: $a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$. The sum of the area under the n parabolas, with $y_i = f(x_i)$, is

$$\frac{h}{3}\left((y_0 + 4y_1 + y_2) + (y_2 + 4y_3 + y_4) + \cdots + (y_{n-2} + 4y_{n-1} + y_n)\right).$$

Collecting like terms gives **Simpson's method** for estimating the definite integral $\int_a^b f(x)\,dx$:



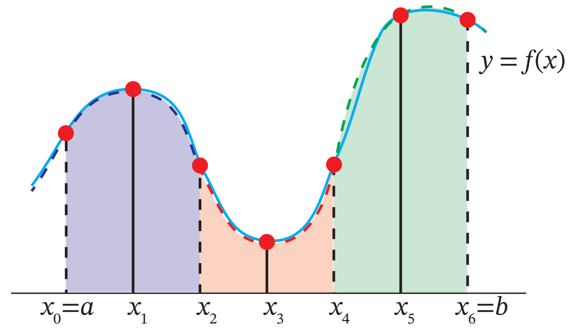


Figure 6.5.7

Formula 6.5.4: *Simpson's Estimate for* $\int_a^b f(x)\,dx$

The Simpson's estimate for $\int_a^b f(x)\,dx$ with $n/2$ equal width parabolic arcs is

$$\int_a^b f(x)\,dx \approx \frac{h}{3}\left(y_0+4y_1+2y_2+4y_3+2y_4+\cdots+2y_{n-2}+4y_{n-1}+y_n\right) \quad \text{where } h=\frac{b-a}{n} \text{ and } n \text{ is even}$$

Except for the first and last terms, the coefficients alternate 4, 2, 4, 2, …, 2, 4. To apply Simpson's estimate, pick an even number n. Then $h=(b-a)/n$. The estimate uses an odd number, $n+1$, of points, x_0, x_1, …, x_n, and $n/2$ parabolas. Example 3 illustrates the method with $n=4$.

Historical Note

While Simpson's method is named for the British mathematician and inventor, Thomas Simpson (1710–1761), this method was discovered a century earlier by the German mathematician Johannes Kepler (1571–1630).

EXAMPLE 3. Use Simpson's method with $n=4$ to estimate $\displaystyle\int_0^1 \frac{dx}{1+x^2}$.

SOLUTION The estimate takes the form $(h/3)\left(y_0+4y_1+2y_2+4y_3+y_4\right)$. with $h=(1-0)/4=1/4$. There are two parabolas, shown in Figure 6.5.8. Because the parabolas look almost like the curve, we expect Simpson's estimate to be even better than the trapezoidal estimate.

The computations are shown in Table 6.5.2.

Combining the data in the table with the factor $h/3=1/12$ provides the estimate

$$\frac{1}{12}(1.00000+3.76471+1.60000+2.56000+0.50000)=\frac{1}{12}(9.42471)=0.78539.$$

◆

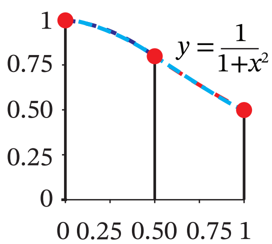


Figure 6.5.8

x_i	$f(x_i)$	coefficient	summand	decimal form
0	$\frac{1}{1+0^2}$	1	$1 \cdot \frac{1}{1+0}$	1.00000
$\frac{1}{4}$	$\frac{1}{1+\left(\frac{1}{4}\right)^2}$	4	$4 \cdot \frac{1}{1+\frac{1}{16}}$	3.76471
$\frac{2}{4}$	$\frac{1}{1+\left(\frac{2}{4}\right)^2}$	2	$2 \cdot \frac{1}{1+\frac{4}{16}}$	1.60000
$\frac{3}{4}$	$\frac{1}{1+\left(\frac{3}{4}\right)^2}$	4	$4 \cdot \frac{1}{1+\frac{9}{16}}$	2.56000
$\frac{4}{4}$	$\frac{1}{1+\left(\frac{4}{4}\right)^2}$	1	$1 \cdot \frac{1}{1+\frac{16}{16}}$	0.50000

Table 6.5.2

NOTE: The estimate found in Example 3 differs from the exact value of the definite integral by only 0.000008.

Comparison of the Four Methods

We know the value of $\int_0^1 dx/(1+x^2)$ is $\pi/4 \approx 0.78539816$, to eight decimal places. Table 6.5.3 compares the estimates made in the three examples to this value. The error is the distance from the approximation to the exact value:

$$\text{Error} = \left|\text{Estimate} - \text{Exact}\right|.$$

Method	Estimate	Error
Left Endpoint Rectangle	0.84529	0.05989
Right Endpoint Rectangle	0.72029	0.06511
Trapezoid	0.78279	0.00261
Simpson's (Parabola)	0.78539	0.000008

Table 6.5.3

Though each method takes about the same amount of work to find the estimate once you know the formulas, the table shows that the Simpson's estimate has the smallest error. The trapezoidal method is next best. The left and right endpoint rectangular estimates have the largest errors. That should not come as a surprise. Parabolas should fit a curve better than sloped lines, and sloped lines should fit better than horizontal line segments. The trapezoidal and Simpson's estimates in Examples 2 and 3 used the same sampling numbers to evaluate the integrand; they differ only in the weights (coefficients) associated with the outputs of the integrand.

The size of the error is connected to the derivatives of the integrand. For a positive number k, let M_k be the largest value of $\left|f^{(k)}(x)\right|$ for x in $[a,b]$. Table 6.5.4 lists the upper bounds for the error when $\int_a^b f(x)\,dx$ is estimated by sections of length $h=(b-a)/n$. They are usually developed in a course on numerical analysis. They can also be obtained by using the Growth Theorem of Section 5.4 and the fundamental theorem of calculus. (See Exercises 44 and 45 in this section and Exercise 76 in the Section 6.S.) They offer a good review of basic ideas.

Table 6.5.4 expresses the bounds on the size of the error for each method in terms of $h = (b-a)/n$ and n.

Method	Bound on Error in Terms of h	Bound on Error in Terms of n
Rectangles	$\frac{1}{2}M_1(b-a)h$	$\frac{1}{2}M_1(b-a)^2/n$
Trapezoids	$\frac{1}{12}M_2(b-a)h^2$	$\frac{1}{12}M_2(b-a)^3/n^2$
Simpson's (Parabolas)	$\frac{1}{180}M_4(b-a)h^4$	$\frac{1}{180}M_4(b-a)^5/n^4$

Table 6.5.4

The coefficients in the error bounds tell us a great deal. For instance, if $M_4 = 0$, then there is no error in Simpson's method. That is, if $f^{(4)}(x) = 0$ for all x in $[a,b]$, then Simpson's method produces an exact answer. As a consequence, for polynomials of degree at most 3, Simpson's approximation is exact. (See Exercises 29 and 30 in this section, and also Exercise 79 in the Section 6.S.)

We have seen that the trapezoidal estimate is exact for polynomials of degree at most one, that is, for functions whose second derivatives are zero. That suggests that its error depends on the size of the second derivative; Table 6.5.4 shows that it is.

The power of h that appears in the error bound is even more important. For instance, if you reduce the width h by a factor of 10 by using 10 times as many sections you expect the error of the rectangular estimates to shrink by a factor of 10, the error in the trapezoidal estimate to shrink by a factor of $10^2 = 100$, and the error in Simpson's estimate by a factor of $10^4 = 10,000$. See Table 6.5.5.

Method	Reduction Factor of h	Expected Reduction Factor of Error
Rectangles	10	10
Trapezoids	10	100
Simpson's (Parabolas)	10	10,000

Table 6.5.5

Summary

Four techniques for estimating definite integrals were suggested by the areas of rectangles, the areas of trapezoids, and the areas under parabolas. The error in each method is influenced by a derivative of the integrand and the distance, $h = (b-a)/n$, between the numbers at which we evaluate the integrand. The estimates have different coefficients to weight the function values $y_i = f(x_i)$ for $i = 0, 1, 2, \dots, n$. In the left-hand rectangular estimate the coefficients are $1, 1, 1, \dots, 1, 0$ because $y_n = f(b)$ is not used. In the right-hand rectangular estimate the coefficients are $0, 1, 1, \dots, 1$. In the trapezoidal estimate, they are $1, 2, 2, \dots, 2, 1$ and in Simpson's estimate they are $1, 4, 2, 4, 2, \dots, 2, 4, 1$. The multiplicative factors associated with each of these four estimates are h, h, $h/2$, and $h/3$.

In this section we estimated definite integrals with the aid of polynomials of degree at most two. Hope that going to polynomials of arbitrarily high degrees would produce better estimates was shattered by **Runge's counterexample**, estimating $\int_{-1}^{1} dx/(1+25x^2)$. It is discussed on the Web and in many numerical analysis courses.

EXERCISES for Section 6.5

In these Exercises, T_n refers to the trapezoidal estimate with n trapezoids (partition with n sections and $n+1$ points), and S_n refers to Simpson's estimate with $n/2$ parabolas (partition with n sections and $n+1$ points).

1. Write out the trapezoid estimate with 6 subintervals for $\int_1^4 5^x\,dx$ but do not carry out the calculations.

2. Write out the Simpson's estimate with 10 subintervals for $\int_0^1 e^{x^2}\,dx$ but do not carry out the calculations.

In Exercises 3 to 10 approximate the definite integrals by the trapezoidal estimate with the indicated trapezoid estimate with the indicated number of subintervals.

3. $\int_0^2 \frac{dx}{1+x^2}$, $n=2$

4. $\int_0^2 \frac{dx}{1+x^2}$, $n=4$

5. $\int_0^2 \sin(\sqrt{x})\,dx$, $n=2$

6. $\int_0^2 \sin(\sqrt{x})\,dx$, $n=4$

7. $\int_1^3 \frac{2^x}{x}\,dx$, $n=3$

8. $\int_1^3 \frac{2^x}{x}\,dx$, $n=6$

9. $\int_1^3 \cos(x^2)\,dx$, $n=2$

10. $\int_1^3 \cos(x^2)\,dx$, $n=4$

In Exercises 11 to 14 use Simpson's estimate to approximate the definite integral with the Simpson's estimate with the indicated number of subintervals.

11. $\int_0^1 \frac{dx}{1+x^3}$, $n=2$

12. $\int_0^1 \frac{dx}{1+x^3}$, $n=4$

13. $\int_0^1 \frac{dx}{1+x^4}$, $n=2$

14. $\int_0^1 \frac{dx}{1+x^4}$, $n=4$

---

15. By a direct computation, show that the trapezoidal estimate is not exact for second-order polynomials.

16. By a direct computation, show that the Simpson's estimate is not exact for fourth-order polynomials.

17. In an interval $[a,b]$ in which $f''(x)$ is positive, do trapezoidal estimates of $\int_a^b f(x)\,dx$ underestimate or overestimate the definite integral? Explain.

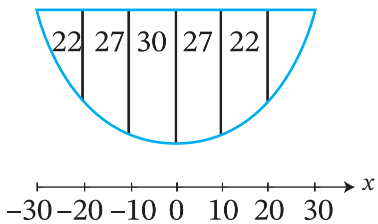
18. The cross section of a ship's hull is shown in Figure 6.5.9(a). (Dimensions are in feet.) Estimate, using $n=6$ subintervals, the hull's area using (a) the trapezoid estimate and (b) Simpson's estimate.

19. Figure 6.5.9(b) shows a map of the outlin of Lake Tahoe. Use Simpson's method and data from the map to estimate the surface area of the lake. Use cross sections parallel to the long sides of the map. (Each little square represents a square mile.)

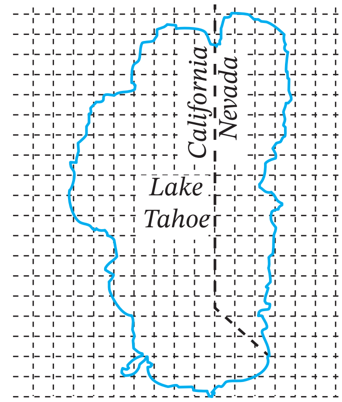
20. A ship is 120 feet long. The cross-sectional area of its hull is given at intervals in Table 6.5.6. Estimate the volume of the hull in cubic feet by (a) a trapezoidal estimate, and (b) a Simpson's estimate.. (What is the largest n that can be used in this problem?)

x	0	20	40	60	80	100	120	feet
area	0	200	400	450	420	300	150	square feet

Table 6.5.6



(a)



(b)

Figure 6.5.9

Exercises 21 and 22 present cases in which the maximum bound on the error is attained.

21. Show that the error for the trapezoidal estimate of $\int_0^1 x^2\,dx$ is exactly $\frac{1}{12}(b-a)M_2h^2$ where $a=0$, $b=1$, $h=1$, and M_2 is the maximum value of $\left|D^2(x^2)\right|$ for x in $[0,1]$.

22. Show that the error for the Simpson estimate of $\int_0^1 x^4\,dx$ is exactly $\frac{1}{180}(b-a)M_4h^4$ where $a=0$, $b=1$, $h=\frac{1}{2}$, and M_4 is the maximum value of $\left|D^4(x^4)\right|$ for x in $[0,1]$.

---

23. Figure 6.5.10(b) shows cross sections of a pond in two directions. Use Simpson's method to estimate the area of the pond using (a) vertical cross sections, three parabolas and, and (b) horizontal cross sections, two parabolas..

24. For trapezoidal estimates, if you double the length of the interval $[a,b]$ and also the number of trapezoids, would you expect the error in the estimates to increase, decrease, or stay about the same? Explain.

25. For Simpson estimates, if you double the length of the interval $[a,b]$ and also the number of parabolas, would you expect the error in the estimates to increase, decrease, or stay about the same? Explain.

26.
(a) Fill in the table below concerning $\int_0^6 x^2\,dx$ and its trapezoidal estimates.

Each trapezoid has width h.

(b) Do the errors in (a) seem to be proportional to h^c for some constant c? If so, what is c?

	$\int_0^6 x^2\,dx$	T_1	T_2	T_3
Value				
Error	—			

27.
(a) Fill in this table concerning $\int_1^7 \frac{dx}{(1+x)^2}$ and its Simpson estimates. Each section has width h.

(b) Do the errors in (a) seem to be proportional to h^c for some constant c? If so, what is c?

	$\int_1^7 \frac{dx}{(1+x)^2}$	S_2	S_4	S_6
Value				
Error	—			

Exercises 28 to 30 provide the basis of Simpson estimates. For convenience we place the origin of the x-axis at the midpoint of the interval for which a single parabola will approximate the function. Because the interval has length $2h$, its ends are $-h$ and h.

28. Let $f(x)$ be a function defined on $[-h,h]$, with $f(-h)=y_1$, $f(0)=y_2$, and $f(h)=y_3$. Show that there is exactly one parabola $P(x)=Ax^2+Bx+C$ that passes through $(-h,y_1)$, $(0,y_2)$, and (h,y_3). (See Figure 6.5.10(a).)

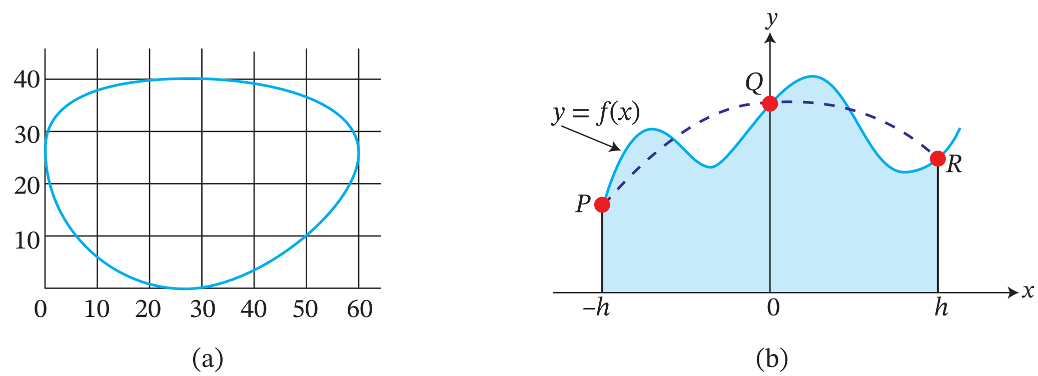


Figure 6.5.10

29. Show that $\int_{-h}^{h} p(x)\,dx = \frac{h}{3}\big(p(-h)+4p(0)+p(h)\big)$ when $p(x)=Ax^2+Bx+C$, for all values of the constants A, B, and C, by computing both sides of the equation. This result, which expresses the fact that that what we now know as Simpson's method is exact for any quadratic function, was known in geometric form to the ancient Greeks.

30. Let $f(x)=x^3$. Show that $\int_{-h}^{h} f(x)\,dx = \frac{h}{3}\big(f(-h)+4f(0)+f(h)\big)$. Combined with Exercise 29, this proves that Simpson's method is exact for all polynomials of degree at most 3.

31. Table 6.5.7 lists a few values of a function f.

(a) Plot the points on the graph of f.

(b) Sketch six trapezoids that can be used to estimate $\int_1^7 f(x)\,dx$.

(c) Find the trapezoidal estimate of $\int_1^7 f(x)\,dx$.

(d) Sketch, by eye, the three parabolas used in Simpson's method to estimate $\int_1^7 f(x)\,dx$.

(e) Find Simpson's estimate of $\int_1^7 f(x)\,dx$.

x	1	2	3	4	5	6	7
$f(x)$	1	2	1.5	1	1.5	3	3

Table 6.5.7

Exercises 32 to 34 describe the **midpoint estimate**.

32. Another way to estimate a definite integral is by a Riemann sum $\sum_{i=1}^{n} f(c_i)h$, where the c_i are the midpoints of the intervals. Find the midpoint estimate with $n = 4$ sections for $\int_0^1 \frac{dx}{1+x^2}$.

33. Using a diagram, show that the midpoint estimate is exact for functions of the form $f(x) = Ax + B$.

34. Use a diagram to show that the midpoint method overestimates $\int_a^b f(x)\,dx$ when $f''(x) < 0$ for $a \le x \le b$.

---

35. A function f is defined on $[a,b]$ and $f(x)$, $f'(x)$, and $f''(x)$ are all positive for x in that interval. Arrange the following quantities in order of size, from smallest to largest. (Some may be equal.) Sketches may help.

(a) the area of the trapezoid with base $[a,b]$ and parallel sides of lengths $f(a)$ and $f(b)$

(b) the area of the midpoint rectangle with base $[a,b]$ and height $f\left(\frac{a+b}{2}\right)$

(c) the area of the right-endpoint rectangle with base $[a,b]$ and height $f(b)$

(d) the area of the left-endpoint rectangle with base $[a,b]$ and height $f(a)$

(e) the average of (c) and (d)

(f) the trapezoid whose base is $[a,b]$ and whose top edge lies on the tangent line at $\left(\frac{a+b}{2}, f\left(\frac{a+b}{2}\right)\right)$

(g) $\int_a^b f(x)\,dx$.

36. If the Simpson estimate with four parabolas estimates a definite integral with an error of 0.35, what error would you expect with (a) with eight parabolas? and (b) with five parabolas?

37. The equation in Exercise 29 is called the **prismoidal formula**. Use it to compute the volume of (a) a sphere of radius a and (b) a right circular cone of radius a and height h.

Exercise 38 provides a review of several ideas as it involves the fundamental theorem of calculus (FTC I), the chain rule, l'Hôpital's rule, and the intermediate value theorem. The midpoint estimate is defined in Exercise 32.

38. Assume that $f''(x)$ is continuous and negative for x in $[-h,h]$. Then the midpoint estimate, M, for $\int_{-h}^{h} f(x)\,dx$ is too large and the trapezoidal estimate, T, is too small. The error of the first is $M - \int_{-h}^{h} f(x)\,dx$ and of the second is $\int_{-h}^{h} f(x)\,dx - T$. Show that $\lim_{h\to 0} \dfrac{M - \int_{-h}^{h} f(x)\,dx}{\int_{-h}^{h} f(x)\,dx - T} = \dfrac{1}{2}$. This suggests that the error in the midpoint estimate when h is small is about half the error of the trapezoidal estimate. However, the midpoint estimate is seldom used because data at midpoints are usually not available and because the Simpson estimate provides a more accurate estimate using the same data. NOTE: An error estimate for this method is obtained in Exercise 10 in Section 12.S.

---

39. Simpson's estimate is not exact for fourth-degree polynomials.

(a) Find the Simpson's estimate of $\int_{0}^{h} x^4\,dx$ with $n = 2$.

(b) What is the ratio between the estimate found in (a) and $\int_{0}^{h} x^4\,dx$?

(c) What does (b) imply about the ratio between Simpson's estimate and $\int_{0}^{h} P(x)\,dx$ for a quartic polynomial?

40. There are other methods for estimating definite integrals. Some old methods, which had been of only theoretical interest because of their messy arithmetic, have with the advent of computers assumed practical importance. This exercise illustrates the simplest of the **Gaussian quadrature** formulas for integrals over $[-1,1]$.

(a) Show that $\int_{-1}^{1} f(x)\,dx = f\left(\dfrac{-1}{\sqrt{3}}\right) + f\left(\dfrac{1}{\sqrt{3}}\right)$ for $f(x) = 1$, x, x^2, and x^3.

(b) Let a and b be two numbers, $-1 \le a < b \le 1$, such that $\int_{-1}^{1} f(x)\,dx = f(a) + f(b)$ for $f(x) = 1$, x, x^2, and x^3. Show that only $a = \dfrac{-1}{\sqrt{3}}$ and $b = \dfrac{1}{\sqrt{3}}$ satisfy the equation.

(c) Show that the Gaussian approximation $f\left(\dfrac{-1}{\sqrt{3}}\right) + f\left(\dfrac{1}{\sqrt{3}}\right)$ has no error when f is a polynomial of degree at most 3.

(d) Use the formula in (a) to estimate $\int_{-1}^{1} \dfrac{dx}{1+x^2}$.

(e) Compare the answer in (d) to the exact value of $\int_{-1}^{1} \dfrac{dx}{1+x^2}$. How large is the error?

41. Let f be a function such that $\left|f^{(2)}(x)\right| \le 10$ and $\left|f^{(4)}(x)\right| \le 50$ for all x in $[1,5]$. If $\int_1^5 f(x)\,dx$ is to be estimated with an error of at most 0.01, how small must h be (a) in the trapezoidal approximation? and (b) in Simpson's approximation?

42. Sam never is at a loss for ideas. Let's see what he's up to now.

SAM: I bet I can find a better way than Simpson's estimate to approximate $\int_{-h}^{h} f(x)\,dx$ using the same three arguments ($-h$, 0, and h).

JANE: How so?

SAM: Look at his formula $\frac{h}{3}(f(-h)+4f(0)+f(h))$, which equals $2h\left(\frac{1}{6}f(-h)+\frac{4}{6}f(0)+\frac{1}{6}f(h)\right)$. The width of the interval is $2h$. That cannot be changed.

JANE: What would you change?

SAM: The weights $\frac{1}{6}$, $\frac{4}{6}$, and $\frac{1}{6}$. I'll use weights w_1, w_2, and w_3 and demand that the estimates I get be exact when $f(x)$ is either constant, x, or x^2.

JANE: Go ahead.

SAM: If $f(x) = c$, a constant, then, because $\int_{-h}^{h} c\,dx = 2hc$, I must have $2hc = 2h(w_1c + w_2c + w_3c)$. That tells me that $w_1 + w_2 + w_3$ must be 1.

JANE: But you need three equations for three unknowns.

SAM: When $f(x) = x$, I get $\int_{-h}^{h} f(x)\,dx = 0$, so $0 = 2h(-w_1h + w_20 + w_3h)$. Now I know that w_1 equals w_3.

JANE: And the third equation?

SAM: With $f(x) = x^2$, I find that $\frac{2}{3}h^3 = 2h^3(w_1 + w_3)$.

JANE: So what are your three w's?

SAM: A little algebra shows they are $1/6$, $4/6$, and $1/6$. What a disappointment. But at least I avoided all the geometry of parabolas. It is really all about assigning proper weights.

Check the missing details and show that Sam is right.

43. Another way to estimate a definite integral is to use Taylor polynomials (introduced in Section 5.5). If the Maclaurin polynomial $P_2(x)$ for $f(x)$ of degree 2 is used to approximate $f(x)$ for x in $[0,h]$, express the possible error in using $\int_0^h P_2(x)\,dx$ to estimate $\int_0^h f(x)\,dx$.

In Section 5.5 we showed why a higher derivative controls the error in using a Taylor polynomial to approximate a function value. Exercises 44 and 45 show why a higher derivative controls the error in using the trapezoidal or Simpson estimate of a definite integral $\int_a^b f(x)dx$. (Exercise 76 in Section 6.S derives the corresponding error estimate for the midpoint estimate.) In each case $h = \frac{b-a}{n}$. The **local error**, $E(t)$, $0 \le t \le h$, is introduced. The local error $E(h)$ is the error in using one trapezoid of width h or one parabola of width $2h$. Once $E(h)$ is controlled by a higher derivative, we multiply by n, where $nh = b-a$, to obtain a measure of the total error in estimating $\int_a^b f(x)\,dx$.

OVERVIEW: Exercises 44 and 45 involve both FTC I and FTC II and provides a review of several basic concepts.

44. *(The error in the trapezoid estimate.)* As usual, let $h = \frac{(b-a}{n}$. We will estimate the error for a single section of width h and then multiply by n to find the error in estimating $\int_a^b f(x)\,dx$. For convenience, we move the graph so the interval (of length h) is $[0,h]$.

(a) Show that the error when using T_1 is $E(h) = \int_0^h f(x)\,dx - \frac{h}{2}(f(0)+f(h))$.

(b) For t in $[0,h]$ let $E(t) = \int_0^t f(x)\,dx - \frac{t}{2}(f(0)+f(t))$. Show that $E(0)=0$, $E'(0)=0$, and $E''(t) = -\frac{t}{2}f''(t)$.

(c) Let M be the maximum of $f''(x)$ on $[a,b]$ and m be the minimum. Show that $\frac{-1}{2}mt \geq E''(t) \geq \frac{-1}{2}Mt$.

(d) Using (b) and (c), show that $\frac{-1}{4}mt^2 \geq E'(t) \geq \frac{-1}{4}Mt^2$.

(e) Show that $\frac{-1}{12}mt^3 \geq E(t) \geq \frac{-1}{12}Mt^3$.

(f) Show that $\frac{-1}{12}mh^3 \geq E(h) \geq \frac{-1}{12}Mh^3$.

(g) Show that $\frac{-1}{12}m(b-a)h^2 \geq \int_a^b f(x)\,dx - T_n \geq \frac{-1}{12}M(b-a)h^2$.

(h) Show that $\int_a^b f(x)\,dx - T_n = \frac{-1}{12}f''(c)(b-a)h^2$ for some number c in $[a,b]$.

(i) Deduce that $\left|\int_a^b f(x)\,dx - T_n\right| \leq \frac{1}{12}M_2(b-a)h^2$, where M_2 is the maximum of $|f''(x)|$ for x in $[a,b]$.

45. *(The error in the Simpson estimate.)* Now n is even and $[a,b]$ is divided into n sections of width $h = \frac{b-a}{n}$. The Simpson estimate is based on $\frac{n}{2}$ intervals of length $2h$. We will place the origin at the midpoint of an interval, so that its ends are $-h$ and h. For $-h \leq t \leq h$, let $E(t) = \int_{-t}^t f(x)\,dx - \frac{t}{3}(f(-t)+4f(0)+f(t))$. We wish to control the size of $E(h)$.

(a) Show that $E'(t) = \frac{2}{3}(f(t)+f(-t)) - \frac{4}{3}f(0) - \frac{t}{3}(f'(t)-f'(-t))$.

(b) Show that $E''(t) = \frac{1}{3}(f'(t)-f'(-t)) - \frac{t}{3}(f''(t)+f''(-t))$.

(c) Show that $E'''(t) = -\frac{t}{3}(f'''(t)-f'''(-t))$.

(d) Show that $E'''(t) = \frac{-2t^2}{3}f^{(4)}(c)$ for some c in $[-h,h]$.

(e) Show that $E(0) = E'(0) = E''(0) = 0$.

(f) Let M_4 be the maximum of $|f^{(4)}(t)|$ on $[a,b]$. Show that $|E(t)| \leq \frac{2t^5}{180}M_4$.

(g) Deduce that $\left|\int_a^b f(x)\,dx - S_n\right| \leq \frac{M_4(b-a)h^4}{180}$.

Historical Note: Technology and Definite Integrals

The left and right endpoint rectangular estimates, trapezoidal estimate and Simpson's estimate are four examples of what is called **numerical integration**. Numerical integration is usually one of the main topics studied in a numerical analysis course. While the fundamental theorem of calculus is useful for evaluating definite integrals, it applies only when an antiderivative is known. Numerical integration is an important tool in estimating definite integrals, particularly when the FTC cannot be applied. Numerical integration can always be used to find out something about the value of a definite integral.

The design of an efficient and accurate general-purpose numerical integration algorithm is harder than it might seem. Effective algorithms typically divide the interval into sections of unequal lengths. The sections can be longer where the function is almost constant. Shorter sections are used where the function changes very rapidly. Large, or unbounded, intervals can lead to another set of difficulties. Some examples of challenging definite integrals include

$$\int_0^2 \sqrt{x(4-x)}\,dx, \qquad \int_{-1}^1 \frac{dx}{x^2+10^{-10}}, \quad \text{and} \quad \int_0^{600\pi} \frac{\sin^2(x)}{\sqrt{x}+\sqrt{x+\pi}}\,dx$$

The HP-34C was, in 1980, the first handheld calculator to perform numerical integration. Now this is a common feature on most scientific calculators. The algorithms used vary greatly, and their details are often corporate secrets. The techniques are similar to those presented in this Section and in Exercise 40.

REFERENCE: William Kahan, "Handheld Calculator Evaluates Integrals," *Hewlett-Packard Journal*, 31(8), Aug. 1980, pp. 23–32.

6.S Chapter Summary

This chapter introduced the second major concept in calculus, the definite integral,

$$\int_a^b f(x)\,dx = \lim_{\max \Delta x_i \to 0} \sum_{i=1}^n f(c_i)\Delta x_i.$$

For a continuous function the limit always exists and $\int_a^b f(x)\,dx$ can be viewed as the net area under the graph of $y = f(x)$ on the interval $[a,b]$. Both the definite integral and an antiderivative of a function f are called integrals. Context tells which is meant. An antiderivative is also called an indefinite integral.

Unlike the derivative, which gives local information about a function, the definite integral gives global information. Whereas the derivative provides local information about the rate of change of a function at a point, the definite integral provides global information about the function. As the first and last of the applications in Table 6.S.1 show, the definite integral of the rate at which some quantity is changing gives the total change.

The first part of the fundamental theorem of calculus (FTC I) gives a way to evaluate many definite integrals. However, finding an antiderivative can be tedious or impossible. For instance, $\exp(x^2)$ does not have an elementary antiderivative. However, continuous functions do have antiderivatives, as slope fields suggest, even if they cannot be written explicitly. Indeed $G(x) = \int_a^x f(t)\,dt$ is an antiderivative of the integrand. FTC II tells how to differentiate functions defined as a definite integral:

$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x).$$

One way to estimate a definite integral is to employ one of the sums of rectangles, $\sum_{i=1}^n f(c_i)\,\Delta x_i$, that appear

Integrand	Definite Integral
$f(x)$	$\int_a^b f(x)\, dx$
velocity	change in position
speed (|velocity|)	distance traveled
cross-sectional length of plane region	area of a plane region
cross-sectional area of solid	volume of solid
rate at which a quantity grows	total growth

Table 6.S.1

in its definition. A more accurate method, which involves the same amount of arithmetic, uses trapezoids. It is

$$\int_a^b f(x)\, dx \approx \frac{h}{2}\big(f(x_0) + 2f(x_1) + 2f(x_2) + \ldots + 2f(x_{n-1}) + f(x_n)\big),$$

where consecutive x_i's are a fixed distance $h = (b-a)/n$ apart. In Simpson's method the graph is approximated by parabolas, n is even, and the estimate is

$$\frac{h}{3}\big(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \ldots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)\big).$$

EXERCISES for Section 6.S

1. State FTC I in words, using no mathematical symbols.

2. State FTC II in words, using no mathematical symbols.

Evaluate the definite integrals in Exercises 3 to 16.

3. $\int_1^2 (2x^3 + 3x - 5)\, dx$

4. $\int_5^7 \frac{3}{x}\, dx$

5. $\int_1^4 \frac{dx}{\sqrt{x}}$

6. $\int_1^4 \frac{x + 2x^3}{\sqrt{x}}\, dx$

7. $\int_0^1 x(3+x)\, dx$

8. $\int_0^2 (2+3x)^2\, dx$

9. $\int_1^2 \frac{(2+3x)^2}{x^2}\, dx$

10. $\int_1^2 e^{2x}\, dx$

11. $\int_0^{\pi} \sin(3x)\, dx$

12. $\int_0^{\pi/4} \sec^2(x)\, dx$

13. $\int_0^{\sqrt{2}/2} \frac{3\, dx}{\sqrt{1-x^2}}\, dx$

14. $\int_0^{\pi/4} \cos(x)\, dx$

15. $\int_0^{\pi/4} \sec(x)\tan(x)\, dx$

16. $\int_{1/2}^{\sqrt{2}/2} \frac{dx}{x\sqrt{x^2-1}}$

In Exercises 17 to 24 find an antiderivative of the function by guess and experiment. Check your answer by differentiating it.

17. $(2x+1)^5$

18. $(x+1)^{-1}$

19. $\ln(x)$

20. $x\sin(x)$

21. $\sin(2x)$

22. xe^{x^2}

23. $\dfrac{1}{(2x+1)^5}$

24. $\dfrac{1}{2x+1}$

Use Simpson's estimate with three parabolas ($n=6$) to approximate the definite integrals in Exercises 25 and 26.

25. $\displaystyle\int_0^{\pi/2} \sin(x^2)\,dx$

26. $\displaystyle\int_1^2 \sqrt{1+x^2}\,dx$

---

27. Use the trapezoidal estimate with $n=6$ to estimate the integral in Exercise 25.

28. Use the trapezoidal estimate with $n=6$ to estimate the integral in Exercise 26.

Exercises 29 and 30 provide some initial background in preparation for the historical discussion in Section 11.1 (on page 587) of Newton's manual calculation of the area under a hyperbola that is accurate to more than 50 decimal places.

29. Let c be a positive constant less than 1 ($0<c<1$).

(a) Show that the area under the curve $y=\dfrac{1}{1+x}$ above the interval $[0,c]$ is $\ln(1+c)$.

(b) Show that the area under the curve $y=\dfrac{1}{1+x}$ above the interval $[-c,0]$ is $-\ln(1-c)$.

30. (a) In his approximation of $\ln(1.1)$ to 53 decimal places, Newton used, in effect, $P_{53}(0.1;0)$ for $f(x)=\ln(1+x)$ to approximate $\ln(1.1)$. What is the bound on the error for the approximation?

(b) Could Newton have used fewer terms to obtain an equally accurate answer? Explain your answer.

---

31. (a) What is the area under $y=\dfrac{1}{x}$ and above $[1,b]$, $b>1$?

(b) Is the area under $y=\dfrac{1}{x}$ and above $[1,\infty)$ finite or infinite?

(c) The region under $y=\dfrac{1}{x}$ and above $[1,b]$ is rotated around the x-axis. What is the volume of the solid produced?

32. The basis for this chapter is that if f is continuous and $x>a$, then $\dfrac{d}{dx}\displaystyle\int_a^x f(t)\,dt=f(x)$.

(a) Review how the equation was obtained.

(b) Use a similar method to show that if $x<b$, then $\dfrac{d}{dx}\displaystyle\int_x^b f(t)\,dt=-f(x)$.

33. Let $f(x)$ and $g(x)$ be differentiable functions with $f(x) \geq g(x)$ for all x in $[a,b]$, $a < b$. Explain each answer. (a) Is $f'(x) \geq g'(x)$ for all x in $[a,b]$? (b) Is $\int\limits_a^b f(x)dx \geq \int\limits_a^b g(x)dx$?

34. Find $D\left(\int\limits_{x^2}^{x^3} e^{-t^2}\,dt\right)$.

35. **JANE:** I'm not happy. The text says that a definite integral measures area. But they never defined area under a curve. I know that the area of a rectangle is width times length. But what is meant by the area under a curve? If they say, "Well, it's the definite integral of the cross sections," that won't do. What if I integrate cross sections that are parallel to the x-axis instead of the y-axis? How do I know I'll get the same answer? Once again, the authors are hoping no one will notice a big gap in their logic.

Is Jane right? Have the authors tried to slip something past the reader?

36. Call the left endpoint estimate of $\int\limits_a^b f(x)\,dx$ with n rectangles L_n, the right endpoint estimate with n rectangles R_n, and the trapezoidal estimate with n trapezoids T_n.
(a) Show that $R_n - T_n = T_n - L_n = \frac{h}{2}(f(b) - f(a))$. (b) Show that $T_n = \frac{1}{2}(L_n + R_n)$.

37. Denote the trapezoidal estimate of $\int\limits_a^b f(x)\,dx$ with n by T_n and the midpoint estimate with n sections by M_n. Show that $\frac{1}{3}T_n + \frac{2}{3}M_n$ equals the Simpson estimate S_{2n} with n parabolas. (Consider a typical interval of length h.)

38. A river flows at the rate of $r(t)$ cubic feet per second.
- (a) Approximate how many cubic feet pass during the short time interval from time t to time $t + \Delta t$ seconds.
- (b) How much passes from time t_1 to time t_2 seconds?

39. Let $f(x) = xe^{-x}$ for $x \geq 0$. For which interval of length 1 is the area below the graph of f and above the interval a maximum?

40. Let $f(x) = \frac{x}{(x+1)^2}$ for $x \geq 0$.
- (a) Graph f, showing any extrema.
- (b) Looking at the graph, estimate for which interval of length one the area below the graph of f and above the interval is a maximum.
- (c) Using calculus, find the interval in (b) that yields the maximum area.

41.
- (a) Estimate $\int\limits_0^1 \frac{\sin(x)}{x}\,dx$ by approximating $\sin(x)$ by the Maclaurin polynomial $P_6(x;0)$.
- (b) Use the Lagrange bound on the error to bound the error in (a).

42. (a) Estimate $\displaystyle\int_1^3 \frac{e^x}{x}\,dx$ by using the Taylor polynomial $P_3(x;2)$ to approximate e^x.

SUGGESTION: To avoid computing e^2, approximate e by 2.71828.

(b) Use the Lagrange bound on the error to bound the error in (a).

43. Assume $f(2) = 0$, $f'(2) = 0$, and $f''(x) \le 5$ for all x in $[0,7]$. Show that $\displaystyle\int_2^3 f(x)\,dx \le 5/6$.

44. Find $\displaystyle\lim_{t\to 0} \frac{\int_0^t \left(e^{x^2}-1\right)\,dx}{\int_0^t \sin(2x^2)\,dx}$.

45. Let $G(t) = \displaystyle\int_0^t \cos^5(\theta)\,d\theta$ for t in $[0,2\pi]$. (a) Sketch a graph of $y = G'(t)$. (b) Sketch a graph of $y = G(t)$.

Usually we use a sum to estimate a definite integral. We can also use a definite integral to estimate a sum. In Exercises 46 and 47, rewrite the sum so that it is a sum estimating a definite integral. Then use the definite integral to estimate the sum.

46. $\displaystyle\frac{1}{100}\sum_{i=1}^{100}\frac{1}{1+i^2}$

47. $\displaystyle\sum_{n=51}^{100}\frac{1}{n}$

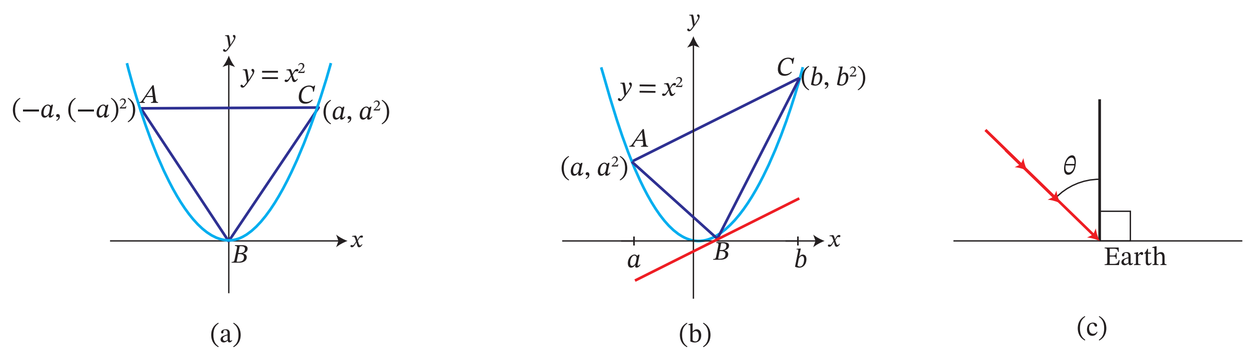


Figure 6.S.1

48. Figure 6.S.1(a) shows a triangle ABC inscribed in the parabola $y = x^2$ with $A = (-a, a^2)$, $B = (0,0)$, and $C = (a, a^2)$. Let $T(a)$ be its area and $P(a)$ the area of the region bounded by AC and the parabola above the interval $[-a, a]$. Find $\displaystyle\lim_{a\to 0}\frac{T(a)}{P(a)}$.

HISTORICAL NOTE: Archimedes established the following much more general result: In Figure 6.S.1(b) the tangent line at B is parallel to AC. He determined for any chord AC the ratio between the area of triangle ABC and the area of the parabolic section.

49. (a) Show that the average value of $\cos(\theta)$ for θ in $\left[0, \frac{\pi}{2}\right]$ is about 0.637.

(b) The average in (a) is much more than half of the maximum value of $\cos(\theta)$. Why is that good news for a farmer or solar engineer on Earth who depends on heat from the sun? (See Figure 6.S.1(c).)

50. Assume f' is continuous on $[0, t]$.

(a) Find the derivative of $F(t) = 2\int_0^t f(x)f'(x)\,dx - f(t)^2$. (b) Give a shorter formula for $F(t)$.

51. Find a simple expression for the function $F(t) = \int_1^t \cos(x^2)\,dx - \int_1^{t^2} \frac{\cos(u)}{2\sqrt{u}}\,du$.

52. A tent has a square base of side b and a pole of length $\frac{b}{2}$ above the center of the base.

(a) Set up a definite integral for the volume of the tent.

(b) Evaluate the integral in (a) by the fundamental theorem of calculus.

(c) Find the volume of the tent by showing that six copies of it fill a cube of side b.

53. **SAM:** The mean value theorem in Section 6.3 was not new.

JANE: What do you mean?

SAM: It's just a special case of the mean value theorem for derivatives in Section 4.1, the one that says $f(b) - f(a) = f'(c)(b-a)$.

JANE: I don't believe it.

SAM: Just apply the one in Section 4.1 to the function $G(x) = \int_a^x f(t)\,dt$. Then $G(b) - G(a) = G'(c)(b-a)$. But the left side is $\int_a^b f(t)\,dt$ and the right side is $f(c)(b-a)$.

JANE: Another page saved.

SAM: And lots of trees.

Is Sam right? Explain.

54. **SAM:** I can get the first FTC, the one about $F(b) - F(a)$, without all that stuff in the second FTC.

JANE: That would be nice.

SAM: I assume F' is continuous and $\int_a^b F'(x)dx$ exists. Now, $F(b) - F(a)$ is the total change in F. Well, bust up $[a, b]$ by $t_0, t_1, \ldots, t_n$ in the usual way. Then the total change is just the sum of the changes in F over each of the n intervals, $[t_{i-1}, t_i]$, $i = 1, 2, \ldots, n$.

JANE: That's a no-brainer, but then what?

SAM: The change in F over the typical interval is $F(t_i) - F(t_{i-1})$. By the mean value theorem for F, that equals $F'(t_i^*)(t_i - t_{i-1})$ for some t_i^* in the i^{th} interval. The rest is automatic.

JANE: I see. You let all the intervals get shorter and shorter and the sums of the $F'(t_i^*)(t_i - t_{i-1})$ approach $\int_a^b F'(x)\,dx$. But they are all already equal to $F(b) - F(a)$.

SAM: Pretty neat, yes?

JANE: Something must be wrong.

Is anything wrong?

Evaluate the following limits.

55. $\lim_{\Delta x\to 0}\left(\frac{1}{\Delta x}\int_5^{7+\Delta x} e^{x^3}\,dx - \frac{1}{\Delta x}\int_5^{7} e^{x^3}\,dx\right)$.

56. $\lim_{\Delta x\to 0}\left(\frac{1}{\Delta x}\int_{5+\Delta x}^{7} e^{x^3}\,dx - \frac{1}{\Delta x}\int_5^{7} e^{x^3}\,dx\right)$.

57. (a) Graph $y = e^x$ for x in $[0,1]$.
(b) Let c be the number such that the area under the graph of $y = e^x$ above $[0,c]$ equals the area under the graph above $[c,1]$. From the graph in (a), decide whether c is bigger or smaller than $\frac{1}{2}$.
(c) Find c.

58. A company is founded with capital investment A. It plans to have its rate of investment proportional to its total investment at any time. Let $f(t)$ denote the rate of investment at time t.

(a) Show that there is a constant k such that $f(t) = k\left(A + \int_0^t f(x)dx\right)$ for any $t \geq 0$. (b) Find a formula for f.

There are two definite integrals in each of Exercises 59 to 62. One can be evaluated by the FTC, the other not. Evaluate the one that can be evaluated by the FTC and approximate the other by Simpson's estimate with $n = 4$.

59. $\int_0^1 (e^x)^2\,dx$; $\int_0^1 e^{x^2}\,dx$.

60. $\int_0^{\pi/4} \sec(x^2)\,dx$; $\int_0^{\pi/4} (\sec(x))^2\,dx$.

61. $\int_1^3 e^{x^2}x\,dx$; $\int_1^3 \frac{e^{x^2}}{x}\,dx$.

62. $\int_{0.2}^{0.4} \frac{dx}{\sqrt{1-x^2}}$; $\int_{0.2}^{0.4} \frac{dx}{\sqrt{1-x^3}}$.

63. If $F'(x) = f(x)$, find an antiderivative for (a) $g(x) = x + f(x)$, (b) $g(x) = 2f(x)$, and (c) $g(x) = f(2x)$.

64. This exercise verifies the claims made in the last paragraph before the Summary of Section 5.8.

(a) Explain why, for each angle θ in $[0,\pi]$, a sector of the unit circle with angle 2θ has area θ.
(b) In Figure 6.S.2, the area of the shaded region is twice the area of region OAP. The area of OAP is the area of a triangle less the area under the hyperbola. Express this area in terms of t.
(c) Verify that $\frac{1}{2}\left(x\sqrt{x^2-1} - \ln(x+\sqrt{x^2-1})\right)$ is an antiderivative of $\sqrt{x^2-1}$ for all $x > 1$.
(d) Show that the area of the shaded region in Figure 6.S.2 is t.

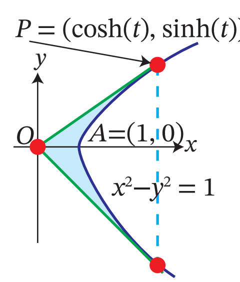


Figure 6.S.2

65. John M. Robson in "The Physics of Fly Casting," *American J. Physics* 58(1990), pp. 234–240, lets the reader fill in calculus steps. For instance, he has $\mu(4z+h)\dot{z}^2 = 2\int_0^t crh\rho\dot{z}^3\,dt + T(0)$ where z is a function of time t, $\dot{z} = \frac{dz}{dt}$, and $\ddot{z} = \frac{d^2z}{dt^2}$. He then states "differentiating this gives $(2\mu - crh\rho)\dot{z}^2 + (4z+h)\mu\ddot{z} = 04$". Is this statement correct?

66. Jane is running from a to b, on the x-axis. When she is at x, her speed is $v(x)$. How long does it take her to go from a to b?

67.
(a) Find all continuous functions $f(t)$, $t \geq 0$, such that $\int_0^{x^2} f(t)\,dt = 3x^3$, $x \geq 0$.
(b) Check that they satisfy the equation.

68. Let $f(x)$ be defined for x in $[0,b]$, $b>0$. Assume that $f(0)=0$ and $f'(x)$ is positive.
(a) Use Figure 6.S.3(a) to show that $\int_0^b f(x)\,dx + \int_0^{f(b)} (\mathrm{inv} f)(x)\,dx = bf(b)$.
(b) As a check on the equation in (a), differentiate both sides with respect to b. You should get a valid equation.
(c) Use (a) to evaluate $\int_0^1 \arcsin(x)\,dx$.

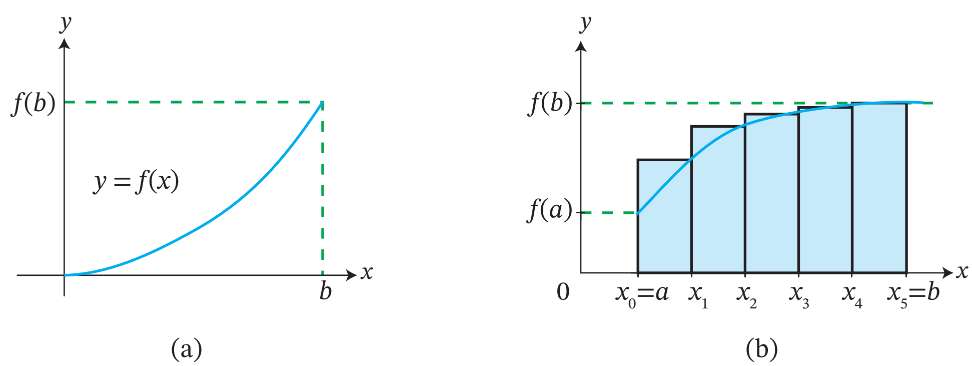


Figure 6.S.3

69.
(a) Verify, without using the FTC, that $\int_0^2 \sqrt{x(4-x)}\,dx = \pi$. (What region has an area given by that integral?)
(b) Approximate the definite integral in (a) by the trapezoidal estimate with four trapezoids.
(c) Approximate the definite integral in (a) by the trapezoidal estimate with eight trapezoids.
(d) Compute the errors for the approximations obtained in both (b) and (c).
(e) Estimate how many trapezoids are needed to have an approximation that is accurate to three decimal places.
(f) Why is the error bound for the trapezoidal estimate of no use in (e)?

70.
(a) Approximate the definite integral in Exercise 69 by Simpson's estimate with two parabolas and again with four parabolas. (These use the same number of subintervals as in Exercise 69.)
(b) Compute the error in each case.
(c) By trial-and-error, estimate how many parabolas are needed to have an estimate accurate to three decimal places. (Use a computational tool to automate the calculations.)
(d) Why is the error bound for the Simpson's estimate of no use in (c)?

71. In his *Principia*, published in 1687, Newton examined the error in approximating an area by rectangles. He considered an increasing differentiable function f defined on the interval $[a,b]$ and drew a figure similar to Figure 6.S.3(b). All rectangles have the same width h. Let R equal the sum of the areas of the rectangles using right endpoints and let L equal the sum of the areas of the rectangles using left endpoints. Let A be the area under the curve $y = f(x)$ and above $[a,b]$.

(a) Why is $R - L = (f(b) - f(a))h$?

(b) Show that any approximating sum for A, formed with rectangles of equal width h and any sampling points, differs from A by at most $(f(b) - f(a))h$.

(c) Let M_1 be the maximum value of $\left|f'(x)\right|$ for x in $[a,b]$. Show that any approximating sum for A formed with equal widths h differs from A by at most $M_1(b-a)h$.

(d) Newton also considered the case where the rectangles do not necessarily have the same widths. Let h be the largest of their widths. What can be said about the error in this case?

72. Assume that f is a continuous function such that $f(x) > 0$ for $x > 0$ and $\int_0^x f(t)\,dt = (f(x))^2$ for $x \geq 0$.

(a) Find $f(0)$. (b) Find $f(x)$ for $x > 0$.

73. A particle moves on a line so that the average velocity over any interval of the form $[a,b]$ is equal to the average of the velocities at a and b. Prove that the velocity $v(t)$ must be of the form $ct + d$ for some constants c and d.

Exercises 74 and 75 present Archimedes' derivations for the area of a disk and the volume of a ball. He viewed the explanations as informal, and also presented rigorous proofs for them.

74. Archimedes pictured a disk as made up of almost isosceles triangles, with a vertex of each triangle at the center of the disk and the base of the triangle part of the boundary of the disk. Using this he conjectured that the area of a disk is one-half the product of its radius and its circumference. Explain why Archimedes' reasoning is plausible.

75. Archimedes pictured a ball as made up of almost pyramids, with the vertex of each pyramid at the center of the ball and its base as part of the surface of the ball. Using this he conjectured that the volume of a ball is one-third the product of its radius and its surface area. Explain why Archimedes' reasoning is plausible.

---

76. Let M_n be the midpoint estimate of $\int_a^b f(x)\,dx$ based on n sections of width $h = \frac{b-a}{n}$. This exercise shows that the bound on the error, $\left|\int_a^b f(x)\,dx - M_n\right|$, is half of the error bound on the trapezoidal estimate. The argument is like that in Exercises 44 and 45 of Section 6.5, and is an application of the Growth Theorem of Section 5.4. Let $E(t) = \int_{-t/2}^{t/2} f(x)\,dx - f(0)t$.

The midpoint estimate for a definite integral is described in Exercises 32 to 34 in Section 6.5.

(a) Show that $E(0) = E'(0) = 0$, and that $E''(t) = \frac{1}{4}\left(f'\left(\frac{t}{2}\right) - f'\left(\frac{-t}{2}\right)\right)$.

(b) Show that $\left|\int_a^b f(x)\,dx - M_n\right| \leq \frac{1}{24}M(b-a)h^2$, where M is the maximum of $|f''(x)|$ for x in $[a,b]$.

77. Assume that $y = f(x)$ is a function such that $f(x) \geq 0$, $f'(x) \geq 0$, and $f''(x) \geq 0$ for all x in $[1,4]$. An estimate of the area under $y = f(x)$ is made by dividing the interval into sections and forming rectangles. The height of each rectangle is the value of $f(x)$ at the midpoint of the corresponding section.

(a) Show that the estimate is less than or equal to the area under the curve.

(b) How does the estimate compare to the area under the curve if, instead, $f''(x) \leq 0$ for all x in $[1,4]$?

78. The definite integral $\int_0^1 \sqrt{x}\,dx$ gives numerical analysts a pain. The integrand is not differentiable at 0. What is worse, the derivatives (first, second, etc.) of $\sqrt{x}$ become arbitrarily large for x near 0. It is instructive, therefore, to see how the error in Simpson's estimate behaves as h is made small.

(a) Use the FTC to show that $\int_0^1 \sqrt{x}\,dx = \frac{2}{3}$.

(b) Fill in the table. (Keep at least seven decimal places in each answer.)

h	Simpson's Estimate	Error	Ratio
$\frac{1}{2}$			—
$\frac{1}{4}$			
$\frac{1}{8}$			
$\frac{1}{16}$			
$\frac{1}{32}$			
$\frac{1}{64}$			

(c) In the typical application of Simpson's method, when h is reduced by a factor of 2, the error is cut by a factor of $2^4 = 16$. (That is, the ratio of the two errors would be $1/16 = 0.0625$.) Fill in the ratios of consecutive errors in the table.

(d) Let $E(h)$ be the error in using Simpson's method to estimate $\int_0^1 \sqrt{x}\,dx$ with sections of length h. Assume that $E(h) = Ah^k$ for constants k and A. Estimate k and A.

79. Since Simpson's method was designed to be exact when $f(x) = Ax^2 + Bx + C$, we expect the error associated with it to involve $f^{(3)}(x)$. By a quirk of good fortune, Simpson's method happens to be exact even when $f(x)$ is a cubic, $Ax^3 + Bx^2 + Cx + D$. This exercise confirms that this is so without using the formula for the error in Simpson's method.

(a) Show that if $f(x) = x^3$, then $\int_c^d x^3\,dx = \frac{d-c}{6}\left(f(c) + 4f\left(\frac{c+d}{2}\right) + f(d)\right)$.

(b) Why does (a) show that Simpson's estimate exact for cubic polynomials?

Exercise 80 is an extension of Exercise 80 in Section 5.S.

80. A producer of wine can choose to store it and sell it at a higher price after it has aged. However, he also must consider storage costs. Assume the revenue received from selling the wine at time t is $V(t)$. If the interest rate on bank balances is r, which we assume is constant, the present value of the sale is $V(t)e^{-rt}$.

The cost $c(t)$ of storing the wine varies with time. Assume the cost of storing the wine during the short interval $[t, t+\Delta t]$ is approximately $c(t)\Delta t$.

(a) What is the present value of storing the wine for the period $[0, x]$?

(b) What is the present value, $P(x)$, of the profit made by selling all the wine at time x? That is, what is the present value of the revenue minus the present value of the storage cost if sold at time x?

(c) Show that $P'(x) = V'(x)e^{-rx} - rV(x)e^{-rx} - c(x)e^{-rx}$.

(d) Show that if $V'(x)e^{-rx} > rV(x)e^{-rx} + c(x)e^{-rx}$, then $P'(x)$ is positive. Should he continue to store the wine?

(e) What is the meaning of each terms in the inequality in (d)? Why does the inequality make economic sense?

---

81. The definition of average value of a function introduced in Section 6.3 is sometimes referred to as the **arithmetic average**. In some applications the geometric average is more appropriate and useful. The **geometric average** of n positive numbers is defined as the n^{th} root of their product.

(a) If the positive numbers are $p_1, p_2, \dots, p_n$, their geometric average G is $\left(p_1 p_2 \cdots p_n\right)^{1/n}$. Show that $\ln(G)$ is the arithmetic average of the n numbers $\ln(p_1), \ln(p_2), \dots, \ln(p_n)$.

(b) Let f be a continuous positive function on $[a, b]$. How would you define the geometric average of f on $[a, b]$?

(c) Check that your definition in (b) is between the minimum and maximum of f on $[a, b]$.

(d) How would you define the geometric average of a continuous positive function defined on $(0, \infty)$?

Exercises 82 to 87 offer an opportunity to practice differentiation skills. In each case, verify that the derivative of the first expression is the second expression.

82. $\ln\left(\dfrac{e^x}{1+e^x}\right)$; $\dfrac{1}{1+e^x}$ To simplify, first take logs.

83. $\dfrac{1}{m}\arctan\left(e^{mx}\right)$; $\dfrac{1}{e^{mx}+e^{-mx}}$ m is a constant.

84. $\ln(\tan(x))$; $\dfrac{1}{\sin(x)\cos(x)}$

85. $\tan\left(\dfrac{x}{2}\right)$; $\dfrac{1}{1+\cos(x)}$

86. $\dfrac{1}{2}\ln\left(\dfrac{1+\sin(x)}{1-\sin(x)}\right)$; $\sec(x) = \dfrac{1}{\cos(x)}$

87. $\arcsin(x) - \sqrt{1-x^2}$; $\sqrt{\dfrac{1+x}{1-x}}$

In Exercises 88 to 90 differentiate the expressions.

88. $\dfrac{\sin(2x)\tan(3x)}{x^3}$

89. $2^{x^2}x^3\cos(4x)$

90. $\dfrac{x^2e^{3x}}{\sqrt{1+x^2}}$

Calculus is Everywhere # 8

Peak Oil Production

This is the second of three CIEs devoted to the models for U.S. oil production. The first, CIE 1 in Chapter 1, uses graphs but no calculus. However, in hindsight it can be seen as an informal argument for the Fundamental Theorem of Calculus. The present CIE uses calculus to develop this idea in more detail. The final CIE related to oil production, CIE 15 at the end of Chapter 12, uses more calculus concepts to justify the model used to determine the phenomenon known as **Hubbert's Peak** that is the focus of this CIE.

The United States in 1956 produced most of the oil it consumed and the rate of production was increasing. It is important to realize that this oil production was almost entirely from the "lower 48" (continental) states; the oil boom in Alaska did not start until 1957 and the Alaskan Pipeline was not completed until 1977.

Even so, in his presentation "Nuclear Energy and the Fossil Fuels", at the Spring Meeting of the Southern District Division of Production in March 1956 M. King Hubbert, a geologist at Shell Oil, predicted that production would peak near 1970 and then decline. (https://www.resilience.org/stories/2006-03-08/nuclear-energy-and-fossil-fuels/) His prediction did not convince geologists, who were reassured by the rising curve in Figure C.8.1.

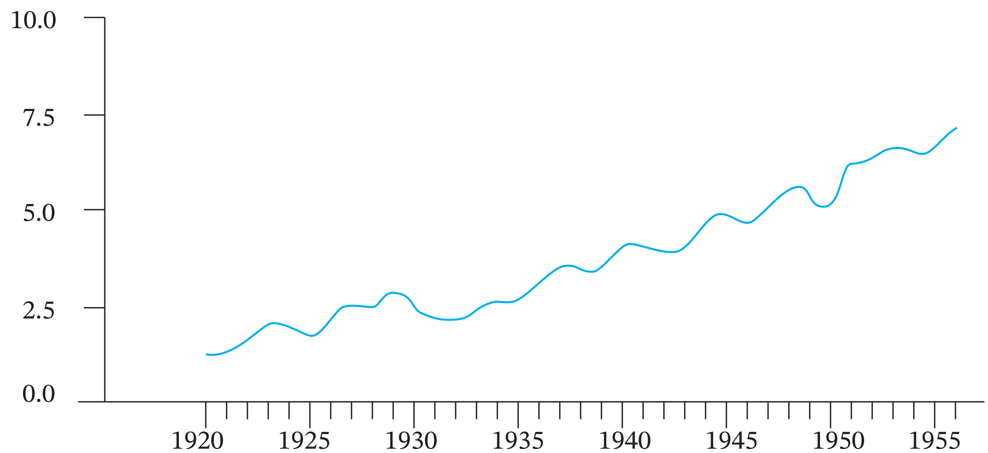


Figure C.8.1

Monthly reports of U. S. field production of crude oil as reported by the U. S. Energy Information Administration from January 1920 to the present are readily available. See, for example, https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=MCRFPUS2&f=M. There was, in fact, a peak of 7.693 million barrels per day in March 1956. But after a short decline, production then increased to 10.044 million barrels per day in November 1970. Production declined throughout the 1970's and made a moderate recovery into the 1980's before plateauing at around 9 million barrels per day throughout 1985 and 1986. Production then declined for most of the next quarter-century, to 3.974 million barrels per day in September 2018, during the Great Recession. Since then, production has grown steadily, except for a brief decline in 2016.

Hubbert's prediction of peak production in 1970 was correct — until February 2018, when U.S. field production of crude oil reached 10.281 million barrels per day, exceeding the 10.044 million barrels per day peak in November 1970. Production continued to grow, exceeding 13 million barrels per day in 2020. In mid-2021 production declined slightly, to about 11 million barrels per day. The forecast is for oil production to grow to nearly 12 million barrels per day in 2022 — if higher prices hold.

Many factors have contributed to the continued growth of US oil production, including new extraction technologies, more efficient refinement, expanded demand, and international competition. All-in-all, Hubbert's prediction was much more accurate than most experts predicted. Why? In part because his argument was so simple, as we are about to see.

We present Hubbert's original reasoning from his presentation in 1956. In it he uses an integral over the entire positive x-axis, a concept we will define in Section 7.8. However, since he is modeling a finite resource that is exhausted in a finite time, his integral is an ordinary definite integral whose upper bound is not known.

He stated two principles when analyzing curves that describe the rate of exploitation of a finite resource:

1. For any production curve of a finite resource of fixed amount, two points on the curve are known at the outset, namely that at $t = 0$ and again at $t = \infty$. The production rate will be zero when the reference time is zero, and the rate will again be zero when the resource is exhausted; that is to say, in the production of any resource of fixed magnitude, the production rate must begin at zero, and then after passing through one or several maxima, it must decline again to zero. (See Figure C.8.2.)

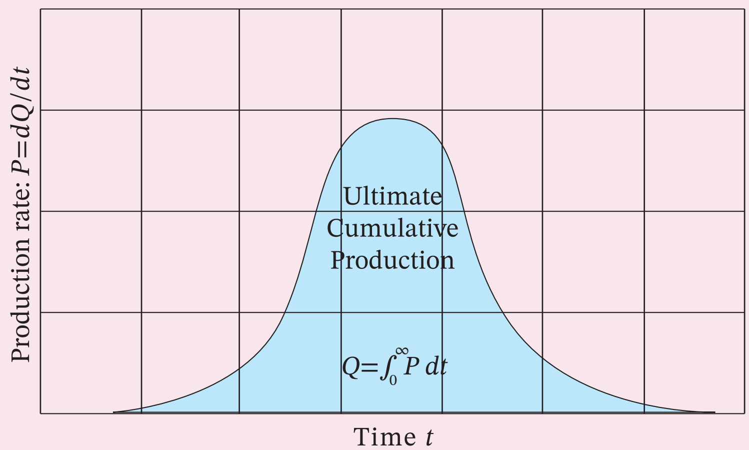


Figure C.8.2

2. The second consideration arises from the fundamental theorem of the integral calculus; namely, if there exists a single-valued function $y = f(x)$, then

$$\int_0^{x_1} y\,dx = A, \tag{C.8.1}$$

where A is the area between the curve $y = f(x)$ and the x-axis from the origin out to the x_1.

In the case of the production curve plotted against time on an arithmetical scale, we have as the ordinate

$$P = \frac{dQ}{dt},$$

where dQ is the quantity of the resource produced in time dt. Likewise, from (C.8.1) the area under the curve up to any time t is given by

$$A = \int_0^t P\,dt = \int_0^t \left(\frac{dQ}{dt}\right) dt = Q(t),$$

where Q is the cumulative production up to the time t. Likewise, the ultimate production will be given by

$$Q_{max} = \int_0^\infty P\,dt,$$

and will be represented on the graph of production-versus-time as the total area beneath the curve.

These basic relationships are indicated in Figure 11 [Figure C.8.2]. The only a priori information concerning the magnitude of the ultimate cumulative production of which we may be certain is that it will be less than, or at most equal to, the quantity of the resource initially present. Consequently, if we knew the quantity initially present, we could draw a family of possible production curves, all of which would exhibit the common property of beginning and ending at zero, and encompassing an area equal to or less than the initial quantity.

That the production of exhaustible resources does behave this way can be seen by examining the production curves of some of the older producing areas.

Hubbert did not use a formula. Instead he employed the key idea in calculus, expressed in terms of production of oil, "The definite integral of the rate of production equals the total production." He looked at the continental US crude-oil production data up to 1956 and extrapolated the curve. This is his reasoning:

Figure C.8.3 shows "a graph of the [United States crude-oil] production up to the present, and two extrapolations into the future. The unit rectangle in this case represents 25 billion barrels so that if the ultimate potential production is 150 billion barrels, then the graph can encompass but six rectangles before returning to zero. Since the cumulative production is already a little more than 50 billion barrels, then only four more rectangles are available for future production. Also, since the production rate is still increasing, the ultimate production peak must be greater than the present rate of production and must occur sometime in the future. At the same time it is possible to delay the peak for more than a few years and still allow time for the unavoidable prolonged period of decline due to the slowing rates of extraction from depleting reservoirs.

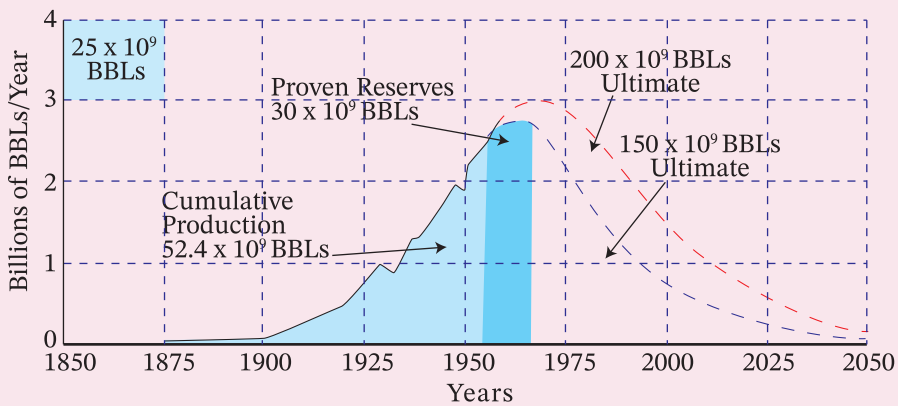


Figure C.8.3

With due regard for these considerations, it is almost impossible to draw the production curve based upon an assumed ultimate production of 150 billion barrels in any manner differing significantly from that shown in Figure C.8.3, according to which the curve must culminate in about 1965 and then must decline at a rate comparable to its earlier rate of growth.

If we suppose the figure of 150 billion barrels to be 50 billion barrels too low — an amount equal to eight East Texas oil fields — then the ultimate potential reserve would be 200 billion barrels. The second of the two extrapolations shown in Figure C.8.3 is based upon this assumption; but it is interesting to note that even then the date of culmination is retarded only until about 1970.

Geologists are now trying to predict when world production of oil will peak. (Hubbert predicted it would occur in 2000.) In 2009 oil was being extracted at the rate of 85 million barrels per day. Some say the peak occurred as early as 2005, but others believe it may not occur until after 2020.

CIE 15 at the end of Chapter 12 presents subsequent work by Hubbert in which he uses calculus to analyze oil use and depletion. An online search for "peak oil model" will quickly produce volumes of additional information, and data, about the modeling of peak oil production.

Chapter 7

Applications of the Definite Integral

This chapter develops four applications of the definite integral. Sections 7.1 and 7.4 describe two geometric applications: finding total area from the lengths of cross sections and finding total volume from the areas of cross sections. Section 7.5 gives an alternate way to compute volumes. Sections 7.6 and 7.7 present two applications in physics: water pressure and the work accomplished by a force.

Section 7.8 generalizes the definite integral to cases when either the integrand becomes infinite or the interval of integration is infinite. Integrating from zero to infinity, for instance, is common in physics and statistics. Advice on drawing and setting up definite integrals, two often overlooked skills, is found in Sections 7.2 and 7.3.

7.1 Computing Area by Parallel Cross Sections

In Section 6.1 we computed the area under $y = x^2$ and above the interval $[a, b]$ and later saw that it equals the definite integral $\int_a^b x^2\,dx$. Now we generalize the idea behind this example.

Area as a Definite Integral of Cross Sections

How can we express the area of the region R shown in Figure 7.1.1(a) as a definite integral? We introduce an axis, as in Figure 7.1.1(b).

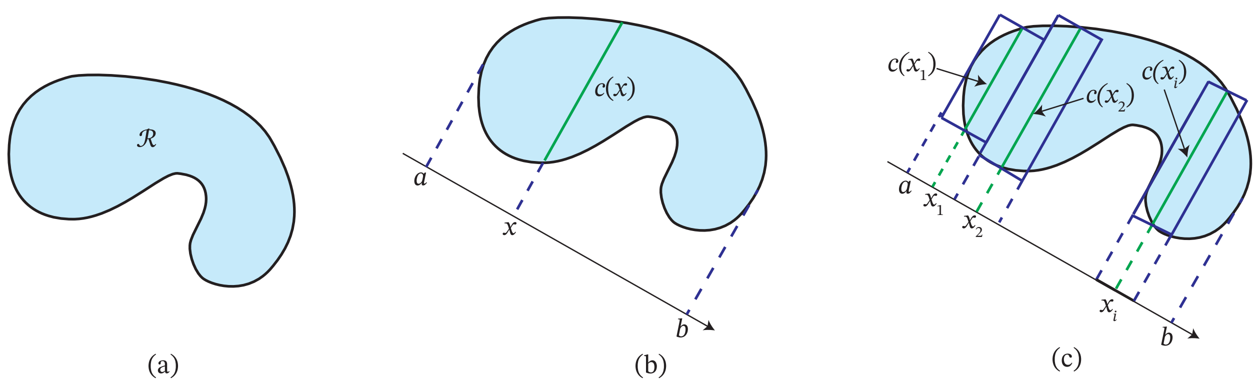


Figure 7.1.1

Assume that each line perpendicular to the axis for x in $[a, b]$ intersect the region R in an interval of length $c(x)$. The interval is called the **cross section** of R at x.

We approximate R by a collection of rectangles, just as we estimated the area of the region under $y = x^2$.

Pick an integer n and divide the interval $[a,b]$ on the axis into n congruent sections. The total length of the interval $[a,b]$ is $b-a$ and each section has width $\Delta x = (b-a)/n$. In the i^{th} section, $i = 1, 2, \ldots, n$, we pick a sampling number x_i. In the sections we form rectangles of width Δx and height $c(x_i)$, as indicated in Figure 7.1.1(c).

Since the i^{th} rectangle has area $c(x_i)\Delta x$, the total area of the rectangles is $\sum_{i=1}^{n} c(x_i)\Delta x$. As n increases, the total provides a better approximation to the area of R. This suggests that

$$\lim_{n\to\infty} \sum_{i=1}^{n} c(x_i)\Delta x = \text{Area of region } R$$

Observation 7.1.1: *Notation for Cross-Sectional Lengths and Sampling Points*

Since c is used for the cross-sectional length, it cannot be used to name the sampling point. Instead, x_i is used to denote the sampling point. This should not cause any confusion since we are not using x_i to describe the endpoints of a partition.

By the definition of a definite integral,

$$\lim_{n\to\infty} \sum_{i=1}^{n} c(x_i)\Delta x = \int_a^b c(x)\, dx.$$

This leads to the following definition.

Definition: *Area as Definite Integral of Cross-Sectional Length*

The area of a region equals the integral of its cross-sectional lengths.

$$\text{Area of } R = \int_a^b c(x)\, dx.$$

The axis need not be the x-axis of the xy-plane. It may be any line in the plane, and could even be the y-axis. Then we would denote the cross-sectional length by $c(y)$.

To compute an area

1. Find the endpoints a and b and the cross-sectional length $c(x)$.
2. Evaluate $\int_a^b c(x)\, dx$ by the fundamental theorem of calculus, if the antiderivative of $c(x)$ is elementary.

Chapter 6 showed how to accomplish Step 2. FTC I is used when the antiderivative is an elementary function. If it is not, then the integral can be approximated numerically. This section is concerned primarily with Step 1, how to find the cross-sectional length $c(x)$ and set up the definite integral.

If R is the region under the graph of $f(x)$ and above the interval $[a,b]$, then the cross-sectional length is simply $f(x)$. We met this special case in Sections 6.2–6.4 with $f(x) = x^2$ and $f(x) = 2^x$.

EXAMPLE 1. Find the area of a disk of radius r.

SOLUTION Introduce an xy-coordinate system with its origin at the center of the disk, as in Figure 7.1.2(a).

A cross section perpendicular to the x-axis is shown in Figure 7.1.2(b). Its length, $|AC|$, is twice $|BC|$. By the Pythagorean Theorem,

$$x^2 + |BC|^2 = r^2.$$

Then $|BC|^2 = r^2 - x^2$ and, because $|BC|$, a length, is positive, $|BC| = \sqrt{r^2 - x^2}$. Since x is between $-r$ and r,

$$\text{Area of disk of radius } r = \int_{-r}^{r} 2\sqrt{r^2 - x^2}\, dx. \tag{7.1.1}$$

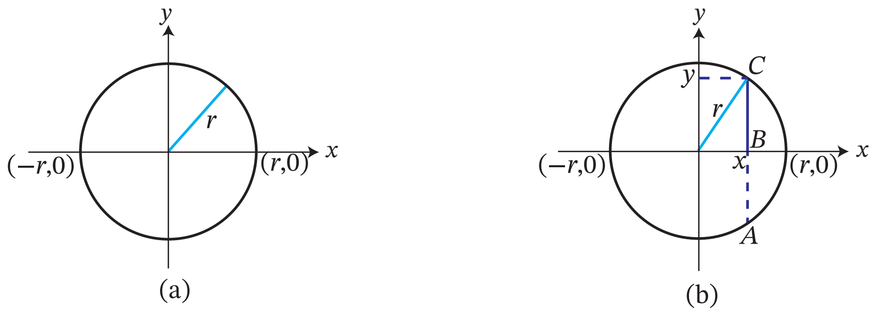


Figure 7.1.2

By symmetry, we can also say that the total area is four times the area of a quadrant:

$$\text{Area of disk of radius } r = 4\int_0^r \sqrt{r^2 - x^2}\, dx. \tag{7.1.2}$$

This completes the setup of the integral for the area of the region.

NOTE: One reason (7.1.2) is preferable to (7.1.1) because it reduces the chance of making an error when working with the subtraction of negative numbers.

The next chapter presents a technique for finding an antiderivative of $\sqrt{r^2 - x^2}$. In the meantime, we use the book's Table of Integrals (in Appendix A). According to formula 48,

$$\int \sqrt{r^2 - x^2}\, dx = \frac{r^2}{2}\left(\arcsin\left(\frac{x}{r}\right) + \frac{x}{r^2}\sqrt{r^2 - x^2}\right).$$

Thus, by FTC I,

$$\begin{aligned}
\int_0^r \sqrt{r^2 - x^2}\, dx &= \frac{r^2}{2}\left(\arcsin\left(\frac{x}{r}\right) + \frac{x}{r^2}\sqrt{r^2 - x^2}\right)\Bigg|_0^r \\
&= \frac{r^2}{2}\left(\arcsin\left(\frac{r}{r}\right) + \frac{r}{r^2}\sqrt{r^2 - r^2}\right) - \frac{r^2}{2}\left(\arcsin\left(\frac{0}{r}\right) + \frac{0}{r^2}\sqrt{r^2 - 0^2}\right) \\
&= \frac{r^2}{2}\left(\frac{\pi}{2}\right) \\
&= \frac{\pi r^2}{4}.
\end{aligned}$$

Thus one quarter of the disk has area $\pi r^2/4$ and the whole disk has area πr^2. ♦

Historical Note: Archimedes, a Parabola, and a Line

Archimedes found the area in the next example, expressing it in terms of the area of a triangle (see Exercise 41 in this section and also Exercise 1 in Section 7.S).

He used geometric properties of a parabola since calculus would not be available for another 1900 years.

REFERENCE: S. Stein, *Archimedes: What did he do besides cry Eureka?*, MAA, 1999.

EXAMPLE 2. Set up a definite integral for the area of a region above the parabola $y = x^2$ and below the line through $(2,0)$ and $(0,1)$ shown in Figure 7.1.3.

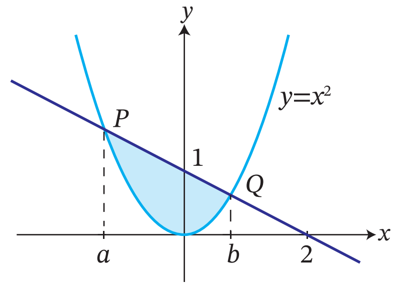


Figure 7.1.3

SOLUTION Since the x-intercept of the line is 2 and its y-intercept is 1, its equation is

$$\frac{x}{2} + \frac{y}{1} = 1.$$

Hence $y = 1 - x/2$. The length $c(x)$ of a cross section taken parallel to the y-axis is therefore

$$c(x) = \left(1 - \frac{x}{2}\right) - x^2 = 1 - \frac{x}{2} - x^2.$$

To find the interval of integration $[a,b]$ we find the x-coordinates of P and Q in Figure 7.1.3 where the line meets the parabola. There

$$x^2 = 1 - \frac{x}{2},$$

so

$$2x^2 + x - 2 = 0. \tag{7.1.3}$$

The solutions to (7.1.3) are

$$x = \frac{-1 \pm \sqrt{17}}{4}.$$

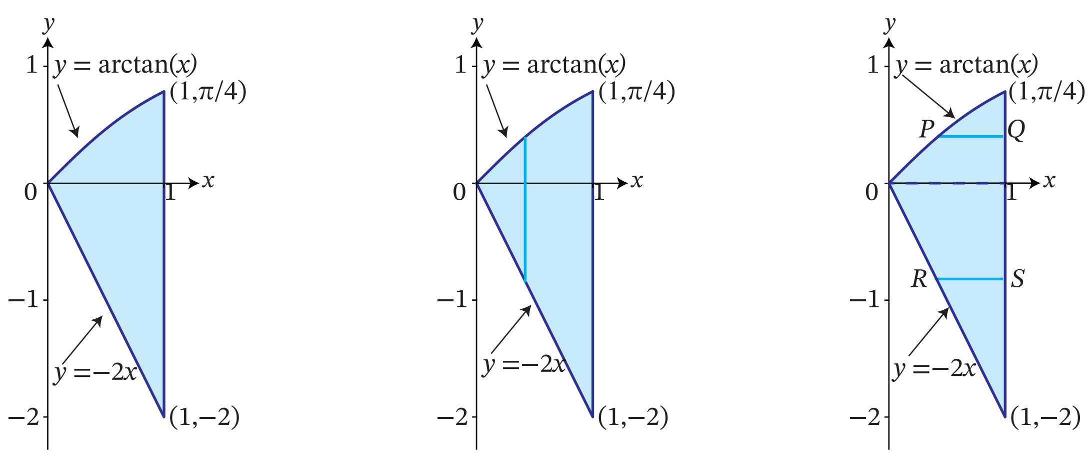
Hence

$$\text{Area} = \int_{(-1-\sqrt{17})/4}^{(-1+\sqrt{17})/4} \left(1 - \frac{x}{2} - x^2\right) dx.$$

The value of the definite integral is found in Exercise 33. ♦

EXAMPLE 3. Find the area of the region in Figure 7.1.4(a) bounded by $y = \arctan(x)$, $y = -2x$, and $x = 1$.

SOLUTION We will find the area two ways, first (a) with cross sections parallel to the y-axis, then (b) with cross sections parallel to the x-axis.



(a) (b) (c)

Figure 7.1.4

(a) A cross section has length $\arctan(x) - (-2x) = \arctan(x) + 2x$. (See Figure 7.1.4(b).) Thus the area is

$$\int_0^1 (\arctan(x) + 2x)\ dx.$$

We begin the evaluation of this integral by noticing that $\int 2x\ dx$ is x^2. To find $\int \arctan(x)\ dx$ we use Formula 90 from the book's Table of Integrals (in Appendix A): $\int \arctan(x)\ dx = x\arctan(x) - \ln(1+x^2)/2$.

By the fundamental theorem of calculus

$$\begin{aligned}
\int_0^1 (\arctan(x) + 2x)\ dx &= \left(x\arctan(x) - \frac{1}{2}\ln(1+x^2) + x^2\right)\Big|_0^1 \\
&= \left(1\arctan(1) - \frac{1}{2}\ln(1+1^2) + 1^2\right) \\
&\quad - \left(0\arctan(0) - \frac{1}{2}\ln(1+0^2) + 0^2\right) \\
&= \left(\frac{\pi}{4} - \frac{1}{2}\ln(2) + 1\right) - 0 \\
&= \frac{\pi}{4} + 1 - \frac{1}{2}\ln(2) \approx 1.43882. \qquad (7.1.4)
\end{aligned}$$

(b) Now we use cross sections parallel to the x-axis, as in Figure 7.1.4(c).

Cross sections above the x-axis meet the curved part of the boundary, while those below the x-axis meet the slanted line.

We determine the cross-sectional length as a function of y by finding the x-coordinates of P and Q, the ends of the cross section above the x-axis. The x-coordinate of Q is 1. If the x-coordinate of P is x then $y = \arctan(x)$ and $x = \tan(y)$. Hence

$$c(y) = 1 - \tan(y), \qquad \text{for } y \geq 0.$$

The length of RS, a cross section below the x-axis, is $1 - (x\text{-coordinate of } R)$. Since R is on the line $y = -2x$, we have $x = -y/2$. Thus

$$c(y) = 1 - (-y/2) = 1 + y/2 \qquad \text{for } -2 \leq y \leq 0.$$

The interval of integration is $[-2, \pi/4]$. Hence

$$\text{Area of } R = \int_{-2}^{\pi/4} c(y)\ dy.$$

We break the integral into two parts,

$$\int_{-2}^{0} \left(1 + \frac{y}{2}\right)\ dy \qquad \text{and} \qquad \int_0^{\pi/4} (1 - \tan(y))\ dy.$$

In Example 3 (in Section 8.5) it will be shown that

$$\int \tan(y)\ dy = \ln(\sec(y)).$$

To check this, differentiate $\ln(\sec(y))$ (see Exercise 31). Also, $\sec(y) > 0$ for $-\pi/2 < y < \pi/2$.

First,

$$\int_{-2}^{0} \left(1+\frac{y}{2}\right)\,dy = \left(y+\frac{y^2}{4}\right)\Big|_{-2}^{0} = \left(0+\frac{0^2}{4}\right) - \left((-2)+\frac{(-2)^2}{4}\right) = 1. \tag{7.1.5}$$

Second,

$$\int_{0}^{\pi/4} (1-\tan(y))\,dy = (y-\ln\sec(y))\Big|_{0}^{\pi/4} = \left(\frac{\pi}{4}-\ln(\sec(\frac{\pi}{4}))\right) - (0-\ln(\sec(0))) = \frac{\pi}{4}-\ln(\sqrt{2}). \tag{7.1.6}$$

Adding (7.1.5) and (7.1.6) gives

$$\text{Area of } R = \frac{\pi}{4} - \ln(\sqrt{2}) + 1 \approx 1.43882. \tag{7.1.7}$$

The expressions for the exact area in (7.1.4) and (7.1.7) may look different but they agree, as is confirmed in Exercise 32. ♦

We could have simplified the solution by observing that the area below the x-axis is a triangle of area 1, but we chose to illustrate the general approach.

Summary

The key idea in this section, that the area of a region equals the integral of cross-sectional length, was already anticipated in Chapter 6. There we saw the special case where the region is bounded by the graph of a function, the x-axis, and two lines perpendicular to the axis. In this section we looked at more general regions.

EXERCISES for Section 7.1

In Exercises 1 to 6 (a) draw the region, (b) determine the lengths of vertical cross sections ($c(x)$), and (c) determine the lengths of horizontal cross sections ($c(y)$).

1. The finite region bounded by $y=\sqrt{x}$ and $y=x^2$.
2. The finite region bounded by $y=x^2$ and $y=x^3$.
3. The finite region bounded by $y=2x$, $y=3x$, and $x=1$.
4. The finite region bounded by $y=x^2$ and $y=2x$.
5. The triangle with vertices $(0,0)$, $(3,0)$, and $(0,4)$.
6. The triangle with vertices $(1,0)$, $(3,0)$, and $(2,1)$.

In Exercises 7 to 12 find the areas. Use the book's Table of Integrals (in Appendix A) to find antiderivatives, if necessary.

7. Under $y=\sqrt{x}$ and above $[1,2]$.
8. Under $y=e^{2x}$ and above $[0,1]$.
9. Under $y=\sin(2x)$ and above $\left[\frac{\pi}{6},\frac{\pi}{3}\right]$.
10. Under $y=\cos(x)$ and above $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$.
11. Under $y=\ln(x)$ and above $[1,e]$.
12. Under $y=\dfrac{1}{\sqrt{1-x^2}}$ and above $\left[0,\frac{1}{2}\right]$.

In Exercises 13 to 20 find the areas of the region between the two curves using cross sections parallel to the x-axis.

13. $y = x^2$, $y = x^3$

14. $y = 2^x$, $y = 2x$

15. $y = \arcsin(x)$, $y = \frac{\pi}{2}x$ (with $x > 0$)

16. $y = \sin(x)$, $y = \cos(x)$ and above $\left[0, \frac{\pi}{4}\right]$

17. $y = 2^x$, $y = 3^x$ (for x in $[0,1]$)

18. $y = x^3$, $y = -x$ (for x in $[1,2]$)

19. $y = x^3$, $y = \sqrt[3]{2x-1}$ (for x in $[1,2]$)

20. $y = 1 + x$, $y = \ln(x)$ (for x in $[1, e]$)

In Exercises 21 to 27 set up a definite integral for the area of the given region.

Do not evaluate the integrals; they will be evaluated in Exercise 29 in this section and in Exercises 35 to 41 in Section 8.S.

21. The region under the curve $y = \arctan(2x)$ and above the interval $\left[\frac{1}{2}, \frac{1}{2\sqrt{3}}\right]$.

22. The region in the first quadrant below $y = -7x + 29$ and above $y = \frac{8}{x^2 - 8}$.

23. The region below $y = 10^x$ and above $y = \log_{10}(x)$ for x in $[1, 10]$.

24. The region under the curve $y = \frac{x}{x^2 + 5x + 6}$ and above the interval $[1, 2]$.

25. The region below $y = \frac{2x+1}{x^2 + x}$ and above the interval $[2, 3]$.

26. The region bounded by $y = \tan(x)$, $y = 0$, $x = 0$, and $x = \frac{\pi}{2}$
(a) by vertical cross sections and (b) by horizontal cross sections.

27. The region bounded by $y = \sin(x)$, $y = 0$, and $x = \frac{\pi}{4}$ (consider only $x \geq 0$)
(a) by vertical cross sections and (b) by horizontal cross sections.

28.
(a) Draw the region inside the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.
(b) Find a definite integral for the area of the ellipse in (a) using horizontal cross sections.
(c) Find a definite integral for the area of the ellipse in (a) using vertical cross sections.

Do not evaluate these definite integrals; the will be evaluated in Exercise 42 in Section 8.S.

29. Cross sections in different directions lead to different definite integrals for the same area. Both integrals must give the same area, but one of them can be easier to evaluate. Identify and evaluate the easier definite integral found (a) in Exercise 26 and (b) in Exercise 27.

30. Set up the definite integral for the area $A(b)$ of the region in the first quadrant under the curve $y = e^{-x}(\cos(x))^2$ and above the interval $[0, b]$.

31. In Example 3 it is asserted that $\int \tan(y)\, dy = \ln(\sec(y))$. Verify this by differentiating.

32. In Example 3 the area of the region bounded by $y = \arctan(x)$, $y = 2x$, and $x = 1$ is found to be $\frac{\pi}{4} + 1 - \frac{1}{2}\ln(2)$ and, also, $\frac{\pi}{4} - \ln(\sqrt{2}) + 1$. Explain why the answers are equal.

33. In Example 2 the area of the region above the parabola $y = x^2$ and below the line through $(2,0)$ and $(0,1)$ is found to be $\int\limits_{(-1-\sqrt{17})/4}^{(-1+\sqrt{17})/4} \left(1 - \frac{x}{2} - x^2\right)\, dx$. Evaluate the integral.

34. Let R be the region bounded by $y = x^3$, $y = x + 2$, and the y-axis.
(a) Find a definite integral for the area of R.
(b) Use a graph or other method to approximate the endpoints of the definite integral found in (a).
(c) Use the estimates in (b) to obtain an estimate of the area of R.

35. Let R be the region to the right of the y-axis bounded by $y = 3$ and $y = e^x/x$.
(a) Graph R.
(b) Find a definite integral for the area of R.
(c) Find approximate values for the limits of integration of the definite integral in (b).
(d) Because the antiderivative of e^x/x is not elementary, it is not easy to estimate the area of R. What methods do we have for estimating the integral? Use one of them to find an approximate value for the area of R.

36. Let a be a positive number. What fraction of the rectangle whose vertices are $(0,0)$, $(a,0)$, (a,a^4), and $(0,a^4)$1/13 is occupied by the region under the curve $y = x^4$ and above $[0,a]$?

37. Let $A(t)$ be the area of the region in the first quadrant between $y = x^2$ and $y = 2x^2$ and inside the rectangle bounded by $x = t$, $y = t^2$, and the coordinate axes. (The shaded region in Figure 7.1.5(a).) If $R(t)$ is the area of the rectangle, find (a) $\lim_{t\to 0} \frac{A(t)}{R(t)}$ and (b) $\lim_{t\to\infty} \frac{A(t)}{R(t)}$.

Figure 7.1.5

38. Figure 7.1.5(b) shows the graph of an increasing function $y = f(x)$ with $f(0) = 0$. Assume that $f'(x)$ is continuous and $f'(0) > 0$. Do not assume that $f''(x)$ exists. Our objective is to investigate

$$\frac{\text{Area of shaded region under the curve}}{\text{Area of triangle } ABC} \quad \text{as } t \text{ decreases toward 0.} \tag{7.1.8}$$

(a) Experiment with various functions, including some trigonometric functions and polynomials.
(b) Make a conjecture about (7.1.8) and explain why it is true.

39. Repeat Exercise 38, but now assume that $f'(0) = 0$, f'' is continuous, and $f''(0) \neq 0$.

40. Let f be an increasing function with $f(0) = 0$, and assume that it has an elementary antiderivative. Then f^{-1} is an increasing function, and $f^{-1}(0) = 0$. Prove that if f^{-1} is elementary, then it also has an elementary antiderivative. (See Figure 7.1.6(a).)

41. Show that the area of the shaded region in Figure 7.1.6(b) is two-thirds the area of the parallelogram $ABCD$. This is an illustration of a theorem of Archimedes concerning sectors of parabolas. He showed that the shaded area is 4/3 the area of triangle BOC. See also Example 2.

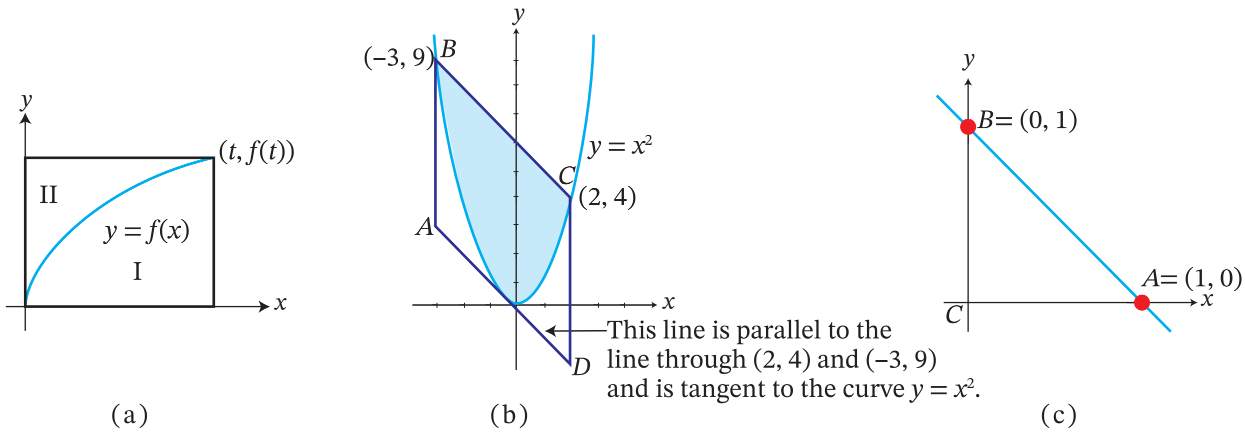


Figure 7.1.6

42. Figure 7.1.6(c) shows a right triangle ABC.
- (a) Find the equation of a line parallel to AB that cuts the triangle into two pieces of equal area.
- (b) Find the equation of a line parallel to BC that cuts the triangle into two pieces of equal area.
- (c) Find the equation of a line parallel to AC that cuts the triangle into two pieces of equal area.
- (d) Do the three lines found in (a), (b), and (c) meet at a single point? If so, find this point.

7.2 Some Pointers on Drawing

None of us was born knowing how to draw three-dimensional solids. As we grew up, we lived in flatland: Earth's surface. Few high school math classes cover solid geometry, so calculus is often the first place where one has to think and sketch in three dimensions. That is why we offer here a few words on how to draw. Too often you cannot work a problem simply because the diagrams you draw confuse even yourself. Our advice is not based on artistic principles. It derives from years attempting to sketch diagrams that do more good than harm.

Observation 7.2.1: *A Few Words of Advice about Mathematical Drawing*

1. *Draw large.* Small diagrams may have no room to place labels or to sketch cross sections.
2. *Draw neatly.* Use a straightedge to make straight lines that are actually straight. Use a compass to make circles that look like circles. Draw lines and curves slowly.
3. *Avoid clutter.* If you end up with too many labels or the cross section does not show up well, add separate diagrams for important parts of the figure.
4. *Practice.*

EXAMPLE 1. Draw a diagram of a ball of radius a that shows the circular cross section made by a plane at a distance x from the center of the ball. Use the diagram to help find the radius of the cross section as a function of x.

This example is continued in Example 1 in Section 7.4.

POOR SOLUTION Is the object in Figure 7.2.1 a potato or a ball? What segment has length r? What is x? What does the cross section look like?

REASONABLE SOLUTION Draw the ball carefully, as in Figure 7.2.2(a). Draw an equator to give it perspective. Add a little shading.

Show a typical cross section at a distance x from the center, as in Figure 7.2.2(b). Shading the cross section helps, too.

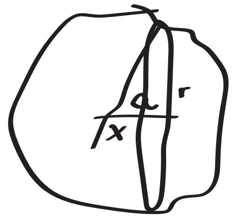


Figure 7.2.1

To find the radius of the cross section r in terms of x sketch a companion diagram. The radius is part of a right triangle. To avoid clutter, draw only the part of interest in a convenient side view, as in Figure 7.2.2(c).

The right triangle in this figure shows that $r^2 + x^2 = a^2$, hence (because r must be positive) $r = \sqrt{a^2 - x^2}$. ♦

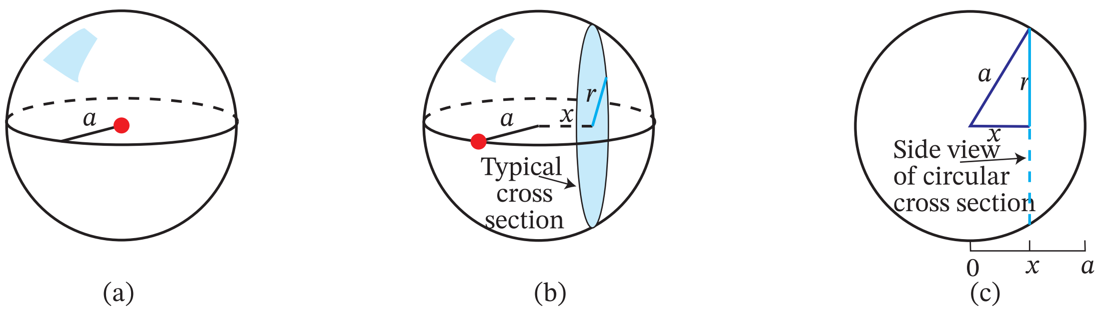


Figure 7.2.2

EXAMPLE 2. A pyramid has a square base with a side of length a. The top of the pyramid is above the center of the base at height h. Draw the pyramid and its cross sections by planes parallel to the base. Then find the area of the cross sections in terms of their distance x from the top.

This example is continued in Example 2 in Section 7.4.

POOR SOLUTION Figure 7.2.3 is too small; there is no room for the symbols. While it's pretty clear which side has length a, to what are the x and h attached? Also, without the hidden edges of the pyramid the shape of the base is not clear.

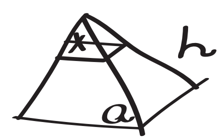


Figure 7.2.3

REASONABLE SOLUTION Draw a large pyramid with a square base, as in Figure 7.2.4(a). The opposite edges of the base are parallel lines. Then show a cross section in perspective and side views, as in Figures 7.2.4(b) and (c).

Figure 7.2.4(c) includes named labels of important points and lengths that are likely to be important in answering the question. Invisible lines are dashed. A vertical segment is added, and labeled, to show h. Right angles are clearly identified. The x-axis is drawn separate from the pyramid.

The use of s is chosen because it suggests its meaning – side.

As x increases, so does s, the width of the square cross section. Thus s is a function of x, which we call $s(x)$. Figure 7.2.4(b) shows that $s(0) = 0$ and $s(h) = 1$. To find $s(x)$ for all x in $[0, h]$ use the similar triangles ABC and ADE, shown in Figure 7.2.4(c). The triangles show that

$$\frac{x}{s} = \frac{h}{a}; \qquad \text{hence} \qquad s = \frac{ax}{h}. \tag{7.2.1}$$

Thus s is a linear function of x. As a check on (7.2.1), $s(0) = 0$ and $s(h) = a$, as expected. So the area A of a cross section is given by

$$A = s^2 = \left(\frac{ax}{h}\right)^2.$$ ♦

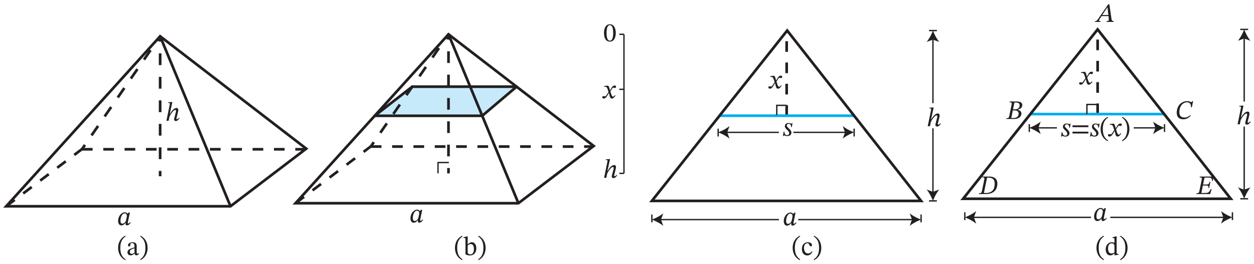


Figure 7.2.4

This example is continued in Example 2 and in Exercise 18, both in Section 7.4.

EXAMPLE 3. A cylindrical drinking glass of height h and radius a is full of water. It is tilted until the water covers exactly half the base.

(a) Draw a diagram of the glass and water.

(b) Using a plane perpendicular to the water surface and parallel to the axis of the cylinder, show a triangular cross section of the water.

(c) Find the area of the triangle in terms of the distance x of the cross section from the axis of the glass.

POOR SOLUTION The diagram in Figure 7.2.5 is too small to clearly present all of the needed information. It is not clear what has length a. The cross section is not clearly identified.

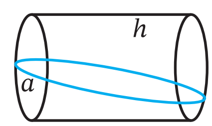


Figure 7.2.5

REASONABLE SOLUTION Draw a neat, large diagram of a slanted cylinder, as in Figure 7.2.6(a). Do not put in too much detail at first. When showing the cross section, draw only the water. Figures 7.2.6 and 7.2.7 show various views in both two and three dimensions. Let u and v be the lengths of the legs of the cross section, as shown in Figure 7.2.7(d).

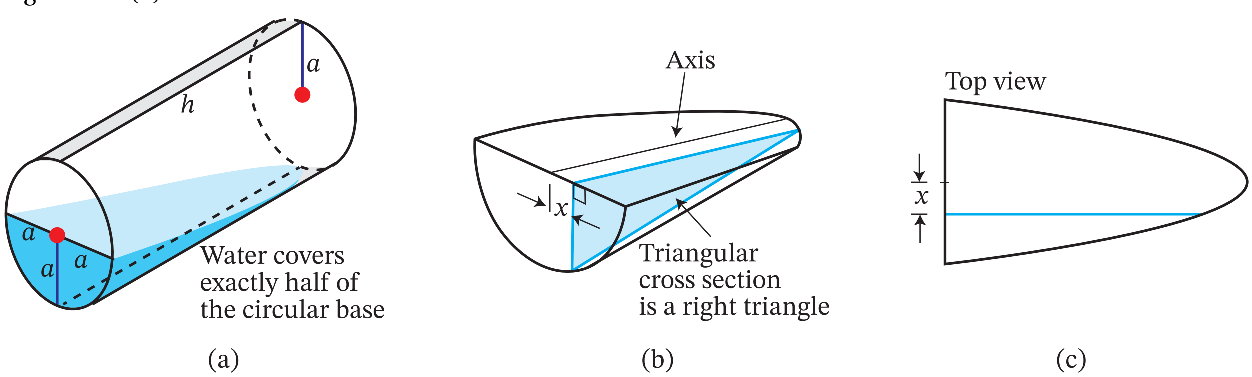


Figure 7.2.6

Comparing Figures 7.2.7(a) and (b), we have, by similar triangles, the relation

$$\frac{u}{a} = \frac{v}{h} \qquad \text{hence} \qquad v = \frac{h}{a}u.$$

Let $A(x)$ be the area of the cross section at a distance x from the center of the base, as shown in Figure 7.2.6(b). If we can find u as a function of x, the cross-sectional area $A(x) = \frac{1}{2}uv$ known in terms of x. Figure 7.2.7(b) suggests how to find u. Copy it and draw in the necessary radius, as in Figure 7.2.7(d). By the Pythagorean Theorem,

$$u = \sqrt{a^2 - x^2}.$$

Thus

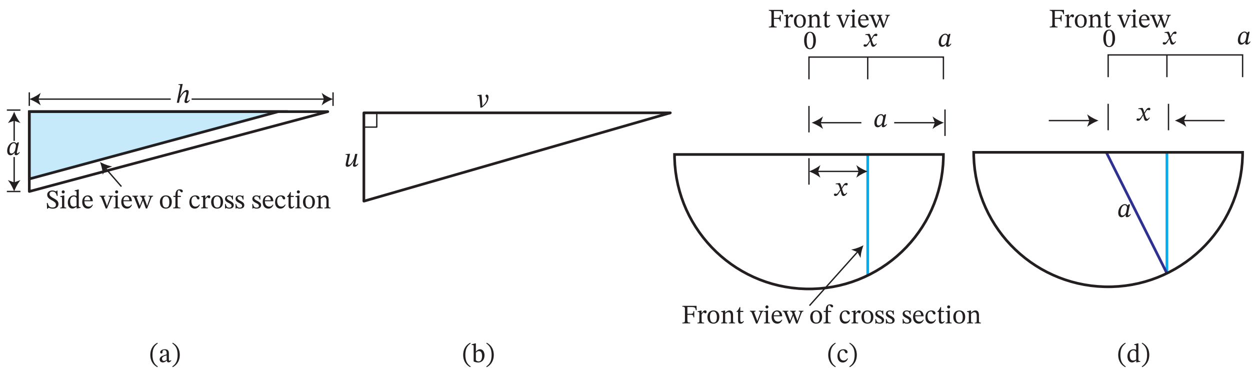


Figure 7.2.7

$$A(x) = \frac{1}{2}uv = \frac{1}{2}u\left(\frac{h}{a}u\right) = \frac{h}{2a}u^2 = \frac{h}{2a}(a^2 - x^2). \qquad (7.2.2)$$

As a check, note that

$$A(a) = \frac{h}{2a}(a^2 - a^2) = 0 \qquad \text{and} \qquad A(0) = \frac{h}{2a}(a^2 - 0^2) = \frac{1}{2}ah.$$

♦

Summary

The theme of this section was the importance of clear diagrams. Circles should look like circles, whether freehand or done with a compass or a jar lid. Straight lines should be straight, whether done freehand or with a ruler.

EXERCISES for Section 7.2

1. Cross sections of the pyramid in Example 2 are made using planes perpendicular to the base and parallel to an edge of the base. What is the area of the cross section made by a plane that is a distance x from the pyramid's top?

(a) Draw a large perspective view of the pyramid.
(b) Copy the diagram in (a) and shade the cross section.
(c) Draw a side view that shows the shape of the cross section.

2. Cross sections of the water in Example 3 are made by using planes parallel to the plane that passes through the horizontal diameter of the base and the axis of the glass. What is the area of the cross section made by a plane that is a distance x from the center of the base?

(a) Draw a large perspective view of the water and glass.
(b) Copy the diagram in (a) and show the typical cross section shaded.
(c) Draw a view that shows the shape of the cross section.
(d) Draw the water and glass from a different view.
(e) Put labels, such as x, a, and h, on the diagrams. (You will need to introduce more labels.)
(f) Find the area of the cross section, $A(x)$, as a function of x.

3. Cross sections of the water in Example 3 are made by using planes perpendicular to the axis of the glass. Make clear diagrams, including perspective and side views, that show the cross sections. Do not find their areas.

4. A wedge is to be cut out of a cylindrical tree of radius a. The first cut is parallel to the ground and stops at the axis of the tree. The second cut makes an angle θ with the first cut and meets it along a diameter.

(a) Draw a cross section that is a triangle.
(b) Find the area of the triangle as a function of x, the distance of the plane from the axis of the tree.
(c) Draw a cross section that is a rectangle.
(d) Find the area of the rectangle as a function of x, the distance of the plane from the axis of the tree.

5. A cylindrical glass is partially full of water. The glass is tilted until the remaining water just covers the base of the glass. The radius of the glass is a and its height is h. Consider parallel planes such that cross sections of the water are rectangles. The amount of water in this tank is found in Exercise 15, in Section 7.4.

(a) Make clear diagrams that show the situation. (Include a top view to show the cross sections.)

(b) Obtain a formula for the area of the cross sections. (The two planes at the same distance x from the axis of the glass cut out cross sections of different areas. Introduce an x-axis with 0 at the center of the base and extending from $-a$ to a in a convenient direction.)

6. Repeat Exercise 5, but this time use parallel planes such that the cross sections are trapezoids. To obtain trapezoidal cross sections, pick the planes parallel to the plane through the axis of the cylinder and the point where the water touches the top of the glass. The amount of water in this tank is found in Exercise 16, in Section 7.4.

7. A right circular cone has a radius a and height h as shown in Figure 7.2.8(a). Make cross sections with planes parallel to the base.

(a) Draw perspective and side views of the cone and typical cross sections.

(b) Drawing as many diagrams as necessary, find the area of the cross section made by a plane at a distance x from the vertex of the cone.

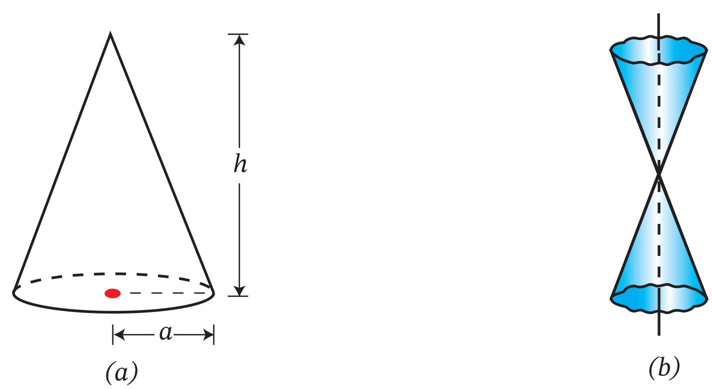


Figure 7.2.8

8. Draw the cross section made by a plane parallel to the axis of the cone. Draw perspective and side views, but do not find a formula for the area of the cross section. (See Exercise 7.)

9. Figure 7.2.8(b) shows an unbounded solid right circular cone. Draw a cross section that is bounded by (a) a circle, (b) an ellipse (that is not a circle), (c) a parabola, and (d) a hyperbola.

10. Draw a cross section of a bounded right circular cylinder that is (a) a circle, (b) an ellipse that is not a circle, and (c) a rectangle.

11. Draw a cross section of a solid cube that is (a) a square, (b) an equilateral triangle, (c) a five-sided polygon, and (d) a regular hexagon.

12. The plane region between the curves $y = x$ and $y = x^2$ is spun around the x-axis to produce a solid resembling the bell of a trumpet.

(a) Draw the plane region.

(b) Draw the solid region produced by spinning this region around the x-axis.

(c) Draw cross sections made by a plane perpendicular to the x-axis; show both perspective and side views.

(d) Find the area of the cross section in terms of the distance x of the plane from the origin to the x-axis.

13. Obtain a circular stick such as a broom handle or a dowel. Saw off a piece, making one cut perpendicular to the axis and the second cut at an angle to the axis. Mark on the surface of the piece you cut out the borders of cross sections that are (a) rectangles and (b) trapezoids.

7.3 Setting Up a Definite Integral

This section presents a shortcut for setting up a definite integral. The formal (complete) and informal (shortcut) approaches are contrasted in the case of setting up the definite integral for area. Then the informal approach will be illustrated as it is applied in several contexts.

The Complete Approach

To show that area is represented by a definite integral, $A = \int_a^b f(x)\ dx$, we partitioned the interval $[a,b]$ by the numbers $x_0 < x_1 < x_2 < \cdots < x_n$ with $x_0 = a$ and $x_n = b$. A sampling number c_i was chosen in each section $[x_{i-1}, x_i]$. For convenience, all the sections were of equal length, $\Delta x = (b-a)/n$. (See Figure 7.3.1(a).)

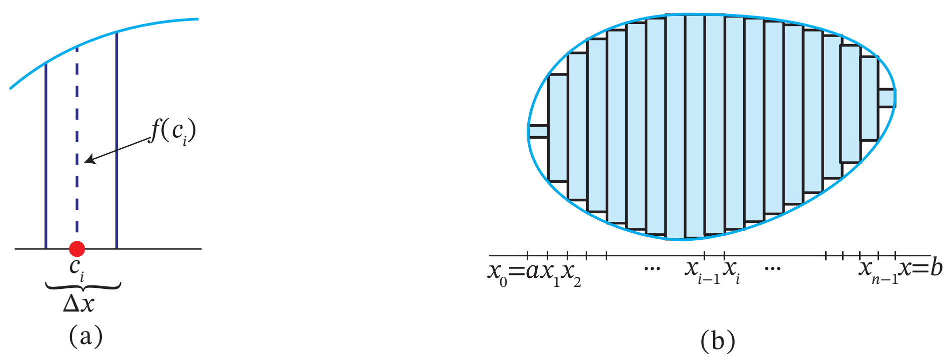


Figure 7.3.1

We then formed the sum

$$\sum_{i=1}^{n} f(c_i)\Delta x. \tag{7.3.1}$$

and noted that it equaled the total area of the rectangular approximation of the region whose boundary is the cyan curve in Figure 7.3.1(b).

As Δx approaches 0, the sum (7.3.1) approaches the area of the region. By the definition of the definite integral, (7.3.1) approaches

$$\int_a^b f(x)\ dx.$$

This process is the justification behind the statement that

$$\text{Area} = \int_a^b f(x)\ dx. \tag{7.3.2}$$

This is the complete, or formal, approach to obtain formula (7.3.2). Next, we develop a more efficient approach that we will used repeatedly.

The Shortcut Approach

The heart of the complete approach is the local estimate $f(c_i)\Delta x$ for the area of a rectangle of height $f(c_i)$ and width Δx, as is shown in Figure 7.3.2.

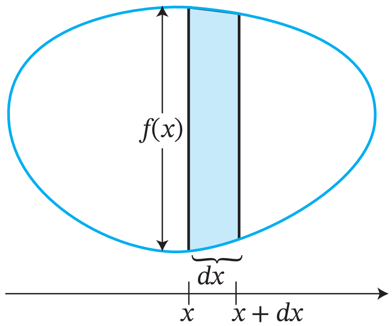


Figure 7.3.2

In the shortcut approach we focus on the local approximation. We do not mention the partition or the sampling numbers. We illustrate this shortcut approach by obtaining (7.3.2) informally. This is not a new method of integration, but a way to save time when setting up an integral.

For a small positive number dx, what would be a good approximation to the area of the region corresponding to the short interval $[x, x+dx]$ of width dx shown in Figure 7.3.2? The area of the rectangle of width dx and height $f(x)$ shown in Figure 7.3.2 would seem to be a plausible approximation. The area of this thin rectangle is

$$f(x)\,dx. \tag{7.3.3}$$

Without further ado, we then write

$$\text{Area} = \int_a^b f(x)\,dx, \tag{7.3.4}$$

which is (7.3.2). The leap from the local approximation (7.3.3) to the definite integral (7.3.4) omits many steps of the complete approach. The informal approach is the method used in most applications of calculus. It is the method most commonly used by engineers, physicists, biologists, economists, and mathematicians to set up integrals in applications.

The Volume of a Ball

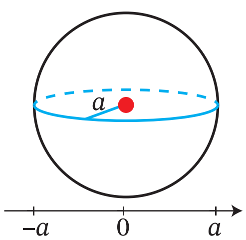


Figure 7.3.3

EXAMPLE 1. Find the volume of a ball of radius a. First use the complete approach. Then use the shortcut approach.

SOLUTION Both approaches require good diagrams. In the complete approach we show an x-axis, a partition into sections of equal lengths, sampling numbers c_i, and the approximating disks. See Figures 7.3.3 and 7.3.4(a). The thickness of a disk is Δx, as shown in the side view of Figure 7.3.4(b), while its radius is labeled r_i, as shown in the end view of Figure 7.3.4(c). The volume of the disk is

$$\pi r_i^2(\Delta x). \tag{7.3.5}$$

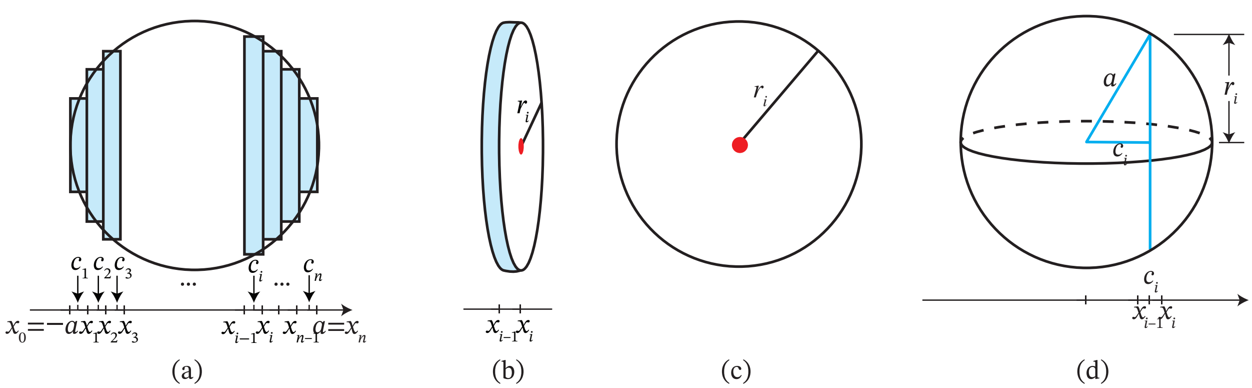


Figure 7.3.4

Figure 7.3.4(d) helps us determine r_i. By the Pythagorean Theorem,

$$r_i^2 = a^2 - c_i^2. \tag{7.3.6}$$

Combining (7.3.1), (7.3.5), and (7.3.6) gives an estimate of the volume of a sphere of radius a:

$$\sum_{i=1}^{n} \pi(a^2 - c_i^2)\Delta x. \tag{7.3.7}$$

By the definition of the definite integral,

$$\lim_{\Delta x \to 0} \sum_{i=1}^{n} \pi(a^2 - c_i^2)\Delta x = \int_{-a}^{a} \pi(a^2 - x^2)\ dx.$$

Hence

$$\text{Volume of ball of radius } a = \int_{-a}^{a} \pi(a^2 - x^2)\ dx.$$

By the fundamental theorem of calculus, the integral equals $4\pi a^3/3$.

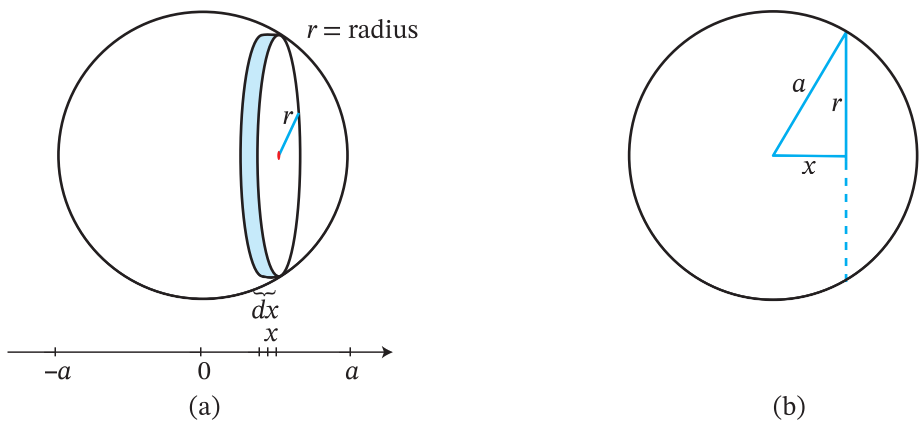


Figure 7.3.5

For the shortcut approach, draw only a short section of an x-axis and label its length dx. Then draw an approximating disk, whose radius is labeled r, as in Figure 7.3.5(a). Since the disk has a base of area πr^2 and thickness dx, its volume is $\pi r^2\ dx$. Moreover, as Figure 7.3.5(b) shows, $r^2 = a^2 - x^2$. Hence the local approximation is

$$\pi(a^2 - x^2)\ dx. \tag{7.3.8}$$

Then, without further ado, without choosing any c_i or showing any approximating sum, we have

$$\text{Volume of ball of radius } a = \int_{-a}^{a} \pi(a^2 - x^2)\ dx.$$

The local approximation (7.3.8) in differential form gives the integrand. The limits of integration are determined separately. ◆

Volcanic Ash

EXAMPLE 2. After the explosion of a volcano, ash gradually falls to the ground. Its depth diminishes with distance from the volcano. Assume that the depth of the ash at a distance x feet from the volcano is Ae^{-kx} feet, where A and k are positive constants. Set up a definite integral for the total volume of ash that falls within a distance b of the volcano.

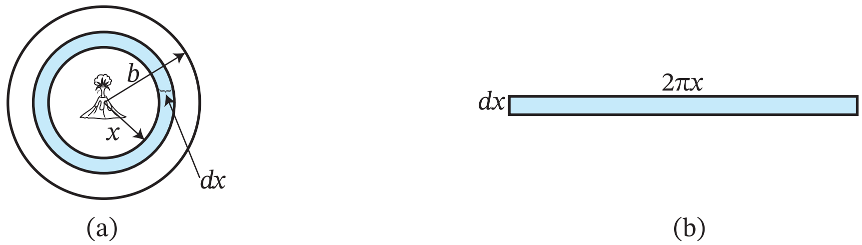


Figure 7.3.6

SOLUTION Estimate the volume of ash that falls on a narrow ring of width dx and inner radius x centered at the volcano. (See Figure 7.3.6(a).) On the ring the depth is almost constant.

Its area is approximately that of a rectangle of length $2\pi x$ and width dx. (See Figure 7.3.6(b).) So the area of the ring is approximately

$$2\pi x\, dx.$$

NOTE: In Exercise 4 it is shown that the exact area of the ring is $2\pi x\, dx + \pi(dx)^2$.

A good estimate of the depth of the ash throughout the ring is Ae^{-kx}. Thus the volume of the ash that falls on the typical ring of inner radius x and outer radius $x + dx$ is approximately

$$Ae^{-kx}(2\pi x)\, dx \text{ cubic feet.} \tag{7.3.9}$$

NOTE: The limits of integration must be determined just as in the formal approach.

Once we have the local estimate (7.3.9), we write the definite integral for the total volume of ash that falls within a distance b of the volcano:

$$\text{Total volume} = \int_0^b Ae^{-kx} 2\pi x\, dx.$$

This completes the shortcut to setting up of the definite integral. To evaluate the integral use a formula from the book's Table of Integrals (in Appendix A) or a technique in Chapter 8. ♦

Kinetic Energy

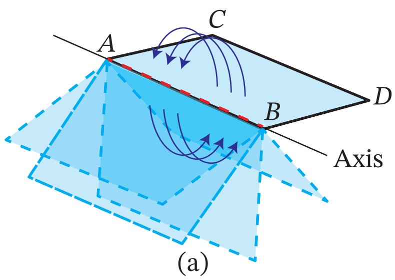
The next example concerns kinetic energy. The kinetic energy of an object with mass m kilograms and velocity v meters per second is defined as

$$\text{Kinetic energy} = \frac{mv^2}{2} \text{joules.}$$

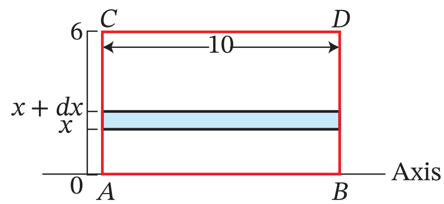
If all parts of an object are not moving at the same speed, an integral is needed to express the total kinetic energy. The next example illustrates how this can be done.

EXAMPLE 3. A thin rectangular piece of sheet metal is spinning around one of its longer edges 3 times per second, as shown in Figure 7.3.7(a). The length of its shorter edge is 6 meters and the length of its longer edge is 10 meters. The density of the sheet metal is 4 kilograms per square meter. Find its kinetic energy.

SOLUTION The farther a mass is from the axis, the faster it moves, and therefore the larger its kinetic energy. To find the total kinetic energy of the rotating piece of sheet metal, divide it into narrow rectangles of length 10 meters and width dx meters parallel to the edge AB, as shown in Figure 7.3.7(b). (Introduce an x-axis parallel to edge AC with the origin corresponding to A.) Since all points of the narrow rectangle move at roughly the same speed, its kinetic energy can be approximated. This provides the key local approximation.



(a)



(b)

Figure 7.3.7

The typical rectangle has area 10 dx square meters and its density is 4 kilograms per square meter. So, the mass of the typical rectangle is

$$4 \cdot 10 \cdot dx \text{ kilograms.}$$

We now approximate its velocity. The rectangle spins 3 times per second around a circle of radius x. In 1 second each point in it covers a distance of about

$$3 \cdot 2\pi x = 6\pi x \text{ meters.}$$

Consequently, the velocity of the rectangle is

$$6\pi x \text{ meters per second.}$$

The local approximation of the rectangle's kinetic energy is therefore

$$\frac{1}{2}\underbrace{(40\ dx)}_{\text{mass}}\ \underbrace{(6\pi x)^2}_{\text{velocity squared}} \quad \text{joules}$$

or

$$720\pi^2 x^2\ dx \text{ joules.} \tag{7.3.10}$$

Having obtained the local approximation (7.3.10), we write the definite integral and conclude that

$$\text{Total kinetic energy of spinning rectangle} = \int_0^6 720\pi^2 x^2\ dx \text{ joules.}$$

♦

Summary

This section described an efficient way to set up a definite integral for a quantity Q. In it we approximate how much of the quantity Q corresponds to a short section $[x, x+dx]$ of the x-axis, say $f(x)\ dx$. Then $Q = \int_a^b f(x)\ dx$, where a and b are appropriate constants.

EXERCISES for Section 7.3

1. In Section 6.4 we showed that if $f(t)$ is the velocity at time t of an object moving along the x-axis, then $\int_a^b f(t)\ dt$ is the change in position during the time interval $[a, b]$. Derive this in the style of this section. Keep in mind that $f(t)$ may be positive or negative.

2. The depth of rain at a distance r feet from the center of a storm is $g(r)$ feet.

(a) Estimate the total volume of rain between a distance r feet and a distance $r + dr$ feet from the center of the storm. Assume that dr is a small positive number.

(b) Using (a), set up a definite integral for the total volume of rain between 1,000 and 2,000 feet from the center of the storm.

3. Consider a disk of radius a with a home base of production at its center. Let $G(r)$ denote the density of foodstuffs (in calories per square meter) produced at a distance r meters from the home base. Then the total number of calories produced in the range is given by what definite integral?

REFERENCE: This analysis of primitive agriculture is taken from *Is There an Optimum Level of Population?*, edited by S. Fred Singer, McGraw-Hill, New York, 1971.

4. In Example 2 the area of the ring with inner radius x and outer radius $x + dx$ was estimated to be about $2\pi x\, dx$.
(a) Using the formula for the area of a disk, show that the area of the ring is $2\pi x\, dx + \pi(dx)^2$.
(b) Show that the ring has the same area as a trapezoid of height dx and bases of lengths $2\pi x$ and $2\pi(x + dx)$.

5. Think of a circular disk of radius a as being composed of concentric circular rings, as in Figure 7.3.8(a).
(a) Use the shortcut to set up a definite integral for the area of the disk.
(b) Evaluate the integral found in (a).

Figure 7.3.8

Exercises 6 to 8 concern the volumes of solids. In each case, (a) draw a good picture of the local approximation of width dx, (b) use the shortcut to set up the appropriate definite integral, and (c) evaluate the integral.

6. A right circular cone of radius a and height h.

7. A pyramid with a square base of side a and of height h. Its top vertex is above one corner of the base.

8. A pyramid with a triangular base of area A and of height h. The triangle can be any shape. See Figure 7.3.8(b).

---

9. At time t hours, $0 \le t \le 24$, a firm uses electricity at the rate of $e(t)$ joules per hour. The cost per joule at time t is $c(t)$ dollars. Assume that both e and c are continuous functions.
(a) Estimate the cost of electricity consumed between times t and $t + dt$, where dt is a small positive number.
(b) Using (a), set up a definite integral for the total cost of electricity for the 24-hour period.

10. The **present value** of one dollar t years from now is $g(t)$ dollars.
(a) What is $g(0)$?
(b) Why is it reasonable to assume that $g(t) \le 1$ and that g is a decreasing function of t?
(c) What is the present value of q dollars t years from now?
(d) Assume that an investment made now will result in an income flow at the rate of $f(t)$ dollars per year t years from now. Assume that f is a continuous function. Estimate informally the present value of the income to be earned between time t and time $t + dt$, where dt is a small positive number.
(e) On the basis of the local estimate made in (d), set up a definite integral for the present value of the income earned during the next b years.

11. Let the number of females in a population in the age range from x years to $x + dx$ years, where dx is a small positive number, be approximately $f(x)\ dx$. Assume that, on average, women of age x produce $m(x)$ offspring during the year before they reach age $x + 1$. Assume that both f and m are continuous functions.

(a) What definite integral represents the number of women between ages a and b years?

(b) What definite integral represents the total number of offspring during the calendar year produced by women whose ages at the beginning of the calendar year were between a and b years?

Exercises 12 to 16 concern **kinetic energy**. A particle of mass m moving with velocity v has kinetic energy $mv^2/2$. (See Example 3.) An object whose density is the same at all its points is called **homogeneous**. If the object is planar and has mass m kilograms and area A square meters, its density is m/A kilograms per square meter.

12. The piece of sheet metal in Example 3 is rotated around the line midway between the edges AB and CD at the rate of 5 revolutions per second.

(a) Using the shortcut approach, obtain a local approximation for the kinetic energy of a narrow strip of the sheet metal.

(b) Using (a), set up a definite integral for the kinetic energy of the piece of sheet metal.

(c) Evaluate the integral in (b).

13. A circular piece of metal of radius 7 meters has a density of 3 kilograms per square meter. It rotates 5 times per second around an axis perpendicular to the circle and passing through the center of the circle.

(a) Devise a local approximation for the kinetic energy of a narrow ring in the circle.

(b) Set up a definite integral for the kinetic energy of the rotating metal.

(c) Evaluate the integral in (b).

14. The density of a rod x centimeters from its left end is $g(x)$ grams per centimeter. The rod has a length of b centimeters. The rod is spun around its left end seven times per second.

(a) Estimate the mass of the rod in the section between x and $x + dx$ centimeters from the left end.

(b) Estimate the kinetic energy of the mass in (a).

(c) Set up a definite integral for the kinetic energy of the rotating rod.

15. A homogeneous square of mass m kilograms and side a meters rotates around an edge five times per second.

(a) Obtain a local estimate of the kinetic energy. What part of the square would you use? Why? Draw it.

(b) What is the local estimate?

(c) What definite integral represents the total kinetic energy of the square?

(d) Evaluate it.

16. Repeat Exercise 15 for a disk of radius a and mass M spinning around a line through its center and perpendicular to it. It is spinning at the rate of ω radians per second. (See Figure 7.3.9.)

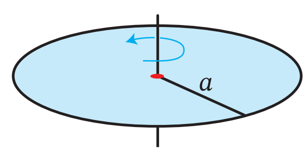


Figure 7.3.9

17. Repeat Exercise 15 for a square spun around a line through its center and parallel to an edge.

Exercises 18 and 19 contain definite integrals that cannot be evaluated by the fundamental theorem of calculus (since the desired antiderivative is not elementary). Use (a) the trapezoidal rule and (b) Simpson's method, each with six sections, to estimate the given definite integral.

18. A homogeneous object of mass M occupies the region under $y = e^{x^2}$ and above $[0, 1]$. It is spun at the rate of ω radians per second around the y-axis. Estimate its kinetic energy.

19. A homogeneous object of mass M occupies the region under $y = \sin(x)/x$ and above $[\pi/2, \pi]$. It is spun around the line $x = 1$ at the rate of ω radians per second. Estimate its kinetic energy.

In Exercises 20 to 23 find a definite integral for the kinetic energy of a plane homogeneous object that occupies the given region, has mass M, and is spun around the y-axis ω radians per second.

Do not attempt to evaluate these integrals. We will learn how to evaluate integrals like these in Chapter 8.

20. The region under $y = e^x$ and above the interval $[1,2]$.

21. The region under $y = \arctan(x)$ and above the interval $[0,1]$.

22. The region under $y = \dfrac{1}{1+x}$ and above $[2,4]$.

23. The region under $y = \sqrt{1+x^2}$ and above $[0,2]$.

---

24. A solid homogeneous right circular cylinder of radius a, height h, and mass M is spun at the rate of ω radians per second around its axis. Find its kinetic energy.

25. A solid homogeneous ball of radius a and mass M is spun at the rate of ω radians per second around a diameter. Find its kinetic energy.

26. (*Actuarial tables*) Let $F(t)$ be the fraction of people born in 1900 who are alive t years later, $0 \le F(t) \le 1$.

- (a) What is $F(150)$, probably?
- (b) What is $F(0)$?
- (c) Sketch the general shape of the graph of $y = F(t)$.
- (d) Let $f(t) = F'(t)$. Is $f(t)$ positive or negative? Assume F is differentiable.
- (e) What fraction of the people born in 1900 die during the time interval $[t, t+dt]$? Express your answer in terms of F.
- (f) Repeat (e), but express your answer in terms of f.
- (g) Evaluate $\int_0^{150} f(t)\,dt$.
- (h) What integral would you call the average life span of people born in 1900? Why?

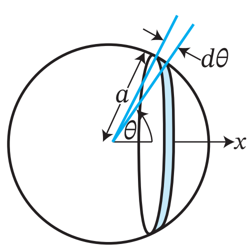


Figure 7.3.10

27. Let $F(t)$ be the fraction of ball bearings that wear out during the first t hours of use, so $F(0) = 0$ and $F(t) \le 1$. Assume F is differentiable.

- (a) As t increases, what happens to $F(t)$?
- (b) Show that during the short interval of time $[t, t+dt]$, the fraction of ball bearings that wear out is approximately $F'(t)\,dt$.
- (c) Assume all wear out in at most 1,000 hours. What is $F(1{,}000)$?
- (d) Use (b) and (c) to devise a definite integral for the average life of the ball bearings.

28. Find the surface area of a sphere of radius a. (See Figure 7.3.10.)

29. (*Poiseuille's law of fluid flows*) Assume a fluid flows through a narrow pipe of radius R with speed $f(r)$ where r is the distance from the center of the pipe. In CIE 17, at the end of this chapter, we will learn that the rate at which the fluid flows through the pipe is proportional to $\displaystyle\int_0^R r f(r)\,dr$.

- (a) What does $f(R) = 0$ mean, in physical terms?
- (b) Assume that $f(r) = R - r$. Show that the rate at which fluid flows through the pipe is proportional to R^3.
- (c) Assume that $f(r) = R^2 - r^2$. Show that the rate at which fluid flows through the pipe is proportional to R^4.

30. The density of Earth at a distance of r miles from its center is $g(r)$ pounds per cubic mile. Set up a definite integral for the total mass of Earth. Assume the radius of the earth to be 4,000 miles.

7.4 Computing Volumes by Parallel Cross Sections

In Section 6.1 we computed areas by integrating lengths of cross sections made by parallel lines. In this section we will use a similar approach, finding volumes by integrating areas of cross sections made by parallel planes. We already saw an example of this method when we represented the volume of a tent as a definite integral.

Cylinders

Let $\mathscr{B}$ be a region in the plane (see Figure 7.4.1(a)) and h a positive number. The **cylinder with base $\mathscr{B}$ and height** h consists of all line segments of length h perpendicular to $\mathscr{B}$, one end of which is in $\mathscr{B}$ and the other end is on a side of $\mathscr{B}$. A cylinder is shown in Figure 7.4.1(b). The top of the cylinder is congruent to $\mathscr{B}$.

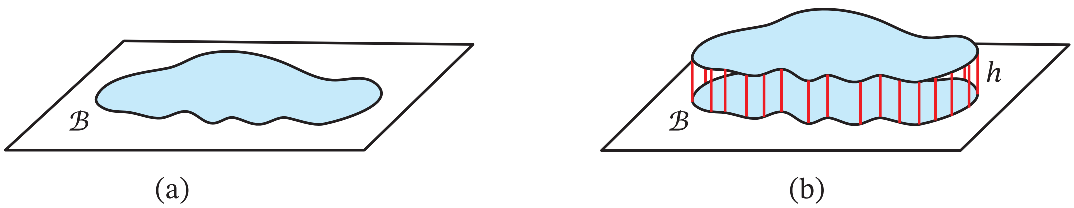


Figure 7.4.1

If $\mathscr{B}$ is a disk, the cylinder is a circular cylinder (see Figure 7.4.2(a)). If $\mathscr{B}$ is a rectangle, the cylinder is a rectangular box (see Figure 7.4.2(b)).

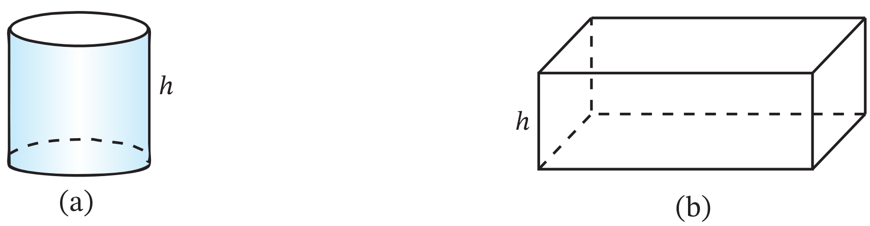


Figure 7.4.2

We will make use of the formula for the volume of a cylinder with base $\mathscr{B}$ and height h:

$$\begin{aligned} V &= \text{Area of Base} \times \text{Height} \\ &= (\text{Area of } \mathscr{B}) \times h. \end{aligned}$$

Volume as the Definite Integral of Cross-Sectional Area

We will use the shortcut approach for setting up a definite integral to see how integration is used to calculate volumes of solids.

Figure 7.4.3 shows the solid region $\mathscr{R}$ that lies between the planes perpendicular to the x-axis at $x = a$ and at $x = b$. We use a cylinder to estimate the volume of the part of $\mathscr{R}$ that lies between two parallel planes a small distance dx apart, shown in perspective in Figure 7.4.3(b). This slab is not usually a cylinder, because the sides are sloped (Figure 7.4.3(c)), but we can approximate it by a cylinder (Figure 7.4.3(d)).

Let x be, say, the left endpoint of an interval of width dx. The plane perpendicular to the x-axis at x intersects $\mathscr{R}$ in a plane cross section of area $A(x)$. The cylinder whose base is the cross section and whose height is dx is a good approximation of the part of $\mathscr{R}$ between the planes corresponding to x and $x + dx$. We therefore have

$$\text{Local Approximation to Volume} = A(x)\ dx.$$

Then, since the volume equals the integral of cross-sectional area,

$$\text{Volume of Solid} = \int_a^b A(x)\,dx.$$

To apply this idea, we compute $A(x)$, for which we need good drawings.

Algorithm: Finding the Volume of a Solid

Given a solid, if we can find a, b and the cross-sectional area $A(x)$ we can write a definite integral for its volume. The steps for finding the volume of a solid are

1. Choose a line to serve as an x-axis. (See Figure 7.4.3(a).)
2. For each plane perpendicular to that axis, find the area of the cross section of the solid made by the plane. Call the area $A(x)$. (See Figure 7.4.3(b).)
3. Determine the limits of integration, a and b, for the region.
4. The volume of this solid is given by the definite integral $\int_a^b A(x)\,dx$. Evaluate it.

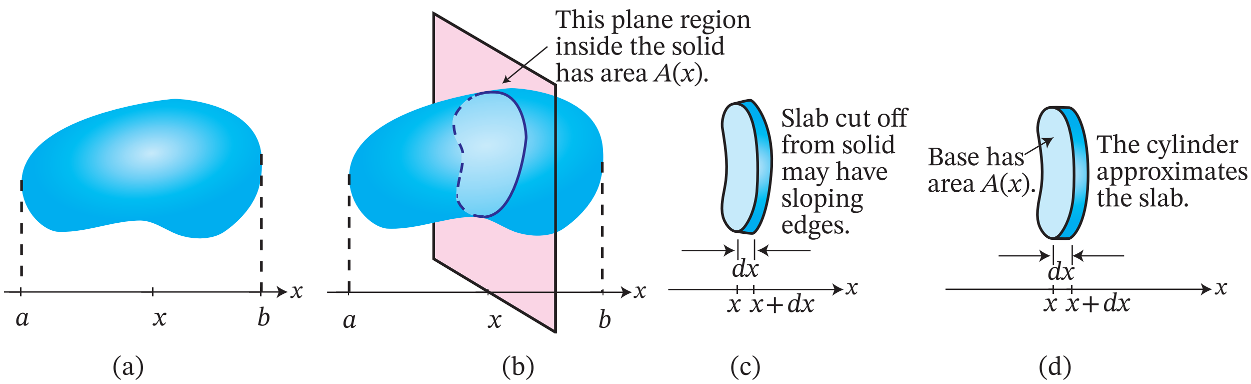


Figure 7.4.3

Most of the effort is usually spent in finding the integrand $A(x)$. In addition to the Pythagorean Theorem and properties of similar triangles, formulas for the areas of familiar plane figures may be needed.

Historical Note: Archimedes and the Volume of a Ball

Archimedes was the first person to find the volume of a ball. He did not express the volume as a number. Rather, in the style of mathematics of the 3$^{\text{rd}}$ century BC, he expressed the volume in terms of the volume of a simpler object. He said that the volume of a ball is two-thirds the volume of the smallest cylinder that contains it. That he considered this one of his greatest accomplishments is evidenced by his request that his tomb be topped with a carving of a ball within a cylinder.

EXAMPLE 1. Find the volume of a ball of radius a.

SOLUTION We sketch a cross section in perspective and in side view (see Figure 7.4.4). The cross section in Figure 7.4.4(a) is a disk of radius r, which depends on x. The area of the cross section is πr^2.

The side view of the situation in Figure 7.4.4(b) is helpful to express r in terms of x. The Pythagorean Theorem

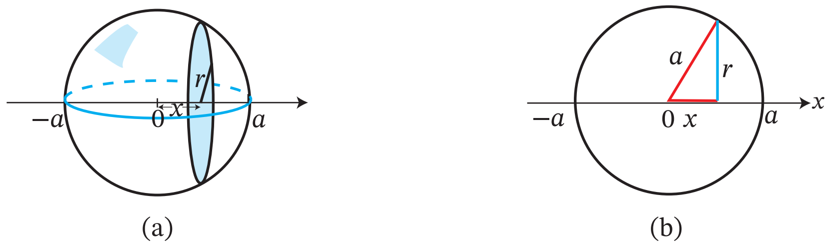


Figure 7.4.4

tells us that $a^2 = x^2 + r^2$, hence $r^2 = a^2 - x^2$. So we have

$$\begin{aligned}
\text{Volume of Ball} &= \int_{-a}^{a} \pi(a^2 - x^2)\,dx && (\text{ setup using informal method })\\
&= \pi\left(a^2x - \frac{x^3}{3}\right)\Bigg|_{-a}^{a} && (\text{ by FTC I })\\
&= \pi\left(\left(a^3 - \frac{a^3}{3}\right) - \left((-a)^3 - \frac{(-a)^3}{3}\right)\right) && (\text{ expanding })\\
&= \frac{4\pi}{3}a^3 && (\text{ simplifying }).
\end{aligned}$$

♦

The next example concerns the solid region discussed in Example 3 of Section 7.2.

EXAMPLE 2. A cylindrical glass of height h and radius a is full of water. It is tilted until the remaining water covers exactly half the base. Find the volume of the remaining water.

SOLUTION We use the triangular cross section shown in Figure 7.4.5.

Introduce the x-axis as in Figure 7.4.5. It was shown in Section 7.2 that the area of the right triangular cross section at x is $(1/2)(h/a)(a^2 - x^2)$. Thus, by similar steps as in Example 1,

$$\begin{aligned}
\text{Volume} &= \int_{-a}^{a} \frac{h}{2a}(a^2 - x^2)\,dx\\
&= \frac{h}{2a}\left(a^2x - \frac{x^3}{3}\right)\Bigg|_{-a}^{a}\\
&= \frac{h}{2a}\left(\left(a^3 - \frac{a^3}{3}\right) - \left(-a^3 + \frac{a^3}{3}\right)\right)\\
&= \frac{h}{2a}\left(\frac{4}{3}a^3\right)\\
&= \frac{2}{3}ha^2.
\end{aligned}$$

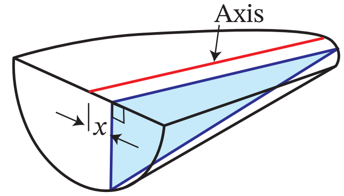


Figure 7.4.5

The volume of the glass is $\pi a^2 h$. From the ratio of these two volumes,

$$\frac{2ha^2/3}{\pi a^2 h} = \frac{2}{3\pi} \approx 0.21221,$$

we see that the water fills about 21% of the glass.

♦

Observation 7.4.1: *Using Symmetry to Simplify Example 2*

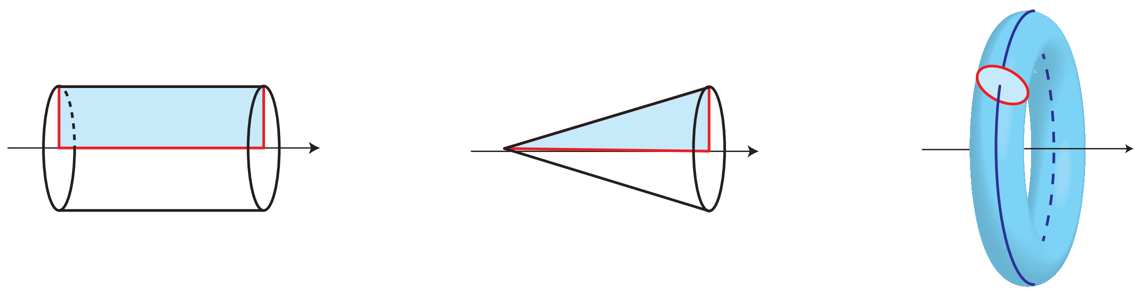
Evaluating the definite integral for the volume in Example 2 could be made easier by noting that the integrand is an even function (the volume to the right of 0 equals the volume to the left of 0). Thus, by FTC I,

$$\text{Volume} = 2\int_0^a \frac{h}{2a}\left(a^2 - x^2\right)\, dx = \frac{h}{a}\left(a^2 x - \frac{x^3}{3}\right)\Big|_0^a = \frac{h}{a}\left(\left(a^3 - \frac{a^3}{3}\right) - (0-0)\right) = \frac{2}{3}ha^2.$$

By avoiding algebra with negative numbers, this approach reduces the chance of making a mistake.

Solids of Revolution

The solid formed by revolving a region $\mathscr{R}$ in the plane about a line in that plane that does not intersect the interior of $\mathscr{R}$ is called a **solid of revolution**.



(a) (b) (c)

Figure 7.4.6

Figure 7.4.6 shows three examples of solids of revolution: (a) a circular cylinder obtained by revolving a rectangle about one of its edges, (b) a cone obtained by revolving a right triangle about one of its two legs, and (c) a torus formed by revolving a disk about a line outside the disk. We will soon see how recognizing a solid of revolution facilitates computations of volume and (surface) area.

The cross sections by planes perpendicular to the line around which the figure is revolved are either disks or washers. A washer is a disk with a round hole. The cross sections in Figures 7.4.6(a) and (b) are disks. In Figure 7.4.6(c) the cross sections perpendicular to the axis of revolution are washers, as shown in Figure 7.4.7(b).

(a) (b)

Figure 7.4.7

EXAMPLE 3. The region under $y = e^{-x}$ and above $[1,2]$ is revolved about the x-axis. Find the volume of the resulting solid of revolution. (See Figures 7.4.8(a) and (b).)

SOLUTION

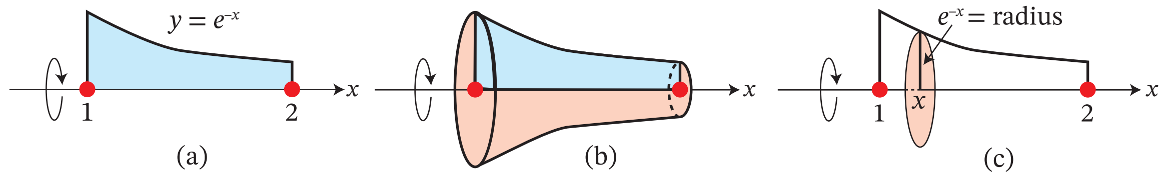


Figure 7.4.8

The cross section by a plane perpendicular to the x-axis is a disk of radius e^{-x}, as shown in Figure 7.4.8(c). The cross-sectional area is

$$\pi\left(e^{-x}\right)^2 = \pi e^{-2x}.$$

The volume of the solid is therefore

$$\int_1^2 \pi e^{-2x}\,dx.$$

Because $d/dx\,(e^{ax}) = ae^{ax}$, an antiderivative of e^{ax} is $(1/a)e^{ax}$. Hence,

$$\begin{aligned}\int_1^2 \pi e^{-2x}\,dx &= \frac{\pi}{-2}e^{-2x}\Big|_1^2 \\ &= \frac{\pi}{-2}\left(e^{-4} - e^{-2}\right) \\ &= \frac{\pi}{2}\left(e^{-2} - e^{-4}\right).\end{aligned}$$

◆

EXAMPLE 4. The region bounded by $y = x^2$, the lines $x = 1$ and $x = \sqrt{2}$, and the x-axis $(y = 0)$. is revolved around the line $y = -1$. Find the volume of the resulting solid.

SOLUTION Figure 7.4.9(a) shows the region being revolved and the line around which it is revolved. Figure 7.4.9(b) shows a perspective view of a cross section.

The cross section is a washer with inner radius 1 and outer radius $1 + x^2$. Its area is therefore $\pi(1+x^2)^2 - \pi(1)^2$. Consequently, since the volume equals the integral of cross-sectional area,

$$\begin{aligned}\text{Volume} &= \int_1^{\sqrt{2}} \left(\pi(1+x^2)^2 - \pi(1)^2\right)\,dx && \text{(setup using informal method)} \\ &= \pi\int_1^{\sqrt{2}} \left(1 + 2x^2 + x^4 - 1\right)\,dx && \text{(expanding powers)} \\ &= \pi\int_1^{\sqrt{2}} \left(2x^2 + x^4\right)\,dx && \text{(more algebra)} \\ &= \pi\left(\frac{2x^3}{3} + \frac{x^5}{5}\right)\Big|_1^{\sqrt{2}} && \text{(by FTC I)} \\ &= \pi\left(\frac{32\sqrt{2}}{15} - \frac{13}{15}\right) && \text{(expanding and simplifying).}\end{aligned}$$

◆

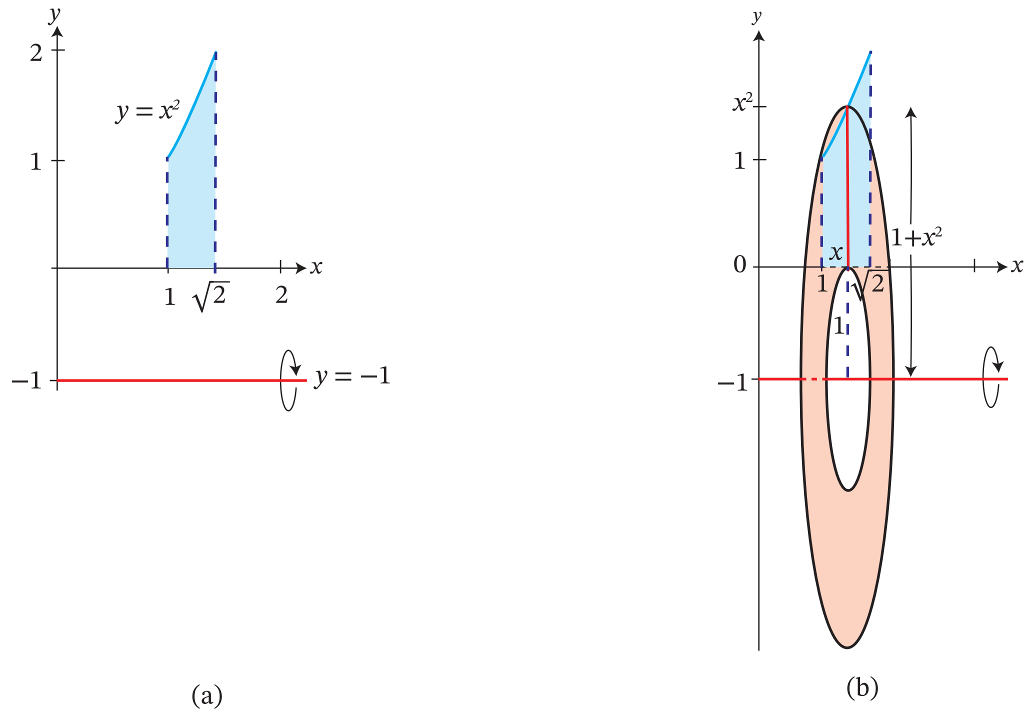


Figure 7.4.9

EXAMPLE 5. Find the volume of the solid formed by revolving the region in Figure 7.4.9(a) around the y-axis ($x = 0$).

SOLUTION The cross sections by planes perpendicular to the y-axis are again washers. For $0 \le y \le 1$ the cross sections are between the vertical lines $x = 1$ and $x = \sqrt{2}$. For $1 \le y \le 2$ they are between the curve and the line $x = \sqrt{2}$. (See Figure 7.4.10(a).)

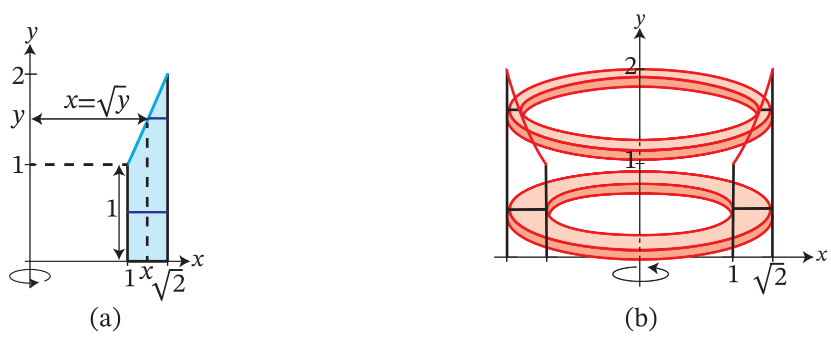


Figure 7.4.10

The cross sections for $0 \le y \le 1$, when rotated about the y-axis, fill out a cylinder whose height is 1 and whose base is a washer of area $\pi(\sqrt{2})^2 - \pi(1)^2 = \pi$. Thus, its volume (height times area of base) is $\pi(1) = \pi$. We do not need an integral for this.

The cross sections for $1 \le y \le \sqrt{2}$ are washers having outer radius $\sqrt{2}$ and inner radius determined by the curve $y = x^2$, as shown in Figure 7.4.10(b). Since $y = x^2$, the inner radius is $x = \sqrt{y}$. The area of a cross sections is

$$\pi(\sqrt{2})^2 - \pi(\sqrt{y})^2.$$

Thus the local estimate of volume is

$$\left(\pi(\sqrt{2})^2 - \pi(\sqrt{y})^2\right)\, dy = (2\pi - \pi y)\, dy.$$

Therefore the total volume is

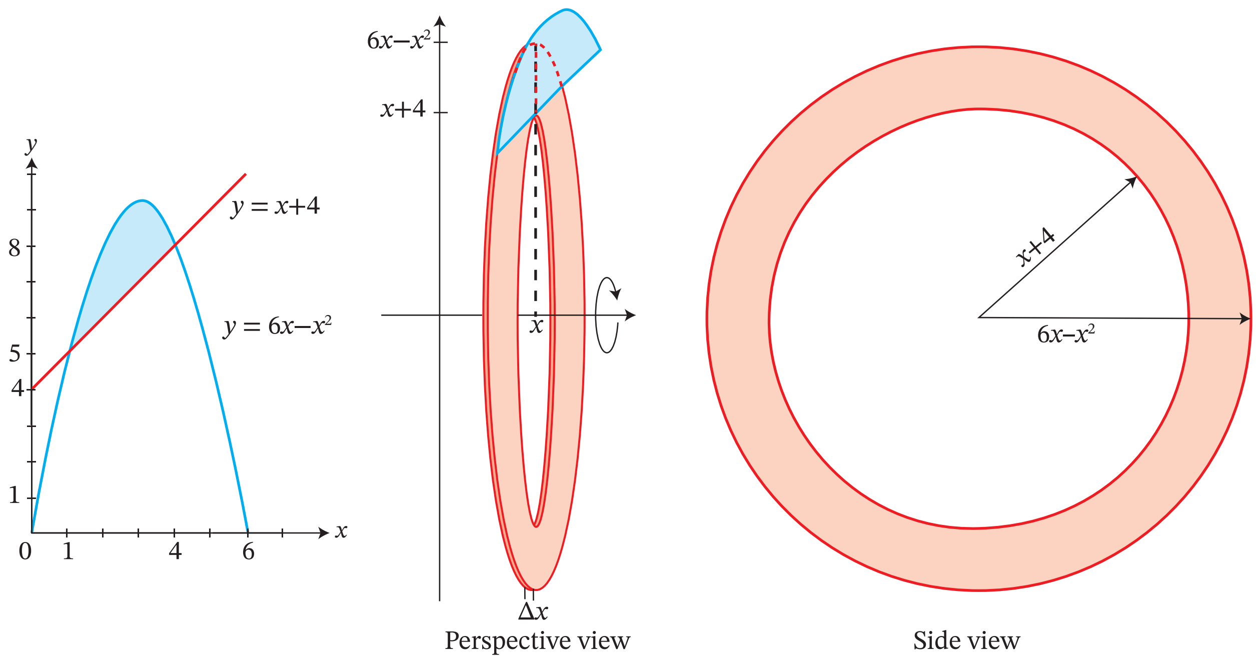
$$\begin{aligned}\int_1^{\sqrt{2}} (2\pi - \pi y)\, dy &= \left(2\pi y - \pi \frac{y^2}{2}\right)\Bigg|_1^{\sqrt{2}} \qquad \text{(by FTC I)}\\ &= \left(2\pi\sqrt{2} - \pi\right) - \left(2\pi - \frac{\pi}{2}\right)\\ &= 2\pi\sqrt{2} - \frac{5}{2}\pi.\end{aligned}$$

Adding this to the volume of the cylinder gives

$$\text{Total Volume} = \left(2\pi\sqrt{2} - \frac{5}{2}\pi\right) + \pi = 2\pi\sqrt{2} - \frac{3}{2}\pi \approx 4.173.$$

◆

EXAMPLE 6. The region bounded by the graphs of $y = x+4$ and $y = 6x - x^2$, shown in Figure 7.4.11(a), is revolved about the x-axis to form a solid of revolution. Express its volume as a definite integral.



Perspective view
Side view
(a) (b) (c)

Figure 7.4.11

SOLUTION We draw a local approximation to a thin slice of the solid (see Figure 7.4.11(b)). The side view in f7.4.11(c) shows the area of a cross section is

$$\pi\left(6x - x^2\right)^2 - \pi\,(x+4)^2.$$

This is the integrand.

Next we determine the interval of integration. The ends of the interval of integration are determined by where the curves cross, which occurs when $x+4=6x-x^2$. Moving all terms to the left-hand side yields $x^2-5x+4=0$, or $(x-1)(x-4)=0$. The endpoints of the interval are $x=1$ and $x=4$. The volume of the solid is given by

$$\int_1^4 \left(\pi\left(6x-x^2\right)^2-\pi\left(x+4\right)^2\right)\,dx.$$

◆

Summary

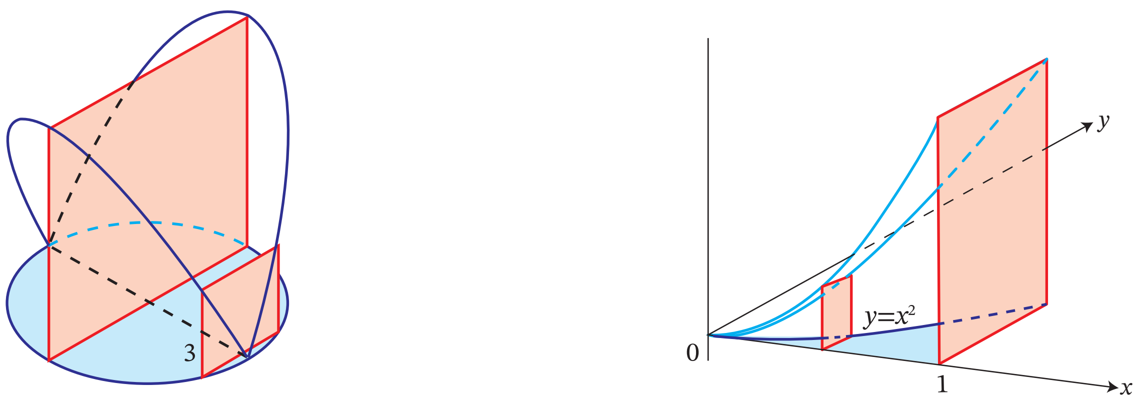
The idea in this section was that volume is the definite integral of cross-sectional area. To implement it we must find that varying area and also the interval of integration. A solid of revolution, where the cross section may be a disk or a washer, is a special case.

EXERCISES for Section 7.4

In Exercises 1 to 8, (a) draw the solid, (b) draw a cross section in perspective and side view, (c) find the area of a cross section, (d) set up the definite integral for the volume, and (e) evaluate the definite integral (if possible).

1. Find the volume of a cone of radius a and height h.

2. The base of a solid is a disk of radius 3. Each plane perpendicular to one diameter meets the solid in a square, one side of which is in the base of the solid. (See Figure 7.4.12(a).) Find its volume.



(a) (b)

Figure 7.4.12

3. The base of a solid is the region bounded by $y=x^2$, the line $x=1$, and the x-axis. Each cross section perpendicular to the x-axis is a square. (See Figure 7.4.12(b).) Find the volume of the solid.

4. Repeat Exercise 3 except that the cross sections perpendicular to the base are equilateral triangles.

5. Find the volume of a pyramid with a square base of side a and height h, using square cross sections parallel to the base. The top of the pyramid is above the center of the base.

6. Repeat Exercise 5 using trapezoidal cross sections perpendicular to the base. To obtain trapezoidal cross sections, pick the planes parallel to the plane through the axis of the cylinder and the point where the water touches the top of the glass.

7. Find the volume of the solid whose base is the disk of radius 5 and whose cross sections perpendicular to a diameter are equilateral triangles. (See Figure 7.4.13(a).)

Figure 7.4.13

8. Find the volume of the pyramid shown in Figure 7.4.13(b) by using cross sections perpendicular to the edge of length c.

In Exercises 9 to 14 set up a definite integral for the volume of the solid formed by revolving the region $\mathscr{R}$ about the given axis. Do not attempt to evaluate these integrals.

9. The region $\mathscr{R}$ is bounded by $y=\sqrt{x}$, $x=1$, $x=2$, and the x-axis and rotated about the x-axis.

10. The region $\mathscr{R}$ is bounded by $y=\dfrac{1}{\sqrt{1+x^2}}$, $x=0$, $x=1$, and the x-axis and rotated about the x-axis.

11. The region $\mathscr{R}$ is bounded by $y=x^{-1/2}$, $y=x^{-1}$, $x=1$, and $x=2$ and rotated about the x-axis.

12. The region $\mathscr{R}$ is bounded by $y=x^2$ and $y=x^3$ and rotated about the y-axis.

13. The region $\mathscr{R}$ is bounded by $y=\tan(x)$, $y=\sin(x)$, $x=0$, and $x=\dfrac{\pi}{4}$ and rotated about the x-axis.

14. The region $\mathscr{R}$ is bounded by $y=\sec(x)$, $y=\cos(x)$, $x=\dfrac{\pi}{6}$, and $x=\dfrac{\pi}{3}$ and rotated about the x-axis.

---

15. A cylindrical drinking glass of height h and radius a, full of water, is tilted until the water just covers the base. Set up a definite integral that represents the amount of water left in the glass. Use rectangular cross sections. Refer to Figure 7.4.14(a) and follow the directions preceding Exercise 1.

This configuration was first encountered in Exercise 5 in Section 7.2.

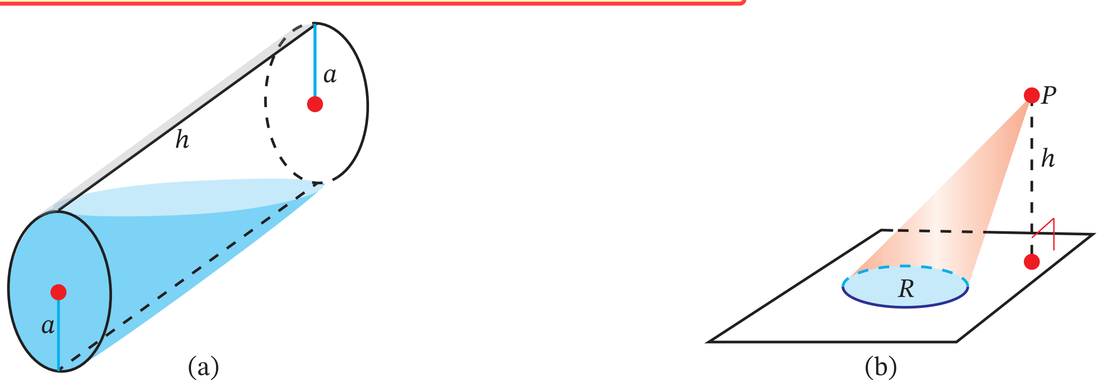


Figure 7.4.14

16. Repeat Exercise 15, but use trapezoidal cross sections.

NOTE: This configuration was first encountered in Exercise 6 in Section 7.2.

17. Find the volume described in Exercise 15 using only common sense (and geometry). Do not use any calculus.

18. A cylindrical drinking glass of height h and radius a, full of water, is tilted until the water remaining covers half the base.

(a) Set up a definite integral for the volume of water in the glass, using cross sections that are parts of disks.

(b) Compare the answer in (a) with the definite integral found in Example 2. Which definite integral looks easier to evaluate?

19. Repeat Exercise 18, but use rectangular cross sections.

20. A solid is formed in the following manner. A plane region $\mathscr{R}$ and a point P not in the plane are given. The solid consists of all line segments joining P to points in $\mathscr{R}$. If $\mathscr{R}$ has area A and P is a distance h from the plane $\mathscr{R}$, show that the volume of the solid is $Ah/3$. (See Figure 7.4.14(b) and use the fact that the areas of similar figures are proportional to the square of their linear dimensions.)

21. A drill of radius 4 inches bores a hole through a wooden ball of radius 5 inches, passing symmetrically through the center of the sphere.

(a) Draw the part of the sphere removed by the drill.

(b) Find $A(x)$, the area of a cross section of the region in (a) made by a plane perpendicular to the axis of the drill and at a distance x from the center of the sphere.

(c) Set up the definite integral for the volume of wood removed.

22. What fraction of the volume of a sphere is contained between two parallel planes that trisect the diameter to which they are perpendicular? (Leave your answer in terms of a definite integral.)

23. The disk bounded by the circle $(x-b)^2 + y^2 = a^2$, where $0 < a < b$, is revolved around the y-axis. Set up a definite integral for the volume of the doughnut (torus) produced.

In Exercises 24 to 27 set up definite integrals for (a) the area of $\mathscr{R}$, (b) the volume formed when $\mathscr{R}$ is revolved around the x-axis, and (c) the volume formed when $\mathscr{R}$ is revolved around the y-axis.

24. The region $\mathscr{R}$ is the region under $y = \tan(x)$ and above the interval $[0, \frac{\pi}{4}]$.

25. The region $\mathscr{R}$ is the region under $y = e^x$ and above the interval $[0, 1]$.

26. The region $\mathscr{R}$ is the region under $y = \dfrac{1}{\sqrt{1-x^2}}$ and above the interval $[0, \frac{1}{2}]$.

27. The region $\mathscr{R}$ is the region under $y = \sin(x)$ and above the interval $[0, \pi]$.

---

28. When a convex region $\mathscr{R}$ of area A situated to the right of the y-axis is revolved around the y-axis, the resulting solid of revolution has volume V. When $\mathscr{R}$ is revolved around the line $x = -k$, where k is positive, the volume of the resulting solid is V^*. Express V^* in terms of k, A, and V.

29. Set up a definite integral for the volume of one octant of the region common to two right circular cylinders of radius 1 whose axes intersect at right angles, as shown in Figure 7.4.15. The volume was found by Archimedes.

30. Archimedes viewed a ball as a cone whose height is the radius of the ball and whose base is the surface of the ball. To understand this viewpoint imagine the cone-like solid whose vertex is at the center of the sphere and whose base is a small patch on the sphere. On that basis he computed that the volume of the ball is one third the product of the radius and the surface area. He then gave a rigorous proof of his conjecture.

Clever Sam, inspired by this, said "I'm going to get the volume of a circular cylinder in a new way. Say its radius is r and height is h. Then I'll view it as a cylinder made up of r by h rectangles, all of which have the axis as an edge. Then I pile them up to make a box whose base is an r by h rectangle and whose height is $2\pi r$ (the circumference of the cylinder's base). So the volume would be $2\pi r$ times rh, or $2\pi r^2 h$. That's twice the usual volume, so the standard formula is wrong." Is Sam right? (Explain.)

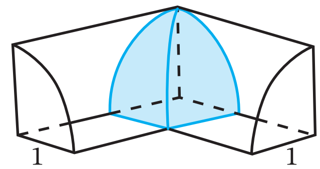


Figure 7.4.15

31. (a) Use Archimedes' observation in Exercise 30 to show that the surface area of a sphere of radius a is $r\pi a^2$.

(b) To obtain this result by integration, first make an estimate of the surface area of the sphere between parallel planes at distances x and $x + dx$ from the center. Show that when this narrow ring is cut out and laid flat, it resembles a rectangle of dimension $2\pi\sqrt{a^2 - x^2}$ by $\dfrac{a\,dx}{\sqrt{a^2 - x^2}}$.)

(c) Use the local estimate obtained in (a) to show that the surface area of the unit sphere is $4\pi a^2$.

32. (See Exercise 31.) Two parallel planes a distance h apart intersect a sphere of radius a. Show that the area of the surface between them is $2\pi ha$. This implies that the corresponding area depends only on h, not on the position of the planes. Also, the interested surface area is proportional to the distance between the two planes.

7.5 Computing Volumes by Shells

Imagine revolving the plane region $\mathscr{R}$ about the line L, as in Figure 7.5.1(a).

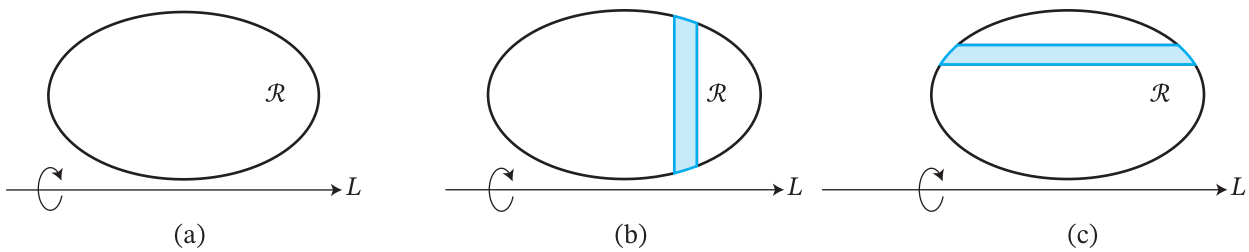


Figure 7.5.1

We may think of $\mathscr{R}$ as being formed from narrow strips perpendicular to L, as in Figure 7.5.1(b). Revolving such a strip around L produces a washer. This is how we found volumes in the last section.

We can also think of $\mathscr{R}$ as being formed from narrow strips parallel to L, as in Figure 7.5.1(c). Revolving such a strip around L produces a solid shaped like a bracelet or part of a drinking straw, as shown, in perspective, in Figure 7.5.2. We will call it a **shell**. (Perhaps tube or pipe might be a better choice, but shell is the accepted term.)

This section describes how to find the volume of a solid of revolution using shells. Sometimes shells provide easier calculations than disks or washers.

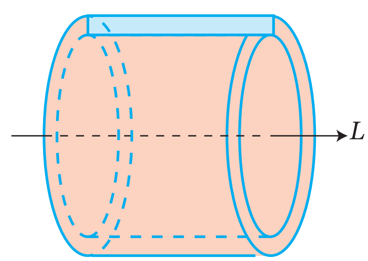


Figure 7.5.2

The Shell Technique

To apply the shell technique we imagine cutting the plane region $\mathscr{R}$ in Figure 7.5.3(a) into a finite number of narrow strips by lines parallel to L. Each strip is approximated by a rectangle of width dx as in Figure 7.5.3(b). Then we approximate the solid of revolution by a collection of tubes (like the parts of a collapsible telescope), as in Figure 7.5.3(c).

The key to this method is estimating the volume of each shell.

Figure 7.5.4(a) shows a local approximation. Its height, $c(x)$, is the width of the cross section of $\mathscr{R}$ corresponding to the value x on a line that we will call the x-axis. The radius of the shell, shown in Figure 7.5.4(b), is $x - k$, where

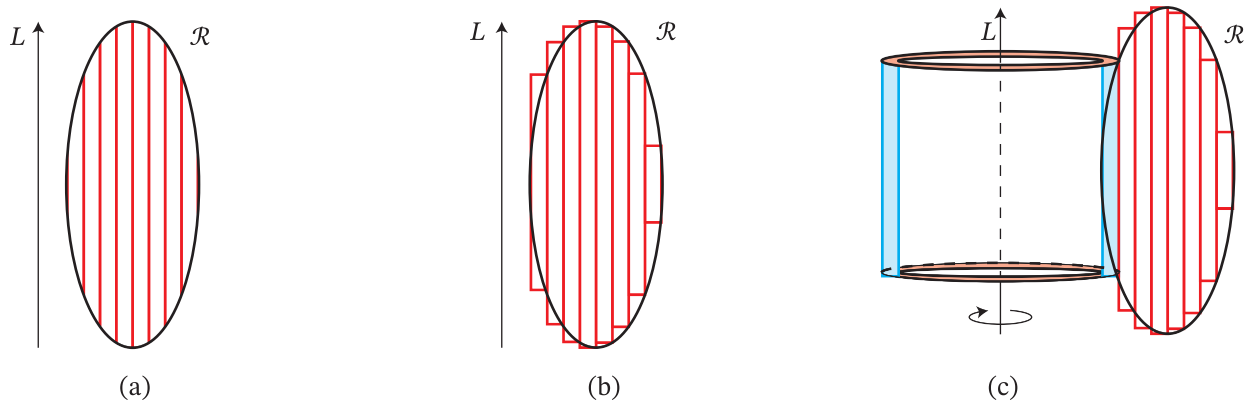


Figure 7.5.3

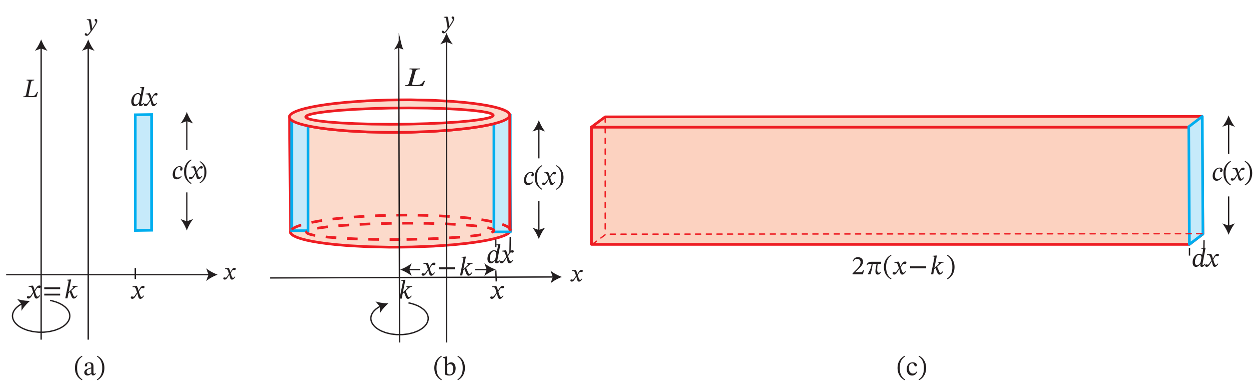


Figure 7.5.4

k is the x-coordinate of the equation of the axis of rotation. Imagine cutting the shell along a direction parallel to L, unrolling it, and then laying it flat like a carpet. When laid flat, the shell resembles a thin slab of thickness dx, width $c(x)$, and length $2\pi(x-k)$, as shown in Figure 7.5.4(c).

Do you understand why the unrolled shell is not a rectangular prism? The exact volume of the shell is found in Exercise 23.

The local approximation to the volume of this shell is

$$2\pi(x-k)c(x)\;dx. \tag{7.5.1}$$

From (7.5.1) we conclude that

Definition: *Shell Method for Volume of Solid of Revolution*

$$\text{Volume of Solid of Revolution} = \int_a^b \underbrace{2\pi(x-k)}_{\text{circumference}}\ \underbrace{c(x)}_{\text{height}}\ \underbrace{dx}_{\text{thickness}}.$$

EXAMPLE 1. The region $\mathscr{R}$ below the line $y=e$, above $y=e^x$, and to the right of the y-axis is revolved around the y-axis to produce a solid $\mathscr{S}$. Set up the definite integrals for the volume of $\mathscr{S}$ using (a) disks and (b) shells.

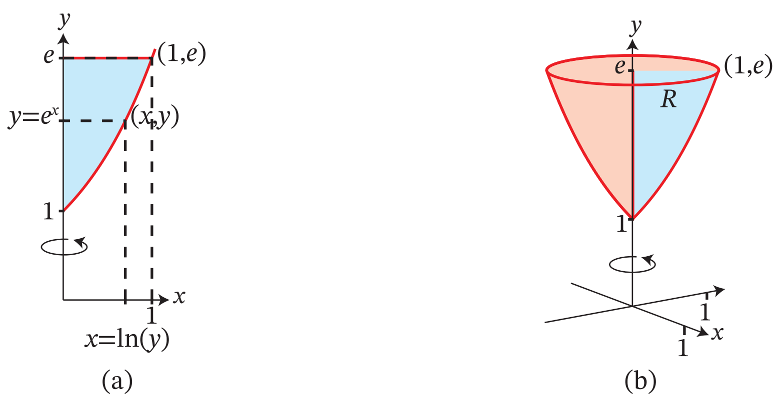


Figure 7.5.5

SOLUTION Figure 7.5.5(a) shows the region $\mathscr{R}$ and Figure 7.5.5(b) shows the solid $\mathscr{S}$.

(a) If we use cross sections perpendicular to the y-axis, as in the preceding section, we find in Exercise 21 that

$$\text{Volume} = \int_1^e \pi(\ln(y))^2\;dy.$$

This integrand has an elementary antiderivative, and we will learn how to find it in Chapter 8. Formula 84 (with $a=1$) in the book's Table of Integrals (in Appendix A) has $\int(\ln(x))^2\;dx = x((\ln(x))^2-2\ln(x)+2)$, which you may check by differentiation. Thus

$$\text{Volume} = \pi\; x((\ln(x))^2-2\ln(x)+2)\Big|_1^e = \pi\left(e(\ln(e)^2-2\ln(e)+2)-1(\ln(1)^2-2\ln(1)+2)\right) = \pi(e-2)\approx 2.257.$$

(b) If we use cross sections parallel to the y-axis, we have a simpler integration. The shell has radius x, height $e-e^x$, and thickness dx as shown in Figure 7.5.6(a).

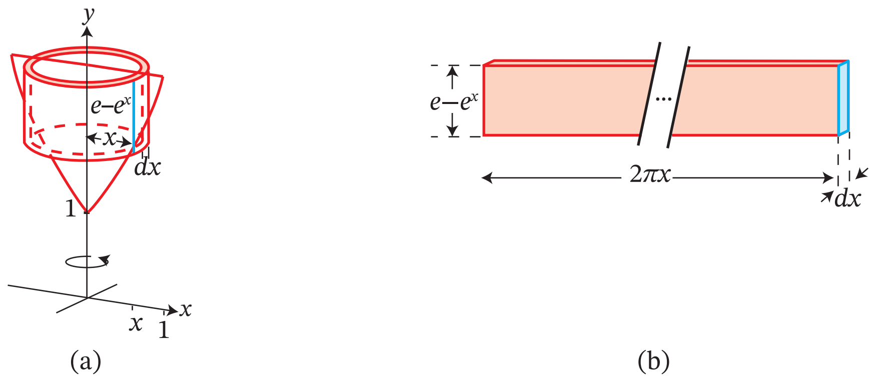


Figure 7.5.6

The local approximation to the volume of the shell is

$$\underbrace{2\pi x}_{\text{circumference}} \underbrace{\left(e-e^x\right)}_{\text{height}} \underbrace{dx}_{\text{thickness}},$$

so the volume of $\mathscr{S}$ is

$$\int_0^1 2\pi x\left(e-e^x\right)\, dx = 2\pi \int_0^1 ex - xe^x\, dx.$$

We would like to have an antiderivatives of ex and xe^x. The first part is trivial, $\int ex\, dx = ex^2/2$. In Chapter 8 we will learn how to find an antiderivative of xe^x. We will find it is easier to find this antiderivative than it is to find $\int \left(\ln(y)\right)^2\, dy$.
Formula 77 in the Table of Integrals (in Appendix A) gives $\int xe^x\, dx = xe^x - e^x$. Then, as expected, and as shown in Exercise 22, once again the volume is found to be $\pi(e-2)$. ♦

It is not unusual to find the calculations with one approach easier than the other. In Example 1 both methods were feasible. In the next example, only the shell technique is tractable.

EXAMPLE 2. The region $\mathscr{R}$ bounded by the line $y = \frac{\pi}{2} - 1$, the y-axis, and the curve $y = x - \sin(x)$ is revolved around the y-axis. Set up definite integrals for the volume of the solid using (a) disks and (b) shells.

SOLUTION The region $\mathscr{R}$ is shown in Figure 7.5.7(a).

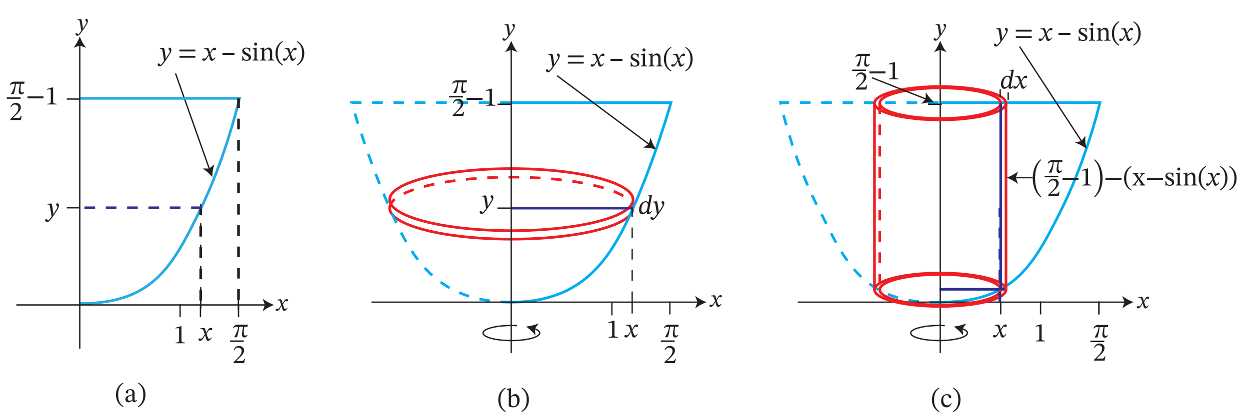


Figure 7.5.7

(a) To use the method of parallel cross sections we need the radius of the disk shown in Figure 7.5.7(b). For a given value of y the radius is the value of x for which $x - \sin(x) = y$. We need to express x as a function of y. This inverse function is not elementary, ending any hopes of using the FTC.

(b) The shell technique goes through smoothly. A shell, shown in Figure 7.5.7(c), has radius x and height $\pi/2 - 1 - (x - \sin(x))$. The volume of the local approximation is

$$\underbrace{2\pi x}_{\text{circumference}} \underbrace{\left(\frac{\pi}{2} - 1 - (x - \sin(x))\right)}_{\text{height}} \underbrace{dx}_{\text{thickness}} .$$

Notice that, when $y = 0$, then $x = 0$, and when $y = \pi/2 - 1$, then $x = \pi/2$. Thus, the total volume of the bowl is

$$\int_0^{\pi/2} 2\pi x \left(\frac{\pi}{2} - 1 - (x - \sin(x))\right) dx.$$

The value of the definite integral is found in Exercise 49 for Section 8.S. ◆

Summary

The volume of a solid of revolution may be found by approximating the solid by concentric thin shells. The volume of a shell is approximately $2\pi R(x)\, c(x)\, dx$, where $R(x)$ is its radius and $c(x)$ is its height. (See Figure 7.5.8.) The shell technique is useful when integration by cross sections is difficult or impossible.

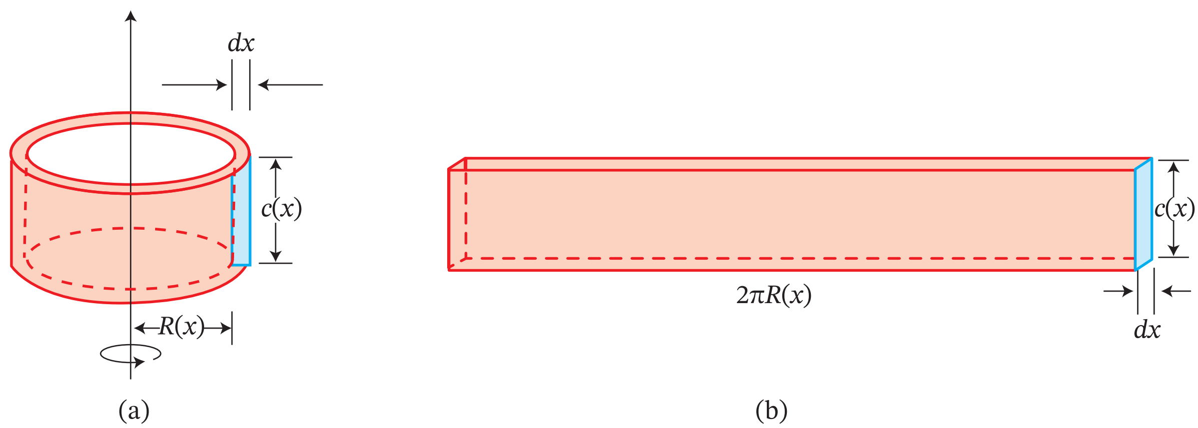


Figure 7.5.8

EXERCISES for Section 7.5

INSTRUCTIONS: While this section is about computing volumes by the method of shells, some of these exercises might not be possible to evaluate by the method of shells. Others might be possible to evaluate by either the method of shells or the method of parallel cross sections. For each problem, unless specified otherwise, evaluate your options and use the one that is most convenient.

In Exercises 1 to 4 draw an approximating cylindrical shell for the solid described, set up a definite integral for its volume, and evaluate the integral.

1. (a) The trapezoid bounded by $y = x$, $x = 1$, $x = 2$, and the x-axis, when it is revolved around the x-axis.
(b) The same trapezoid, revolved around the y-axis.

2. (a) The trapezoid in Exercise 1, revolved about the line $y = -3$.
(b) The trapezoid in Exercise 1, revolved about the line $x = -3$.

3. The triangle with vertices $(0,0)$, $(1,0)$, and $(0,2)$ is revolved around the y-axis.

4. The triangle in Exercise 3 is revolved about the x-axis.

---

5. Find a definite integral for the volume of the solid produced by revolving about the y-axis the finite region bounded by $y = x^2$ and $y = x^3$.

6. Repeat Exercise 5, except revolve the region around the x-axis.

7. Set up a definite integral for the volume of the solid produced by revolving about the x-axis the finite region bounded by $y = \sqrt{x}$ and $y = \sqrt[3]{x}$.

8. Repeat Exercise 7, except revolve the region about the y-axis.

9. Find a definite integral for the volume of the right circular cone of radius a and height h by the shell method.

10. Let $\mathscr{R}$ be the region bounded by $y = x + x^3$, $x = 1$, $x = 2$, and the x-axis. Set up a definite integral for the volume of the solid produced by revolving $\mathscr{R}$ about (a) the x-axis and (b) the line $x = 3$.

11. Set up a definite integral for the volume of the solid produced by revolving the region $\mathscr{R}$ in Exercise 10 about (a) the y-axis and (b) the line $y = -2$.

12. Set up a definite integral for the volume of the solid of revolution formed by revolving the region bounded by $y = 2 + \cos(x)$, $x = \pi$, $x = 10\pi$, and the x-axis around (a) the y-axis and (b) the x-axis.

13. The region below $y = \cos(x)$, above the x-axis, and between $x = 0$ and $x = \dfrac{\pi}{2}$ is revolved around the x-axis. Find a definite integral for the volume of the solid of revolution (a) using parallel cross sections and (b) using concentric shells.

14. Let $\mathscr{R}$ be the region below $y = \dfrac{1}{(1+x^2)^2}$ and above $[0,1]$. Set up a definite integral for the volume of the solid produced by revolving $\mathscr{R}$ about the y-axis.

15. The region between $y = e^{x^2}$, the x-axis, $x = 0$, and $x = 1$ is revolved about the y-axis.
(a) Set up a definite integral for the area of the region.
(b) Set up a definite integral for the volume of the solid produced.

NOTE: The FTC is of no use in evaluating the area of the region in (a), but the volume of the solid in (b) is easily evaluated using the FTC.

16. Let $\mathscr{R}$ be the region below $y = \ln(x)$ and above $[1, e]$. Find a definite integral for the volume of the solid produced by revolving $\mathscr{R}$ about (a) the x-axis and (b) the y-axis.

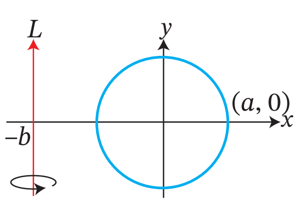


Figure 7.5.9

17. Set up a definite integral for the volume of the doughnut (torus) produced by revolving the disk of radius a about a line L at a distance $b > a$ from its center. (See Figure 7.5.9.)

18. Let $\mathscr{R}$ be the region below $y = \dfrac{1}{\sqrt{1+x^2}}$ and above $\left[\sqrt{3}, \sqrt{8}\right]$. Set up a definite integral for the volume of the solid produced by revolving $\mathscr{R}$ about the (a) the x-axis and (b) the y-axis.

19. Let $\mathscr{R}$ be the region below $y = \dfrac{1}{x^2 + 4x + 1}$ and above $[0,1]$. Find a definite integral for the volume of the solid produced by revolving $\mathscr{R}$ about the line $x = -2$.

20. The region below $y = e^x \dfrac{1+\sin(x)}{x}$ and above $[0, 10\pi]$ is revolved about the y-axis to produce a solid of revolution. Try to set up definite integrals for the volume of this solid by shells and by parallel cross sections. Which approach do you think is easier to set up and to evaluate?

Exercises 21 and 22 complete Example 1. The solid $\mathscr{S}$ is formed by revolving around the y-axis the region below $y = e$, above $y = e^x$, and to the right of the y-axis.

21. The volume of $\mathscr{S}$ using cross sections perpendicular to the y-axis was found to be $\displaystyle\int_1^e \pi\left(\ln(y)\right)^2\, dy$.

(a) Verify that $x\left((\ln(x))^2 - 2\ln(x) + 2\right)$ is an antiderivative of $(\ln(x))^2$. (b) Find the volume of $\mathscr{S}$.

22. The volume of $\mathscr{S}$ using cross sections parallel to the y-axis was found to be $\displaystyle\int_0^1 2\pi x\left(e - e^x\right)\, dx$.

(a) Verify that $xe^x - e^x$ is an antiderivative of xe^x. (b) Find the volume of $\mathscr{S}$.

---

23. In Figure 7.5.4, when we unrolled the shell as a carpet we pictured it as a rectangular solid whose faces meet at right angles. However, since the inner radius is $x - k$ and the outer radius is $x - k + \Delta x$ the circumference of the inside of the shell is less than the outer circumference. Show that the volume of the unrolled shell is $2\pi\left(x - k + \dfrac{\Delta x}{2}\right)c(x)\Delta x$.

NOTE: This result means that if we used $x - k + \Delta x/2$ as a sampling number in the interval $[x-k, x-k+\Delta x]$ instead of x, the local approximation to the volume of the shell would be exact.

The **kinetic energy** of a particle of mass m grams moving at a velocity of v centimeters per second is $mv^2/2$ ergs. Exercises 24 and 25 ask for the kinetic energy of rotating objects.

24. A solid cylinder of radius r and height h centimeters has a uniform density of g grams per cubic centimeter. It is rotating at the rate of two revolutions per second around its axis.

(a) Find the speed of a particle at a distance x from the axis.
(b) Find a definite integral for the kinetic energy of the rotating cylinder.

25. A solid ball of radius r centimeters has a uniform density of g grams per cubic centimeter. It is rotating around a diameter at the rate of three revolutions per second around its axis.

(a) Find the speed of a particle at a distance x from the diameter.
(b) Find a definite integral for the kinetic energy of the rotating ball.

---

26. When a region $\mathscr{R}$ in the first quadrant is revolved around the y-axis, a solid of volume 24 is produced. When $\mathscr{R}$ is revolved around the line $x = -3$, a solid of volume 84 is produced. What is the area of $\mathscr{R}$?

27. Let $\mathscr{R}$ be a region in the first quadrant. When it is revolved around the x-axis, a solid of revolution, $\mathscr{S}_1$, is produced. When it is revolved around the y-axis, another solid of revolution, $\mathscr{S}_2$, is produced. Give an example of a region $\mathscr{R}$ with the property that the volume of $\mathscr{S}_1$ cannot be evaluated by the FTC, but the volume of $\mathscr{S}_2$ can be evaluated by the FTC.

7.6 Water Pressure Against a Flat Surface

This section shows how to use integration to compute the force of water against a submerged flat surface.

Physical Background

Imagine the portion of Earth's atmosphere directly above one square inch at sea level. That air forms a column some hundred miles high which weighs about 14.7 pounds. It exerts a pressure of 14.7 pounds per square inch (14.7 psi).

The pressure does not crush us because the cells in our body are at the same pressure. If we were to go into a vacuum, we would explode. (This is why astronauts wear pressurized suits.)

The pressure inside a flat tire is 14.7 psi. When you pump up a bicycle tire so that the gauge reads 60 psi, the pressure is actually $60 + 14.7 = 74.7$ psi. The tire must be strong enough to avoid bursting.

One cubic foot of water weighs 62.6 pounds, so one cubic inch weighs $62.6/12^3 = 0.036227$ pounds. In other words, the density of water is 0.036227 pounds per cubic inch.

Imagine diving into a lake and descending 33 feet (10 meters). Extending the 100-mile-high column 33 feet into the water adds $(33)(12)(0.036227) = 14.7$ pounds of water. The pressure is now twice 14.7, or 29.4 psi. You cannot escape the pressure by turning, since at a given depth the pressure is the same in all directions.

Pressure and force are closely related. If the force is the same throughout a region, then the pressure is total force divided by area,

$$\text{Pressure} = \frac{\text{Force}}{\text{Area}}.$$

Equivalently,

$$\text{Force} = \text{Pressure} \times \text{Area}.$$

Thus, when the pressure is constant in a plane region it is easy to find the total force against it: multiply the pressure and the area of the region.

If the pressure varies in the region we make use of integration to find the total force.

Using an Integral to Find the Force of Water

We will see how to find the total force on a flat submerged object due to the water. We will disregard the pressure due to the atmosphere. (See Figure 7.6.1(a).)

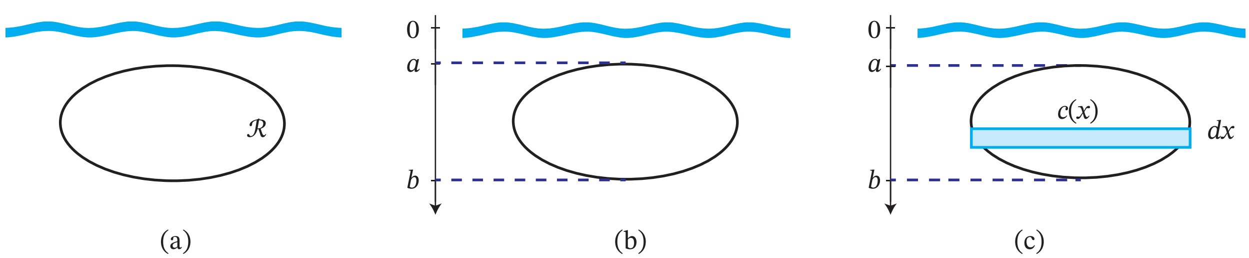


Figure 7.6.1

At a depth of h inches, water exerts a pressure of about $0.03623h$ psi. Therefore the water exerts a force on a flat horizontal object of area A square inches at a depth of h inches equal to $0.03623hA$ pounds.

To deal with a tilted submerged surface $\mathscr{R}$ takes more calculation, so we will only consider vertical surfaces. Since the pressure is not constant over the surface shown in Figure 7.6.1(b), introduce a vertical x-axis, pointed down, with its origin O at a distance k below the water's surface. $\mathscr{R}$ lies between lines corresponding to $x = a$ and $x = b$. The depth of the water corresponding to x is $x + k$. (If the origin is at the water's surface, then $k = 0$.)

We will find the local approximation of the force by considering a horizontal strip corresponding to the interval $[x, x+dx]$ of the x-axis, as in Figure 7.6.1(c). Letting $c(x)$ denote the cross-sectional length, we see that the force of the water on this strip is approximately

$$\underbrace{(0.03623)}_{\text{density of } H_2O}\ \underbrace{(x+k)}_{\text{depth}}\ \underbrace{c(x)\ dx}_{\text{area of strip}} \text{ pounds.}$$

Consequently

$$\text{Force against } \mathscr{R} = 0.03623 \int_a^b (x+k)c(x)\ dx \text{ pounds.}$$

EXAMPLE 1. A circular tank is submerged in water. An end is a disk 10 inches in diameter. The top of the tank is 12 inches below the surface of the water and parallel to it. Find the force against one end.

SOLUTION

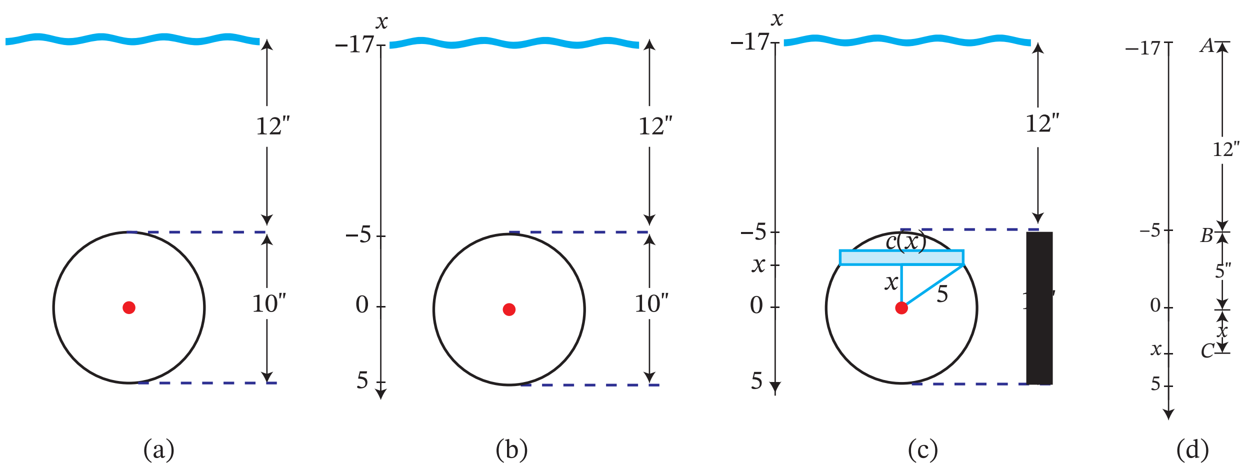


Figure 7.6.2

The end of the tank is shown in Figure 7.6.2(a). Introduce a vertical x-axis with its origin O level with the center of the disk. (See Figure 7.6.2(b).) To find the cross section $c(x)$ we use Figure 7.6.2(c).

This placement of O will make it easier to compute the cross-sectional lengths.

By the Pythagorean Theorem applied to the right triangle in Figure 7.6.2(c) we have

$$\left(\frac{c(x)}{2}\right)^2 + |x|^2 = 5^2.$$

Therefore $(c(x))^2 + 4x^2 = 100$ and, because $c(x) > 0$,

$$c(x) = \sqrt{100 - 4x^2}.$$

We will find the depth as a function of x. From Figure 7.6.2(d), the depth $|AC|$ equals $|AB| + |BC| = 12 + (x - (-5)) = 17 + x$. We have

As a check, if $x = 0$, the depth is 17 inches.

$$\text{Local Approximation of Force} = \underbrace{(0.03623)(x+17)}_{\text{pressure}}\ \underbrace{\sqrt{100-4x^2}\ dx}_{\text{area}}.$$

From this we obtain

$$\begin{aligned}\text{Total Force} &= \int_{-5}^{5} (0.03623)(x+17)\sqrt{100-4x^2}\,dx \text{ pounds}\\ &= 0.03623\int_{-5}^{5} x\sqrt{100-4x^2}\,dx + 0.03623\int_{-5}^{5} 17\sqrt{100-4x^2}\,dx \text{ pounds.}\end{aligned}$$

The first integral is 0 because the integrand, $x\sqrt{100-4x^2}$, is an odd function and the interval of integration is symmetric about $x = 0$. The integrand in the second integral is even, so, after factoring out the 17, the remaining integral is evaluated by recognizing it as the area of a quarter disk:

$$\int_{-5}^{5} \sqrt{100-4x^2}\,dx = 2\int_{0}^{5} \sqrt{4\left(25-x^2\right)}\,dx = 4\int_{0}^{5} \sqrt{25-x^2}\,dx = 4\begin{pmatrix}\text{Area of one quarter}\\ \text{of disk of radius 5}\end{pmatrix} = 4\left(\frac{1}{4}\pi 5^2\right) = 5^2\pi = 25\pi.$$

Thus,

$$\text{Total Force} = (0.03623)(17)(25\pi) \text{ pounds} \approx 48 \text{ pounds.}$$

♦

EXAMPLE 2. Figure 7.6.3(a) shows a submerged vertical equilaterial triangle of side h. Find the force of water against it.

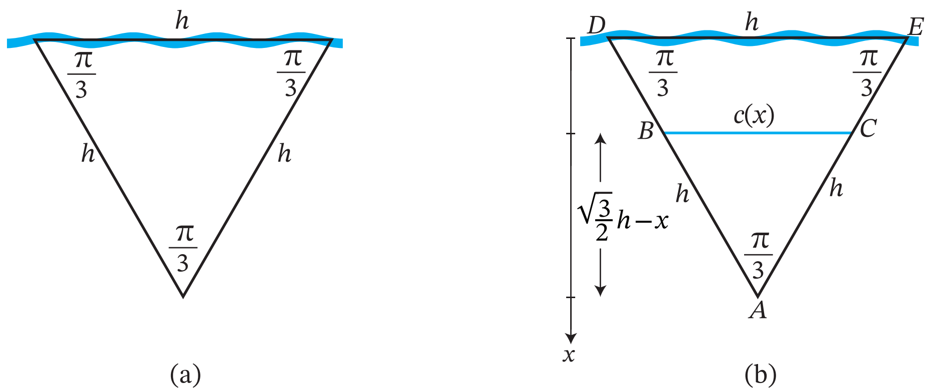


Figure 7.6.3

SOLUTION We place the origin of the vertical axis at the surface of the water (see Figure 7.6.3(b)). To set up an integral we compute $c(x)$. The length of an altitude in the triangle, $\sqrt{3}h/2$, is marked on the x-axis.

The similar triangles $\triangle ABC$ and $\triangle ADE$ give

$$\frac{c(x)}{h} = \frac{\frac{\sqrt{3}}{2}h - x}{\frac{\sqrt{3}}{2}h}.$$

Thus, $c(x) = h - 2x/\sqrt{3}$.

NOTE: c is linear, $c(0) = h$, and $c\left(\sqrt{3}h/2\right) = 0$ which agree with Figure 7.6.3(b).

The local approximation of force is therefore

$$\underbrace{0.03623x}_{\text{pressure}}\underbrace{\left(h - \frac{2x}{\sqrt{3}}\right)\,dx}_{\text{area}}.$$

Hence

$$\text{Total Force} = \int_0^{\frac{\sqrt{3}}{2}h} 0.03623\, x\left(h - \frac{2x}{\sqrt{3}}\right)dx \quad = 0.03623 \int_0^{\frac{\sqrt{3}}{2}h}\left(hx - \frac{2x^2}{\sqrt{3}}\right)dx$$
$$= 0.03623\left(h\frac{x^2}{2} - \frac{2}{\sqrt{3}}\frac{x^3}{3}\right)\Bigg|_0^{\frac{\sqrt{3}}{2}h} \quad = 0.00453\frac{h^3}{8} \text{ pounds.}$$

◆

Summary

We introduced the notion of water pressure defined as force divided by area or force per unit area. If the pressure is constant over a flat region, the force is the product, pressure times area. When $p(x)$ is the pressure and $c(x)$ is the length of a horizontal cross section, then $p(x)c(x)\,dx$ is a local approximation to the force. The water pressure $p(x)$ is 0.03623 times the water's depth at location x.

NOTE: The dimensions are in inches and the force is in pounds.

EXERCISES for Section 7.6

A cubic inch of water weighs about 0.03623 pounds. All dimensions are in inches. In Exercises 1 to 4 find a definite integral for the force of water on the surface.

1. The triangular surface in Figure 7.6.4(a).

2. The circular surface in Figure 7.6.4(b).

3. The triangular surface in Figure 7.6.4(c).

4. The trapezoidal surface in Figure 7.6.4(d).

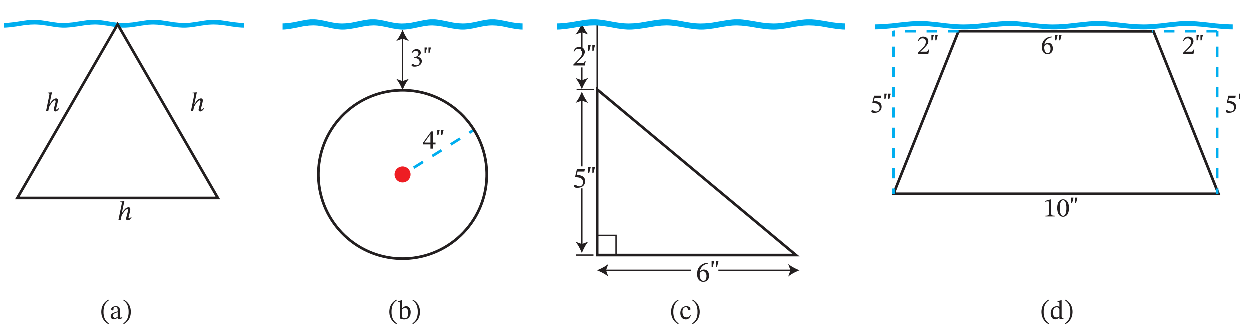


Figure 7.6.4

In Exercises 5 and 6 the surfaces are tilted like the bottoms of swimming pools. Find the force of the water against them.

5. The surface is an a by b rectangle inclined at an angle of 30° ($\pi/6$ radians) to the horizontal. The top edge of the surface is at a depth k. (See Figure 7.6.5.)

6. The surface is a disk of radius r tilted at an angle of 45° ($\pi/4$ radians) to the horizontal. Its top is at the surface of the water.

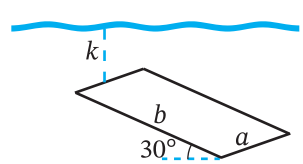


Figure 7.6.5

7. A vertical disk is totally submerged. Show that the force of the water against it is the same as the product of its area and the pressure at its center.

8. If the region in Exercise 7 is not vertical, is the same conclusion true?

9. Let $\mathscr{R}$ be a convex plane region. $\mathscr{R}$ is called **centrally symmetric** if it contains a point P such that P is the midpoint of every chord of $\mathscr{R}$ that passes through it. For instance, a parallelogram is centrally symmetric but no triangle is. Assume that a centrally symmetric region is placed vertically in water and is completely submerged. Show that the force against it equals the product of its area and the pressure at P.

10. Why is finding volume by shells essentially the same as finding the force against a submerged object?

7.7 Work

In this section we treat the work accomplished by a force operating along a line, for example the work done in stretching a spring.

Physical Background

If the force has the constant value F and it operates over a distance s in the direction of the force, then the work W accomplished is

$$\text{Work} = \text{Force} \cdot \text{Distance} \qquad \text{or} \qquad W = F \cdot s.$$

If force is measured in newtons and distance in meters, work is measured in newton-meters or joules:

$$1 \text{ joule} = 1 \text{ newton meter} = 1 \text{ watt second} = 0.7376 \text{ ft lb}.$$

For example, the force needed to lift a mass of m kilograms at Earth's surface is about $9.8m$ newtons.

A weightlifter who raises 100 kilograms a distance of 0.5 meter accomplishes $9.8(100)(0.5) = 490$ joules of work. A weightlifter who carries a barbell from one place to another in the weightlifting room, without raising or lowering it accomplishes no work because the barbell is not moved in the direction of the force.

The Stretched Spring

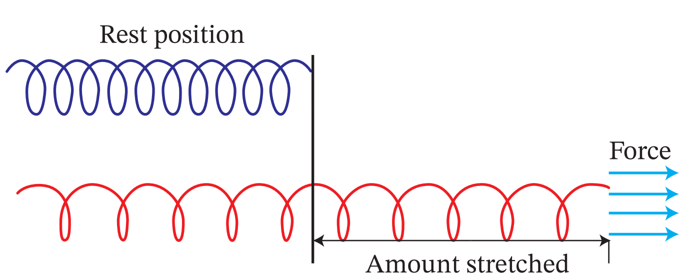


Figure 7.7.1

As you stretch a spring (or rubber band) from its rest position, the farther you stretch it the harder you have to pull. According to Hooke's law, the force you must exert is proportional to the distance that the spring is stretched, as shown in Figure 7.7.1. In symbols, $F = kx$, where F is the force and x is the distance from the rest position.

Because the force is not constant, we cannot compute the work accomplished by multiplying force times distance. We need an integral, as the next example illustrates.

EXAMPLE 1. A spring is stretched 0.5 meter longer than its rest length. The force required to keep it at that length is 3 newtons. Find the total work accomplished in stretching the spring 0.5 meter from its rest position.

SOLUTION Let us estimate the work involved in stretching the spring from x to $x+dx$. (See Figure 7.7.2.)

The distance dx is small. As the end of the spring is stretched from x to $x+dx$, the force is almost constant. Since the force is proportional to x, it is kx for some constant k. We know that the force, F, is 3 when $x=0.5$. The equation $F=kx$ implies $3=k(0.5)$, from which it follows that $k=6$.

$$F = kx \qquad \text{gives} \qquad 3 = k(0.5) \qquad \text{which implies} \qquad k = 6.$$

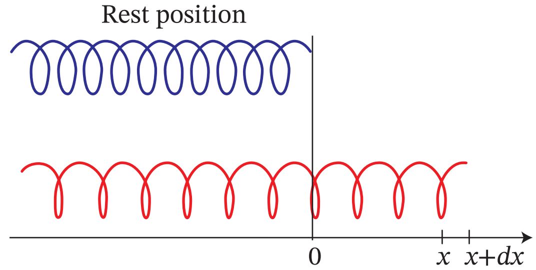


Figure 7.7.2

The work accomplished in stretching the spring from x to $x+dx$ is then approximately

$$\underbrace{kx}_{\text{force}} \cdot \underbrace{dx}_{\text{distance}} \text{ joule.}$$

Hence the total work is

$$\int_a^b kx\,dx = \int_0^{0.5} 6x\,dx = 3x^2\Big|_0^{0.5} = 0.75 \text{ joule.}$$

♦

Work in Launching a Rocket

When an object is rocketed into space, that the force of gravity diminishes with distance from the center of the earth is critical. The force of gravity that Earth exerts on an object diminishes as the object gets farther away from Earth. The work required to lift an object 1 foot at sea level is greater than the work required to lift it 1 foot at the top of Mt. Everest. However, the difference in altitudes is so small in comparison to the radius of the earth that the difference in work is negligible.

The force of gravity on a one-pound mass is proportional to the reciprocal of the square of the distance of the mass from the center of Earth.

That is, there is a constant k such that the gravitational force, $F(r)$, at distance r from the center of the earth is given by

$$F(r) = \frac{k}{r^2}.$$

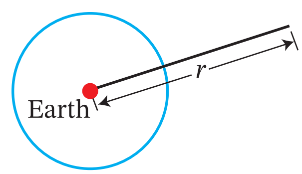


Figure 7.7.3

NOTE: In Figure 7.7.3, r is the distance to the center of Earth, not the distance to the surface. Earth's surface is about 4,000 miles from its center.

EXAMPLE 2. How much work is required to lift a 1 pound payload from Earth's surface to the moon, which is about 240,000 miles away?

SOLUTION The work necessary to lift an object a distance x against a constant vertical force F is the product of force times distance:

$$\text{Work} = F \cdot x.$$

The gravitational pull of the earth on the payload changes with distance from the center of the earth. As a result, an integral will be needed to express the total work.

The weight is 1 pound at the surface of Earth. Thus, the farther it is from the center of Earth the less it weighs for the force of Earth on the mass is inversely proportional to the square of its distance from Earth's center. Thus the force on the payload is k/r^2 pounds, where k is a constant, that will be determined in a moment and r is the distance in miles from the payload to the center of the earth. When $r = 4{,}000$ (miles), the force is 1 pound, so

$$1 \text{ pound} = \frac{k}{(4{,}000 \text{ miles})^2}.$$

From this it follows that $k = 4{,}000^2$, and therefore the gravitational force on a 1-pound mass is $(4{,}000/r)^2$ pounds. As the payload recedes from Earth it weight less and is easier to lift. as recorded in Figure 7.7.4(a).

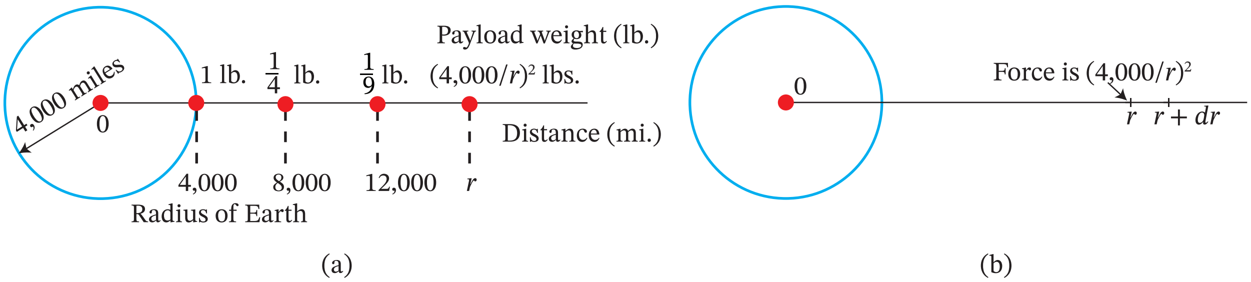


Figure 7.7.4

Figure 7.7.4(b) illustrates that the work done in lifting the payload from a distance r to a distance $r + dr$ from the center of the earth is approximately

$$\underbrace{\left(\frac{4{,}000}{r}\right)^2}_{\text{force}} \underbrace{(dr)}_{\text{distance}} \text{ miles-pounds.}$$

Hence the work required to move the 1 pound mass from Earth's surface the moon is given by

$$\begin{aligned}
\int_{4{,}000}^{240{,}000} \left(\frac{4{,}000}{r}\right)^2 dr = -\frac{4{,}000^2}{r}\bigg|_{4{,}000}^{240{,}000} \quad &= -4{,}000^2\left(\frac{1}{240{,}000} - \frac{1}{4{,}000}\right) \\
= -\frac{4{,}000}{60} + 4{,}000 \quad &\approx 3{,}933 \text{ mile-pounds} \\
= 2.8154 \times 10^7 \text{ joules.}
\end{aligned}$$

♦

Summary

The work accomplished by a constant force F that moves an object a distance x in the direction of the force is the product Fx, force times distance. The work by a variable force, $F(x)$, moving an object over the interval $[a, b]$ is $\int_a^b F(x)\, dx$.

EXERCISES for Section 7.7

1. A spring is stretched 0.20 meters from its rest length. The force required to keep it at that length is 5 newtons. Assuming that the force of the spring is proportional to the distance it is stretched, find the work accomplished in stretching the spring (a) 0.20 meters from its rest length and (b) 0.30 meters from its rest length.

2. A spring is stretched 3 meters from its rest length. The force required to keep it at that length is 24 newtons. Assume that the force of the spring is proportional to the distance it is stretched. Find the work accomplished in stretching the spring (a) 3 meters from its rest length and (b) 4 meters from its rest length.

3. Suppose a spring does not obey Hooke's law. Instead, the force it exerts when stretched x meters from its rest length is $F(x) = 3x^2$ newtons. Find the work done in stretching it 0.80 meters from its rest length.

4. Suppose a spring does not obey Hooke's law. Instead, the force it exerts when stretched x meters from its rest length is $F(x) = 2\sqrt{x}$ newtons. Find the work done in stretching it 0.50 meters from its rest length.

5. What fraction of the total work done in lifting the payload in Example 2 to the moon is done during the first 4,000 miles of its journey to the moon?

6. If a mass that weighs 1 pound at Earth's surface is launched from a position 20,000 miles from the center of Earth, how much work would be required to send it to the moon, 240,000 miles from the center of the earth? How does this compare with the work required to launch this payload from Earth's surface to the moon?

7. Assume that the force of gravity obeys an inverse cube law, so that the force on a 1 pound payload a distance r miles from the center of Earth ($r \geq 4{,}000$) is $\left(\dfrac{4{,}000}{r}\right)^3$ pounds. How much work would be required to lift a 1 pound payload from the surface of the Earth to the moon?

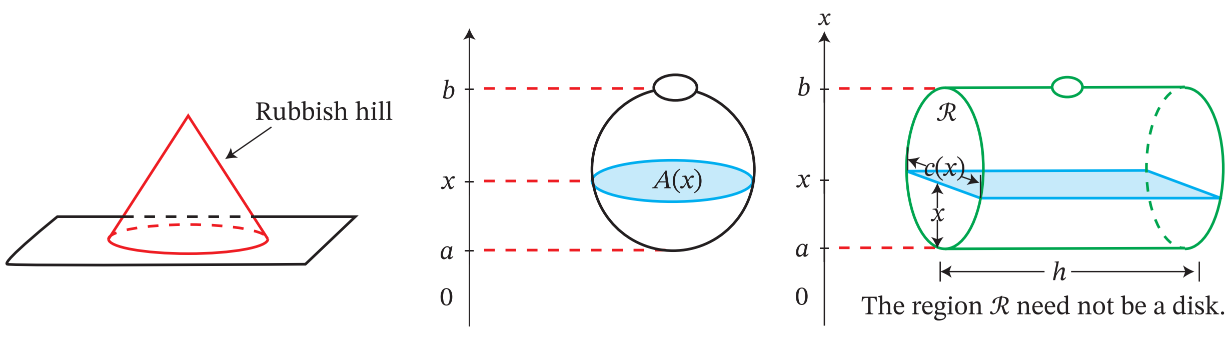
8. Geologists, when considering the origin of mountain ranges, estimate the energy required to lift a mountain up from sea level. Assume that two mountains are composed of the same type of matter, which weighs k pounds per cubic foot. Both are right circular cones in which the height is equal to the radius. One mountain is twice as high as the other. The base of each is at sea level. If the work required to lift the matter in the smaller mountain above sea level is W, what is the corresponding work for the larger mountain?

9. Assume that Mt. Everest has a shape of a right circular cone of height 30,000 feet and radius 150,000 feet, with uniform density of 200 pounds per cubic foot.

(a) How much work was required to lift the material in Mt. Everest if it was initially all at sea level?
(b) How does this work compare with the energy of a 1 megaton hydrogen bomb?

One megaton is the energy in a million tons of TNT, about 3×10^{14} foot-pounds.

10. A town in a flat valley made a conical hill out of its rubbish, as shown in Figure 7.7.5(a). The work required to lift all the rubbish was W. Happy with the result, the town decided to make another hill with the same slope and twice the volume. How much work will be required to build it? Explain.



(a) (b) (c)

Figure 7.7.5

11. The container in Figure 7.7.5(b) is full of water. The cross-sectional area at height x is $A(x)$ for $a < x < b$. (Do not assume the tank is spherical.) All the water is pumped out of an opening at the top of the container. Develop a definite integral for the work accomplished.

RECALL: Water weighs 64.2 pounds per cubic foot.

12. A cylindrical tank with ends $\mathscr{R}$ (which do not have to be disks) and length h feet is completely full of water. The tank is laid on its side. An opening is created in the middle of the top side of the tank, as shown in Figure 7.7.5(c). Assume the area of $\mathscr{R}$ is given by $\int_a^b c(x)\,dx$ where $c(x)$, $a \le x \le b$, is the width of the cross section of $\mathscr{R}$. Develop a definite integral for the total work accomplished when all the water is pumped out of the opening at the top.

In Exercise 13 to 18 a and b are constants. In each case verify that the derivative of the first expression is the second expression.

13. $\ln\left(x+\sqrt{a^2+x^2}\right)$; $\dfrac{1}{\sqrt{a^2+x^2}}$

14. $\dfrac{1}{2ab}\ln\left|\dfrac{a+bx}{a-bx}\right|$; $\dfrac{1}{a^2-b^2x^2}$

15. $x-\ln\left(1+e^x\right)$; $\dfrac{1}{1+e^x}$

16. $\dfrac{e^{ax}}{a^2+1}(a\sin(x)-\cos(x))$; $e^{ax}\sin(x)$

17. $\dfrac{1}{b}\ln\left|\dfrac{x}{ax+b}\right|$; $\dfrac{1}{x(ax+b)}$

18. $\dfrac{x^4}{8}-\left(\dfrac{x^3}{4}-\dfrac{3x}{8}\right)\sin(2x)-\dfrac{3}{8}\left(x^2\cos(2x)+\sin^2(x)\right)$; $x^3\sin^2(x)$

7.8 Improper Integrals

This section develops the analog of a definite integral when the interval of integration is infinite or the integrand becomes arbitrarily large in the interval of integration. This type of integral appears in many applications. The definition of a definite integral does not cover it.

Improper Integrals: Interval Unbounded

A question about areas will introduce the notion of an improper integral. Figure 7.8.1 shows the region under $y = 1/x$ and above the unbounded interval $[1,\infty)$.

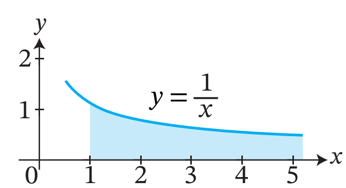


Figure 7.8.1

We would like to say that the area in Figure 7.8.1 is $\int_1^\infty dx/x$. Unfortunately, the symbol $\int_1^\infty f(x)\,dx$, with an unbounded interval of integration, has not yet been given any meaning.

Recall that the definition of the definite integral $\int_a^b f(x)\,dx$ on a bounded interval $[a,b]$ involves a limit of sums of the form $\sum_{i=1}^n f(c_i)(x_i - x_{x-1})$, where each $x_i - x_{i-1}$ is the length of an interval $[x_{i-1}, x_i]$. If you cut the interval $[1,\infty)$ into a finite number of intervals, then at least one section has infinite length, and the sum is meaningless.

It does make sense, however, to find the area of that part of the region in Figure 7.8.1 from $x=1$ to $x=b$, where $b>1$, and find what happens to these areas as $b\to\infty$. To do this, first calculate the definite integral over the interval $[1,b)$:

$$\int_1^b \frac{dx}{x} = \ln(x)\Big|_1^b = \ln(b)-\ln(1) = \ln(b).$$

Then, look at the values of these integrals in the limit as $b\to\infty$:

$$\lim_{b\to\infty}\int_1^b \frac{dx}{x} = \lim_{b\to\infty}\ln(b) = \infty.$$

In this sense, we will say the area of the region in Figure 7.8.1 is infinite.

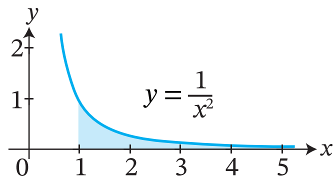


Figure 7.8.2

Next, examine the area of the shaded region in Figure 7.8.2. This is the region under $y = 1/x^2$ and above the interval $[1, \infty)$. In this case we find

$$\int_1^b \frac{dx}{x^2} = -\frac{1}{x}\bigg|_1^b = -\frac{1}{b} - \left(-\frac{1}{1}\right) = 1 - \frac{1}{b} \quad \text{and} \quad \lim_{b\to\infty}\int_1^b \frac{dx}{x^2} = \lim_{b\to\infty}\left(1 - \frac{1}{b}\right) = 1.$$

The area of the region in Figure 7.8.2 is finite; in fact, this area is 1.

Though the regions in Figures 7.8.1 and 7.8.2 look similar, one has an infinite area, and the other a finite area. These examples motivate the following two definitions.

Definition: *Convergent Improper Integral* $\int_a^\infty f(x)\, dx$.

Let f be continuous for $x \geq a$. If $\lim_{b\to\infty}\int_a^b f(x)\, dx$ exists, it is denoted $\int_a^\infty f(x)\, dx$ and the integral is called a **convergent improper integral**. The integral is said to converge and its value is equal to that limit.

We saw that $\int_1^\infty dx/x^2$ converges, with value 1.

Definition: *Divergent Improper Integral* $\int_a^\infty f(x)\, dx$.

Let f be a continuous function for $x \geq a$. If $\lim_{b\to\infty}\int_a^b f(x)\, dx$ does not exist, then $\int_a^\infty f(x)\, dx$ is called a **divergent improper integral**. The integral is said to diverge.

The improper integral $\int_1^\infty dx/x$ diverges because $\int_1^b dx/x \to \infty$ as $b \to \infty$.

An improper integral $\int_a^\infty f(x)\, dx$ can diverge without being infinite. For $\int_0^\infty \cos(x)\, dx$ we have

$$\int_0^b \cos(x)\, dx = \sin(x)\bigg|_0^b = \sin(b).$$

As $b \to \infty$, $\sin(b)$ does not approach a limit, nor does it become arbitrarily large. As $b \to \infty$, $\sin(b)$ oscillates in the range -1 to 1 infinitely often. Thus $\int_0^\infty \cos(x)\, dx$ is divergent.

The improper integral $\int_{-\infty}^b f(x)\, dx$ is defined similarly:

Definition: *Other Improper Integrals* $\int_{-\infty}^{b} f(x)\,dx$ *and* $\int_{-\infty}^{\infty} f(x)\,dx$

1. Let $f(x)$ be continuous for $x \le b$. If $\lim_{a\to-\infty}\int_a^b f(x)\,dx$ exists, then $\int_{-\infty}^{b} f(x)\,dx$ is a **convergent improper integral** and $\int_{-\infty}^{b} f(x)\,dx = \lim_{a\to-\infty}\int_a^b f(x)\,dx$.
2. Let $f(x)$ be continuous for $x \le b$. If $\lim_{a\to-\infty}\int_a^b f(x)\,dx$ does not exist, then $\int_{-\infty}^{b} f(x)\,dx$ is a **divergent improper integral**.
3. Let $f(x)$ be continuous for all x $(-\infty < x < \infty)$. For an improper integral over the entire x-axis, $\int_{-\infty}^{\infty} f(x)\,dx$ is a **convergent improper integral** if and only if both

$$\int_{-\infty}^{0} f(x)\,dx \qquad \text{and} \qquad \int_{0}^{\infty} f(x)\,dx \tag{7.8.1}$$

are convergent. In that case, we will say

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\infty} f(x)\,dx.$$

And, if at least one of the improper integrals in (7.8.1) is divergent, $\int_{-\infty}^{\infty} f(x)\,dx$ will be called **divergent**.

EXAMPLE 1. Is the area of the region bounded by the curve $y = \dfrac{1}{1+x^2}$ and the x-axis finite or infinite?

SOLUTION

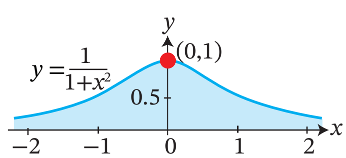


Figure 7.8.3

The region is graphed in Figure 7.8.3. The area equals $\int_{-\infty}^{\infty} dx/(1+x^2)$, if this improper integral converges.

To determine whether this improper integral converges or diverges rewrite the improper integral as the sum of two improper integrals with one finite endpoint:

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^2} = \int_{-\infty}^{0} \frac{dx}{1+x^2} + \int_{0}^{\infty} \frac{dx}{1+x^2}$$

Because the integrand, $1/(1+x^2)$, is an even function, the two improper integrals either both converge (with the same value) or both diverge.

The convergence of $\int_0^\infty dx/(1+x^2)$ is determined by considering the limit of associated definite integrals over bounded intervals $[0, b]$:

$$\begin{aligned}\int_0^\infty \frac{dx}{1+x^2} &= \lim_{b\to\infty} \int_0^b \frac{dx}{1+x^2} \\ &= \lim_{b\to\infty} (\arctan(b) - \arctan(0)) \\ &= \frac{\pi}{2} - 0 = \frac{\pi}{2}.\end{aligned}$$

Because this limit exists the definite integral converges, and has value $\pi/2$.

Hence,

$$\int_{-\infty}^\infty \frac{dx}{1+x^2} = \frac{\pi}{2} + \frac{\pi}{2} = \pi,$$

the integral is convergent and the total area under $y = 1/(1+x^2)$ is π. ♦

Observation 7.8.1: *Shortcut Notation to Evaluate a Convergent Improper Integral* $\int_a^\infty f(x)\, dx$

If $\int_a^\infty f(x)\, dx$ is convergent and $F(x)$ is an antiderivative of $f(x)$, then $\int_a^\infty f(x)\, dx = \lim_{b\to\infty} F(b) - F(a)$. We write $\int_a^\infty f(x)\, dx = F(x)\Big|_a^\infty$ where it is understood that "$F(\infty)$" is short for the limit "$\lim_{b\to\infty} F(b)$".

CAUTION: This notation should be used only when the improper integral is known to be convergent.

Comparison Test for the Convergence of $\int_a^\infty f(x)\, dx$, $f(x) \geq 0$

The integral $\int_0^\infty e^{-x^2}\, dx$ is important in statistics. Is it convergent or divergent? We cannot evaluate $\int_0^b e^{-x^2}\, dx$ by the FTC since e^{-x^2} does not have an elementary antiderivative. There is a way of showing that $\int_0^\infty e^{-x^2}\, dx$ converges without finding its exact value. The essential idea is described in Theorem 7.8.2.

Theorem 7.8.2: Comparison Test for the Convergence of Improper Integrals

Let $f(x)$ and $g(x)$ be continuous functions for $x \geq a$. Assume that $0 \leq f(x) \leq g(x)$ and that $\int_a^\infty g(x)\, dx$ converges. Then $\int_a^\infty f(x)\, dx$ converges and $\int_a^\infty f(x)\, dx \leq \int_a^\infty g(x)\, dx$.

The parallel result in cases where $g(x) \leq f(x) \leq 0$ is: If $\int_a^\infty g(x)\, dx$ converges, so does $\int_a^\infty f(x)\, dx$.

While we omit a proof of Theorem 7.8.2, the following observation makes it believable, and suggests how the result could be proven.

Observation 7.8.3: *Geometric Interpretation of Comparison Test for the Convergence of Improper Integrals*

Geometrically, Theorem 7.8.2 says if the area under the red curve ($y = g(x)$) is finite, so is the area under the cyan curve ($y = f(x)$).

The converse of Theorem 7.8.2 is not necessarily true. If the area under the lower curve is finite, the area under the upper curve might not be finite.

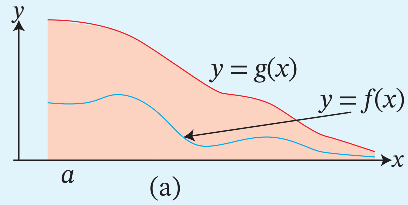


Figure 7.8.4

EXAMPLE 2. Show that $\displaystyle\int_0^\infty e^{-x^2}\,dx$ converges and put a bound on its value.

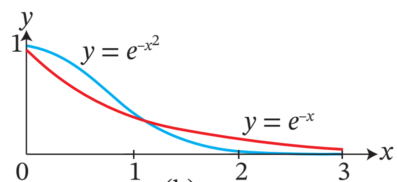


Figure 7.8.5

SOLUTION Since e^{-x^2} does not have an elementary antiderivative, we compare $\int_0^\infty e^{-x^2}\,dx$ to an improper integral that we know converges.

For $x \geq 1$, $x^2 \geq x$ so $e^{-x^2} \leq e^{-x}$. (See Figure 7.8.5.) We know

$$\int_1^b e^{-x}\,dx = -e^{-x}\Big|_1^b = e^{-1} - e^{-b}, \qquad \text{and} \qquad \lim_{b\to\infty}\int_1^b e^{-x}\,dx = \frac{1}{e},$$

so the improper integral $\int_1^\infty e^{-x}\,dx$ converges.

The comparison test for convergence of improper integrals, Theorem 7.8.2, tells us that $\int_1^\infty e^{-x^2}\,dx$ is also convergent. While Theorem 7.8.2 does not provide a way to determine the value of this improper integral, it does provide an upper bound: $\int_1^\infty e^{-x^2}\,dx \leq \int_1^\infty e^{-x}\,dx = 1/e$. Thus

$$\int_0^\infty e^{-x^2}\,dx = \int_0^1 e^{-x^2}\,dx + \int_1^\infty e^{-x^2}\,dx \leq \int_0^1 e^{-x^2}\,dx + \frac{1}{e}.$$

Since $e^{-x^2} \leq 1$ for $0 < x \leq 1$, we conclude that $\displaystyle\int_0^\infty e^{-x^2}\,dx \leq 1 + \frac{1}{e}. \approx 1.3679.$

In fact, $\int_0^\infty e^{-x^2}\,dx = \sqrt{\pi}/2 \approx 0.8862$, as will be found in Exercise 38 of Section 17.3. ♦

Comparison Test for the Divergence of $\displaystyle\int_a^\infty f(x)\,dx$.

Theorem 7.8.4: Comparison Test for Divergence of Improper Integrals

Let $f(x)$ and $g(x)$ be continuous functions for $x \geq a$. Assume that $0 \leq g(x) \leq f(x)$ and that $\displaystyle\int_a^\infty g(x)\,dx$ is divergent. Then $\displaystyle\int_a^\infty f(x)\,dx$ is also divergent.

Observation 7.8.5: *Geometric Interpretation of Comparison Test for Divergence of Improper Integrals*

Figure 7.8.6 shows why Theorem 7.8.4 is true.

The area under $f(x)$ is larger than the area under $g(x)$. When the area under $g(x)$ is infinite, the area under f must also be infinite. That is, $\int_a^\infty f(x)\,dx$ is a divergent improper integral.

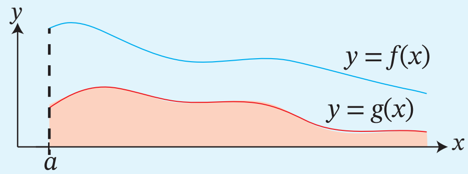


Figure 7.8.6

EXAMPLE 3. Show that $\int_1^\infty \frac{x^2+1}{x^3}\,dx$ diverges.

SOLUTION For $x > 0$,

$$\frac{x^2+1}{x^3} > \frac{x^2}{x^3} = \frac{1}{x}.$$

Since $\int_1^\infty dx/x = \infty$, it follows that $\int_1^\infty (x^2+1)/x^3\,dx = \infty$. ♦

Convergence of $\int_a^\infty f(x)\,dx$ When $\int_a^\infty |f(x)|\,dx$ Converges

Is $\int_0^\infty e^{-x}\sin(x)\,dx$ convergent or divergent? Because $\sin(x)$ takes on both positive and negative values, the integrand is not always positive, nor is it always negative. It is not possible to compare it with $\int_0^\infty e^{-x}\,dx$.

The next theorem provides a way to establish the convergence of $\int_a^\infty f(x)\,dx$ when $f(x)$ is a function that takes on both positive and negative values. It says that if $\int_a^\infty |f(x)|\,dx$ converges, so does $\int_a^\infty f(x)\,dx$. We show that the negative and positive parts of the function both have convergent integrals.

Theorem 7.8.6: Absolute Convergence Test for Improper Integrals

If $f(x)$ is continuous for $x \geq a$ and $\int_a^\infty |f(x)|\,dx$ converges to the number L, then $\int_a^\infty f(x)\,dx$ converges to a number between L and $-L$.

Proof of Theorem 7.8.6

We introduce two functions, $g(x)$, which is nonnegative, and $h(x)$, which is nonpositive. That they are both continuous is shown in Exercise 43. They enable us to use the comparison tests. Let

$$g(x) = \begin{cases} f(x) & \text{if } f(x) \text{ is positive} \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad h(x) = \begin{cases} f(x) & \text{if } f(x) \text{ is negative} \\ 0 & \text{otherwise} \end{cases}$$

Then $f(x) = g(x) + h(x)$ and $g(x)$ and $h(x)$ are continuous for $x \geq a$. We show that $\int_a^\infty g(x)\,dx$ and $\int_a^\infty h(x)\,dx$ converge.

Figure 7.8.7 shows the graphs of $y = f(x)$ along with the graphs of $y = g(x)$, $y = h(x)$, $y = |f(x)|$, and $y = -|f(x)|$. The area of the shaded region under the graph of $y = |f(x)|$ is $\int_a^\infty |f(x)|\,dx$.

Because $\int_a^\infty |f(x)|\,dx$ converges, has value L, and $0 \leq g(x) \leq |f(x)|$, then $\int_a^\infty g(x)\,dx$ converges, and the value of the integral is a number A between 0 and L we conclude that

$$0 \leq A \leq \int_a^\infty |f(x)|\,dx = L.$$

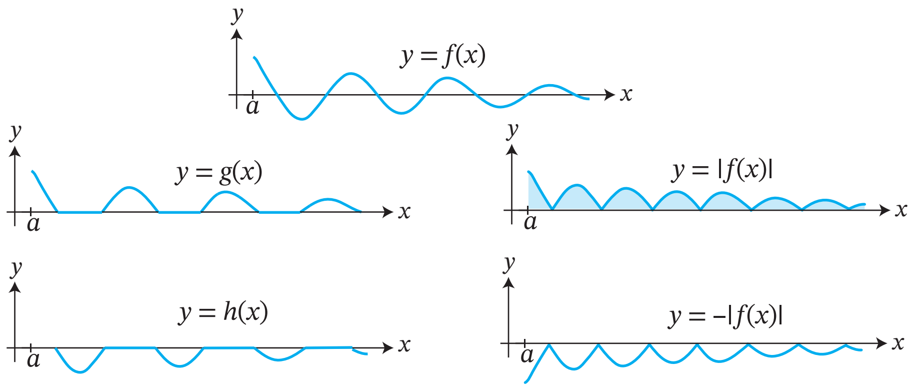


Figure 7.8.7

Since $\int_a^\infty -|f(x)|\ dx$ converges, has value $-L$, and $-|f(x)| \leq h(x) \leq 0$, we conclude that $\int_a^\infty h(x)\ dx$ converges and its value is a number B between $-L$ and 0:

$$-L = -\int_a^\infty |f(x)|\ dx \leq B \leq 0.$$

Thus $\int_a^\infty f(x)\ dx = \int_a^\infty (g(x)+h(x))\ dx$ converges to $A+B$, which is a number somewhere in the interval $[-L, L]$. ●

EXAMPLE 4. Show that $\displaystyle\int_0^\infty e^{-x}\sin(x)\ dx$ converges.

SOLUTION Since $|\sin(x)| \leq 1$, we have $|e^{-x}\sin(x)| \leq e^{-x}$. From Example 2 we know $\int_0^\infty e^{-x}\ dx$ converges. Thus $\int_0^\infty e^{-x}\sin(x)\ dx$ converges. ◆

Improper Integrals: Integrand Unbounded

The second type of improper integral concerns functions which become infinite in an interval $[a, b]$. For any partition of $[a, b]$, the approximating sum $\sum_{i=1}^n f(c_i)(x_i - x_{i-1})$ can be made arbitrarily large when c_i is chosen so that $f(c_i)$ is very large. (See Exercise 37.) The next example shows how to get around this difficulty.

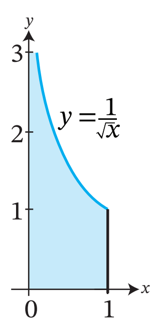


Figure 7.8.8

EXAMPLE 5. Determine the area of the region bounded by $y = \dfrac{1}{\sqrt{x}}$, $x = 1$, and the coordinate axes shown in Figure 7.8.8.

SOLUTION We cannot write Area $= \int_0^1 1/\sqrt{x}\ dx$ because the integral $\int_0^1 1/\sqrt{x}\ dx$ is not defined. Its integrand is unbounded in $[0, 1]$. However $\int_t^1 1/\sqrt{x}\ dx$ is defined if $t > 0$ and

$$\int_t^1 \frac{1}{\sqrt{x}}\ dx = 2\sqrt{x}\ \Big|_t^1 = 2 - 2\sqrt{t},$$

Thus

$$\lim_{t\to 0^+} \int_t^1 \frac{dx}{\sqrt{x}} = \lim_{t\to 0^+}(2 - 2\sqrt{t}) = 2.$$

The area of the shaded region in Figure 7.8.8 is 2. ◆

NOTE: In Exercise 30 the same value for the shaded area is obtained by taking horizontal cross sections and evaluating an improper integral from 0 to ∞.

The reasoning in Example 5 motivates the definition of the second type of improper integral, in which the integrand rather than the interval is unbounded.

Definition: *Convergent and Divergent Improper Integrals* $\int_a^b f(x)\,dx$

Let f be continuous at every number in $[a,b]$ except at a.

1. If $\lim_{t\to a^+}\int_t^b f(x)dx$ exists, it is denoted $\int_a^b f(x)\,dx$ and is called a **convergent improper integral**.
2. If $\lim_{t\to a^+}\int_t^b f(x)dx$ does not exist, then $\int_a^b f(x)\,dx$ is called a **divergent improper integral**.
3. If f is not defined at b, define $\int_a^b f(x)\,dx$ as $\lim_{t\to b^-}\int_a^t f(x)dx$, if this limit exists.

As Example 5 showed, the improper integral $\int_0^1 dx/\sqrt{x}$ is convergent and has the value 2.

If a function $f(x)$ is not defined for more than one number, break the domain of $f(x)$ into intervals $[a,b]$ for which $\int_a^b f(x)\,dx$ is either improper or proper (an ordinary definite integral).

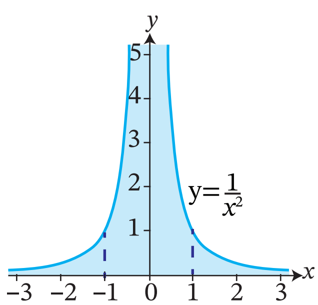


Figure 7.8.9

For instance, the improper integral $\int_{-\infty}^{\infty} dx/x^2$ is troublesome because $\lim_{x\to 0^-} 1/x^2 = \infty$, $\lim_{x\to 0^+} 1/x^2 = \infty$, and the range extends infinitely to the left and also to the right. (See Figure 7.8.9.) To treat the integral, we can write it as the sum of four improper integrals:

$$\int_{-\infty}^{\infty}\frac{1}{x^2}\,dx = \int_{-\infty}^{-1}\frac{1}{x^2}\,dx + \int_{-1}^{0}\frac{1}{x^2}\,dx + \int_{0}^{1}\frac{1}{x^2}\,dx + \int_{1}^{\infty}\frac{1}{x^2}\,dx.$$

and treat each separately. All the integrals on the right must converge for $\int_{-\infty}^{\infty} dx/x^2$ to converge. Only the first and last do, so $\int_{-\infty}^{\infty} dx/x^2$ diverges.

Summary

We introduced two types of integrals that are not definite integrals, but are defined as limits of definite integrals.

The improper integral $\int_a^{\infty} f(x)\,dx$ is defined as $\lim_{b\to\infty}\int_a^b f(x)\,dx$. If $f(x)$ is continuous in $[a,b]$ except at a, then $\int_a^b f(x)\,dx$ is defined as $\lim_{t\to a^+}\int_t^b f(x)\,dx$. The first type is far more common in applications.

We also developed two comparison tests for convergence or divergence of $\int_a^{\infty} f(x)\,dx$, where the integrand keeps a constant sign. If $f(x)$ has both positive and negative values, we showed that if $\int_a^{\infty} |f(x)|\,dx$ converges, so does $\int_a^{\infty} f(x)\,dx$.

EXERCISES for Section 7.8

In Exercises 1 to 21 determine whether the improper integral converges or diverges. Evaluate the convergent ones if possible. Some may require using the book's Table of Integrals (in Appendix A).

1. $\displaystyle\int_1^\infty \frac{dx}{x^3}$

2. $\displaystyle\int_1^\infty \frac{dx}{\sqrt[3]{x}}$

3. $\displaystyle\int_0^\infty e^{-x}\,dx$

4. $\displaystyle\int_0^\infty \frac{dx}{x+100}$

5. $\displaystyle\int_0^\infty \frac{x^3\,dx}{x^4+1}$

6. $\displaystyle\int_1^\infty x^{-1.01}\,dx$

7. $\displaystyle\int_0^\infty \frac{dx}{(x+2)^3}$

8. $\displaystyle\int_0^\infty \sin(2x)\,dx$

9. $\displaystyle\int_1^\infty x^{-0.99}\,dx$

10. $\displaystyle\int_0^\infty \frac{e^{-x}\sin(x^2)}{x+1}\,dx$

11. $\displaystyle\int_0^\infty \frac{dx}{x^2+4}$

12. $\displaystyle\int_0^\infty \frac{x^2\,dx}{2x^3+5}$

13. $\displaystyle\int_0^\infty \frac{dx}{(x+1)(x+2)(x+3)}$

14. $\displaystyle\int_0^\infty \frac{\sin(x)}{x^2}\,dx$

15. $\displaystyle\int_1^\infty \frac{\ln(x)}{x}\,dx$

16. $\displaystyle\int_0^\infty e^{-2x}\sin(3x)\,dx$

17. $\displaystyle\int_0^1 \frac{dx}{\sqrt[3]{x}}$

18. $\displaystyle\int_1^\infty \frac{dx}{\sqrt[3]{x}}$

19. $\displaystyle\int_0^1 \frac{dx}{(x-1)^2}$

20. $\displaystyle\int_0^\infty \frac{e^{-x}}{\sqrt{x}}\,dx$

21. $\displaystyle\int_0^1 \frac{dx}{\sqrt{x}\sqrt{1-x}}$

In Exercises 22 to 24 determine all constants p for which the given integral (a) is a convergent improper integral, (b) is a divergent improper integral, and (c) is not an improper integral..

22. $\displaystyle\int_0^1 \frac{dx}{x^p}$

23. $\displaystyle\int_1^\infty \frac{dx}{x^p}$

24. $\displaystyle\int_0^\infty \frac{dx}{x^p}$

---

25. Let $\mathscr{R}$ be the region between the curves $y=\dfrac{1}{x}$ and $y=\dfrac{1}{x+1}$ to the right of the line $x=1$. Is the area of $\mathscr{R}$ finite or infinite? If it is finite, evaluate it.

26. Let $\mathscr{R}$ be the region between the curves $y=\dfrac{1}{x}$ and $y=\dfrac{1}{x^2}$ to the right of $x=1$. Is the area of $\mathscr{R}$ finite or infinite? If it is finite, evaluate it.

27. Describe how you would go about estimating $\displaystyle\int_0^\infty e^{-x^2}\,dx$ with an error less than 0.02. Do not evaluate.

28. Describe how you would go about estimating $\displaystyle\int_0^\infty \frac{dx}{\sqrt{1+x^4}}$ with an error less than 0.01. Do not evaluate.

29. Example 4 showed that $\displaystyle\int_0^\infty e^{-x}\sin(x)\,dx$ is convergent. Find its value.

30. In Example 5 the region bounded by $y = \frac{1}{\sqrt{x}}$, $x = 1$, and the coordinate axes was found to have area 2. Confirm this by using horizontal cross sections and evaluating an improper integral from 0 to ∞.

31. The function $f(x) = \frac{\sin(x)}{x}$ for $x \neq 0$ and $f(0) = 1$ occurs in communication theory. Show that the energy $E = \int_{-\infty}^{\infty} \left(f(x)\right)^2 \, dx$ of the signal represented by f is finite.

32. Let $f(x)$ be a positive function and let $\mathscr{R}$ be the region under $y = f(x)$ and above $[1,\infty]$. Assume that the area of $\mathscr{R}$ is infinite. Does it follow that the volume of the solid of revolution formed by revolving $\mathscr{R}$ about the x-axis is infinite?

33. Let $f(x)$ be a positive function and let R be the region under $y = f(x)$ and above $[1,\infty]$. Assume that the area of R is finite. Does it follow that the volume of the solid of revolution formed by revolving R about the x-axis is finite?

34. (a) Sketch the graph of $y = \frac{1}{x}$, for $x > 0$. (b) Is the region bounded by the y-axis, $y = 1$, and the graph of $y = \frac{1}{x}$ congruent to the region below the graph of $y = \frac{1}{x}$ and above $[1,\infty)$? (c) Determine the convergence or divergence of $\int_0^1 \frac{dx}{x}$ and $\int_1^{\infty} \frac{dx}{x}$.

35. (a) Sketch the graph of $y = \frac{1}{x^2}$ for $x > 0$. (b) Is the region bounded by the y-axis, $y = 1$, and the graph of $y = \frac{1}{x^2}$ congruent to the region below the graph of $y = \frac{1}{x^2}$ and above $[1,\infty)$? (c) Determine the convergence or divergence of $\int_0^1 \frac{dx}{x^2}$ and $\int_1^{\infty} \frac{dx}{x^2}$.

36. In the study of harmonic oscillators the integral $\int_{-\infty}^{\infty} \frac{dx}{(1+kx^2)^3}$ occurs, where k is a positive constant. Show this improper integral converges.

37. Consider the improper integral $\int_0^1 \frac{dx}{x^2}$. Partition the interval $[0,1]$ into n equal-width pieces; $x_i = \frac{i}{n}$ for $i = 0$, 1, $\dots$, n. Show that (a) the approximating sum $S_n = \sum_{i=1}^{n} \frac{1}{c_i^2} \Delta x_i = \sum_{i=1}^{n} \frac{n}{i^2}$ and (b) $\lim_{n\to\infty} S_n$ does not exist.

38. Plankton are small football-shaped organisms. The resistance they meet when falling through water is proportional to $\int_0^{\infty} \frac{dx}{\sqrt{(a^2+x)(b^2+x)(c^2+x)}}$, where a, b, and c describe the dimensions of the plankton. Is the integral convergent or divergent? Explain.

39. In R. P. Feynman, *Lectures on Physics*, Addison-Wesley, Reading, MA, 1963, one finds: "...the expression becomes $\frac{U}{V} = \frac{(kT)^4}{\hbar^3 \pi^2 c^3} \int_0^{\infty} \frac{x^3 \, dx}{e^x - 1}$. This integral is just some number that we can get, approximately, by drawing a curve and taking the area by counting squares. It is roughly 6.5. The mathematicians among us can show that the integral is exactly $\frac{\pi^4}{15}$." Show that the improper integral is convergent.

40. Assume that $f(x)$ is continuous and nonnegative and that $\int_1^\infty f(x)\,dx$ is convergent.

(a) Show by sketching a graph that $\lim_{x\to\infty} f(x)$ may not exist.
(b) Show that if we add the condition that f is a decreasing function, then $\lim_{x\to\infty} f(x) = 0$.

41. Show that Theorem 7.8.2 implies Theorem 7.8.4.

42. If A is in $[0,L]$ and B is in $[-L,0]$, why is $A+B$ in $[-L,L]$?

This result is used in the proof of Theorem 7.8.6.

43. Here is the standard proof of Theorem 7.8.6, the absolute convergence test:

> Assume that $\int_0^\infty |f(x)|\,dx$ converges. Define $g(x) = f(x)+|f(x)|$. Because $0 \le g(x) \le 2|f(x)|$, $\int_0^\infty g(x)\,dx$ converges. That is, $\int_0^\infty (f(x)+|f(x)|)\,dx$ converges. It follows, since $f(x) = (f(x)+|f(x)|) - |f(x)|$, that $\int_0^\infty f(x)\,dx$ converges.

(a) Explain why each step in this argument is valid.
(b) State the advantages and disadvantages of the standard proof and the proof in the text.

7.S Chapter Summary

There were two themes in this chapter. One was to make a large, clear drawing when setting up a definite integral. The other is make a local approximation of the total quantity whether the quantity was area, volume, force of water, work, or something altogether different. If the local approximation is $f(x)\,dx$, the total quantity is represented by a definite integral $\int_a^b f(x)\,dx$ (or an improper integral). Table 7.S.1 summarizes some applications of the definite integral.

Sec	Application	Memory Aid	General Formula
§ 7.1	Area by slicing (horizontal or vertical)	Area of narrow strip $\approx c(x)\,\Delta x$ (See Figure 7.S.1(a).)	Area of region $= \int_a^b c(x)\,dx$
§ 7.4	Volume of solid (cross sections perpendicular to axis)	Volume of thin disk $\approx A(x)\,\Delta x$ (See Figure 7.S.1(b).)	Volume of region $= \int_a^b A(x)\,dx$
§ 7.5	Solid of revolution (cross sections parallel to axis)	Volume of thin shell $\approx 2\pi R(x)c(x)\,\Delta x$ (See Figure 7.S.1(c).)	Volume of region $= \int_a^b 2\pi R(x)c(x)\,dx$
§ 7.6	Force of water	Force on thin slab $\approx \rho D(x)c(x)\,\Delta x$ (See Figure 7.S.1(d).)	Force on object $= \int_a^b \rho D(x)c(x)\,dx$
§ 7.7	Work to move an object	Work to move thin slice $\approx F(x)\,\Delta x$ (See Figure 7.S.1(e).)	Work to move object $= \int_a^b F(x)\,dx$

Table 7.S.1

The final section, on improper integrals, showed how to deal with integrals over infinite intervals and integrands that become infinite. In both cases, the critical idea is to introduce an appropriate limit.

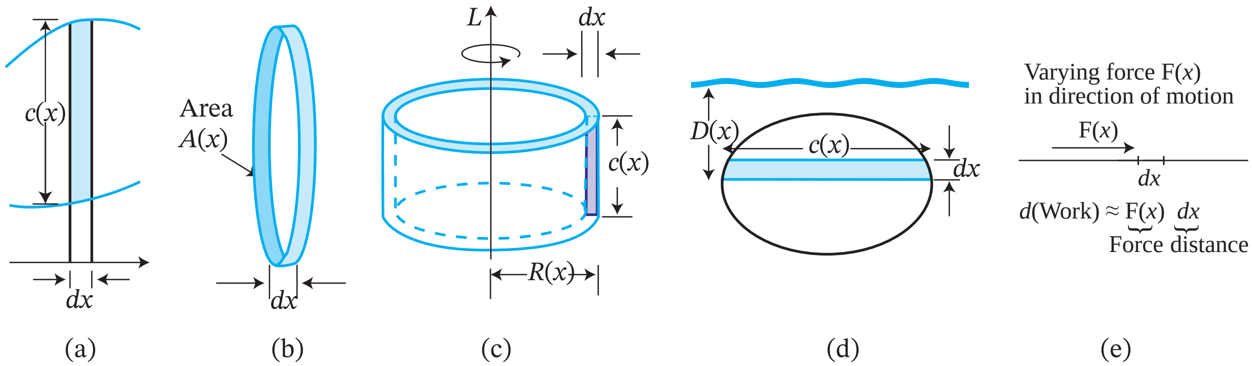


Figure 7.S.1

EXERCISES for Section 7.S

1. Let $P=(a,a^2)$ and $Q=(b,b^2)$ be on the parabola $y=x^2$.

(a) Show that the tangent to the parabola at the midpoint between P and Q, $R=\left(\frac{a+b}{2},\left(\frac{a+b}{2}\right)^2\right)$, is parallel to the chord PQ.

(b) Show that the area of the parabola below the chord is $\frac{(b-a)^3}{6}$.

(c) Show that the area of triangle PQR is $\frac{(b-a)^3}{4}$.

Archimedes proved that the area of the parabolic section under PQ is $\frac{4}{3}$ times the area of triangle $\triangle PQR$.

REFERENCE: S. Stein, *Archimedes: What did he do besides cry Eureka?*, MAA, Washington, DC, 1999 (pp. 51–60).

2. (a) The exponential function is increasing for all x. Use this to show that $e^x>1$ for $x>0$.

(b) Suppose $f(t)>g(t)$ for $t>a$. Explain why $\int_a^x f(t)\,dt > \int_a^x g(t)\,dt$ for $x>a$.

(c) Use (b) to show that $e^x>1+x$ for $x>0$.

(d) Use (b) and (c) to show that $e^x>1+x+\frac{x^2}{2}$ for $x>0$.

3. (a) Extend the argument in Exercise 2 to show that $e^x>\sum_{i=0}^{n+1}\frac{x^i}{i!}$.

(b) Use (a) to show that for any fixed positive integer n, $\lim_{x\to\infty}\frac{x^n}{e^x}=0$.

4. For a hydrogen atom in its ground state, the average distance between hydogen's sole electron and the nucleus is $\frac{4}{a_0^3}\int_0^\infty e^{-2r/a_0}r^3\,dr$. Here, a_0 is the **Bohr radius** (5.29177×10^{-11} m), which is the most probable distance between the electron and nucleus for the ground state hydrogen atom. Show that this improper integral is convergent, and has value $\frac{3}{2}a_0$.

REFERENCE: P. Atkins, J. de Paula, J. Keller, *Phys Chem*, 11e (2017), Oxford Univ. Press, ISBN: 978-0198769866.

5. If $\int_0^\infty f(x)\,dx$ is convergent, does it follow that

NOTE: Compare with Exercise 18 in Section 11.S.

(a) $\lim_{x\to\infty} f(x) = 0$? (b) $\lim_{x\to\infty} \int_x^{x+0.1} f(t)\,dt = 0$? (c) $\lim_{x\to\infty} \int_x^{2x} f(t)\,dt = 0$? (d) $\lim_{x\to\infty} \int_x^{\infty} f(t)\,dt = 0$?

6. Here is a proposed way to find the surface area of a sphere: Approximate the surface area of the sphere of radius a shown in Figure 7.S.2(a) as follows. To approximate the surface area between x and $x+dx$, let us try using the area of the narrow curved part of the cylinder used to approximate the volume between x and $x+dx$, shaded in Figure 7.S.2(b). The local approximation can be pictured, when unrolled and laid flat, as a rectangle of width dx and length $2\pi r$. The surface area of a sphere is $\int_{-a}^{a} 2\pi r\,dx = 4\pi \int_0^a \sqrt{a^2-x^2}\,dx$. But $\int_0^a \sqrt{a^2-x^2}\,dx = \pi a^2/4$, since it equals the area of a quadrant of a disk. Hence the area of the sphere is then $\pi^2 a^2$." This does not agree with the correct value, $4\pi a^2$, which was discovered by Archimedes in the third century B.C. What is wrong?

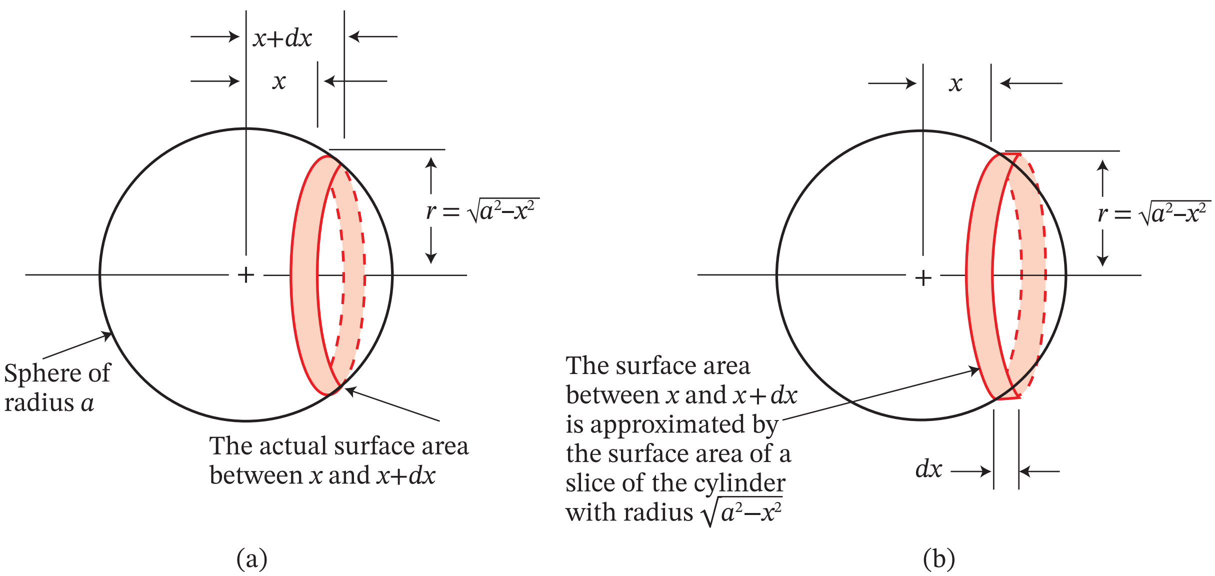


Figure 7.S.2

7. Determine if $\int_0^\infty \frac{x\,dx}{\sqrt{1+x^4}}$ converges or diverges.

8. The probability that ball bearing A survives at least until time t will be denoted as $F(t)$. For ball bearing B let $G(t)$ be the probability that it survives at least until time t.

(a) Show that the probability that A lasts at least as long as B is $-\int_0^\infty F(t)G'(t)\,dt$.

(b) Similarly, the probability that B lasts at least as long as A is $-\int_0^\infty G(t)F'(t)\,dt$. Assume that the probability that A and B last exactly the same time is 0. Why should $-\int_0^\infty F(t)G'(t)\,dt - \int_0^\infty G(t)F'(t)\,dt = 1$? Show that it does equal 1.

9. (a) Draw the curve $y = \frac{e^x}{x}$ for $x > 0$, showing any asymptotes or critical points.

(b) Find the number t such that the area below $y = \frac{e^x}{x}$ and above the interval $[t, t+1]$ is a minimum.

10. Assume that f is continuous on $(-\infty, \infty)$, that $f(x+1) = -f(x)$, $\lim_{x\to 0} \frac{f(x)}{x} = 1$, and $f(x)$ is positive for x in $(0,1)$.

(a) Is $\int_{-\infty}^{\infty} f(x)\, dx$ convergent?

(b) Is $\int_{-\infty}^{\infty} \frac{f(x)}{x}\, dx$ convergent?

(c) Show that $\int_{-\infty}^{\infty} \frac{f(x)}{x}\, dx = \int_{-\infty}^{\infty} \frac{f(x)}{x+1}\, dx$.

(d) Show that $\int_{-\infty}^{\infty} \frac{f(x)}{x(x+1)}\, dx = 0$.

(e) Show that $\int_{-\infty}^{\infty} \frac{\sin(\pi x)}{x(x+1)}\, dx = 0$.

In Exercises 11 to 13 a, b, c, m, and p are constants. Verify that the second expression is the derivative of the first one.

11. $\frac{x}{a} - \frac{1}{ap}\ln\left(a + be^{px}\right)$; $\frac{1}{a + be^{px}}$.

12. $\frac{1}{\sqrt{-c}}\arcsin\left(\frac{-2cx - b}{\sqrt{b^2 - 4ac}}\right)$; $\frac{1}{\sqrt{a + bx + cx^2}}$. Assume $b^2 - 4ac > 0$ and c is any negative number.

13. $\frac{1}{\sqrt{c}}\ln\left(\sqrt{a + bx + cx^2} + x\sqrt{c} + \frac{b}{2\sqrt{c}}\right)$, $\frac{1}{\sqrt{a + bx + cx^2}}$, for any positive number c.

Calculus is Everywhere # 9

Escape Velocity

In Example 2 in Section 7.7 we saw that the total work required to lift a 1-pound payload from the surface of the earth to the moon is 3,933 mile-pounds. Since the radius of the earth is about 4,000 miles, the work required to launch a payload on an endless journey is given by the improper integral

$$\int_{4,000}^{\infty} \left(\frac{4,000}{r}\right)^2 dr = 4,000 \text{ mile-pounds.}$$

Because the integral is convergent, only a finite amount of energy is needed to send a payload on an endless journey, as the Voyager spacecraft continue to demonstrate. It takes only a little more energy than is required to lift the payload to the moon.

That the work required for the endless journey is finite raises the question: What initial velocity will cause the payload to rise forever away from Earth?

We supply kinetic energy to the payload. The force of gravity slows the payload and reduces its kinetic energy. If the kinetic energy was ever zero, then the velocity of the payload would be zero and the payload would start to fall back to Earth.

As we will show, the kinetic energy of the payload is reduced by exactly the amount of work done on the payload by gravity. If v_{esc} is the minimal velocity needed for the payload to escape and not fall back, then

$$\frac{1}{2} m v_{\text{esc}}^2 = 4,000 \text{ mile-pounds,} \tag{C.9.1}$$

where m is the mass of the payload, which can be solved for v_{esc}, the **escape velocity**.

To solve (C.9.1) for v_{esc}, we calculate the mass of a payload that weighs 1 pound at the surface of the earth. The weight of 1 pound is the gravitational force of the earth pulling on it. Newton's second law of motion implies that

$$\text{Force} = \text{Mass} \times \text{Acceleration.} \tag{C.9.2}$$

This equation will be our main tool.

The acceleration of an object at Earth's surface is 32 feet per second per second, or 0.0061 miles per second per second. For the 1-pound payload (C.9.2) becomes

$$1\text{pound} = m(0.0061 \frac{\text{miles}}{\text{sec}^2}). \tag{C.9.3}$$

Combining (C.9.1) and (C.9.3) gives

$$\frac{1}{2} \frac{1}{0.0061} (v_{\text{esc}})^2 = 4,000$$

or

$$(v_{\text{esc}})^2 = (8,000)(0.0061) \approx 48.8 \frac{\text{miles}^2}{\text{sec}^2}.$$

Hence $v_{\text{esc}} \approx 7$ miles per second, which is about 25,000 miles per hour. To put this into perspective, the fastest any human has traveled is 24,790 mi/hr, 39,897 km/hr, when the Apollo 10 astronauts returned to Earth in 1969.

To justify the claim that the change in kinetic energy equals the work done by the force let $v(r)$ be the velocity of the payload when it is r miles from Earth's center. Let $F(r)$ be the force on the payload when it is r miles from Earth's center. Since the Earth's gravitational force is in the opposite direction from the motion, we will define $F(r)$ to be negative.

Let a and b be numbers, $4,000 \le a < b$. (See Figure C.9.1.) We wish to show that

$$\underbrace{\frac{1}{2}m(v(b))^2 - \frac{1}{2}m(v(a))^2}_{\text{change in kinetic energy}} = \underbrace{\int_a^b F(r)\,dr}_{\text{work done by gravity}}, \tag{C.9.4}$$

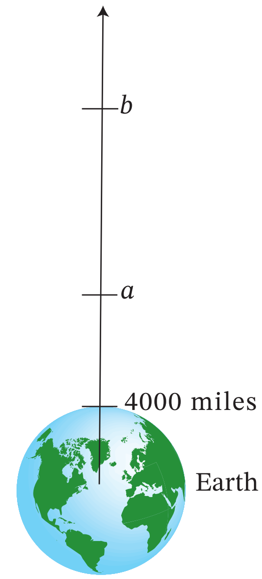


Figure C.9.1

where m is the payload mass. Both sides of (C.9.4) are negative.

Equation (C.9.4) resembles the Fundamental Theorem of Calculus. If we could show that $m(v(r))^2/2$ is an antiderivative of $F(r)$, then (C.9.4) would follow immediately. Let us find the derivative of $m(v(r))^2/2$ with respect to r and show that it equals $F(r)$:

$$\begin{aligned}
\frac{d}{dr}\left(\frac{1}{2}m(v(r))^2\right) &= mv(r)\frac{dv}{dr} \\
&= mv(r)\frac{dv/dt}{dr/dt} && (\text{ chain rule; } t \text{ is time }) \\
&= mv(r)\frac{d^2r/dt^2}{v(r)} && (\ v(r) = dr/dt\ ) \\
&= m\frac{d^2r}{dt^2} && (\text{mass} \times \text{acceleration}) \\
&= F(r) && (\text{ Newton's second law of motion }).
\end{aligned}$$

Hence (C.9.4) is valid and we have justified our calculation of escape velocity.

Incidentally, the escape velocity is $\sqrt{2}$ times the velocity required for a satellite to orbit the earth (and not fall into the atmosphere and burn up).

EXERCISES for CIE C.9

1. Repeat the derivation of the escape velocity using CGS units. That is, assume the radius of the earth is 6,371 kilometers and the acceleration of gravity is 9.80665 meters per second per second at Earth's surface.

2. Determine the escape velocity from the moon.

3. Determine the escape velocity from the sun.

Calculus is Everywhere # 10

An Improper Integral in Economics

What is the present value of one dollar t years in the future? What is the present value of a business in terms of its future profit? People in business frequently face the question, “How much money do I need today to have one dollar t years in the future?”

We determine the present value of a business that depends on its future rate of profit.

t can be any positive number, not necessarily an integer.

Assume that the annual interest rate r remains constant and that 1 dollar deposited today is worth e^{rt} dollars t years from now. This assumption corresponds to continuously compounded interest or to natural growth. Thus A dollars today will be worth Ae^{rt} dollars t years from now. What is the present value of the promise of 1 dollar t years from now? In other words, what amount A invested today will be worth 1 dollar t years from now? To find out, solve the equation $Ae^{rt} = 1$ for A. The solution is

$$A = e^{-rt}. \tag{C.10.1}$$

That is, the present value of \$1 t years from now is e^{-rt} dollars.

Assume that the profit flow t years from now is at the rate $f(t)$, which may vary as a continuous function of time. The profit in the small interval of time dt, from time t to time $t + dt$, would be approximately $f(t)dt$. The total future profit, $F(T)$, from now, when $t = 0$, to time T in the future is therefore

$$F(T) = \int_0^T f(t)dt. \tag{C.10.2}$$

The present value of the future profit is not given by (C.10.2). We need to approximate the present value of the profit earned in a short interval of time from t to $t + dt$. According to (C.10.1), it is approximately

$$e^{-rt} f(t)\, dt.$$

Hence the **present value of future profit** from $t = 0$ to $t = T$ is given by

$$\int_0^T e^{-rt} f(t)dt.$$

To see what influence the interest rate r has, denote by $P(r)$ the present value of all future revenue when the interest rate is r:

$$P(r) = \int_0^\infty e^{-rt} f(t)dt. \tag{C.10.3}$$

If the interest rate r increases then according to (C.10.3) the present value of a business decreases. Investors need to make assumptions about future interest rates when valuing their investments.

A proponent of a project may argue that the interest rate r will be low in the future. An opponent may predict that it will be high. Neither knows for certain what the interest rates will be. The prediction is important in a **cost-benefit analysis**.

Equation (C.10.3) assigns to a profit function f (which is a function of time t) a present-value function P, which is a function of r, the interest rate. In the theory of differential equations, P is called the **Laplace transform** of f. It can replace a differential equation by a simpler equation. Examples of the Laplace transform will be encountered in Exercises 51 to 59 in Section 8.3 and Exercises 65 to 67 in Section 8.S.

EXERCISES for CIE C.10

In Exercise 1 to 5 $f(t)$ is defined on $[0,\infty)$ and is continuous. Assume that for $r>0$, $\int_0^\infty e^{-rt}f(t)\,dt$ converges and that $e^{-rt}f(t)\to 0$ as $t\to\infty$. Let $P(r)=\int_0^\infty e^{-rt}f(t)\,dt$. Find $P(r)$, the Laplace transform of $f(t)$, in Exercises 1 to 4.

1. $f(t)=1$

2. $f(t)=t$

3. $f(t)=\sin(t)$

4. $f(t)=e^t$ (assume $r>1$)

---

5. Which is worth more today, \$100, eight years from now or \$80, five years from now?
(a) Assume $r=4\%$. (b) Assume $r=8\%$. (c) For which interest rate are the two equal?

Chapter 8

Computing Antiderivatives

In Chapter 7 we saw several uses for definite integrals in geometry and physics. Other applications can be found in many fields, including economics, engineering, biology, and statistics. Definite integrals are usually evaluated using the fundamental theorem of calculus or estimated numerically, as in Section 6.5.

To evaluate $\int_a^b f(x)\, dx$ by the fundamental theorem of calculus (FTC I) we must find an antiderivative $F(x)$ of the integrand $f(x)$. Then $\int_a^b f(x)\, dx$ is $F(b) - F(a)$. This chapter describes techniques for finding antiderivatives.

The problem of finding an antiderivative differs from that of finding a derivative in two ways. The first difference is that the antiderivatives of some elementary functions, such as e^{x^2}, are not elementary, while the derivative of an elementary functions is always an elementary function. Second, a slight change in the form of a function can cause a great change in the form of its antiderivative. For instance,

$$\int \frac{dx}{x^2+1} = \arctan(x) + C \qquad \text{while} \qquad \int \frac{x\, dx}{x^2+1} = \frac{1}{2}\ln(x^2+1) + C,$$

as may be checked by differentiating each alleged integral. This is in sharp contrast with differentiation, where a slight change in the form of an elementary function produces only a slight change in the form of its derivative.

There are at least three different ways to find an antiderivative: by hand, using techniques described in this chapter, by an integral table, such as this book's Table of Integrals (in Appendix A), or by computer, calculator, or other online reference source.

Section 8.1 illustrates a few shortcuts, describes how to use integral tables, and discusses the strengths and weaknesses of computer-based evaluation of integrals. Section 8.2 presents a method known as substitution; this is probably the most frequently used technique for finding an antiderivative and Section 8.3 describes integration by parts. Both techniques have applications other than finding antiderivatives. The integration of rational functions is discussed in Section 8.4 . Section 8.5 describes how to integrate some special integrands. The final section in the chapter, Section 8.6, offers an opportunity to practice all of the techniques that have been presented.

As you gain experience with these methods you will find that you develop an intuition about which integrands have elementary antiderivatives, and which method – or methods – to apply to find it. In this sense you appreciate that there is an art to evaluating integrals.

8.1 Shortcuts, Tables, and Technology

In this section we list the antiderivatives of some common functions and give some common shortcuts for evaluating some definite integrals. Then we describe how to use integral tables and the computation of antiderivatives by computers.

Some Common Integrands

Every derivative formula provides a corresponding formula for an antiderivative. For instance, since the derivative of $x^3/3$ is x^2, it follows that an antiderivative of x^2 is $x^3/3$ and the most general antiderivative of x^2 is $x^3/3+C$:

$$\int x^2\,dx = \frac{x^3}{3}+C.$$

Table 8.1.1 lists a few formulas. Each antiderivative can be checked by showing the derivative of the right-hand side is the integrand on the left-hand side.

f(x)	$F(x)=\int f(x)\,dx$	Remarks
x^a	$\frac{x^{a+1}}{a+1}+C$	for $a\neq -1$
$\frac{1}{x}$	$\ln\lvert x\rvert+C$	this is $\int x^a\,dx$ for $a=-1$
$\frac{f'(x)}{f(x)}$	$\ln\lvert f(x)\rvert+C$	if $f(x)>0$, the absolute value can be omitted
$\left(f(x)\right)^a f'(x)$	$\frac{\left(f(x)\right)^{a+1}}{a+1}+C$	for $a\neq -1$
e^{ax}	$\frac{e^{ax}}{a}+C$	
$\sin(ax)$	$\frac{-1}{a}\cos(ax)+C$	remember the negative sign
$\cos(ax)$	$\frac{1}{a}\sin(ax)+C$	
$\frac{1}{\sqrt{a^2-x^2}}$	$\arcsin\left(\frac{x}{a}\right)+C$	
$\frac{1}{a^2+x^2}$	$\frac{1}{a}\arctan\left(\frac{x}{a}\right)+C$	
$\frac{1}{\lvert x\rvert\sqrt{x^2-a^2}}$	$\frac{1}{a}\operatorname{arcsec}\left(\frac{x}{a}\right)+C$	

Table 8.1.1

EXAMPLE 1. Find $\int (2x^4-3x+2)\,dx$.

Antiderivative of x^a, a a nonnegative integer

SOLUTION

$$\begin{aligned}\int (2x^4-3x+2)\,dx &= \int 2x^4\,dx-\int 3x\,dx+\int 2\,dx && (\text{ sum property of integrals })\\ &= 2\int x^4\,dx-3\int x\,dx+2\int 1\,dx && (\text{ constant multiple property of integrals })\\ &= 2\frac{x^5}{5}-3\frac{x^2}{2}+2x+C.\end{aligned}$$

Note that while each term contributes a constant to the final answer, all three of these constants are combined into a single constant of integration. ♦

EXAMPLE 2. Find $\int \sqrt{x}\, dx$.

Antiderivative of x^a, a a rational number

SOLUTION

$$\int \sqrt{x}\, dx = \int x^{1/2}\, dx = \frac{x^{1/2+1}}{\frac{1}{2}+1} + C = \frac{2}{3}x^{3/2} + C.$$

♦

EXAMPLE 3. Find $\int \frac{1}{x^3}\, dx$.

Antiderivative of x^a, a a negative number

SOLUTION

$$\int \frac{1}{x^3}\, dx = \int x^{-3}\, dx = \frac{x^{-3+1}}{-3+1} + C = -\frac{1}{2}x^{-2} + C = -\frac{1}{2x^2} + C.$$

♦

EXAMPLE 4. Find $\int (3\cos(x) - 4\sin(2x) + \frac{1}{x^2})\, dx$.

SOLUTION

$$\begin{aligned}\int (3\cos(x) - 4\sin(2x) + \frac{1}{x^2})\, dx &= 3\int \cos(x)\, dx - 4\int \sin(2x)\, dx + \int \frac{1}{x^2} dx\\ &= 3\sin(x) + 2\cos(2x) - \frac{1}{x} + C.\end{aligned}$$

♦

EXAMPLE 5. Find $\int \frac{4x^3}{x^4+1}\, dx$.

Antiderivative of $f'(x)/f(x)$

SOLUTION Observe that the numerator is the derivative of the denominator. Hence

$$\int \frac{4x^3}{x^4+1}\, dx = \ln|x^4+1| + C$$

Since x^4+1 is always positive, the absolute value sign is not needed, and $\int 4x^3/(x^4+1)\, dx = \ln(x^4+1) + C$. ♦

EXAMPLE 6. Find $\int \frac{x}{1+x^2}\, dx$.

Multiplying the integrand by a constant

SOLUTION If the numerator were exactly $2x$, the numerator would be the derivative of the denominator and we would have an integral of the form $\int (f'(x)/f(x))\, dx$. An antiderivative would be $\ln|f(x)| = \ln(1+x^2)$.

The numerator can be multiplied by 2 if we simultaneously divide by 2:

$$\begin{aligned}\int \frac{x}{1+x^2}\, dx = \int \frac{1}{2}\frac{2x}{1+x^2}\, dx &= \frac{1}{2}\int \frac{2x}{1+x^2}\, dx\\ &= \frac{1}{2}\ln(1+x^2) + C.\end{aligned}$$

♦

Special Shortcuts

We present three shortcuts for evaluating some special but common definite integrals. They are easy to apply and can save a good deal of work.

SHORTCUT 8.1.1: IF f IS AN ODD FUNCTION, THEN $\displaystyle\int_{-a}^{a} f(x)\,dx = 0$.

Explanation. For an odd function $f(-x) = -f(x)$. In Figure 8.1.1, the shaded area to the left of the y-axis equals the shaded area to the right — but they have opposite signs. That is, as integrals, $\int_{-a}^{0} f(x)\,dx = -\int_0^a f(x)\,dx$. Therefore, the definite integral over any symmetric interval is 0.

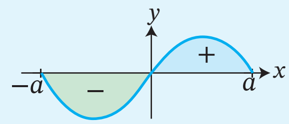


Figure 8.1.1

EXAMPLE 7. Find $\displaystyle\int_{-2}^{2} x^3\sqrt{4-x^2}\,dx$.

SOLUTION The function $f(x) = x^3\sqrt{4-x^2}$ is odd, as can be checked. By Shortcut 8.1.1, $\int_{-2}^{2} x^3\sqrt{4-x^2} = 0$. ◆

SHORTCUT 8.1.2: $\displaystyle\int_0^a \sqrt{a^2-x^2}\,dx = \frac{1}{4}\pi a^2$

Explanation. The graph of $y = \sqrt{a^2-x^2}$ is part of a circle of radius a. The region bounded by the graph of $y = \sqrt{a^2-x^2}$ above the interval $[0, a]$ is a quarter circle with radius a. (See Figure 8.1.2.)
Therefore, the definite integral $\int_0^a \sqrt{a^2-x^2}\,dx$ is a quarter of the area of that circle.

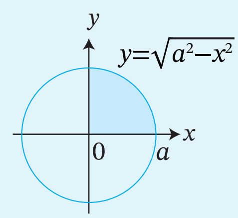


Figure 8.1.2

EXAMPLE 8. Find $\displaystyle\int_0^1 \sqrt{1-x^2}\,dx$

SOLUTION Use Shortcut 8.1.2, with $a = 1$, to get $\int_0^1 \sqrt{1-x^2}\,dx = \pi/4$. ◆

SHORTCUT 8.1.3: IF f IS AN EVEN FUNCTION, $\displaystyle\int_{-a}^{a} f(x)\,dx = 2\int_0^a f(x)\,dx$.

Explanation. For an even function $f(-x) = f(x)$. In Figure 8.1.3, the shaded area to the left of the y-axis equals the shaded area to the right — and they have same sign. That is, as integrals, $\int_{-a}^{0} f(x)\,dx = \int_0^a f(x)\,dx$.
Therefore, the definite integral over any symmetric interval is twice the integral over either half of the interval.

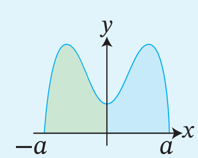


Figure 8.1.3

EXAMPLE 9. Find $\displaystyle\int_{-1}^{1} \sqrt{1-x^2}\, dx$.

SOLUTION Since $\sqrt{1-x^2}$ is an even function, by Shortcut 8.1.3 $\int_{-1}^{1} \sqrt{1-x^2}\, dx = 2\int_0^1 \sqrt{1-x^2}\, dx$. So, by Shortcut 8.1.2, with $a=1$,

$$\int_{-1}^{1} \sqrt{1-x^2}\, dx = 2\int_{0}^{1} \sqrt{1-x^2}\, dx = 2\cdot\frac{\pi}{4} = \frac{\pi}{2}.$$

♦

Using an Integral Table

An integral table lists antiderivatives of functions. While most entries are indefinite integrals, there are also many definite integrals. Several books list hundreds, even thousands, of integrals.

Observation 8.1.1: *Two Observations about Integral Tables*

1. Many integral tables use log to denote ln; it is understood that e is the base.
2. Most integral tables omit the constant of integration ($+C$). The reader must know when and where, and how, to insert an appropriate constant of integration.

Browse through an integral table to see how the formulas are grouped. First might come the forms used most frequently. Then come forms containing $ax+b$, forms containing $a^2 \pm x^2$, forms containing ax^2+bx+c, and so on, running through different algebraic forms. There are separate sections with trigonometric, logarithmic, and exponential integrands. The Table of Integrals provided with this textbook (in Appendix A) is similarly grouped.

An online search for "integral table" yields more sources; some will direct you to printed books and others will be to online lists — some of which have sophisticated search features.

EXAMPLE 10. Use the Table of Integrals to evaluate the indefinite integral $\displaystyle\int \frac{dx}{x\sqrt{3x+2}}$. (Assume $x > -2/3$.)

SOLUTION From Formula 38 we know

$$\int \frac{dx}{x\sqrt{ax+b}} = \frac{1}{\sqrt{b}} \ln\left|\frac{\sqrt{ax+b}-\sqrt{b}}{\sqrt{ax+b}+\sqrt{b}}\right| \qquad (b>0).$$

We use this after replacing a by 3 and b by 2. Thus

$$\int \frac{dx}{x\sqrt{3x+2}} = \frac{1}{\sqrt{2}} \ln\left|\frac{\sqrt{3x+2}-\sqrt{2}}{\sqrt{3x+2}+\sqrt{2}}\right| + C.$$

♦

EXAMPLE 11. Use the Table of Integrals to integrate $\displaystyle\int \frac{dx}{x\sqrt{3x-2}}$. (Assume $x > 2/3$.)

SOLUTION From Formula 39 we know

$$\int \frac{dx}{x\sqrt{ax+b}} = \frac{2}{\sqrt{-b}} \arctan\left(\sqrt{\frac{ax+b}{-b}}\right) \qquad (b<0).$$

Using this, with $a=3$ and $b=-2$, to get

$$\int \frac{dx}{x\sqrt{3x-2}} = \frac{2}{\sqrt{2}} \arctan\left(\sqrt{\frac{3x-2}{2}}\right) + C.$$

♦

Though the integrands in Examples 10 and 11 are similar, their antiderivatives are not.

When using an integral table, be cautious and keep a cool head, matching patterns carefully, including any conditions on the variables and their coefficients. Some formulas are expressed in terms of an integral of a different integrand, so you will have to search through the table more than once. (Exercises 35 and 36 illustrate this.)

Computers, Calculators, and Other Automated Integrators

Using an integral table is an exercise in pattern matching, where you hunt for the formula that fits a particular integral. Computers are good at pattern matching, so it is not surprising that for many years computers have been used to find antiderivatives. MACSYMA is one of the earliest computer-based programs that perform the basic operations of calculus: limits, derivatives, integrals. Today, the most widely used computer algebra systems are Maple, Mathematica, and Sage.

The technology is now included in some handheld calculators. Calculus users do not need to rely as much on formal integration techniques or tables of integrals. What is essential is to understand what an integral is, what it can represent, and how to use it.

Summary

Integration, or antidifferentiation, is hard – at least in comparison with differentiation. The differentiation rules for sums, differences, products, quotients, and compositions of functions allow one to find the derivative of almost any function, this is not the case for integration.

Finding antiderivatives by hand typically requires reorganizing the integrand so it can be recognized either as one of the common integrals or as an entry in a table of integrals. While symbolic computation tools on computers, calculators, and even smartphones can find many antiderivatives, the algorithms they use sometimes produce results that do not resemble what we would find. It is often simpler and quicker to use one of the techniques that will be described in the remaining sections of this chapter.

For definite integrals, there are shortcuts when the integrand is an even or odd function, or when the graph of the integrand is a part of a circle or other region whose area is known.

EXERCISES for Section 8.1

In Exercises 1 to 14 find the integrals. Use Table 8.1.1 at the beginning of the section.

1. $\int 5x^3\,dx$ **2.** $\int (8+11x)\,dx$ **3.** $\int x^{1/3}\,dx$ **4.** $\int \sqrt[3]{x^2}\,dx$ **5.** $\int \frac{6\,dx}{x^2}$

6. $\int \frac{dx}{x^3}$ **7.** $\int 5e^{-2x}\,dx$ **8.** $\int \frac{5\,dx}{1+x^2}$ **9.** $\int \frac{6\,dx}{|x|\sqrt{x^2-1}}$ **10.** $\int \frac{5\,dx}{\sqrt{1-x^2}}$

11. $\int \frac{4x^3\,dx}{1+x^4}$ **12.** $\int \frac{e^x\,dx}{1+e^x}$ **13.** $\int \frac{\sin(x)\,dx}{1+\cos(x)}$ **14.** $\int \frac{dx}{1+3x}$

In Exercises 15 to 20, rewrite the integrand by algebra so that it becomes easy to find the antiderivative using one or more entries from Table 8.1.1.

15. $\int \frac{1+2x}{x^2}\,dx$ **16.** $\int \frac{1+2x}{1+x^2}\,dx$ **17.** $\int (x^2+3)^2\,dx$

18. $\int (1+e^x)^2\,dx$ **19.** $\int (1+3x)x^2\,dx$ **20.** $\int \frac{1+\sqrt{x}}{x}\,dx$

21. This problem presents a shortcut for evaluating $\int_0^{\pi/2} \sin^2(\theta)\,d\theta$.

(a) Why is $\int_0^{\pi/2} \sin^2(\theta)\,d\theta + \int_0^{\pi/2} \cos^2(\theta)\,d\theta = \frac{\pi}{2}$?

(b) Why would you expect $\int_0^{\pi/2} \cos^2(\theta)\,d\theta$ to equal $\int_0^{\pi/2} \sin^2(\theta)\,d\theta$?

(c) Conclude that $\int_0^{\pi/2} \sin^2(\theta)\,d\theta = \frac{\pi}{4}$.

The integrals in Exercises 22 to 28 can be evaluated using one (or more) of the shortcuts.

22. $\int_{-1}^{1} x^5\sqrt{1+x^2}\,dx$

23. $\int_{-\pi/2}^{\pi/2} \sin(3x)\cos(5x)\,dx$

24. $\int_{-1}^{1} x^5\sqrt[4]{1-x^2}\,dx$

25. $\int_{-\pi}^{\pi} \sin^3(x)\,dx$

26. $\int_{0}^{5} \sqrt{25-x^2}\,dx$

27. $\int_{-3}^{3} \sqrt{9-x^2}\,dx$

28. $\int_{-3}^{3} \left(x^3\sqrt{9-x^2}+10\sqrt{9-x^2}\right)dx$

In Exercises 29 to 34 find the antiderivative using a table of integrals, such as this book's Table of Integrals (in Appendix A).

29. (a) $\int \frac{dx}{(3x+2)^2}$ and (b) $\int \frac{dx}{x(3x+2)}$.

30. (a) $\int \frac{dx}{x\sqrt{3x+4}}$ and (b) $\int \frac{dx}{x^2\sqrt{3x+4}}$.

31. (a) $\int \frac{dx}{x\sqrt{3x-4}}$ and (b) $\int \frac{dx}{x^2\sqrt{3x-4}}$.

32. (a) $\int \frac{dx}{4x^2+9}$ and (b) $\int \frac{dx}{4x^2-9}$.

33. (a) $\int \frac{dx}{x^2+8x+7}$ and (b) $\int \frac{dx}{x^2+2x+5}$.

34. (a) $\int \frac{dx}{\sqrt{11-x^2}}$ and (b) $\int \frac{dx}{\sqrt{11+x^2}}$.

---

35. Using the Table of Integrals in Appendix A, find $\int \frac{x\,dx}{\sqrt{2x^2+x+5}}$.

36. Using the Table of Integrals in Appendix A, find (a) $\int \frac{dx}{\sqrt{3x^2+x+2}}$ and (b) $\int \frac{dx}{\sqrt{-3x^2+x+2}}$.

8.2 Substitution Method

This section describes the substitution method, which aims to change an integrand into one that can be integrated easily. Several examples will illustrate the technique, which is really just the chain rule in disguise. Sometimes we can use a substitution to transform an integral not listed in an integral table to one that is listed. After the examples, we show why the substitution method is valid.

EXAMPLE 1. Find $\int \sin(x^2)\,2x\,dx$.

SOLUTION Because $2x$ is the derivative of x^2, this suggests the substitution $u = x^2$. If we make the substitution $u = x^2$ then $du = D(x^2)\,dx = 2x\,dx$ and

$$\int \sin(x^2)\,2x\,dx = \int \sin(u)\,du.$$

We know $\int \sin(u)\,du = -\cos(u) + C$. Replacing u by x^2 in $-\cos(u)$ yields $-\cos(x^2)$. Thus

$$\int \sin(x^2)2x\,dx = -\cos(x^2) + C.$$

◆

This result can be checked by differentiating the right-hand side, which involves using the chain rule, and verifying that the result is the original integrand:

$$\frac{d}{dx}\left(-\cos(x^2)+C\right) = -(-\sin(x^2))\frac{d(x^2)}{dx} = \sin(x^2)2x.$$

In contrast, $\int \sin(x^2)\ dx$ is not elementary. The presence of $2x$, the derivative of x^2, made it possible to find $\int \sin(x^2)\ 2x\ dx$.

Description of the Substitution Method

In Example 1, the integrand $f(x)$ could be written in the form

$$f(x) = \underbrace{g(h(x))}_{\text{function of } h(x)} \cdot \underbrace{h'(x)}_{\text{derivative of } h(x),}$$

for some function $h(x)$. That is, the expression $f(x)\ dx$ could be written as

$$f(x)\ dx = \underbrace{g(h(x))}_{\text{function of } h(x)} \cdot \underbrace{h'(x)}_{\text{derivative of } h(x),} \quad dx$$

So the substitution of u for $h(x)$ and du for $h'(x)\ dx$ transforms $\int f(x)\ dx$ to an integral involving u instead of x, $\int g(u)\ du$.

If you can find an antiderivative $G(u)$ of $g(u)$, replace u by $h(x)$. The resulting function, $G(h(x))$, is an antiderivative of $f(x)$. (This will be justified at the end of the section.)

The process of using substitution to evaluate an indefinite integral can be summarized as

$$\int f(x)\ dx = \int g(h(x))\ h'(x)\ dx = \int g(u)\ du = G(u) + C = G(h(x)) + C.$$

EXAMPLE 2. Find $\displaystyle\int (1+x^3)^5\ x^2\ dx$.

SOLUTION The derivative of $1+x^3$ is $3x^2$, which differs from the x^2 in the integrand only by the constant factor 3. So let $u = 1+x^3$. Hence

$$du = 3x^2\ dx \qquad \text{and} \qquad \frac{du}{3} = x^2\ dx.$$

Then

$$\int (1+x^3)^5 x^2\ dx = \int u^5 \frac{du}{3} = \frac{1}{3}\int u^5\ du = \frac{1}{3}\frac{u^6}{6} + C = \frac{(1+x^3)^6}{18} + C.$$

◆

If the factor x^2 were not present in the integrand, we could still compute $\int (1+x^3)^5\ dx$ by multiplying out $(1+x^3)^5$, giving a polynomial of degree 15.

As Example 2 shows, the exact derivative of $h(x)$ as a factor is not needed. A constant times it will do.

Similarly, $\int x^2/\sqrt{1+x^3}\ dx$ can be found by substitution (use $u = 1+x^3$), but $\int dx/\sqrt{1+x^3}$ is not elementary. The presence of x^2 makes a great difference.

Substitution in a Definite Integral

The substitution technique, or change of variables, extends to definite integrals, $\int_a^b f(x)\ dx$, with one important proviso. When making the substitution from x to u, be sure to replace the interval $[a,b]$ by the interval whose endpoints are $u(a)$ and $u(b)$. An example will illustrate the change in the limits of integration. The technique is justified in Theorem 8.2.2.

EXAMPLE 3. Evaluate $\int_1^2 3(1+x^3)^5\, x^2\, dx$.

SOLUTION Let $u = 1+x^3$. Then $du = 3x^2\, dx$. As x goes from 1 to 2, $u = 1+x^3$ goes from $1+1^3 = 2$ to $1+2^3 = 9$. Thus

$$\int_1^2 3(1+x^3)^5 x^2\, dx = \int_2^9 u^5\, du = \left.\frac{u^6}{6}\right|_2^9 = \frac{9^6-2^6}{6}.$$

After the substitution in the integrand and the limits of integration there is no need to bring back x. ◆

The remaining examples present integrals needed in Section 8.4. They also show how some formulas in integral tables are obtained.

EXAMPLE 4. Integral tables include formulas for (a) $\int \frac{dx}{ax+b}$ and (b) $\int \frac{dx}{(ax+b)^n}$, $n \neq 1$.Obtain them by using the substitution $u = ax+b$.

SOLUTION

(a) Let $u = ax+b$. Hence $du = a\, dx$ and therefore $dx = du/a$. Thus, by Formula 29 in the Table of Integrals (in Appendix A),

$$\int \frac{dx}{ax+b} = \int \frac{du/a}{u} = \frac{1}{a}\int \frac{du}{u}$$
$$= \frac{1}{a}\ln|u| + C = \frac{1}{a}\ln|ax+b| + C.$$

(b) The same substitution $u = ax+b$ gives

$$\int \frac{dx}{(ax+b)^n} = \int \frac{du/a}{u^n} = \frac{1}{a}\int u^{-n}\, du = \frac{1}{a}\frac{u^{-n+1}}{(-n+1)} + C$$
$$= \frac{(ax+b)^{-n+1}}{a(-n+1)} + C = \frac{-1}{a(n-1)(ax+b)^{n-1}} + C.$$

◆

In the next example we use u instead of x, in preparation for Example 6.

EXAMPLE 5. Find $\int \frac{du}{4u^2+9}$.

SOLUTION $\int du/(4u^2+9)$ resembles $\int du/(u^2+1)$. This suggests rewriting $4u^2$ as $9t^2$, so we could then factor the 9 out of $9t^2+9$, getting $9(t^2+1)$. Here are the details.

Introduce t so $4u^2 = 9t^2$. To do this let $2u = 3t$, so $u = 3t/2$. Then $du = \frac{3}{2}dt$. Also, $t = 2u/3$. The substitution gives

$$\int \frac{du}{4u^2+9} = \int \frac{\frac{3}{2}\, dt}{9t^2+9} = \frac{3}{2}\cdot\frac{1}{9}\int \frac{dt}{t^2+1}$$
$$= \frac{1}{6}\arctan(t) + C = \frac{1}{6}\arctan\left(\frac{2u}{3}\right) + C.$$

◆

The next example uses a substitution together with completing the square. To complete the square in the quadratic expression x^2+bx+c add and subtract $(b/2)^2$ to get the simpler form v^2+k where k is a constant:

$$x^2+bx+c = x^2+bx+\left(\frac{b}{2}\right)^2+c-\left(\frac{b}{2}\right)^2 = \left(x+\frac{b}{2}\right)^2+c-\frac{b^2}{4}.$$

We square half the coefficient of b: $(b/2)^2$. To complete the square in ax^2+bx+c, where a is not 1, factor a out first:

$$ax^2+bx+c = a\left(x^2+\frac{b}{a}x+\frac{c}{a}\right)$$

and then complete the square in $x^2+(b/a)x+c/a$.

EXAMPLE 6. Find $\int \frac{dx}{4x^2+8x+13}$.

SOLUTION Complete the square in the denominator:

We subtract $4(1^2)$, not 1^2.

$$\begin{aligned} 4x^2+8x+13 &= 4\left(x^2+2x+\boxed{}\right)+13-4\left(\boxed{}\right) \\ &= 4(x^2+2x+1^2)+13-4(1^2) \\ &= 4(x+1)^2+9. \end{aligned}$$

Thus the integral is

$$\int \frac{dx}{4(x+1)^2+9}.$$

Let $u=x+1$, hence $du=dx$ and we have

$$\int \frac{dx}{4(x+1)^2+9} = \int \frac{du}{4u^2+9}.$$

We found in Example 5 that

$$\int \frac{du}{4u^2+9} = \frac{1}{6}\arctan\left(\frac{2u}{3}\right)+C.$$

Putting both ideas together

$$\begin{aligned} \int \frac{dx}{4x^2+8x+13} &= \int \frac{dx}{4(x+1)^2+9} = \int \frac{du}{4u^2+9} \\ &= \frac{1}{6}\arctan\left(\frac{2u}{3}\right)+C = \frac{1}{6}\arctan\left(\frac{2(x+1)}{3}\right)+C. \end{aligned}$$

REMINDER: As always, it is a good idea to check the final antiderivative by differentiating.

◆

The integral

$$\int \frac{2ax+b}{ax^2+bx+c}\,dx \tag{8.2.1}$$

has the form $\int f'(x)/f(x)\,dx$, so it is $\ln|az^2+b+c|+C$. We use this idea in the next example.

EXAMPLE 7. Find $\int \frac{x}{4x^2+8x+13}\,dx$.

SOLUTION If $8x+8$ were in the numerator, we would have an integral of the form $\int f'(x)/f(x)\,dx$, for $8x+8$ is the derivative of the denominator. We use algebra to get $8x+8$ into the numerator. Write $x=(8x+8)/8-8/8=\frac{1}{8}(8x+8)-1$. Then we have

Example 6 in Section 8.2 is used here.

$$\begin{aligned}
\int \frac{x\,dx}{4x^2+8x+13} &= \int \frac{\frac{1}{8}(8x+8)-1}{4x^2+8x+13}\,dx \\
&= \frac{1}{8}\int \frac{8x+8}{4x^2+8x+13}\,dx - \int \frac{dx}{4x^2+8x+13} \\
&= \frac{1}{8}\ln\left|4x^2+8x+13\right| - \frac{1}{6}\arctan\left(\frac{2(x+1)}{3}\right) + C.
\end{aligned}$$

♦

The techniques of completing the square, substitution, and rewriting x in the numerator, illustrated in Examples 6 and 7, show how to integrate any integrand of the form $1/(ax^2+bx+c)$ or $x/(ax^2+b+c)$.

Why Substitution Works

Theorem 8.2.1: Substitution in an Indefinite Integral

Assume that f and g are continuous functions and $u = h(x)$ is differentiable. Suppose that $f(x)$ can be written as $g(u)\frac{du}{dx}$ and that G is an antiderivative of g. Then $G(h(x))$ is an antiderivative of $f(x)$.

Proof of Theorem 8.2.1

Differentiate $G(h(x))$ and check that the result is $f(x)$.

$$\begin{aligned}
\frac{d}{dx}G(h(x)) &= \frac{dG}{du}\frac{du}{dx} && (\text{ chain rule }) \\
&= g(u)\frac{du}{dx} && (\text{ definition of } G\text{ }) \\
&= f(x) && (\text{ assumption }).
\end{aligned}$$

●

Theorem 8.2.2: Substitution in a Definite Integral

The assumptions in Theorem 8.2.1, together with the assumption that $h(x)$ is monotonic, imply that

$$\int_a^b f(x)\,dx = \int_{h(a)}^{h(b)} g(u)\,du.$$

Proof of Theorem 8.2.2

Let $F(x) = G(h(x))$, where G is defined in the previous proof.

$$\begin{aligned}
\int_a^b f(x)\,dx &= F(b) - F(a) && (\text{ FTC I }) \\
&= G(h(b)) - G(h(a)) && (\text{ definition of } F\text{ }) \\
&= \int_{h(a)}^{h(b)} g(u)\,du && (\text{ FTC I, again })
\end{aligned}$$

To understand the changes to the limits of integration: as x goes from a to b, $h(x)$ goes from $h(a)$ to $h(b)$. ●

Summary

This section introduced the most commonly used integration technique, substitution. If $f(x)\,dx$ can be written as $g(h(x))\,d(h(x))$ for a function $h(x)$ then $\int f(x)dx = \int g(u)\,du$ and $\int_a^b f(x)\,dx = \int_{h(a)}^{h(b)} g(u)\,du$.

We hope that finding $\int g(u)\,du$ is easier than finding $\int f(x)\,dx$. If it is not, try another substitution or a method presented in the rest of the chapter. There is no routine method for antidifferentiation of elementary functions. Practice pays off in spotting which technique is most promising and also being able to transform an integral into one listed in a table.

EXERCISES for Section 8.2

In Exercises 1 to 14 use the given substitution to find an antiderivative or the definite integral.

1. $\int (1+3x)^5 3\,dx$, $u = 1+3x$

2. $\int e^{\sin(\theta)}\cos(\theta)\,d\theta$, $u = \sin(\theta)$

3. $\int \sin(2x)\,dx$, $u = 2x$

4. $\int_{\sqrt{8}}^{\sqrt{15}} x\sqrt{1+x^2}\,dx$, $u = 1+x^2$

5. $\int_0^1 \frac{x}{\sqrt{1+x^2}}\,dx$, $u = 1+x^2$

6. $\int \frac{e^{2x}}{(1+e^{2x})^2}\,dx$, $u = 1+e^{2x}$

7. $\int_{-1}^{2} e^{3x}\,dx$, $u = 3x$

8. $\int_{\pi/6}^{\pi/4} \tan(\theta)\sec^2(\theta)\,d\theta$, $u = \tan(\theta)$

9. $\int \frac{1}{\sqrt{1-9x^2}}\,dx$, $u = 3x$

10. $\int \frac{t\,dt}{\sqrt{2-5t^2}}$, $u = 2-5t^2$

11. $\int_2^3 \frac{e^{1/x}}{x^2}\,dx$, $u = \frac{1}{x}$

12. $\int_{\pi^2/16}^{\pi^2/4} \frac{\sin(\sqrt{x})}{\sqrt{x}}\,dx$, $u = \sqrt{x}$

13. $\int \frac{(\ln(x))^4}{x}\,dx$, $u = \ln(x)$

14. $\int \frac{\sin(\ln(x))}{x}\,dx$, $u = \ln(x)$

An antiderivative can be verified by checking that its derivative is the integrand. That is, if $\int f(x)\,dx = F(x)$, then $F'(x) = f(x)$. In Exercises 15 to 21 use differentiation to verify the given antiderivative (from one of the examples in this section).

15. Verify the result from Example 1 that $\int \sin(x^2)\,2x\,dx = -\cos(x^2) + C$.

16. Verify the result from Example 2 that $\int (1+x^3)^5\,x^2\,dx = \frac{1}{18}(1+x^3)^6 + C$.

17. Verify the result from Example 4(a) that $\int \frac{dx}{ax+b} = \frac{1}{a}\ln|ax+b| + C$.

18. Verify the result from Example 4(b) that $\int \frac{dx}{(ax+b)^n} = \frac{1}{a(-n+1)(ax+b)^{n-1}} + C$.

19. Verify the result from Example 5 that $\int \frac{dx}{4x^2+9} = \frac{1}{6}\arctan\left(\frac{2x}{3}\right) + C$.

20. Verify the result from Example 6 that $\int \frac{dx}{4x^2+8x+13} = \frac{1}{6}\arctan\left(\frac{2(x+1)}{3}\right) + C$.

21. Verify the result from Example 7 that $\int \frac{x\,dx}{4x^2+8x+13} = \frac{1}{8}\ln\left|4x^2+8x+13\right| - \frac{1}{6}\arctan\left(\frac{2(x+1)}{3}\right) + C$.

In Exercises 22 to 47 use a substitution to find the antiderivative.

22. $\int (1-x^2)^5 x\,dx$

23. $\int \frac{x\,dx}{(x^2+1)^3}$

24. $\int x\sqrt[3]{1+x^2}\,dx$

25. $\int \frac{\sin(\theta)}{\cos^2(\theta)}\,d\theta$

26. $\int \frac{e^{\sqrt{t}}}{\sqrt{t}}\,dt$

27. $\int e^x \sin(e^x)\,dx$

28. $\int \sin(3\theta)\,d\theta$

29. $\int \frac{dx}{\sqrt{2x+5}}$

30. $\int (x-3)^{5/2}\,dx$

31. $\int \frac{dx}{(4x+3)^3}$

32. $\int \frac{2x+3}{x^2+3x+2}\,dx$

33. $\int \frac{2x+3}{(x^2+3x+5)^4}\,dx$

34. $\int \frac{x^3}{\sqrt{1-x^8}}\,dx$

35. $\int \frac{dx}{\sqrt{x}(1+\sqrt{x})^3}$

36. $\int x^4\sin(x^5)\,dx$

37. $\int \frac{\cos(\ln(x))\,dx}{x}$

38. $\int \frac{x}{1+x^4}\,dx$

39. $\int \frac{x^3}{1+x^4}\,dx$

40. $\int \frac{x\,dx}{(1+x)^3}$

41. $\int \frac{x^2\,dx}{(1+x)^3}$

42. $\int \frac{\ln(3x)\,dx}{x}$

43. $\int \frac{\ln(x^2)\,dx}{x}$

44. $\int \frac{(\arcsin(x))^2}{\sqrt{1-x^2}}\,dx$

45. $\int \frac{dx}{\arctan(2x)(1+4x^2)}$

46. $\int \frac{dx}{9x^2+1}$

47. $\int \frac{dx}{9x^2+25}$

In Exercises 48 and 49 complete the square.

48. (a) $x^2+6x+10$ and (b) $4x^2+6x+11$.

49. (a) $x^2+\frac{5}{3}x+4$ and (b) $3x^2+5x+12$.

In Exercises 50 to 53 evaluate the indefinite integral by completing the square in the denominator. Note the there are only two denominators, and they appear very similar.

50. $\int \frac{dx}{x^2+2x+5}$

51. $\int \frac{dx}{2x^2+2x+5}$

52. $\int \frac{x}{x^2+2x+5}\,dx$

53. $\int \frac{x}{2x^2+2x+5}\,dx$

In Exercises 54 to 59 find the area of the region under the graph of the given function and above the interval.

54. $f(x)=x^2e^{x^3}$, $[1,2]$

55. $f(x)=\sin^3(\theta)\cos(\theta)$, $\left[0,\frac{\pi}{2}\right]$

56. $f(x)=\tan^5(\theta)\sec^2(\theta)$, $\left[0,\frac{\pi}{3}\right]$

57. $f(x)=\frac{x^2-x}{(3x+1)^2}$, $[1,2]$

58. $f(x)=\frac{(\ln(x))^3}{x}$, $[1,e]$

59. $f(x)=\frac{x^2+3}{(x+1)^4}$, $[0,1]$

In Exercises 60 to 63 use a substitution to evaluate the integral. (Assume $a\neq 0$.)

60. $\int \frac{x^2}{ax+b}\,dx$

61. $\int \frac{x}{(ax+b)^2}\,dx$

62. $\int \frac{x^2}{(ax+b)^2}\,dx$

63. $\int \frac{x}{ax+b}\,dx$

---

64. Use a substitution to show that if f is an odd function then $\int\limits_{-a}^{a} f(x)\,dx=0$.

65. Use a substitution to show that if f is an even function, then $\int\limits_{-a}^{a} f(x)\,dx = 2\int\limits_{0}^{a} f(x)\,dx$.

66. (a) Graph $y=\frac{\ln(x)}{x}$. (b) Find the area under the curve in (a) and above the interval $[e,e^2]$.

67. **SAM:** Jane, what did you find for the antiderivative of $\int 2\cos(\theta)\sin(\theta)\ d\theta$?

JANE: I found $\int 2\cos(\theta)\sin(\theta)\ d\theta = \sin^2(\theta) + C$.

SAM: That's too bad.

JANE: Why?

SAM: Because I found $\int 2\cos(\theta)\sin(\theta)\ d\theta = -\cos^2(\theta) + C$.

JANE: I'm pretty sure of my answer. I used the substitution $u = \sin(\theta)$.

SAM: Well, that's the problem. I used $u = \cos(\theta)$.

Who is right? Explain.

68. **JANE:** $\int_0^{\pi} \cos^2(\theta)\ d\theta$ is obviously positive.

SAM: No, it's zero. Just make the substitution $u = \sin(\theta)$; hence $du = \cos(\theta)\ d\theta$. Then I get

$$\int_0^{\pi} \cos^2(\theta)\ d\theta = \int_0^{\pi} \cos(\theta)\cos(\theta)\ d\theta = \int_0^0 \sqrt{1-u^2}\ du = 0.$$

Simple.

(a) Who is right? What is the mistake?

(b) Use the identity $\cos^2(\theta) = \frac{1}{2}(1+\cos(2\theta))$ to evaluate the integral using the FTC.

69. **JANE:** $\int_{-2}^{1} 2x^2\ dx$ is obviously positive.

SAM: Why are you so sure of this?

JANE: After all, the integrand is never negative and $-2 < 1$. It equals the area under $y = 2x^2$ and above $[-2, 1]$'.

SAM: You're wrong again. It's negative. Here are my computations. Let $u = x^2$; hence $du = 2x\ dx$. Then

$$\int_{-2}^{1} 2x^2\ dx = \int_{-2}^{1} x \cdot 2x\ dx = \int_4^1 \sqrt{u}\ du = -\int_1^4 \sqrt{u}\ du,$$

which is obviously negative.

Who is right? Explain.

8.3 Integration by Parts

In the last section we saw that integration by substitution is based on the chain rule. In this section we learn that the technique known as integration by parts is based on the product rule for derivatives.

The Basis for Integration by Parts

If u and v are differentiable functions then, by the product rule,

$$(uv)' = u'\ v + u\ v'.$$

This tells us that uv is an antiderivative of $u'\ v + u\ v'$:

$$uv = \int (u'\ v + u\ v')\ dx,$$

Then, since the antiderivative of a sum is the sum of antiderivatives,

$$uv = \int u'\, v\, dx + \int u\, v'\, dx,$$

which can be rearranged as

$$\int u\, v'\, dx = u\, v - \int u'\, v\, dx \tag{8.3.1}$$

This equation tells us that if we can integrate $u'\, v$, then we can integrate $u\, v'$, and vice versa. This fact could be useful in cases where $\int u'\, v\, dx$ is easier to find than $\int u\, v'\, dx$. The technique based on (8.3.1) is called **integration by parts**.

Inroducing the differentials $du = u'\, dx$ and $dv = v'\, dx$ in (8.3.1), leads us to the most common formula for integration by parts.

Theorem 8.3.1: Integration by Parts

When u and v are differentiable functions,

$$\int u\, dv = uv - \int v\, du \tag{8.3.2}$$

Examples of Integration by Parts

EXAMPLE 1. Find $\displaystyle\int xe^{3x}\, dx$.

SOLUTION Let's see what happens if we let $u = x$. Because $u\, dv$ must equal $xe^{3x}\, dx$, we must choose $dv = e^{3x}\, dx$. That is, $v' = e^{3x}$. Then, differentiating u gives $du = dx$ and integrating v' gives $v = \int e^{3x}\, dx = e^{3x}/3$. The integration by parts formula (8.3.2) tells us that:

$$\int \underbrace{x}_{u}\ \underbrace{e^{3x}\, dx}_{dv} = \underbrace{x}_{u}\ \underbrace{\frac{e^{3x}}{3}}_{v} - \int \underbrace{\frac{e^{3x}}{3}}_{v}\ \underbrace{dx}_{du} = \frac{xe^{3x}}{3} - \frac{e^{3x}}{9} = e^{3x}\left(\frac{x}{3} - \frac{1}{9}\right) + C.$$

To check this result, verify that the derivative of $e^{3x}(x/3 - 1/9) + C$ is xe^{3x}. ♦

Integration by parts worked in this example because $v\, du = e^{3x}\, dx$ is immediately recognized as the differential of $e^{3x}/3$. This was facilitated by the facts that the derivative of $u = x$ is simpler than u and it is easy to integrate $v' = e^{3x}$ to find v.

EXAMPLE 2. Find $\displaystyle\int x\ln(x)\, dx$.

SOLUTION Setting $dv = \ln(x)\, dx$ is not wise, since $v = \int \ln(x)\, dx$ is not immediately apparent. Setting $u = \ln(x)$ is promising because $du = d(\ln(x)) = dx/x$ is easier to handle than $\ln(x)$. This forces dv to be $x\, dx$. Then we have

$$u = \ln(x) \qquad dv = x\, dx$$
$$du = \frac{dx}{x} \qquad v = \frac{x^2}{2}.$$

Thus, by integration by parts,

$$\int x\ln(x)\, dx = \int \underbrace{\ln(x)}_{u}\ \underbrace{x\, dx}_{dv} = \underbrace{\ln(x)}_{u}\ \underbrace{\frac{x^2}{2}}_{v} - \int \underbrace{\frac{x^2}{2}}_{v}\ \underbrace{\frac{dx}{x}}_{du} = \frac{x^2\ln(x)}{2} - \int \frac{x\, dx}{2} = \frac{x^2\ln(x)}{2} - \frac{x^2}{4} + C.$$

As before, this antiderivative can be checked by differentiation. ♦

Observation 8.3.2: *Guidelines for Applying Integration by Parts*

To utilize (8.3.2), that is, to "integrate by parts" requires appropriate selection of u and dv. These three general objectives should be kept in mind while selecting u and dv:

1. v can be found by integrating dv, and is not too messy.
2. du should not be messier than u.
3. $\int v\,du$ should be easier to evaluate than the original integral, $\int u\,dv$

EXAMPLE 3. Find $\displaystyle\int \arctan(x)\,dx$.

SOLUTION The derivative of $\arctan(x)$ is $1/(1+x^2)$, a simpler function than $\arctan(x)$. This suggests the following integration by parts:

$$\begin{aligned} u &= \arctan(x) & dv &= dx \\ du &= \frac{dx}{1+x^2} & v &= x \end{aligned}$$

which allows us to write

$$\int \underbrace{\arctan(x)}_{u}\ \underbrace{dx}_{dv} = \underbrace{\arctan(x)}_{u}\ \underbrace{x}_{v} - \int \underbrace{x}_{v}\ \underbrace{\frac{dx}{1+x^2}}_{du} = x\arctan(x) - \int \frac{x}{1+x^2}\,dx.$$

Because the numerator in the new integrand is a constant times the derivative of the denominator, we have

$$\int \frac{x}{1+x^2}\,dx = \frac{1}{2}\int \frac{2x}{1+x^2}\,dx = \frac{1}{2}\ln(1+x^2).$$

Therefore

$$\int \arctan(x)\,dx = x\arctan(x) - \frac{1}{2}\ln(1+x^2) + C.$$

◆

The ideas used in Example 3 can be used to integrate any inverse trigonometric function. For practice, find $\int \arcsin(x)\,dx$. Compare your result with Formula 86 in the Table of Integrals (in Appendix A). (Actually, since the natural logarithm is the inverse of the exponential function, we see that the same idea was also used in Example 2.)

The next Example illustrates the importance of the guidelines, particularly the last one about $\int v\,du$ being easier than the original integral.

EXAMPLE 4. Find $\displaystyle\int x\sin(x)\,dx$.

SOLUTION We present two different choices of u and dv when using integration by parts to evaluate this antiderivative. We could choose $u = \sin(x)$ and $dv = x\,dx$ or we could choose $u = x$ and $dv = \sin(x)\,dx$.

Approach 1: Choose $u = \sin(x)$ and $dv = x\,dx$. Then

$$\begin{aligned} u &= \sin(x) & dv &= x\,dx \\ du &= \cos(x)\,dx & v &= \frac{x^2}{2}. \end{aligned}$$

Note that $du = \cos(x)\,dx$ is not any worse than $u = \sin(x)$. Thus,

$$\int \underbrace{\sin(x)}_{u}\,\underbrace{(x\,dx)}_{dv} = \underbrace{\sin(x)}_{u}\ \underbrace{\frac{x^2}{2}}_{v} - \int \underbrace{\frac{x^2}{2}}_{v}\ \underbrace{\cos(x)\,dx}_{du}.$$

The problem of finding $\int x\sin(x)\,dx$ has been replaced with the problem of finding $\int x^2\cos(x)\,dx/2$ — except that this new problem is harder than the original problem. That is not progress.

Approach 2: Choose $u = x$ and $dv = \sin(x)\,dx$. Then

$$\begin{aligned} u &= x & dv &= \sin(x)\,dx \\ du &= dx & v &= -\cos(x). \end{aligned}$$

This time integration by parts replaces the original integral with one requiring the antiderivative of $-\cos(x)$ — which is an easier problem than the original problem.

$$\begin{aligned} \int \underbrace{x}_{u}\ \underbrace{(\sin(x)\,dx)}_{dv} &= \underbrace{x}_{u}\ \underbrace{(-\cos(x))}_{v} - \int \underbrace{-\cos(x)}_{v}\ \underbrace{dx}_{du} \\ &= -x\cos(x) + \int \cos(x)dx = -x\cos(x) + \sin(x) + C. \end{aligned}$$

♦

EXAMPLE 5. Find $\displaystyle\int x^2 e^{3x}\,dx$.

SOLUTION If we let $u = x^2$, then $du = 2x\,dx$. This is a good start, for it lowers the exponent of x. For this to work we must also have $dv = e^{3x}\,dx$:

$$\begin{aligned} u &= x^2 & dv &= e^{3x}\,dx \\ du &= 2x\,dx & v &= \frac{1}{3}e^{3x}. \end{aligned}$$

Thus

$$\begin{aligned} \int \underbrace{x^2}_{u}\ \underbrace{e^{3x}\,dx}_{dv} &= \underbrace{x^2}_{u}\ \underbrace{\frac{1}{3}e^{3x}}_{v} - \int \underbrace{\frac{1}{3}e^{3x}}_{v}\underbrace{2x\,dx}_{du} && \text{(applying integration by parts)} \\ &= \frac{x^2}{3}e^{3x} - \frac{2}{3}\int xe^{3x}\,dx && \text{(simplification)} \\ &= \frac{x^2}{3}e^{3x} - \frac{2}{3}\left(e^{3x}\left(\frac{x}{3}-\frac{1}{9}\right)+C\right) && \text{(by Example 1)} \\ &= e^{3x}\left(\frac{x^2}{3} - \frac{2}{3}\left(\frac{x}{3}-\frac{1}{9}\right)\right) - \frac{2}{3}C && \text{(factoring the common } e^{3x}\text{)} \\ &= e^{3x}\left(\frac{x^2}{3} - \frac{2x}{9} + \frac{2}{27}\right) - \frac{2C}{3} && \text{(expanding and collecting like terms)} \end{aligned}$$

Lastly, we may rename $-2C/3$, the arbitrary constant, with something simpler, such as K, obtaining

$$\int x^2 e^{3x}\,dx = e^{3x}\left(\frac{x^2}{3} - \frac{2x}{9} + \frac{2}{27}\right) + K.$$

♦

Notice that the solution of Example 1 involves integration by parts, we actually applied integration by parts twice to find an antiderivative of x^2e^{3x}, As a result, the idea in Example 5 applies to any integral of the form $\int P(x)g(x)\,dx$, where $P(x)$ is a polynomial and $g(x)$ is a function, such as $\sin(x)$, $\cos(x)$, or e^x, that can be repeatedly integrated without affecting the complexity of this term. Choose $u = P(x)$ because the degree of u' is one lower than the degree of u. By applying integration by parts several times, eventually the polynomial $P(x)$ is reduced to a constant.

Definite Integrals and Integration by Parts

Integration by parts of a definite integral $\int_a^b f(x)\,dx$, where $f(x) = u(x)v'(x)$, is completed as follows:

Theorem 8.3.3: Integration by Parts for a Definite Integral

When u and v are differentiable functions on the closed interval between a and b,

$$\int_a^b u(x)v'(x)\,dx = u(x)v(x)\Big|_a^b - \int_a^b v(x)u'(x)\,dx$$

EXAMPLE 6. Find the area under the curve $y = \arctan(x)$ and above $[0,1]$.

SOLUTION The area is $\int_0^1 \arctan(x)\,dx$. (See Figure 8.3.1.) By Example 3,

$$\int \arctan(x)\,dx = x\arctan(x) - \frac{1}{2}\ln(1+x^2) + C.$$

Since only one antiderivative is needed in order to apply the fundamental theorem of calculus, we may choose $C = 0$. Then

$$\begin{aligned}\int_0^1 \arctan(x)\,dx &= x\arctan(x)\Big|_0^1 - \frac{1}{2}\ln(1+x^2)\Big|_0^1\\ &= 1\arctan(1) - 0\arctan(0) - \frac{1}{2}\ln(1+1^2) + \frac{1}{2}\ln(1+0^2)\\ &= \frac{\pi}{4} - \frac{1}{2}\ln(2) \approx 0.439.\end{aligned}$$

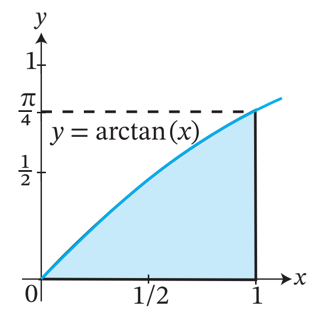


Figure 8.3.1

♦

Reduction Formulas

Formulas 52, 59, and 62 (and others) in the Table of Integrals (in Appendix A) express the integral of a function that involves the power of some expression in terms of the integral of a function that involves a lower power of the same expression. These are **reduction formulas** or **recursion formulas**. Usually they are obtained by integration by parts.

An example of a reduction formula is

$$\int \sin^n(x)\,dx = -\frac{\sin^{n-1}(x)\cos(x)}{n} + \frac{n-1}{n}\int \sin^{n-2}(x)\,dx \tag{8.3.3}$$

for integer values of $n \geq 2$.

EXAMPLE 7. Use the reduction formula (8.3.3) to evaluate $\int \sin^5(x)\,dx$.

SOLUTION With $n = 5$, (8.3.3) gives

$$\int \sin^5(x)\,dx = -\frac{\sin^4(x)\cos(x)}{5} + \frac{4}{5}\int \sin^3(x)\,dx. \tag{8.3.4}$$

Use (8.3.3) again, this time with $n = 3$, to get

$$\int \sin^3(x)\,dx = -\frac{\sin^2(x)\cos(x)}{3} + \frac{2}{3}\int \sin(x)\,dx. \tag{8.3.5}$$

This time, the remaining integral is one that can be evaluated directly. Thus, the full evaluation of this antiderivative is:

$$\begin{aligned}
\int \sin^5(x)\,dx &= -\frac{\sin^4(x)\cos(x)}{5} + \frac{4}{5}\int \sin^3(x)\,dx && (\text{ apply (8.3.3) with } n=5\,) \\
&= -\frac{\sin^4(x)\cos(x)}{5} + \frac{4}{5}\left(\frac{-\sin^2(x)\cos(x)}{3} + \frac{2}{3}\int \sin(x)\,dx\right) && (\text{ apply (8.3.3) with } n=3\,) \\
&= -\frac{\sin^4(x)\cos(x)}{5} + \frac{4}{5}\left(\frac{-\sin^2(x)\cos(x)}{3} - \frac{2}{3}\cos(x)\right) + C && (\,\sin(x)\,dx = -\cos(x)\,) \\
&= -\frac{1}{5}\sin^4(x)\cos(x) - \frac{4}{15}\sin^2(x)\cos(x) - \frac{8}{15}\cos(x) + C && (\text{ simplification }).
\end{aligned}$$

♦

Observation 8.3.4: *Antiderivatives of Powers of* $\sin(x)$

In general, an integer power of $\sin(x)$ can be integrated by repeated application of (8.3.3). Every time (8.3.3) is used, the exponent of $\sin(x)$ decreases by 2. Repeated application of (8.3.3) will reduce the exponent to either 1 or 0. The resulting integrand is a multiple of $\sin(x)$ or a constant, either of which is easily integrated.

The next example shows how (8.3.3) can be obtained by integration by parts.

EXAMPLE 8. Obtain the reduction formula (8.3.3).

SOLUTION Write $\int \sin^n(x)\,dx$ as $\int \sin^{n-1}(x)\sin(x)\,dx$. Then let $u = \sin^{n-1}(x)$ and $dv = \sin(x)\,dx$. Thus

$$\begin{aligned}
u &= \sin^{n-1}(x) & dv &= \sin(x)\,dx \\
du &= (n-1)\sin^{n-2}(x)\cos(x)\,dx & v &= -\cos(x).
\end{aligned}$$

Integration by parts yields

$$\int \underbrace{\sin^{n-1}(x)}_{u}\underbrace{\sin(x)\,dx}_{dv} = \underbrace{(\sin^{n-1}(x))}_{u}\underbrace{(-\cos(x))}_{v} - \int \underbrace{(-\cos(x))}_{v}\underbrace{(n-1)\sin^{n-2}(x)\cos(x)\,dx}_{du}.$$

The integral on the right-hand side of this equation is easily evaluated by replacing $\cos^2(x)$ with $1-\sin^2(x)$:

$$\begin{aligned}
-\int (n-1)\cos^2(x)\sin^{n-2}(x)\,dx &= -(n-1)\int (1-\sin^2(x))\sin^{n-2}(x)\,dx \\
&= -(n-1)\int \sin^{n-2}(x)\,dx + (n-1)\int \sin^n(x)\,dx.
\end{aligned}$$

Thus

$$\begin{aligned}
\int \sin^n(x)\,dx &= -\sin^{n-1}(x)\cos(x) - \left(-(n-1)\int \sin^{n-2}(x)\,dx + (n-1)\int \sin^n(x)\,dx\right) \\
&= -\sin^{n-1}(x)\cos(x) + (n-1)\int \sin^{n-2}(x)\,dx - (n-1)\int \sin^n(x)\,dx.
\end{aligned}$$

Collecting the terms involving $\int \sin^n(x)\,dx$, we obtain

$$n\int \sin^n(x)\,dx = -\sin^{n-1}(x)\cos(x) + (n-1)\int \sin^{n-2}(x)\,dx,$$

from which (8.3.3) follows. ♦

The reduction formula for $\int \cos^n(x)\, dx$ is obtained similarly.

EXAMPLE 9. Obtain the reduction formula for $\int \frac{dx}{\left(x^2+c\right)^n}$ where n is a positive integer.

SOLUTION The only choice that comes to mind for integration by parts is

$$\begin{aligned} u &= \frac{1}{(x^2+c)^n} & dv &= dx \\ du &= \frac{-2nx}{(x^2+c)^{n+1}} & v &= x. \end{aligned}$$

Then, integration by parts gives

$$\int \frac{dx}{\left(x^2+c\right)^n} = \frac{x}{\left(x^2+c\right)^n} + 2n \int \frac{x^2}{\left(x^2+c\right)^{n+1}}\, dx.$$

It looks as though we have created a more complicated integrand. However, in the numerator of the integrand on the right-hand side, write x^2 as x^2+c-c. We then have

$$\int \frac{dx}{\left(x^2+c\right)^n} = \frac{x}{\left(x^2+c\right)^n} + 2n \int \frac{x^2+c}{\left(x^2+c\right)^{n+1}}\, dx - 2nc \int \frac{dx}{(x^2+c)^{n+1}}$$

Canceling out x^2+c in the first integrand on the right gives an equation that could be rewritten to express $\int dx/\left(x^2+c\right)^{n+1}$ in terms of $\int dx/\left(x^2+c\right)^n$. Here is the algebra.

In order to make the equations easier to read we introduce I_n to represent $\int dx/\left(x^2+c\right)^n$ for any positive integer n. Thus the last equation is

$$I_n = \frac{x}{(x^2+c)^n} + 2nI_n - 2ncI_{n+1}.$$

Do not despair. While this equation suggests we are going the wrong way, expressing I_n in terms of I_n and I_{n+1}, a little algebra will rewrite this equation so it expresses I_{n+1} in terms of I_n. Moving I_{n+1} to the left-hand side and I_n from the left-hand side to the right-hand side of the equation, we obtain

$$2ncI_{n+1} = \frac{x}{(x^2+c)^n} + (2n-1)I_n$$

hence

$$I_{n+1} = \frac{x}{2nc(x^2+c)^n} + \frac{2n-1}{2nc} I_n,$$

We have a reduction formula that lowers the index from $n+1$ to n.

If we prefer to lower the index from n to $n-1$ instead of from $n+1$ to n, we simply replace n by $n-1$ (so that $n+1$ gets replaced by n) to obtain

$$I_n = \frac{x}{2(n-1)c(x^2+c)^{n-1}} + \frac{2n-3}{2(n-1)c} I_{n-1},$$

◆

An Unusual Example

In the final Example one integration by parts appears at first to be useless, but two applications in succession provide a way to evaluate the integral.

EXAMPLE 10. Find $\int e^x \cos(x)\, dx$

SOLUTION There are two reasonable choices for applying integration by parts: $u = e^x$, $dv = \cos(x)\, dx$ or $u = \cos(x)$, $dv = e^x\, dx$. In neither case is du simpler, but watch what happens when integration by parts is applied twice.

With the first choice,

$$\begin{aligned} u &= e^x & dv &= \cos(x)\, dx \\ du &= e^x\, dx & v &= \sin(x), \end{aligned}$$

and integration by parts gives

$$\int \underbrace{e^x}_{u} \underbrace{\cos(x)\, dx}_{dv} = \underbrace{e^x}_{u} \underbrace{\sin(x)}_{v} - \int \underbrace{\sin(x)}_{v} \underbrace{e^x\, dx}_{du}. \tag{8.3.6}$$

It may seem that nothing useful has been accomplished because we cannot evaluate the integral on the right. Let us integrate it by parts. We will use U and V instead of u and v to distinguish this computation from the first one:

$$\begin{aligned} U &= e^x & dV &= \sin(x)\, dx \\ dU &= e^x\, dx & V &= -\cos(x). \end{aligned}$$

This yields

$$\int \underbrace{e^x}_{U} \underbrace{\sin(x)\, dx}_{dV} = \underbrace{e^x}_{U} \underbrace{(-\cos(x))}_{V} - \int \underbrace{(-\cos(x))}_{V} \underbrace{e^x\, dx}_{dU} = -e^x \cos(x) + \int e^x \cos(x)\, dx. \tag{8.3.7}$$

Combining (8.3.6) and (8.3.7) yields

$$\int e^x \cos(x)\, dx = e^x \sin(x) - \left(-e^x \cos(x) + \int e^x \cos(x)\, dx \right) = e^x(\sin(x) + \cos(x)) - \int e^x \cos(x)\, dx.$$

Bringing $-\int e^x \cos x\, dx$ to the left side of the equation gives

$$2 \int e^x \cos(x)\, dx = e^x(\sin(x) + \cos(x)),$$

and, remembering to include a constant of integration because this is an indefinite integral, we conclude that

$$\int e^x \cos(x)\, dx = \frac{1}{2} e^x (\sin(x) + \cos(x)) + C.$$

♦

Summary

Integration by parts is described by the formulas

$$\int u\, dv = uv - \int v\, du \qquad \text{and} \qquad \int_a^b u\, dv = uv\Big|_a^b - \int_a^b v\, du$$

To split an integrand into the parts u and dv, try to choose them so that

1. v, an antiderivative of dv, can be found and it is not too complicated
2. du is simpler than u
3. $\int v\, du$ is easier to evaluate that the original integral, $\int u\, dv$.

Sometimes integration by parts must be used more than once, for instance, in finding $\int e^x \cos(x)\, dx$.

EXERCISES for Section 8.3

Use integration by parts to evaluate the integrals in Exercises 1 to 20.

1. $\int xe^{2x}\,dx$ **2.** $\int (x+3)e^{-x}\,dx$ **3.** $\int x\sin(2x)\,dx$ **4.** $\int (x+3)\cos(2x)\,dx$

5. $\int x\ln(3x)\,dx$ **6.** $\int (2x+1)\ln(x)\,dx$ **7.** $\int_1^2 x^2e^{-x}\,dx$ **8.** $\int_0^1 x^2e^{2x}\,dx$

9. $\int_0^1 \arcsin(x)\,dx$ **10.** $\int_0^{1/2} \arctan(2x)\,dx$ **11.** $\int x^2\ln(x)\,dx$ **12.** $\int x^3\ln(x)dx$

13. $\int_2^3 (\ln(x))^2\,dx$ **14.** $\int_2^3 (\ln(x))^3\,dx$ **15.** $\int_1^e \frac{\ln(x)\,dx}{x^2}$ **16.** $\int_e^{e^2} \frac{\ln(x)\,dx}{x^3}$

17. $\int e^{3x}\cos(2x)\,dx$ **18.** $\int e^{-2x}\sin(3x)\,dx$ **19.** $\int \frac{\ln(1+x^2)}{x^2}\,dx$ **20.** $\int x\ln(x^2)\,dx$

In Exercises 21 to 24 evaluate the integral two ways: (a) by substitution and (b) by integration by parts.

21. $\int x\sqrt{3x+7}\,dx$ **22.** $\int \frac{x\,dx}{\sqrt{2x+7}}$ **23.** $\int x(ax+b)^3\,dx$ **24.** $\int \frac{x\,dx}{\sqrt[3]{ax+b}}$ $(a\neq 0)$

---

25. Use differentiation to verify (8.3.3).

26. Use the recursion in Example 8 to find (a) $\int \sin^2(x)\,dx$, (b) $\int \sin^4(x)\,dx$, and (c) $\int \sin^6(x)\,dx$.

27. Use the recursion in Example 8 to find (a) $\int \sin^3(x)\,dx$ and (b) $\int \sin^5(x)\,dx$.

28. (a) Graph $y=e^x\sin(x)$ for x in $[0,\pi]$, showing extrema and inflection points.
(b) Find the area of the region below the graph and above $[0,\pi]$.

29. (a) Graph $y=e^{-x}\sin(x)$ for x in $[0,\pi]$, showing extrema and inflection points.
(b) Find the area of the region below the graph and above $[0,\pi]$.

30. Find the volume of the solid produced when the region under $y=xe^x$ and above $[0,1]$ is revolved around the x-axis to produce a solid.

31. Find the volume of the solid produced when the region under $y=\ln(x)$ and above $[1,3]$ is revolved around the x-axis to produce a solid.

32. In Exercise 71 in Section 6.4 it is claimed that $\frac{e^x}{x}$ does not have an elementary antiderivative. From this we can show other functions also lack elementary antiderivatives. (a) Show, using substitution or integration by parts, that $\int \frac{e^x}{x}\,dx$ equals $\ln(x)e^x - \int \ln(x)e^x\,dx$, $\frac{e^x}{x} + \int \frac{e^x}{x^2}\,dx$, and $\int \frac{du}{\ln(u)}$ (where $u=e^x$). (b) Deduce that $\ln(x)e^x$, $\frac{e^x}{x^2}$, and $\frac{1}{\ln(x)}$ do not have elementary antiderivatives.

In Exercises 33 to 36 evaluate the integrals. Make a substitution before integrating by parts. In Exercises 35 and 36 the notation $\exp(u)$ is used for e^u.

33. $\int \sin\left(\sqrt{x}\right)\,dx$ **34.** $\int \sin\left(\sqrt[3]{x}\right)\,dx$ **35.** $\int \exp\left(\sqrt{x}\right)\,dx$ **36.** $\int \exp\left(\sqrt[3]{x}\right)\,dx$

---

37. Explain how you would go about finding $\int x^{10}(\ln(x))^{18}\,dx$. Explain why your approach would work, but do not evaluate this definite integral. Include only enough calculation to convince a reader that it would succeed.

NOTE: Don't say, "I'd use integral tables or a computer."

38. Assume that $\int \frac{\sin(x)}{x}\,dx$ is not elementary. Deduce that $\int \cos(x)\,\ln(x)\,dx$ is not elementary.

39. Assume that $\int x\tan(x)\,dx$ is not elementary. Deduce that $\int \left(\frac{x}{\cos(x)}\right)^2 dx$ is not elementary.

40. Let I_n denote $\int_0^{\pi/2} \sin^n(\theta)\,d\theta$, where n is a nonnegative integer.

(a) Evaluate I_0 and I_1.
(b) Using the recursion in Example 8, show that $I_n = \frac{n-1}{n} I_{n-2}$ for $n \geq 2$.
(c) Use (b) to evaluate I_2 and I_3.
(d) Use (c) to evaluate I_4 and I_5.
(e) Explain why $I_n = \frac{2\cdot 4\cdot 6\cdots(n-1)}{3\cdot 5\cdot 7\cdots n}$ when n is odd.
(f) Explain why $I_n = \frac{1\cdot 3\cdot 5\cdots(n-1)}{2\cdot 4\cdot 6\cdots n}\cdot\frac{\pi}{2}$ when n is even.
(g) Explain why $\int_0^{\pi/2} \sin^n(\theta)\,d\theta = \int_0^{\pi/2} \cos^n(\theta)\,d\theta$.

41. Find $\int \ln(x+1)\,dx$ (a) using $u=\ln(x+1)\,dx$, $dv=dx$, $v=x$ and (b) using $u=\ln(x+1)\,dx$, $dv=dx$, $v=x+1$. (c) Which approach is easier? Why?

42. If n is a positive integer and a is a constant, obtain a formula that expresses $\int x^n e^{-ax}\,dx$ in terms of $\int x^{n-1}e^{-ax}$.

43. Find $\int x\sin(ax)\,dx$.

44. Let a be a constant and n a positive integer. (a) Express $\int x^n\sin(ax)\,dx$ in terms of $\int x^{n-1}\cos(ax)\,dx$. (b) Express $\int x^n\cos(ax)\,dx$ in terms of $\int x^{n-1}\sin(ax)\,dx$. (c) Why do (a) and (b) enable us to find $\int x^n\sin(ax)\,dx$?

45. (a) Express $\int(\ln(x))^n\,dx$ in terms of $\int(\ln(x))^{n-1}\,dx$ and (b) use (a) to find $\int(\ln(x))^3\,dx$.

46. (a) Show how the integral $\int \frac{dx}{(ax^2+bx+c)^{n+1}}$ can be reduced to an integral of the form $A\int \frac{du}{(u^2+p)^{n+1}}$.
(b) Apply the reduction formula found in Example 9 to write the reduction formula for $\int \frac{du}{(u^2+p)^{n+1}}$.
(c) Use (a) and (b) to find a reduction formula for $\int \frac{dx}{(x^2+bx+c)^{n+1}}$.
(d) How does your answer in (c) compare with Formula 52 in the Table of Integrals (in Appendix A)? Explain any differences.

In Exercises 47 to 50 obtain reduction formulas for each integral. (Assume n is a positive integer and $a \neq 0$.)

47. $\int x^n e^{ax}\,dx$ **48.** $\int(\ln(x))^n\,dx$ **49.** $\int x^n\sin(x)\,dx$ **50.** $\int \cos^n(ax)\,dx$

Laplace Transform Let $f(t)$ be a continuous function defined for $t \geq 0$. Assume that, for positive numbers s, $\int_0^\infty e^{-st} f(t)\,dt$ converges and that $e^{-st} f(t) \to 0$ as $t \to \infty$. Define $P(s)$ to be $\int_0^\infty e^{-st} f(t)\,dt$. The function P is called the **Laplace transform** of the function f. It is an important tool for solving differential equations, and appears in the CIE 10 (in Chapter 7). In Exercises 51 to 59 find the Laplace transform of each function.

51. $f(t) = 1$

52. $f(t) = t$

53. $f(t) = t^2$

54. $f(t) = t^n$ (n is a positive integer)

55. $f(t) = e^{at}$ ($s > a$)

56. $f(t) = \sin(t)$

57. $f(t) = \cos(t)$

58. $f(t) = \sin(at)$ ($a \neq 0$)

59. $f(t) = \cos(at)$ ($a \neq 0$)

---

60. Let $P(x)$ be a polynomial.

(a) Check by differentiation that $(P(x) - P'(x) + P''(x) - \cdots)e^x$ is an antiderivative of $P(x)e^x$. (The signs alternate and the derivatives are taken to successively higher orders until they are constant, with value 0.)

(b) Use (a) to find $\int (3x^3 - 2x - 2)e^x\,dx$.

(c) Apply integration by parts to $\int P(x)e^x\,dx$ to show how the formula in (a) could be obtained.

61. In Example 10, $\int e^x \cos(x)\,dx$ was evaluated by applying integration by parts twice, each time differentiating an exponential and antidifferentiating a trigonometric function. What happens when integration by parts is applied (twice, if necessary) when a trigonometric function is differentiated and an exponential is antidifferentiated? That is, evaluate $\int e^x \cos(x)\,dx$ by applying integration by parts with $u = \cos(x)$ and $dv = e^x\,dx$.

62. Find $\int_{-1}^{1} x^3 \sqrt{1 + x^{20}}\,dx$.

63. Find $\int_{-\pi/4}^{\pi/4} \tan(x)(1 + \cos(x))^{3/2}\,dx$.

---

64. We have learned that integration by parts can be used to find $\int e^x \sin(x)\,dx$ but this is long, tedious, and error-prone. In this problem, find $\int e^x \sin(x)\,dx$ by assuming that the integral has the form $ae^x \sin(x) + be^x \cos(x)$. Find the equations a and b must satisfy. Then solve for the values of the constants a and b for which $\int e^x \sin(x)\,dx = ae^x \sin(x) + be^x \cos(x) + C$. (Are these values of a and b unique?)

65. (a) What does the graph of $y = \cos(ax)$ look like when $a = 1$? when $a = 2$? when $a = 3$? when a is a large constant? Include graphs and a written description in your answers.

(b) Let $f(x)$ be a function with a continuous derivative. Assume that $f(x)$ is positive. What does the graph of $y = f(x)\cos(ax)$ look like when a is large? Express your response in terms of the graph of $y = f(x)$. Include a sketch of $y = f(x)\cos(ax)$ to give an idea of its shape.

(c) On the basis of (b), what do you think happens to $\int_0^1 f(x)\cos(ax)\,dx$ as $a \to \infty$? Give an explanation.

(d) Use integration by parts to justify your answer in (c).

66. In Exercise 46, a reduction formula for $\int \frac{dx}{(ax^2 + bx + c)^{n+1}}$ was found. Why does the Table of Integrals (in Appendix A) not include a reduction formula for $\int \frac{x\,dx}{(ax^2 + bx + c)^{n+1}}$?

8.4 Integrating Rational Functions: The Algebra

Recall that a rational function is a polynomial or the quotient of two polynomials. Every rational function, no matter how complicated, has an elementary antiderivative that is the sum of some or all of:

- rational functions (including polynomials),
- logarithms of linear or quadratic polynomials: $\ln(ax+b)$ or $\ln(ax^2+bx+c)$, and
- arctangents of linear or quadratic polynomials: $\arctan(ax+b)$ or $\arctan(ax^2+bx+c)$.

The reason is algebraic. It can be shown that every rational function can be expressed as the sum of constant multiples of the following four types of rational functions,

$$\text{polynomials}, \quad \frac{1}{(ax+b)^n}, \quad \frac{1}{(ax^2+bx+c)^n}, \quad \text{and} \quad \frac{x}{(ax^2+bx+c)^n} \tag{8.4.1}$$

where a, b, and c are constants and n is a positive integer. In Sections 8.2 and 8.3, and their Exercises, we saw how to integrate each of these functions.

Our objective in this section is to see how to express a rational function $f(x)$ as a sum of the functions in (8.4.1), that is, to find the **partial fraction representation** or **partial fraction representation** of $f(x)$. For instance, the techniques and methods we will learn would allow us to find the following partial fraction representation:

$$\frac{2x^3+1}{2x^2-x+1} = x + \frac{1}{2} + \frac{-x/2+1/2}{2x^2-x+1}.$$

Reducible and Irreducible Polynomials

A polynomial $a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, where a_n is not 0, has degree n. Polynomials of degree one are called linear and those of degree two, quadratic. A polynomial of degree zero is a constant. If all a_i are zero, we have the zero polynomial, which is not assigned a degree.

A polynomial of degree at least one is **reducible** if it is a product of nonconstant polynomials of lower degree. Otherwise, it is **irreducible**.

Every polynomial of degree one, $ax+b$, is clearly irreducible. Polynomials of degree two, ax^2+bx+c, are irreducible if and only if its discriminant, b^2-4ac, is negative. (See Exercises 42 and 43.) However, the situation is much simpler for all polynomials of degree three and higher.

Theorem 8.4.1: Factoring Polynomials

Every polynomial of degree three or higher is reducible. That is, every polynomial of degree three or higher is the product of polynomials of degree one or two.

This is far from obvious. For instance, x^4+1 looks like it cannot be factored, but

$$x^4+1 = (x^2+\sqrt{2}x+1)(x^2-\sqrt{2}x+1).$$

On the other hand,

$$x^4-1 = (x^2+1)(x^2-1) = (x^2+1)(x+1)(x-1).$$

These two examples illustrate Theorem 8.4.1. A full proof of this fact is reserved for a more advanced course.

Proper and Improper Rational Functions

Let $A(x)$ and $B(x)$ be polynomials. The rational function $A(x)/B(x)$ is **proper** if the degree of $A(x)$ is less than the degree of $B(x)$. Otherwise it is **improper**. Every **improper rational function** is the sum of a polynomial and a **proper rational function**. The next example illustrates why this is true, using long division.

RECALL: In arithmetic, the fraction m/n is called proper if $|m|$ is strictly less than $|n|$.

EXAMPLE 1. Express the improper rational function $\frac{2x^3+1}{2x^2-x+1}$ as a polynomial plus a proper rational function.

SOLUTION

We carry out a long division

$$\begin{array}{rrrrrl}
 & & & x & +1/2 & \leftarrow \text{quotient} \\
2x^2-x+1) & 2x^3 & +0x^2 & +0x & +1 & \\
 & 2x^3 & -x^2 & +x & & \\
\hline
 & & x^2 & -x & +1 & \\
 & & x^2 & -x/2 & +1/2 & \\
\hline
 & & & -x/2 & +1/2 & \leftarrow \text{remainder}
\end{array}$$

Repeatedly divide until the degree of the remainder is less than the degree of the divisor, or the remainder is 0.

Thus

$$2x^3+1 = (2x^2-x+1)(x+\frac{1}{2}) + \left(-\frac{x}{2}+\frac{1}{2}\right)$$

and we write

$$\underbrace{\frac{2x^3+1}{2x^2-x+1}}_{\text{improper}} = \underbrace{x+\frac{1}{2}}_{\text{polynomial}} + \underbrace{\frac{\frac{-x}{2}+\frac{1}{2}}{2x^2-x+1}}_{\text{proper}}.$$

◆

To check the result of Example 1, rewrite the right-hand side as a single fraction.

To integrate a rational function we first check that it is proper. If it is improper, we carry out a long division to represent it as the sum of a polynomial and a proper rational function. Since we already know how to integrate a polynomial, the rest of this section considers only proper rational functions.

Partial Fractions

As mentioned in the introduction, every rational function is the sum of simple rational functions, ones we know how to integrate. The following algorithm can be used to find the partial fraction representation of any rational function.

Algorithm: Finding a Partial Fraction Representation of a Rational Function

Use this method to find the simple rational functions whose sum is the given proper rational function $\frac{A(x)}{B(x)}$.

Step 1: **Factor the denominator $B(x)$.**
Write $B(x)$ as a product of first-degree polynomials $(px+q)$ and second-degree polynomials (ax^2+bx+c) that are irreducible.

Step 2: **List summands of the form $\frac{k_i}{(px+q)^i}$.**
If $px+q$ appears exactly n times in the factorization of $B(x)$, then the partial fraction representation needs to include the following n terms:

$$\frac{k_1}{px+q} + \frac{k_2}{(px+q)^2} + \cdots + \frac{k_n}{(px+q)^n},$$

where the constants $k_1, k_2, \ldots, k_n$ are to be determined later.

Step 3: **List summands of the form $\frac{r_jx+s_j}{(ax^2+bx+c)^j}$.**

If ax^2+bx+c appears exactly m times in the factorization of $B(x)$, then the partial fraction representation needs to include the following m terms:

$$\frac{r_1x+s_1}{ax^2+bx+c}+\frac{r_2x+s_2}{(ax^2+bx+c)^2}+\cdots+\frac{r_mx+s_m}{(ax^2+bx+c)^m},$$

where the constants $r_1, r_2, \ldots, r_m$ and $s_1, s_2, \ldots, s_m$ are to be determined later.

Step 4: **Solve for the values of the constants.**
Find the values of the constants (k_i, r_j, and s_j) so that the sum of the rational functions in Steps 2 and 3 equals $A(x)/B(x)$.

TERMINOLOGY: The functions in Steps 2 and 3 are called the **partial fractions** of $A(x)/B(x)$.

In practice the denominator $B(x)$ often appears in factored form. If it does not, finding the factorization can be a challenge. To find first-degree factors, look for a root of $B(x)=0$. If r is a root of $B(x)$, then $x-r$ is a factor. Divide $x-r$ into $B(x)$, getting a quotient $Q(x)$. Therefore $B(x)=(x-r)Q(x)$. Repeat the process, continuing as long as you can find roots. If a root is approximated to several decimal places, the integration will be approximate and it might be simpler (and more accurate) to approximate the definite integral.

After finding all the linear factors what's left has to be the product of second-degree polynomials. If there is only one such factor, then we have the complete factorization. But, if the degree is 4 or 6 or higher, the task is best left to a computer or avoided.

Steps 2 and 3 refer to the number of times a factor occurs in the denominator. If you factor $2x^2+4x+2$, you may obtain $(x+1)(2x+2)$. Because $2x+2$ is a constant times $x+1$ the factorization may be written as $2(x+1)^2$, and $x+1$ is a repeated factor. We say that $x+1$ appears exactly two times in the factorization of $2x^2+4x+2$. Always collect factors that are constants times each other.

Step 4, finding the unknown constants, may take a lot of work. If there are only linear factors without repetition, the method illustrated in Example 3 is quick. Clearing denominators and comparing the corresponding coefficients of the polynomials on both sides of the resulting equation will determine them. The number of unknown constants always equals the degree of the denominator $B(x)$. If $B(x)$ has repeated linear or second-degree factors and the degree of $B(x)$ is large, consider using a computing tool to find approximations to the coefficients.

EXAMPLE 2. What is the form of the partial fraction representation of $f(x)=\dfrac{x^{10}+x+3}{(x+1)^2(2x+2)^3(x-1)^2(x^2+x+3)^2}$?

SOLUTION The degree of the denominator is 11 and the degree of the numerator is 10. Thus $f(x)$ is a proper rational function; there is no need to divide the numerator by the denominator.

The factor $2x+2$ is $2(x+1)$. Therefore $(x+1)^2(2x+2)^3$ should be written as $8(x+1)^5$. The discriminant of x^2+x+3 is $(1)^2-4(1)(3)=-11<0$, which implies that x^2+x+3 is irreducible. Therefore the partial fraction representation of $f(x)$ has the form

$$\begin{aligned}&\frac{k_1}{x+1}+\frac{k_2}{(x+1)^2}+\frac{k_3}{(x+1)^3}+\frac{k_4}{(x+1)^4}+\frac{k_5}{(x+1)^5}\\&\quad+\frac{k_6}{x-1}+\frac{k_7}{(x-1)^2}\\&\quad+\frac{r_1x+s_1}{x^2+x+3}+\frac{r_2x+s_2}{(x^2+x+3)^2}.\end{aligned}$$

As a check, note that the number of unknown constants equals the degree of the denominator of $f(x)$. ◆

Finding the constants in Example 2 by hand would be a major task to undertake. It would involve solving a system of 11 equations for the 11 unknown constants. This is a challenge to complete by hand, but is straightforward for modern mathematical computational tools.

Denominator Has Only Linear Factors, Each Appearing Only Once

The next example illustrates this case.

EXAMPLE 3. Express $\frac{1}{(2x+1)(x+3)}$ in the form $\frac{k_1}{2x+1}+\frac{k_2}{x+3}$.

SOLUTION This requires finding values for the constants k_1 and k_2 that satisfy

$$\frac{1}{(2x+1)(x+3)} = \frac{k_1}{2x+1}+\frac{k_2}{x+3} \tag{8.4.2}$$

for all values of x where the rational expression on the left-hand side is defined (that is, all values of x except $x=-1/2$ and $x=-3$).

The constants k_1 and k_2 in (8.4.2) will be found in two ways.

To find k_1, multiply both sides of (8.4.2) by the denominator of the term that contains k_1, $2x+1$, getting

$$\frac{1}{x+3} = k_1 + \frac{k_2(2x+1)}{x+3}. \tag{8.4.3}$$

Equation (8.4.3) is valid for all values of x except $x=-3$, in particular for the value of x that makes $2x+1=0$, namely $x=-1/2$. Evaluating (8.4.2) when $x=-1/2$ we get

$$\frac{1}{\frac{-1}{2}+3} = k_1 + 0$$

which means k_1 is $2/5$.

Similarly, to find k_2 multiply both sides of (8.4.2) by $(x+3)$, obtaining

$$\frac{1}{2x+1} = \frac{k_1(x+3)}{2x+1} + k_2.$$

Replace x by -3, so the factor $x+3$ reduces to zero, to obtain

$$\frac{1}{2(-3)+1} = 0 + k_2.$$

Thus k_2 is $-1/5$.

Since $k_1=2/5$ and $k_2=-1/5$, (8.4.2) takes the form

$$\frac{1}{(2x+1)(x+3)} = \frac{2/5}{2x+1} - \frac{1/5}{x+3}.$$

To verify this representation, check it by multiplying both sides by $(2x+1)(x+3)$, getting

$$1 = \frac{2}{5}(x+3) - \frac{1}{5}(2x+1) = \frac{2}{5}x + \frac{6}{5} - \frac{2}{5}x - \frac{1}{5} = \frac{5}{5}.$$

For a quicker, but not complete check, replace x in (8.4.2) by a convenient number and see if the resulting equation is correct. Try it with, say, $x=0$.

It checks!

Another way to solve for the unknown constants is to clear the denominator and equate coefficients of like powers of x. For instance, let us find k_1 and k_2 in (8.4.2). We obtain

$$1 = k_1(x+3) + k_2(2x+1).$$

Collecting coefficients, we have

$$1 = (k_1+2k_2)x + (3k_1+k_2). \tag{8.4.4}$$

Comparing coefficients on both sides of (8.4.4) yields

$$\begin{aligned} 0 &= k_1 + 2k_2 && (\text{ equating coefficients of } x\ ) \\ 1 &= 3k_1 + k_2 && (\text{ equating constant terms }). \end{aligned}$$

There are many ways to solve this system of two equations for k_1 and k_2. One is to use the first equation to express k_1 in terms of k_2: $k_1 = -2k_2$. Replacing k_1 by $-2k_2$ in the second equation produces:

$$1 = 3(-2k_2) + k_2 = -5k_2$$

from which it is seen that $k_2 = -1/5$. Then $k_1 = 2/5$.

In this method the number of equations always equals the number of unknowns, which is equal to the degree of the denominator. If the degree is large, it may not be realistic to do the calculations by hand. ◆

Denominator with Repeated Linear Factors

If the denominator is a repeated linear factor, there are two options. One is to use the method just described, which requires solving a system of linear equations. The other is to use a substitution to simplify the denominator.

To illustrate the second method, find the partial fraction representation of

$$\frac{7x+6}{(x+2)^2}.$$

Let $u = x+2$, hence $x = u-2$. Then

$$\begin{aligned} \frac{7x+6}{(x+2)^2} &= \frac{7(u-2)+6}{u^2} && (\text{ substitute } x = u-2\ ) \\ &= \frac{7u}{u^2} - \frac{8}{u^2} && (\text{ distribution and simplification }) \\ &= \frac{7}{u} - \frac{8}{u^2} && (\text{ simplification }) \\ &= \frac{7}{x+2} - \frac{8}{(x+2)^2} && (\text{ replacing } u \text{ with } x+2\ ). \end{aligned}$$

This method for representing

$$\frac{A(x)}{(ax+b)^n}$$

is practical if the degree of $A(x)$ is small. Here $u = ax+b$, so $x = \frac{1}{a}(u-b)$. If the degree of $A(x)$ is not small, expressing a power of $u-b$, $(u-b)^m$, in terms of powers of u would best be done by the Binomial Theorem:

$$(a+b)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} a^{n-k} b^k.$$

The next Example illustrates one way of dealing with a denominator with both first and second degree factors.

EXAMPLE 4. Obtain the partial-fraction representation of $\dfrac{x^2}{x^4-1}$.

SOLUTION Factor the denominator: $x^4 - 1 = (x^2+1)(x+1)(x-1)$. There are constants c_1, c_2, c_3, and c_4 such that

$$\frac{x^2}{x^4-1} = \frac{c_1}{x+1} + \frac{c_2}{x-1} + \frac{c_3 x + c_4}{x^2+1}.$$

Clear the denominator, but do not expand the right-hand side:

$$x^2 = c_1(x-1)(x^2+1) + c_2(x+1)(x^2+1) + (c_3 x + c_4)(x-1)(x+1). \tag{8.4.5}$$

Instead, substitute $x = 1$ and $x = -1$ into (8.4.5) to obtain, respectively:

$$\begin{aligned} 1 &= 0 + 4c_2 + 0 && \text{(substituting } x = 1 \text{ into (8.4.5))} \\ 1 &= -4c_1 + 0 + 0 && \text{(substituting } x = -1 \text{ into (8.4.5)).} \end{aligned}$$

Already we see that $c_1 = -1/4$ and $c_2 = 1/4$.

Setting $x = 0$ compares the constant terms on both sides of (8.4.5).

Next, substitute 0 for x into (8.4.5), obtaining

$$0 = -c_1 + c_2 - c_4 \quad \text{(substituting } x = 0 \text{ in (8.4.5)).}$$

Hence $c_4 = 1/2$.

It remains to determine c_3. We could substitute another number, say $x = 2$, or compare coefficients in (8.4.5). Let us compare coefficients of the highest degree, x^3. Without multiplying (8.4.5) out in full, we can read off the coefficient of x^3 on both sides by sight, getting

$$0 = c_1 + c_2 + c_3.$$

Since $c_1 = -1/4$, and $c_2 = 1/4$ it follows that $c_3 = 0$. Hence

$$\frac{x^2}{x^4 - 1} = \frac{-1/4}{x+1} + \frac{1/4}{x-1} + \frac{1/2}{x^2+1}.$$

◆

Example 4 used a combination of two methods: substituting convenient values of x and equating coefficients. We could have compared coefficients, giving equations corresponding to the constant terms and to each power of x from x^1 up to x^3. That would give four equations in four unknowns. The Exercises suggest how to solve such equations, if they must be solved by hand.

Summary

To integrate a rational function requires expressing the integrand in its partial fraction representation.

The first step is to check that the integrand is a proper rational function, that is, the degree of the numerator is less than the degree of the denominator. If the integrand is not proper, use long division to express the integrand as the sum of a polynomial and a proper rational function. A flowchart in Figure 8.4.1 summarizes the process.

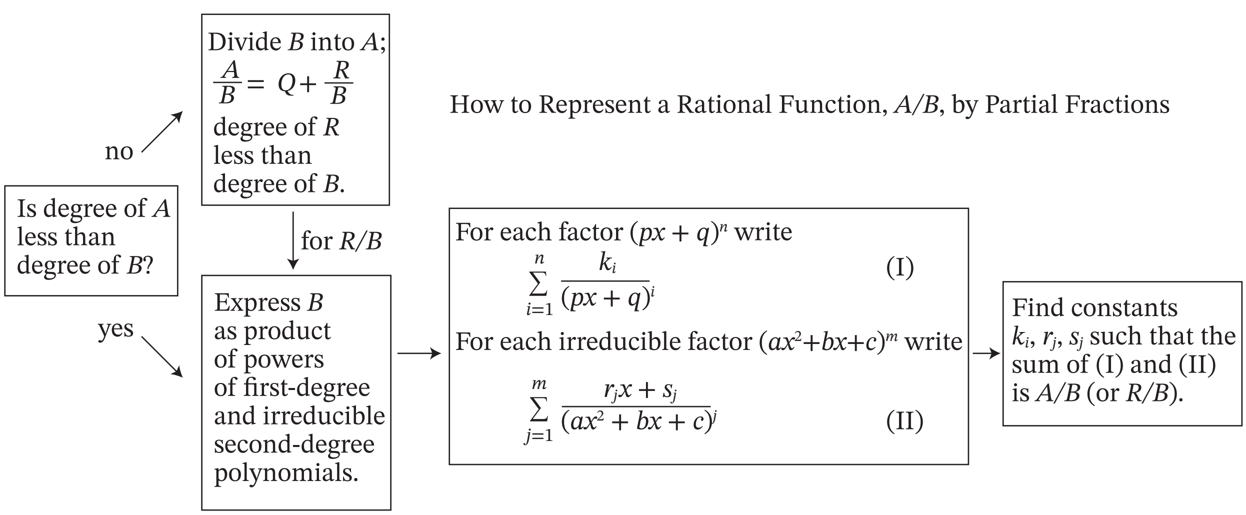


Figure 8.4.1

EXERCISES for Section 8.4

In Exercises 1 to 10 give the form of the partial fraction representation. Each expression is already proper.

1. $\dfrac{3x^3+5x+2}{(x-1)(x-2)(x-3)(x-4)}$

2. $\dfrac{x^2-5x+3}{(x+1)^2(2x+3)}$

3. $\dfrac{2x^2+x+1}{(x+1)^3}$

4. $\dfrac{3x}{(x+1)(2x+2)}$

5. $\dfrac{x^2-x+3}{(x+1)(x^2+1)}$

6. $\dfrac{2x^2+3x+5}{(x-1)(x^2+x+1)}$

7. $\dfrac{5x^3+x^2+1}{(x^2+x+1)^2}$

8. $\dfrac{x^3+x+1}{(x^2+x+1)^3}$

9. $\dfrac{x^2+x+2}{x^3-x}$

10. $\dfrac{x^2+x+2}{x^4-1}$

---

11. The rational function $\dfrac{1}{a^2-x^2}$, where a is a nonzero constant, appears in many applications. Find its partial fraction representation.

Exercises 12 to 15 contain improper rational functions. Express each as the sum of a polynomial and a proper rational function.

12. $\dfrac{x^2}{x^2+x+1}$

13. $\dfrac{x^3}{(x+1)(x+2)}$

14. $\dfrac{x^5-2x+1}{(x+1)(x^2+1)}$

15. $\dfrac{x^5+x}{(x+1)^2(x-2)}$

In Exercises 16 to 19 find the partial fraction representation.

16. $\dfrac{5}{x^2-1}$

17. $\dfrac{x+3}{(x+1)(x+2)}$

18. $\dfrac{1}{(x-1)^2(x+2)}$

19. $\dfrac{6x^2-2}{(x-1)(x-2)(2x-3)}$

---

20. Show that $\dfrac{5e^{3x}+2e^{2x}+e^x+6}{e^{2x}+e^x+5}$ has an elementary antiderivative, but do not find it.

21. Solve Example 3 by clearing the denominator in (8.4.2) to get $1 = k_1(x+3)+k_2(2x+1)$. Replace x by any number you please. That gives an equation in k_1 and k_2. Then replace x by another number of your choice, to obtain a second equation in k_1 and k_2. Solve the equations. Why are $x=-3$ and $x=-1/2$ the nicest choices?

22. Express each polynomial as a product of first degree polynomials.
(a) x^2+2x+1, (b) x^2+5x-3, (c) x^2-4x-6, and (d) $2x^2+3x-4$.

In Exercises 23 to 28 determine if the polynomial is irreducible.

23. $3x^2+2x+1$

24. $4x^2+4x+1$

25. $x^2+2x+10$

26. $x^2-2x+10$

27. x^2+x+1

28. $2x^2+4x+1$

---

29. For which values of b are the following expression reducible? irreducible? (a) $3x^2+bx+2$ (b) $3x^2+bx-2$

30. For which values of c are the following expression reducible? irreducible? (a) $3x^2+5x+c$ (b) $3x^2-5x+c$

In Exercises 31 to 40 find the partial fraction representation of the rational function.

31. $\dfrac{5x^2-x-1}{x^2(x-1)}$

32. $\dfrac{2x^2+3}{x(x+1)(x+2)}$

33. $\dfrac{5x^2-2x-2}{x(x^2-1)}$

34. $\dfrac{5x^2+9x+6}{(x+1)(x^2+2x+2)}$

35. $\dfrac{5x^2+2x+3}{x(x^2+x+1)}$

36. $\dfrac{x^3-3x^2+3x-3}{x^2-3x+2}$

37. $\dfrac{3x^3+2x^2+3x+1}{x(x^2+1)}$

38. $\dfrac{x^5+2x^4+4x^3+2x^2+x-2}{x^4-1}$

39. $\dfrac{5x^2+6x+10}{(x-2)(x^2+3x+4)}$

40. $\dfrac{3x^2-x-2}{(x+1)(2x^2+x+1)}$

---

41. **Sam:** I found this formula in my integral tables: $\displaystyle\int \frac{dx}{a^2-b^2x^2} = \frac{1}{2ab}\ln\left|\frac{a+bx}{a-bx}\right|$ (a, b constants).
Jane: What's your point?
Sam: My instructor said you won't get any logs other than logs of linear and quadratic polynomials.
Jane: Maybe the table is wrong.
Sam: I took the derivative. It's correct. Can I sue my instructor for misleading the young?

Does Sam have a case against his instructor? Explain.

We did not discuss the problem of factoring a polynomial $B(x)$ into linear and irreducible quadratic polynomials. Exercises 42 to 45 concern this problem when the degree of $B(x)$ is 2.

42. (a) Show that if $b^2-4ac>0$, then $ax^2+bx+c=a(x-r_1)(x-r_2)$, where r_1 and r_2 are the distinct roots of $ax^2+bx+c=0$.
(b) Show that if $b^2-4ac=0$, then $ax^2+bx+c=a(x-r)(x-r)$, with r the only root of $ax^2+bx+c=0$.

These two results show that if $b^2-4ac\geq 0$, then ax^2+bx+c is reducible. Compare with Exercise 43.

43. (a) Show that if ax^2+bx+c is reducible, then it can be written in the form $a(x-s_1)(x-s_2)$ for some real numbers s_1 and s_2.
(b) Deduce that s_1 and s_2 are the roots of $ax^2+bx+c=0$.
(c) Deduce that $b^2-4ac\geq 0$.

This exercises shows that if ax^2+bx+c is reducible, then $b^2-4ac\geq 0$. Compare with Exercise 42.

44. Factor (a) x^2+6x+5, (b) x^2-5, and (c) $2x^2+6x+3$.

45. (a) Show that x^2+3x-5 is reducible.
(b) Using (a), find $\displaystyle\int \frac{dx}{x^2+3x-5}$ by partial fractions.
(c) Find $\displaystyle\int \frac{dx}{x^2+3x-5}$ by using an integral table.

---

46. Compute as simply as possible. (a) $\displaystyle\int \frac{x^3}{x^4+1}\,dx$, (b) $\displaystyle\int \frac{x}{x^4+1}\,dx$, and (c) $\displaystyle\int \frac{dx}{x^4+1}$.

47. Show that a rational function of e^x has an elementary antiderivative. That is, any function of the form $\dfrac{P(e^x)}{Q(e^x)}$ where P and Q are polynomials has an elementary antiderivative.

48. If ax^2+bx+c is irreducible must ax^2-bx+c also be irreducible? Must ax^2+bx-c? See Exercises 29 and 30.

49. Explain why every polynomial of odd degree has at least one linear factor.

NOTE: Therefore, every polynomial of odd degree greater than one is reducible.

50. In arithmetic every fraction can be written as an integer plus a proper fraction. For instance, $\frac{25}{3} = 8 + \frac{1}{3}$. Why?

51. In arithmetic, the analog of the partial fraction representation is this: Every fraction can be written as the sum of an integer and fractions of the form c/p^n, where p is a prime and $|c|$ is less than p. Check that this is true for $\frac{53}{18}$.

52. Let a be a solution of the equation $P(x) = 0$, where $P(x)$ is a polynomial. Prove that $x - a$ must be a factor of $P(x)$. In other words, show why the remainder in long division is 0 when $P(x)$ is divided by $x - a$.

NOTE: This is the basis for the Factor Theorem.

53. (a) Use the quadratic formula to find the roots of $x^2 + 7x + 9 = 0$.
(b) With the aid of the Factor Theorem (Exercise 52), write x^2+7x+9 as the product of two linear polynomials.
(c) Check the factorization found in (b) by multiplying it out.

54. Assume $x - c$ is a factor of $Q(x)$ and not of $P(x)$. Also assume $(x-c)^2$ is not a factor of $Q(x)$. The term $\frac{A}{x-c}$ therefore appears in the partial fraction representation of $\frac{P(x)}{Q(x)}$. Show that $A = \frac{P(c)}{Q'(c)}$.

Historical Note: Integrating in the Real World

As noted in Chapter 6, sometimes computers and calculators can give the exact (symbolic) value of a definite integral by first finding an antiderivative. In practical applications, however, formal antidifferentiation is not that important.

Say that you wanted to compute the definite integral

$$\int_1^2 \frac{x+3}{x^3 + x^2 + x + 2}\, dx.$$

One approach is to find an antiderivative of each term in the partial fraction representation of the integrand. But modern computational tools can evaluate it accurately to as many decimal places as we may want.

For example, Simpson's rule with only 8 sections gives 0.514393 as an approximate value with an error less than 0.00001.

In other situations some of the coefficients in either the numerator or denominator of the integrand may be given as decimal approximations. Then it often is easier and more appropriate to use a computational method to obtain a numerical answer.

8.5 Special Techniques

This chapter has presented three techniques for computing integrals. The first, substitution, and the second, integration by parts, are used most often. The method of partial fractions applies to special rational functions and is used in solving some differential equations (see Chapter 13). In this section we compute some special integrals such as $\int \sin(mx)\cos(nx)\, dx$, $\int \sin^2(\theta)\, d\theta$, and $\int \sec(\theta)\, d\theta$ that arise in numerous applications. We also describe substitutions for several special classes of integrands. Additional classes of integrands are explored in the Exercises.

Computing $\int \sin(mx)\sin(nx)\, dx$ (m and n are positive integers)

The integrals

$$\int \sin(mx)\sin(nx)\, dx, \qquad \int \cos(mx)\sin(nx)\, dx, \qquad \text{and} \qquad \int \cos(mx)\cos(nx)\, dx$$

are needed in the study of Fourier series (discussed in Section 12.7) , an important tool in the study of heat, sound, and signal processing. They can be computed with the aid of the identities

$$\begin{aligned}
\sin(A)\sin(B) &= \frac{1}{2}\cos(A-B) - \frac{1}{2}\cos(A+B)\\
\sin(A)\cos(B) &= \frac{1}{2}\sin(A+B) + \frac{1}{2}\sin(A-B)\\
\cos(A)\cos(B) &= \frac{1}{2}\cos(A-B) + \frac{1}{2}\cos(A+B).
\end{aligned}$$

NOTE: These identities can be checked using the identities for $\sin(A \pm B)$ and $\cos(A \pm B)$.

EXAMPLE 1. Find $\displaystyle\int_0^{\pi/4} \sin(3x)\sin(2x)\,dx$.

SOLUTION For this problem we use the first identity, with $A = 3x$ and $B = 2x$, as follows:

$$\begin{aligned}
\int_0^{\pi/4} \sin(3x)\sin(2x)\,dx &= \int_0^{\pi/4} \left(\frac{1}{2}\cos(x) - \frac{1}{2}\cos(5x)\right)dx && \text{(trigonometric identity)}\\
&= \left(\frac{1}{2}\sin(x) - \frac{1}{10}\sin(5x)\right)\Bigg|_0^{\pi/4} && \text{(FTC I)}\\
&= \left(\frac{\sqrt{2}}{4} + \frac{\sqrt{2}}{20}\right) - \left(\frac{0}{2} - \frac{0}{10}\right) && \text{(evaluate at endpoints, and subtract)}\\
&= \frac{3\sqrt{2}}{10} \approx 0.424.
\end{aligned}$$

◆

Computing $\int \sin^2(x)\,dx$ and $\int \cos^2(x)\,dx$

The integrals can be computed with the aid of the identities

$$\sin^2(x) = \frac{1-\cos(2x)}{2} \qquad \text{and} \qquad \cos^2(x) = \frac{1+\cos(2x)}{2}.$$

Note that these **double-angle formulas** are just special cases of the first and third identity when $A = B = x$.

EXAMPLE 2. Find an antiderivative of $\sin^2(x)$.

SOLUTION Once the double-angle formula for $\sin^2(x)$ is applied to the integrand, it is straightforward to find an antiderivative of $\sin^2(x)$. Here are all of the steps:

$$\begin{aligned}
\int \sin^2(x)\,dx &= \int \frac{1-\cos(2x)}{2}\,dx && \text{(double-angle formula)}\\
&= \int \frac{dx}{2} - \int \frac{\cos(2x)}{2}\,dx && \text{(linearity of antiderivatives)}\\
&= \frac{x}{2} - \frac{\sin(2x)}{4} + C && \text{(FTC I).}
\end{aligned}$$

◆

Computing $\int \tan(\theta)\, d\theta$ and $\int \tan^2(\theta)\, d\theta$

EXAMPLE 3. Find $\int \tan(\theta)\, d\theta$.

SOLUTION We rewrite the integrand in a form where the trigonometric functions can be eliminated with a substitution. In the present case, this is accomplished by writing $\tan(\theta) = \sin(\theta)/\cos(\theta)$ and using the substitution $u = \cos(\theta)$ and $du = -\sin(\theta)\, d\theta$ as follows:

$$\int \tan(\theta)\, d\theta = \int \frac{\sin(\theta)}{\cos(\theta)}\, d\theta = \int \frac{-du}{u} = -\ln|u| + C = -\ln|\cos(\theta)| + C.$$

◆

Observation 8.5.1: *Another Antiderivative of* $\tan(x)$

Most integral tables have a different antiderivative of $\tan(x)$:

$$\int \tan(\theta)\, d\theta = \ln|\sec(\theta)| + C.$$

Exercise 50 shows this formula for the antiderivative of $\tan(\theta)$ is consistent with Example 2.

Remember that it is sometimes easier to find an antiderivative of an expression that is more complicated. For example, finding $\int \tan^2(\theta)\, d\theta$ is easier than finding $\int \tan(x)\, dx$. Using the identity $\tan^2(\theta) = \sec^2(\theta) - 1$, we obtain

$$\int \tan^2(\theta)\, d\theta = \int (\sec^2(\theta) - 1)\, d\theta = \tan(\theta) - \theta + C.$$

The next integral is critical to the understanding Mercator maps, which will be discussed in the CIE about Mercator maps, CIE 12 at the end of Chapter 9.

Computing $\int \sec(\theta)\, d\theta$

EXAMPLE 4. Find $\int \sec(\theta)\, d\theta$, assuming $0 \le \theta < \pi/2$.

SOLUTION We use trigonometric defintions and identities to rewrite the integrand in a form where substitution can be used:

$$\int \sec(\theta)\, d\theta = \int \frac{1}{\cos(\theta)}\, d\theta = \int \frac{\cos(\theta)}{\cos^2(\theta)}\, d\theta = \int \frac{\cos(\theta)}{1 - \sin^2(\theta)}\, d\theta.$$

The substitution $u = \sin(\theta)$ and $du = \cos(\theta)\, d\theta$ transforms this last integral into the integral of a rational function:

$$\begin{aligned}
\int \frac{du}{1-u^2} &= \frac{1}{2} \int \left(\frac{1}{1+u} + \frac{1}{1-u} \right) du && \text{(partial fraction representation)} \\
&= \frac{1}{2} \left(\ln(1+u) - \ln(1-u) \right) + C && \text{(FTC I)} \\
&= \frac{1}{2} \ln\left(\frac{1+u}{1-u} \right) + C && \text{(property of logarithms).}
\end{aligned}$$

NOTE: Absolute values are not needed here because $(1+u)/(1-u)$ is positive for $-1 < u < 1$.

Since $u = \sin(\theta)$,

$$\frac{1}{2}\ln\left(\frac{1+u}{1-u}\right) = \frac{1}{2}\ln\left(\frac{1+\sin(\theta)}{1-\sin(\theta)}\right).$$

Thus,

$$\int \sec(\theta)\ d\theta = \frac{1}{2}\ln\left(\frac{1+\sin(\theta)}{1-\sin(\theta)}\right) + C.$$

◆

Observation 8.5.2: *Another Antiderivative of* $\sec(\theta)$

Most integral tables have a different antiderivative of $\sec(\theta)$:

$$\int \sec(\theta)\ d\theta = \ln|\sec(\theta) + \tan(\theta)| + C.$$

Exercise 51 shows that this formula for the antiderivative of $\sec(\theta)$ is consistent with Example 4.

In contrast to Example 4, $\int \sec^2(\theta)\ d\theta$ is easy, since it is simply $\tan(\theta) + C$.

The Substitution $u = \sqrt{ax+b}$

Another large class of integrands for which antiderivatives can be found are those that involve $\sqrt{ax+b}$. The next example shows a typical example using the substitution $u = \sqrt{ax+b}$. We then describe some integrands for which this substitution is appropriate.

EXAMPLE 5. Find $\displaystyle\int_4^7 x^2\sqrt{3x+4}\ dx$.

SOLUTION Let $u = \sqrt{3x+4}$, so $u^2 = 3x+4$. Then $x = (u^2-4)/3$ and $dx = (2u/3)\ du$. As x assumes values between 4 and 7, u ranges from $\sqrt{3(4)+4} = \sqrt{16} = 4$ to $\sqrt{3(7)+4} = \sqrt{25} = 5$. Thus

$$\begin{aligned}
\int_4^7 x^2\sqrt{3x+4}\ dx &= \int_4^5 \underbrace{\left(\frac{u^2-4}{3}\right)^2}_{x^2}\ \underbrace{u}_{\sqrt{3x+4}}\ \underbrace{\frac{2u}{3}\ du}_{dx} && \text{(rationalizing substitution: } u=\sqrt{3x+4}\text{)}\\
&= \frac{2}{27}\int_4^5 (u^2-4)^2u^2\ du && \text{(algebraic manipulations)}\\
&= \frac{2}{27}\int_4^5 (u^6-8u^4+16u^2)\ du && \text{(algebraic expansion)}\\
&= \frac{1,214,614}{2,835} && \text{(FTC I)}\\
&= \frac{1,214,614}{2,835} \approx 428.435.
\end{aligned}$$

Some arithmetic details have been omitted, but are easily reproducible by the interested reader. Note that the denominator of the exact answer is the product of 27, 7, 5, and 3: $2835 = 27\cdot 7\cdot 5\cdot 3$. ◆

The substitution $u = \sqrt[n]{ax+b}$ will be useful in the search for an antiderivative of any "rational function of x and $\sqrt[n]{ax+b}$" where the integer n is greater than or equal to 2. Example 5 illustrated this substitution with $n = 2$.

Definition: *Polynomial and Rational Functions in Two Variables*

A **polynomial in two variables** x **and** y is a sum of terms of the form $cx^m y^n$, where m and n are nonnegative integers and c is a real number. For instance, he expression $2x^3 - \sqrt{2}xy^7 + xy$ is a polynomial in x and y.

The quotient of two polynomials of x and y is called a **rational function of** x **and** y.

For example, the integrand in the Example 5 is a rational function of x and $\sqrt{3x+4}$.

Let $R(x, y)$ be a rational function of x and y. Let n be an integer greater than or equal to 2. Replacing y by $\sqrt[n]{ax+b}$ creates what is called a "rational function of x and $\sqrt[n]{ax+b}$," denoted $R(x, \sqrt[n]{ax+b})$. For instance, if

$$R(x, y) = \frac{x+y^2}{2x-y},$$

then replacing y by $\sqrt[3]{4x+5}$ yields

$$R(x, \sqrt[3]{4x+5}) = \frac{x+\sqrt[3]{4x+5}^2}{2x-\sqrt[3]{4x+5}},$$

a rational function of x and $\sqrt[3]{4x+5}$.

In general, to integrate $R(x, \sqrt[n]{ax+b})$, let $u = \sqrt[n]{ax+b}$. Then $u^n = ax+b$, $x = (u^n - b)/a$ and $dx = nu^{n-1}\, du/a$. The integrand is now a rational function of u that can be evaluated by partial fractions, if necessary.

There are many other substitutions that can be used to evaluate indefinite or definite integrals. We conclude this section by considering four specific forms that arise with some regularity. It is good to be able to recognize these forms, and to know how to find the corresponding antiderivative.

Three Trigonometric Substitutions

A rational function of x and $\sqrt{a^2-x^2}$, where a is a constant, is transformed into a rational function of $\cos(\theta)$ and $\sin(\theta)$ by the substitution $x = a\sin(\theta)$. Similar substitutions are possible for integrands involving $\sqrt{a^2+x^2}$ or $\sqrt{x^2-a^2}$. In each case, one of the trigonometric identities $1-\sin^2(\theta) = \cos^2(\theta)$, $\sec^2(x) = \tan^2(\theta)+1$, or $\sec^2(\theta)-1 = \tan^2(\theta)$ converts a sum or difference of squares into a perfect square.

Observation 8.5.3: *Three Trigonometric Substitutions*

To find an antiderivative when the integrand is a rational function in one of the following forms:

Case 1: $R(x, \sqrt{a^2-x^2})$:
let $x = a\sin(\theta)$, then $dx = a\cos(\theta)$, and $\sqrt{a^2-x^2} = a\cos(\theta)$ $\quad$ ($a > 0$, $-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$)

Case 2: $R(x, \sqrt{a^2+x^2})$:
let $x = a\tan(\theta)$, then $dx = a(\sec(\theta))^2$, and $\sqrt{a^2+x^2} = a\sec(\theta)$ $\quad$ ($a > 0$, $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$)

Case 3: $R(x, \sqrt{x^2-a^2})$:
let $x = a\sec(\theta)$, then $dx = a\sec(\theta)\tan(\theta)$, and $\sqrt{x^2-a^2} = \pm a\tan(\theta)$ $\quad$ ($a > 0$, $0 \le \theta \le \pi$, $\theta \ne \frac{\pi}{2}$)

To derive the result in, for example, Case 1, observe that replacing x in $\sqrt{a^2-x^2}$ by $a\sin(\theta)$ leads to

$$\sqrt{a^2-x^2} = \sqrt{a^2-(a\sin(\theta))^2} = \sqrt{a^2(1-\sin^2(\theta))} = \sqrt{a^2\cos^2(\theta)} = a\cos(\theta).$$

The assumption that a is positive and θ is in the first or fourth quadrant allows the square root to be simplified without any absolute values.

In Case 3, $a > 0$ is assumed, but x can be positive or negative. Whenever $x < 0$, θ is an angle in the second quadrant, hence $\tan(\theta) < 0$ and thus, since $\sqrt{c^2} = -c$ when $c < 0$,

$$\sqrt{x^2-a^2} = \sqrt{(a\sec(\theta))^2-a^2} = a\sqrt{\sec^2(\theta)-1} = a\sqrt{\tan^2(\theta)} = a(-\tan(\theta)).$$

In the following Examples and in several Exercises involving indefinite integrals with Case 3 integrands, it will be assumed that x varies through nonnegative values, so that θ remains in the first quadrant; in these cases $\sqrt{\sec^2(\theta)-1}=\tan(\theta)$.

For definite integrals involving square roots (or an n^{th} root where n is an even integer) it is necessary to ensure that the argument of the square root is not negative. Here is how this goes for the three cases considered here.

Case 1: for $\sqrt{a^2-x^2}$ to be meaningful $|x|$ must be no larger than a: $-a \le x \le a$.

Case 2: the quantity $\sqrt{a^2+x^2}$ is meaningful for all values of x: $-\infty < x < \infty$.

Case 3: for $\sqrt{x^2-a^2}$ to be meaningful $|x|$ must be at least as large as a: $x > a$ or $x < -a$.

NOTE: Instead of memorizing these results, remember the general process and apply it to the specific integrand in each problem.

EXAMPLE 6. Compute $\int \sqrt{1+x^2}\,dx$.

SOLUTION The appearance of $\sqrt{1+x^2}$ in the integrand make this a Case 2 problem. With the substitution $x=\tan(\theta)$, the integrand simplifies as follows: $\sqrt{1+\tan^2(\theta)}=\sec(\theta)$ and $dx=\sec^2(\theta)\,d\theta$. (Figure 8.5.1 shows the geometry.) Thus

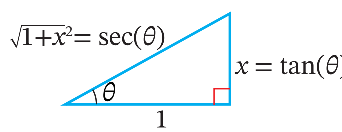


Figure 8.5.1

$$\int \sqrt{1+x^2}\,dx=\int \sec(\theta)\sec^2(\theta)\,d\theta=\int \sec^3(\theta)\,d\theta.$$

By Formula 67 from the Table of Integrals (in Appendix A),

$$\int \sec^3(\theta)\,d\theta=\frac{\sec(\theta)\tan(\theta)}{2}+\frac{1}{2}\ln|\sec(\theta)+\tan(\theta)|+C.$$

To express the antiderivative in terms of x rather than θ, we need to express $\tan(\theta)$ and $\sec(\theta)$ in terms of x. Starting with the definition $x=\tan(\theta)$, we find $\sec(\theta)$ by means of the relation $\sec(\theta)=\sqrt{1+\tan^2(\theta)}=\sqrt{1+x^2}$, as in Figure 8.5.1. Thus

$$\int \sqrt{1+x^2}\,dx=\frac{x\sqrt{1+x^2}}{2}+\frac{1}{2}\ln\left(\sqrt{1+x^2}+x\right)+C.$$

◆

EXAMPLE 7. Compute $\int_4^5 \frac{dx}{\sqrt{x^2-9}}$.

SOLUTION The appearance of $\sqrt{x^2-9}$ in the integrand make this a Case 3 problem. Let $x=3\sec(\theta)$, which implies that $dx=3\sec(\theta)\tan(\theta)\,d\theta$. (See Figure 8.5.2.) As x goes from 4 to 5, the equation $x=3\sec(\theta)$ shows that $\sec(\theta)$ goes from $4/3$ to $5/3$. Thus θ goes from $\alpha=\operatorname{arcsec}(4/3)$ to $\beta=\operatorname{arcsec}(5/3)$. Thus,

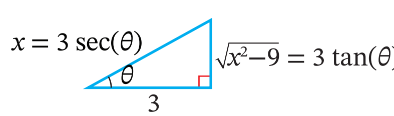


Figure 8.5.2

$$\begin{aligned}
\int_4^5 \frac{dx}{\sqrt{x^2-9}} &= \int_\alpha^\beta \frac{3\sec(\theta)\tan(\theta)\,d\theta}{\sqrt{9\sec^2(\theta)-9}} = \int_\alpha^\beta \frac{\sec(\theta)\tan(\theta)\,d\theta}{\tan(\theta)} && \text{(substitution)}\\
&= \int_\alpha^\beta \sec(\theta)\,d\theta && \text{(simplification)}\\
&= \ln|\sec(\theta)+\tan(\theta)|\Big|_\alpha^\beta && \text{(FTC I)}\\
&= \ln\left|\sec(\beta)+\tan(\beta)\right| - \ln|\sec(\alpha)+\tan(\alpha)|\\
&= \ln\left(\frac{5}{3}+\frac{4}{3}\right) - \ln\left(\frac{4}{3}+\frac{\sqrt{7}}{3}\right) && \text{(see Figure 8.5.3(a) and (b))}\\
&= \ln(3) - \ln\left(\frac{4+\sqrt{7}}{3}\right) && \text{(simplification of fractions)}\\
&= \ln\left(\frac{9}{4+\sqrt{7}}\right) \approx 0.30325 && \text{(property of logarithms).}
\end{aligned}$$

♦

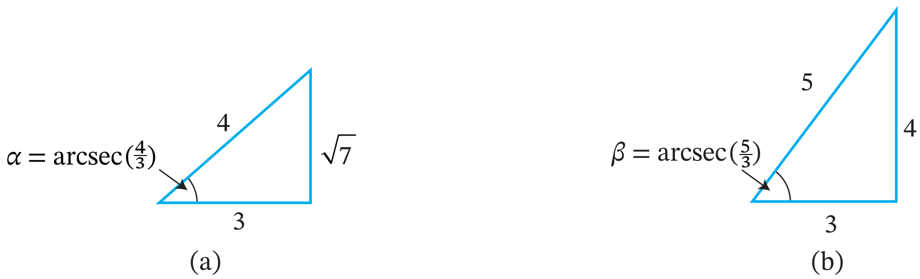


Figure 8.5.3

A Half-Angle Substitution for $R(\cos(\theta), \sin(\theta))$

A rational function of $\cos(\theta)$ and $\sin(\theta)$ is transformed into a rational function u by the substitution $u = \tan(\theta/2)$. This is sometimes useful after a trigonometric substitution has been used, leaving the integrand in terms of $\cos(\theta)$ and $\sin(\theta)$. The substitution with $u = \tan(\theta/2)$ then yields an integral that can be treated by partial fractions. (See Exercises 60 and 61.)

Summary

In this section, we discussed how to evaluate some special integrals and integration techniques. These were applied to find antiderivatives for products and powers of trigonometric functions, in particular:

- $\int \sin(mx)\sin(nx)\,dx$, $\int \sin(mx)\cos(nx)\,dx$, and $\int \cos(mx)\cos(nx)\,dx$;
- $\int \sin^2(x)\,dx$ and $\int \cos^2(x)\,dx$; and
- $\int \sec(\theta)\,d\theta$, $\int \tan(\theta)\,d\theta$, and $\int \tan^2(\theta)\,d\theta$.

The integration of higher powers of trigonometric functions is discussed in the Exercises.

We also pointed out that the substitution $u = \sqrt{ax+b}$ transforms an integrand that is a rational function of x and $\sqrt{ax+b}$, $R(x, \sqrt{ax+b})$, into a rational function of u. Other rational functions, $R(x, \sqrt{a^2-x^2})$, $R(x, \sqrt{x^2-a^2})$ and $R(x, \sqrt{a^2+x^2})$, can be transformed into rational functions of $\cos(\theta)$ and $\sin(\theta)$ by one of three trigonometric substitutions. And, as shown in Exercises 60 to 62, the substitution $u = \tan(\theta/2)$, transforms the integrand from a rational function of $\sin(\theta)$ and $\cos(\theta)$, $R(\cos(\theta), \sin(\theta))$, into a rational function of u, which can then be evaluated by partial fractions.

EXERCISES for Section 8.5

Use the ideas introduced in Examples 1 to 4 to find the requested antiderivative in Exercises 1 to 14.

1. $\displaystyle\int \sin(5x)\sin(3x)\,dx$

2. $\displaystyle\int \sin(5x)\cos(2x)\,dx$

3. $\displaystyle\int \cos(3x)\sin(2x)\,dx$

4. $\displaystyle\int \cos(2\pi x)\sin(5\pi x)\,dx$

5. $\displaystyle\int \sin^2(3x)\,dx$

6. $\displaystyle\int \cos^2(5x)\,dx$

7. $\displaystyle\int (3\sin(2x) + 4\sin^2(5x))\,dx$

8. $\displaystyle\int (5\cos(2x) + \cos^2(7x))\,dx$

9. $\displaystyle\int (3\sin^2(\pi x) + 4\cos^2(\pi x))\,dx$

10. $\displaystyle\int \sec(3\theta)\,d\theta$

11. $\displaystyle\int \tan(2\theta)\,d\theta$

12. $\displaystyle\int \sec^2(4x)\,dx$

13. $\displaystyle\int \tan^2(5x)\,dx$

14. $\displaystyle\int \frac{dx}{\cos^2(3x)}$

---

15. Show that $\sin(A)\sin(B) = \dfrac{1}{2}\cos(A-B) - \dfrac{1}{2}\cos(A+B)$ for all values of A and B.

16. Show that $\sin(A)\cos(B) = \dfrac{1}{2}\sin(A+B) + \dfrac{1}{2}\sin(A-B)$ for all values of A and B.

Exercises 17 to 19 develop the formulas that are the foundation for Fourier series, discussed in more detail in Section 12.7. Assume m and k are distinct positive integers and L is a positive number.

17. Show that (a) $\displaystyle\int_{-L}^{L} \sin\left(\frac{k\pi x}{L}\right)\sin\left(\frac{k\pi x}{L}\right)\,dx = L$ and (b) $\displaystyle\int_{-L}^{L} \sin\left(\frac{k\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)\,dx = 0$.

18. Show that (a) $\displaystyle\int_{-L}^{L} \cos\left(\frac{k\pi x}{L}\right)\cos\left(\frac{k\pi x}{L}\right)\,dx = L$ and (b) $\displaystyle\int_{-L}^{L} \cos\left(\frac{k\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)\,dx = 0$.

19. Show that $\displaystyle\int_{-L}^{L} \sin\left(\frac{k\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)\,dx = 0$.

In Exercises 20 to 29, use the substitution $u = \sqrt[n]{ax+b}$ to evaluate each integral.

20. $\displaystyle\int x^2\sqrt{2x+1}\,dx$

21. $\displaystyle\int \frac{x^2\,dx}{\sqrt[3]{x+1}}$

22. $\displaystyle\int \frac{dx}{\sqrt{x+3}}$

23. $\displaystyle\int \frac{\sqrt{2x+1}}{x}\,dx$

24. $\displaystyle\int x\sqrt[3]{3x+2}\,dx$

25. $\displaystyle\int \frac{\sqrt{x}+3}{\sqrt{x}-2}\,dx$

26. $\displaystyle\int \frac{x\,dx}{\sqrt{x}+3}$

27. $\displaystyle\int x(3x+2)^{5/3}\,dx$

28. $\displaystyle\int \frac{dx}{\sqrt[3]{x}+\sqrt{x}}$

29. $\displaystyle\int (x+2)\sqrt[5]{x-3}\,dx$

In Exercises 30 to 40, find the integrals using trigonometric substitutions. (Assume a is a positive constant.)

30. $\int \sqrt{4-x^2}\,dx$

31. $\int x^3\sqrt{1-x^2}\,dx$

32. $\int \frac{dx}{\sqrt{9+x^2}}$

33. $\int \frac{x^2\,dx}{\sqrt{x^2-9}}$

34. $\int \frac{\sqrt{4+x^2}}{x}\,dx$

35. $\int \frac{dx}{\sqrt{a^2-x^2}}$

36. $\int \sqrt{a^2-x^2}\,dx$

37. $\int \sqrt{a^2+x^2}\,dx$

38. $\int \sqrt{a^2-x^2}\,dx$

39. $\int \frac{dx}{\sqrt{25x^2-16}}$

40. $\int_{\sqrt{2}}^{2} \sqrt{x^2-1}\,dx$

Exercises 41 to 44 concern the recursion formulas for $\int \tan^n(\theta)\,d\theta$ and $\int \sec^n(\theta)\,d\theta$. The simplest cases, $n=1$ and $n=2$, were discussed earlier in Examples 3 and 4, and adjoining observations.

41. Obtain the recursion formula $\int \tan^n(\theta)\,d\theta = \frac{\tan^{n-1}(\theta)}{n-1} - \int \tan^{n-2}(\theta)\,d\theta$ by writing $\tan^n(\theta) = \tan^{n-2}(\theta)\tan^2(\theta)$ and then writing $\tan^2(\theta)$ as $\sec^2(\theta)-1$.

42. Obtain the recursion formula $\int \sec^n(\theta)\,d\theta = \frac{\sec^{n-2}(\theta)\tan(\theta)}{n-1} + \frac{n-2}{n-1}\int \sec^{n-2}(\theta)\,d\theta$ by writing $\sec^n(\theta)$ as $\sec^{n-2}(\theta)\sec^2(\theta)$ and integrating by parts. After the integration, replace $\tan^2(\theta)$ with $\sec^2(\theta)-1$ in the integrand.

43. Evaluate (a) $\int \tan^3(\theta)\,d\theta$ and (b) $\int \tan^4(\theta)\,d\theta$.

44. Evaluate (a) $\int \sec^3(\theta)\,d\theta$, (b) $\int \frac{d\theta}{\cos^4(\theta)}$, and (c) $\int \sec^2(2x)\,dx$..

---

45. Evaluate (a) $\int \csc(\theta)\,d\theta$ and (b) $\int \csc^2(\theta)\,d\theta$.

46. Evaluate (a) $\int \cot(\theta)\,d\theta$ and (b) $\int \cot^2(\theta)\,d\theta$.

---

47. To evaluate $\int \sin^n(\theta)\cos^m(\theta)\,d\theta$, where m and n are nonnegative integers and m is odd, write the integral as $\int \sin^n(\theta)\cos^{m-1}(\theta)\cos(\theta)\,d\theta$. Then, replace $\cos^{m-1}(\theta)$ with $(1-\sin^2(\theta))^{(m-1)/2}$ and use the substitution $u=\sin(\theta)$.

Using this technique, find (a) $\int \sin^3(\theta)\cos^3(\theta)\,d\theta$, (b) $\int_0^{\pi/2} \sin^4(\theta)\cos^3(\theta)\,d\theta$, and (c) $\int \cos^5(\theta)\,d\theta$.

48. How would you integrate $\int \sin^n(\theta)\cos^m(\theta)\,d\theta$, where m and n are nonnegative integers and n is odd? Illustrate your technique by three examples. (See Exercise 47.)

49. The techniques in Exercises 47 and 48 apply to $\int \sin^n(\theta)\cos^m(\theta)\,d\theta$ only when at least one of m and n is odd. If both are even, first use the identities $\sin^2(\theta) = \frac{1}{2}(1-\cos(2\theta))$ and $\cos^2(\theta) = \frac{1}{2}(1+\cos(2\theta))$, giving a polynomial in $\cos(2\theta)$. If $\cos(2\theta)$ appears only to odd powers, the technique of Exercise 47 suffices. To treat an even power of $\cos(2\theta)$, use the identity $\cos^2(2\theta) = \frac{1}{2}(1+\cos(4\theta))$ and continue.

Using this method find (a) $\int \cos^2(\theta)\sin^4(\theta)\,d\theta$ and (b) $\int_0^{\pi/4} \cos^2(\theta)\sin^2(\theta)\,d\theta$.

Exercises 50 to 52 explore some alternate antiderivatives of $\tan(\theta)$ and $\sec(\theta)$.

50. In Example 3, it is shown that $\int \tan(\theta)\ d\theta = -\ln|\cos(\theta)|$ is an antiderivative of $\tan(\theta)$. Show that $\ln|\sec(\theta)|$ is also an antiderivative of $\tan(\theta)$.

51. (a) Verify that $\int \sec(\theta)\ d\theta = \ln|\sec(\theta)+\tan(\theta)|+C$ (for $0 \le \theta < \pi/2$) by differentiating $\ln|\sec(\theta)+\tan(\theta)|$.

(b) Does (a) contradict the formula for $\int \sec(\theta)\ d\theta$ found in Example 4?

52. In 1645, Henry Bond conjectured from experimental data that $\int_0^\theta \sec(t)\ dt = \ln\left(\tan\left(\frac{\theta}{2}+\frac{\pi}{4}\right)\right)$ Verify Bond's conjecture by checking that the formula holds for $\theta = 0$ and checking that the right-hand side is an antiderivative of $\sec(\theta)$. NOTE: Bond's conjecture is related to the discussion of Mercator's projection in CIE 12 after Chapter 9.

---

53. The region R under $y = \sin(x)$ and above $[0,\pi]$ is revolved about the x-axis to produce a solid S.

(a) Draw the region R and the solid S.

(b) Set up a definite integral for the area of R.

(c) Set up a definite integral for the volume of S.

(d) Evaluate the integrals in (b) and (c).

In Exercises 54 to 57 transform the integral into an integral of a rational function of $\cos(\theta)$ and $\sin(\theta)$. Do not evaluate the integrals that you find.

54. $\int \frac{x+\sqrt{9-x^2}}{x^3}\ dx$ **55.** $\int \frac{x^3\sqrt{5-x^2}}{1+\sqrt{5x^2}}\ dx$ **56.** $\int \frac{x^2+\sqrt{x^2-9}}{x}\ dx$ **57.** $\int \frac{x^3\sqrt{5+x^2}}{x+2}\ dx$

---

58. Let n be an integer greater than or equal to 2. Let $R(x,y) = \frac{x+y^2}{2x-y}$.

(a) Find $R(x, \sqrt[3]{4x+5})$.

(b) Use the substitution $u = \sqrt[3]{4x+5}$ to show that $\int \frac{x+(4x+5)^{2/3}}{2x-(4x+5)^{1/3}}\ dx = \frac{3}{8}\int \frac{(u^3+4u^2-5)u^2}{u^3-2u-5}\ du$.

NOTE: The partial fraction representation of this integrand is complicated; do not evaluate the integral.

59. Transform the integrals into integrals of rational functions of u. Leave each answer as a definite integral with an integrand that is a rational function. (a) $\int \frac{\sqrt[3]{x+2}}{x^2+(x+2)^{2/3}}\ dx$ and (b) $\int \frac{\sqrt{x}+x+x^{3/2}}{\sqrt{x}+2}\ dx$.

Exercises 60 to 62 concern $\int R(\cos(\theta),\sin(\theta))\ d\theta$.

60. Let $-\pi < \theta < \pi$ and $u = \tan(\theta/2)$. (See Figure 8.5.4.) The following steps show that the substitution transforms $\int R(\cos\theta,\sin\theta)d\theta$ into the integral of a rational function of u (that can be integrated by partial fractions).

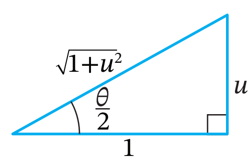


Figure 8.5.4

(a) Show that $\cos\left(\frac{\theta}{2}\right) = \frac{1}{\sqrt{1+u^2}}$ and $\sin\left(\frac{\theta}{2}\right) = \frac{u}{\sqrt{1+u^2}}$.

(b) Using (a), show that $\cos(\theta) = \frac{1-u^2}{1+u^2}$.

(c) Show that $\sin(\theta) = \frac{2u}{1+u^2}$.

(d) Show that $d\theta = \frac{2\ du}{1+u^2}$.

(e) Combining (b), (c), and (d) shows that $u = \tan(\theta/2)$ transforms $\int R(\cos(\theta), \sin(\theta))\, d\theta$ into an integral of a rational function of u: $\int R\left(\frac{1-u^2}{1+u^2}, \frac{2u}{1+u^2}\right) \frac{2}{1+u^2}\, du$.

61. Use the substitution $u = \tan(\theta/2)$ to transform the given integral into integrals of rational functions.

Do not evaluate the resulting integral with a rational integrand.

(a) $\int \frac{1+\sin(\theta)}{1+\cos^2(\theta)}\, d\theta$

(b) $\int \frac{5+\cos(\theta)}{\sin^2(\theta)+\cos(\theta)}\, d\theta$

(c) $\int_0^{\pi/2} \frac{5\, d\theta}{2\cos(\theta)+3\sin(\theta)}$

62. Evaluate $\int_0^{\pi/2} \frac{d\theta}{4\sin(\theta)+3\cos(\theta)}$.

---

63. Explain why a rational function of $\tan(\theta)$ and $\sec(\theta)$ has an elementary antiderivative.

64. Not every rational function of $\sqrt{x+a}$, $\sqrt{x+b}$, and $\sqrt{x+c}$ has an elementary antiderivative. For instance, $\int \frac{dx}{\sqrt{x}\sqrt{x+1}\sqrt{x-1}} = \int \frac{dx}{\sqrt{x^3-x}}$ is not an elementary function. Show that every rational function of x, $\sqrt{x+a}$, and $\sqrt{x+b}$ has an elementary antiderivative. (Use the substitution $u = \sqrt{x+a}$.)

65. Every rational function of x and $\sqrt[n]{\frac{ax+b}{cx+d}}$ has an elementary antiderivative. Explain why.

Exercise 66 is known as the **tractrix problem**. While typically discussed in a differential equations course, only integration is needed to find the solution.

66. A point P is dragged across the xy-plane by a string SP of length a. Let S start at the origin and move to the right along the positive x axis. Assume, as in Figure 8.5.5, P starts at $(0, a)$.

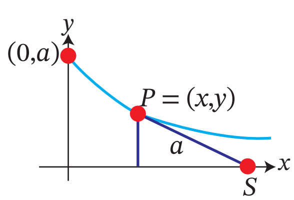


Figure 8.5.5

(a) Find an equation involving $\frac{dy}{dx}$ that is satisfied by the function describing the tractrix.

(b) Rewrite the equation in the form $\frac{dx}{dy}$ equal to an expression involving y and a (but not x).

(c) Find x explicitly in terms of y.

The tractrix can also be visualized as the path of the rear wheel of a scooter when the front wheel follows a straight path. The case when the front wheel follows a circular path is analyzed in CIE 20 at the end of Chapter 15.

8.6 What to do When Confronted with an Integral

Since the exercises in each section of this chapter focus on the techniques of that section, it is usually clear what technique to use on a given integral. But what if an integral is met where there is no clue how to evaluate it? This section suggests what to do.

Table 8.6.1 summarizes the techniques and shortcuts in this chapter. Other specialized techniques are developed in the Exercises in Section 8.5.

General	Substitution	Section 8.2
	Integration by Parts	Section 8.3
	Partial Fractions	Sections 8.4 and 8.2
Special	if f is odd, then $\int_{-a}^{a} f(x)\,dx = 0$	Section 8.1
	if f is even, then $\int_{-a}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx$	Section 8.1
	$\int_{0}^{a} \sqrt{a^2-x^2}\,dx = \frac{\pi a^2}{4}$	Section 8.1
	$\int \sin(mx)\sin(nx)\,dx$, etc.	Section 8.5
	$\int \sin^2(\theta)\,d\theta$, etc.	Section 8.5
	$\int \tan(\theta)\,d\theta$, $\int \sec(\theta)\,d\theta$, etc.	Section 8.5
	$\int R(x, \sqrt[n]{ax+b})\,dx$	Section 8.5
	$\int R(x, \sqrt{a^2-x^2})\,dx$, etc.	Section 8.5
	$\int R(\cos(\theta), \sin(\theta))\,dx$, etc.	Section 8.5

Table 8.6.1

A few examples will illustrate how to choose a method for computing an antiderivative. The more integrals you evaluate, the more quickly you will be able to choose an appropriate technique.

EXAMPLE 1. Evaluate $\int \frac{x\,dx}{1+x^4}$.

SOLUTION Since the integrand is a rational function of x, partial fractions would work. This requires factoring x^4+1 and then representing $x/(1+x^4)$ as a sum of partial fractions. With some effort it can be found that

$$x^4+1 = (x^2+\sqrt{2}x+1)(x^2-\sqrt{2}x+1)$$

The constants A, B, C, and D will have to be found such that

$$\frac{x}{1+x^4} = \frac{Ax+B}{x^2+\sqrt{2}x+1} + \frac{Cx+D}{x^2-\sqrt{2}x+1}$$

The method would work but would be tedious.

Try another attack. Beccause the numerator x is almost the derivative of x^2, the substitution $u = x^2$ is worth trying. With $u = x^2$ we find $du = 2x\,dx$ and

$$\int \frac{x\,dx}{1+x^4} = \int \frac{\frac{1}{2}du}{1+u^2}, = \frac{1}{2}\arctan(u) + C = \frac{1}{2}\arctan(x^2) + C.$$

◆

REMINDER: Every antiderivative can be verified by differentiation.

EXAMPLE 2. Evaluate $\int \frac{1+2x}{1+x^2}\,dx$.

SOLUTION The integrand is a rational function of x, but partial fractions will not help, since the integrand is already in its partial-fraction form.

If we break the integrand into two summands

$$\int \frac{1+x}{1+x^2}\,dx = \int \frac{dx}{1+x^2} + \int \frac{2x\,dx}{1+x^2},$$

the integrals can be found quickly. The first is $\arctan(x)$, and, after using the substitution $u = x^2$, the second is seen to be $\ln(1+x^2)$. Thus

$$\int \frac{1+2x}{1+x^2}\,dx = \arctan(x) + \ln(1+x^2) + C.$$

◆

EXAMPLE 3. Evaluate $\int \frac{e^{2x}}{1+e^x}\,dx$.

SOLUTION The integral may look so peculiar that it may not even be elementary. However, e^x is a simple function, with $d(e^x) = e^x\,dx$. This suggests trying the substitution $u = e^x$, so $du = e^x\,dx$. Thus, expressing dx entirely in terms of u and du:

$$dx = \frac{du}{e^x} = \frac{du}{u}.$$

But what can be done to express e^{2x} in terms of u? Because $e^{2x} = (e^x)^2 = u^2$ there is no difficulty:

$$\int \frac{e^{2x}}{1+e^x}\,dx = \int \frac{u^2}{1+u}\,\frac{du}{u} = \int \frac{u\,du}{1+u}.$$

We observe that the same substitution could have been used differently in this example:

$$\int \frac{e^{2x}}{1+e^x}\,dx = \int \frac{e^x(e^x\,dx)}{1+e^x} = \int \frac{u\,du}{1+u}.$$

Of course, the result is the same.

The integrand is not a proper rational function. One way to deal with this is to use long division. Here is another approach:

$$\begin{aligned}\int \frac{u\,du}{1+u} = \int \frac{u+1-1}{1+u}\,du \quad &= \int \left(1 - \frac{1}{1+u}\right)du \\ = u - \ln(|1+u|) + C &= e^x - \ln(1+e^x) + C.\end{aligned}$$

◆

EXAMPLE 4. Evaluate $\int \frac{x^3\,dx}{(1-x^2)^5}$.

SOLUTION Partial fractions would work for this problem, but the denominator, when factored, would be $(1+x)^5(1-x)^5$. There would be ten unknown constants to find. Maybe it is worth spending some time looking for an easier approach.

Since the denominator is the obstacle, try $u = x^2$ or $u = 1 - x^2$ to see if the integrand gets simpler. Try $u = x^2$ first. Assume that we are interested only in getting an antiderivative for positive x: if $u = x^2$, then $du = 2x\,dx$ and so

$$dx = \frac{du}{2x} = \frac{du}{2\sqrt{u}}.$$

Then

$$\int \frac{x^3\,dx}{(1-x^2)^5} = \int \frac{u^{3/2}}{(1-u)^5}\frac{du}{2\sqrt{u}} = \frac{1}{2}\int \frac{u\,du}{(1-u)^5}.$$

The same substitution $u = x^2$ could be carried out by writing $x^3\,dx$ as $x^2\,x\,dx$, and so

$$\int \frac{x^3\,dx}{(1-x^2)^5} = \int \frac{x^2\,x\,dx}{(1-x^2)^5} = \int \frac{u(du/2)}{(1-u)^5} = \frac{1}{2}\int \frac{u\,du}{(1-u)^5}.$$

In either case, making the additional substitution $v = 1-u$ leads to

$$\frac{1}{2}\int \frac{u\,du}{(1-u)^5} = \frac{1}{2}\int \frac{(1-v)\,(-dv)}{v^5} = \frac{1}{2}\int \frac{1}{v^4} - \frac{1}{v^5}\,dv,$$

which is easily evaluated.

Observe that the two substitutions $u = x^2$ and $v = 1-u$ are equivalent to the single substitution $w = 1-x^2$. That suggests we could also have used the substitution $w = 1-x^2$. Then $dw = -2x\,dx$ and

$$\int \frac{x^3\,dx}{(1-x^2)^5} = \int \frac{x^2(x\,dx)}{(1-x^2)^5} = \int \frac{(1-w)(-dw/2)}{w^5} = \int \frac{1}{2}\left(w^{-4} - w^{-5}\right)\,dw,$$

which is equivalent to the integral found by using two substitutions.

In any case,

$$\begin{aligned}
\int \frac{x^3\,dx}{(1-x^2)^5} &= \frac{1}{2}\int w^{-4} - w^{-5}\,dw, & &= \frac{1}{2}\left(\frac{w^{-3}}{-3} - \frac{w^{-4}}{-4}\right) + C \\
&= \frac{1}{2}\left(\frac{-1}{3(1-x^2)^3} + \frac{1}{4(1-x^2)^4}\right) + C & &= \frac{1}{24}\left(\frac{3}{(1-x^2)^4} - \frac{4}{(1-x^2)^3}\right) + C \\
&= \frac{1}{24}\,\frac{3-4(1-x^2)}{(1-x^2)^4} + C & &= \frac{1}{24}\,\frac{4x^2-1}{(1-x^2)^4} + C.
\end{aligned}$$

With all of steps and algebra involved with this calculation, it is probably a good idea to double-check that this answer is correct. (Thankfully, it is.) ♦

EXAMPLE 5. Evaluate $\int x^3 e^{x^2}\,dx$.

SOLUTION Integration by parts may come to mind, since if $u = x^3$, then $du = 3x^2\,dx$ is simpler. However, dv is then $e^{x^2}\,dx$ and v is nonelementary. This is a dead end. (See Exercise 71.)

Try integration by parts with $u = e^{x^2}$ and $dv = x^3\,dx$. What will $v\,du$ be? We have $v = x^4/4$ and $du = 2xe^{x^2}\,dx$, which is worse than the original $u\,dv$. The exponent of x in $v\,du$ is 5, more than the original exponent, 3.

Try $u = x^2$ and $dv = xe^{x^2}\,dx$. Consequently, $du = 2x\,dx$ and $v = e^{x^2}/2$. Integration by parts yields

$$\begin{aligned}
\int x^3 e^{x^2}\,dx &= \int \underbrace{x^2}_{u}\,\underbrace{xe^{x^2}\,dx}_{dv} & &(\text{ integration by parts: } u = x^2,\ dv = xe^{x^2}\ ) \\
&= \underbrace{x^2}_{u}\,\underbrace{\frac{e^{x^2}}{2}}_{v} - \int \underbrace{\frac{e^{x^2}}{2}}_{v}\,\underbrace{2x\,dx}_{du} & &(\text{ substitution: } u = x^2\ ) \\
&= \frac{x^2 e^{x^2}}{2} - \frac{e^{x^2}}{2} + C.
\end{aligned}$$

Another approach is to use the substitution $u = x^2$ first, followed by a different integration by parts. ♦

EXAMPLE 6. Evaluate $\int \frac{1-\sin(\theta)}{\theta+\cos(\theta)} d\theta$. See also Exercise 72.

SOLUTION Notice that the numerator is the derivative of the denominator. This suggests using the substitution $u = \theta + \cos(\theta)$. Then $du = (1-\sin(\theta))\, du$, and the original problem becomes $\int du/u$, which evaluates to $\ln|u| + C$. Thus, the original integral is $\ln|\theta+\cos(\theta)| + C$. ♦

EXAMPLE 7. Evaluate $\int \frac{1-\sin(\theta)}{\cos(\theta)}\, d\theta$.

SOLUTION Break the integrand into two summands.

$$\int \frac{1-\sin(\theta)}{\cos(\theta)}\, d\theta = \int \left(\frac{1}{\cos(\theta)} - \frac{\sin(\theta)}{\cos(\theta)}\right) d\theta = \int (\sec(\theta) - \tan(\theta))\ d\theta = \ln|\sec(\theta)+\tan(\theta)| + \ln|\cos(\theta)| + C.$$

Since $\ln(A) + \ln(B) = \ln(AB)$, the answer can be simplified to

$$\ln\left(|\sec(\theta)+\tan(\theta)|\,|\cos(\theta)|\right) + C.$$

But $\sec(\theta)\cos(\theta) = 1$ and $\tan(\theta)\cos(\theta) = \sin(\theta)$. The result becomes even simpler,

$$\int \frac{1-\sin(\theta)}{\cos(\theta)}\, d\theta = \ln(1+\sin(\theta)) + C.$$

♦

EXAMPLE 8. Evaluate $\int \frac{\ln(x)}{x}\, dx$.

SOLUTION Integration by parts, with $u = \ln(x)$ and $dv = dx/x$, may come to mind. Then $du = dx/x$ and $v = \ln(x)$. We have

$$\int \underbrace{\ln(x)}_{u} \underbrace{\frac{dx}{x}}_{dv} = \underbrace{(\ln(x))}_{u}\underbrace{(\ln(x))}_{v} - \int \underbrace{\ln(x)}_{v}\underbrace{\frac{dx}{x}}_{du}.$$

Moving the last term on the right-hand side to the left-hand side produces the equation

$$2\int \ln(x)\frac{dx}{x} = (\ln x)^2,$$

from which it follows that

$$\int \frac{\ln(x)}{x}\, dx = \frac{1}{2}(\ln(x))^2 + C.$$

Integration by parts worked, but is not the easiest method. Since $1/x$ is the derivative of $\ln(x)$, the substitution $u = \ln(x)$ gives $du = dx/x$ and we have

$$\int \frac{\ln(x)}{x}\, dx = \int u\, du = \frac{1}{2}u^2 + C = \frac{1}{2}(\ln(x))^2 + C.$$

♦

EXAMPLE 9. Evaluate $\int\limits_0^{3/5} \sqrt{9-25x^2}\, dx$.

SOLUTION This integral reminds us of $\int_0^a \sqrt{a^2-x^2}\, dx = \pi a^2/4$, the area of a quadrant of a circle of radius a. This suggests a substitution u such that $25x^2 = 9u^2$ or $u = 5x/3$, and therefore $dx = 3/\ du$. Substituting,

$$\int\limits_0^{3/5} \sqrt{9-25x^2}\, dx = \int\limits_0^{1} \sqrt{9-9u^2}\, \frac{3}{5}\, du = \frac{9}{5}\int\limits_0^{1} \sqrt{1-u^2}\, du = \frac{9}{5}\cdot\frac{\pi}{4} = \frac{9\pi}{20}.$$

♦

EXAMPLE 10. Evaluate $\int \sin^5(2x)\cos(2x)\,dx$.

SOLUTION We could try integration by parts with $u = \sin^5(2x)$ and $dv = \cos(2x)\,dx$. (See Exercise 73.)

However, $\cos(2x)$ is almost the derivative of $\sin(2x)$. Make the substitution $u = \sin(2x)$ so $du = 2\cos(2x)\,dx$. Then $\cos(2x)\,dx = du/2$ and

$$\int \sin^5(2x)\cos(2x)\,dx = \int u^5\,\frac{du}{2} = \frac{1}{2}\frac{u^6}{6} + C = \frac{1}{12}\sin^6(2x) + C.$$

◆

EXAMPLE 11. Evaluate $\int_{-3}^{3} x^3\cos(x)\,dx$.

SOLUTION Since the integrand is of the form $P(x)\cos(x)$, where P is a polynomial, repeated integration by parts would work. On the other hand, x^3 is an odd function and $\cos(x)$ is an even function. The integrand is therefore an odd function and the integral over $[-3,3]$ is 0. ◆

EXAMPLE 12. Evaluate $\int \sin^2(3x)\,dx$.

SOLUTION One could rewrite the integral as $\int \sin(3x)\sin(3x)\,dx$ and use integration by parts. It is easier to use the identity $\sin^2(\theta) = (1-\cos(2\theta))/2$, as follows:

$$\int \sin^2(3x)\,dx = \int \frac{1-\cos(6x)}{2}\,dx = \int \frac{1}{2}\,dx - \int \frac{\cos(6x)}{2}\,dx = \frac{x}{2} - \frac{\sin(6x)}{12} + C.$$

◆

EXAMPLE 13. Evaluate $\int_{1}^{2} \frac{x^3-1}{(x+2)^2}\,dx$.

SOLUTION Partial fractions would work, after the division of x^3-1 by x^2+4x+4. The substitution $u = x+2$ is easier because it makes the denominator simply u^2. With $u = x+2$, then $du = dx$ and $x = u-2$. Thus

$$\begin{aligned}
\int_{1}^{2} \frac{x^3-1}{(x+2)^2}\,dx &= \int_{3}^{4} \frac{(u-2)^3-1}{u^2}\,du && = \int_{3}^{4} \frac{u^3-6u^2+12u-8-1}{u^2}\,du \\
&= \int_{3}^{4} \left(u - 6 + \frac{12}{u} - \frac{9}{u^2}\right) du && = \left(\frac{u^2}{2} - 6u + 12\ln|u| + \frac{9}{u}\right)\Bigg|_{3}^{4} \\
&= \left(8 - 24 + 12\ln(4) + \frac{9}{4}\right) - \left(\frac{9}{2} - 18 + 12\ln(3) + 3\right) && = -\left(\frac{13}{4}\right) + 12\ln(4) - 12\ln(3) \\
&= 12\ln\left(\frac{4}{3}\right) - \frac{13}{4}.
\end{aligned}$$

◆

Summary

Practice is the best way to improve integration skills. Reading worked examples is a first step to mastering integration, but does not offer the challenge of having to decide which approach is promising and which will only lead to a dead end. The more you practice the more comfortable you will be when facing an integral.

Many integrals can be evaluated in different ways, but one method is usually easier than the others.

It is also important to recognize integrals that do not have an elementary antiderivative.

EXERCISES for Section 8.6

The integrals in Exercises 1 to 59 are elementary. List the technique or techniques that could be used to evaluate the integral. If there is a preferred technique, state what it is (and why). Do not evaluate the integrals.

1. $\int \frac{1+x}{x^2}\,dx$

2. $\int \frac{x^2}{1+x}\,dx$

3. $\int \frac{dx}{x^2+x^3}$

4. $\int x^{10}e^x\,dx$

5. $\int \arctan(2x)\,dx$

6. $\int \arcsin(2x)\,dx$

7. $\int \frac{x+1}{x^2+x^3}\,dx$

8. $\int \frac{\ln(x)}{x^2}\,dx$

9. $\int \frac{\sec^2(\theta)\,d\theta}{\tan(\theta)}$

10. $\int \frac{\tan(\theta)\,d\theta}{\sin^2(\theta)}$

11. $\int \frac{x^3}{\sqrt[3]{x+2}}\,dx$

12. $\int \frac{x^2}{\sqrt[3]{x^3+2}}\,dx$

13. $\int \frac{2x+1}{(x^2+x+1)^5}\,dx$

14. $\int \sqrt{\cos(\theta)}\sin(\theta)\,d\theta$

15. $\int \frac{d\theta}{\sec^2(\theta)}$

16. $\int \tan^2(\theta)\,d\theta$

17. $\int e^{\sqrt{x}}\,dx$

18. $\int \sin\left(\sqrt{x}\right)\,dx$

19. $\int \frac{dx}{(x^2-4x+3)^2}$

20. $\int \frac{x+1}{x^5}\,dx$

21. $\int \frac{x^5}{x+1}\,dx$

22. $\int \frac{\ln(x)}{x(1+\ln(x))}\,dx$

23. $\int \frac{e^{3x}\,dx}{1+e^x+e^{2x}}$

24. $\int \frac{\cos(x)\,dx}{(3+\sin(x))^2}$

25. $\int \ln\left(e^x\right)\,dx$

26. $\int \ln\left(\sqrt[3]{x}\right)\,dx$

27. $\int \frac{\sin(\ln(x))}{x}\,dx$

28. $\int \frac{x+2}{x^4-1}\,dx$

29. $\int \frac{dx}{\sqrt{x}(3+\sqrt{x})^2}$

30. $\int \frac{dx}{(3+\sqrt{x})^3}$

31. $\int (1+\tan(\theta))^3\sec^2(\theta)\,d\theta$

32. $\int \frac{e^{2x}+1}{e^x-e^{-x}}\,dx$

33. $\int \frac{e^x+e^{-x}}{e^x-e^{-x}}\,dx$

34. $\int \frac{(x+3)(\sqrt{x+2}+1)}{\sqrt{x+2}-1}\,dx$

35. $\int \frac{\sqrt[3]{x+2}-1}{\sqrt{x+2}+1}\,dx$

36. $\int \frac{dx}{x^2-9}$

37. $\int \frac{x+7}{(3x+2)^{10}}\,dx$

38. $\int \frac{x^3\,dx}{(3x+2)^7}$

39. $\int \frac{2^x+3^x}{4^x}\,dx$

40. $\int \frac{2^x}{1+2^x}\,dx$

41. $\int \frac{(x+\arcsin(x))\,dx}{\sqrt{1-x^2}}$

42. $\int \frac{x+\arctan(x)}{1+x^2}\,dx$

43. $\int x^3\sqrt{1+x^2}\,dx$

44. $\int x(1+x^2)^{3/2}\,dx$

45. $\int \frac{x\,dx}{\sqrt{x^2-1}}$

46. $\int \frac{x^3}{\sqrt{x^2-1}}\,dx$

47. $\int \frac{x\,dx}{(x^2-9)^{3/2}}$

48. $\int \frac{\arctan(x)}{1+x^2}\,dx$

49. $\int \frac{\arctan(x)}{x^2}\,dx$

50. $\int \frac{x^4-1}{x+2}\,dx$

51. $\int \cos(x)\ln(\sin(x))\,dx$

52. $\int \frac{x\,dx}{\sqrt{x^2+4}}$

53. $\int \frac{dx}{x^2+x+5}$

54. $\int \frac{x\,dx}{x^2+x+5}$

55. $\int \frac{x+3}{(x+1)^5}\,dx$

56. $\int \frac{x^5+x+\sqrt{x}}{x^3}\,dx$

57. $\int (x^2+9)^{10}x\,dx$

58. $\int (x^2+9)^{10}x^3\,dx$

59. $\int \frac{x^4\,dx}{(x+1)^2(x-2)^3}$

In Exercises 60 to 62, (a) find at least 2 positive integers n for which the integrals are elementary and (b) evaluate the integrals listed in (a).

60. $\int \sqrt{1+x^n}\,dx$

61. $\int (1+x^2)^{1/n}\,dx$

62. $\int (1+x)^{1/n}\sqrt{1-x}\,dx$

63. Find $\int \frac{dx}{\sqrt{x+2}-\sqrt{x-2}}$.

64. Find $\int \sqrt{1-\cos(x)}\,dx$.

In Exercises 65 to 70, evaluate the integrals.

65. $\int \frac{x\,dx}{\left(\sqrt{9-x^2}\right)^5}$

66. $\int \frac{dx}{\sqrt{9-x^2}}$

67. $\int \frac{dx}{x\sqrt{x^2+9}}$

68. $\int \frac{x\,dx}{\sqrt{x^2+9}}$

69. $\int \frac{dx}{x+\sqrt{x^2+25}}$

70. $\int (x^3+x^2)\sqrt{x^2-5}\,dx$

71. (a) Evaluate $\int x^3 e^{x^2}$ using the substitution $u=x^2$ followed by integration by parts.
(b) How does this approach compare with the one used in Example 5?

72. In Example 6 it was found that $\int \frac{1-\sin(\theta)}{\theta+\cos(\theta)}\,d\theta = \ln|\theta+\cos\theta|+C$. Check this by differentiation.

73. (a) Use integration parts to evaluate $\int \sin^5(2x)\cos(2x)\,dx$.
(b) How does this approach compare with the one used in Example 10?

8.S Chapter Summary

The previous section reviewed the techniques discussed in the chapter. Here are some general comments on finding antiderivatives.

While the derivative of an elementary function is again elementary, that is not necessarily so with antiderivatives. It is not easy to predict whether an antiderivative will be elementary. For instance, $\ln(x)$ and $\ln(x)/x$ have elementary antiderivatives but $x/\ln(x)$ does not. Also, $x\sin(x)$ does, but s not. Remembering that some elementary functions lack elementary antiderivatives can save time and reduce frustration.

The substitution method will come in handy most often to reduce an integral to an easier one or to something listed in an integral table. Integrands that involve a product or quotient of functions are often evaluated using integration by parts.

A common partial fraction representation is

$$\frac{1}{a^2-x^2} = \frac{1}{2a}\left(\frac{1}{a-x}+\frac{1}{a+x}\right).$$

Though every rational function has an elementary antiderivative, finding it can be a daunting task except for the simplest denominators. Factoring the denominator into first and second degree polynomials may be difficult. Finding the unknown coefficients in the representation could require long computations. It may be simpler to use Simpson's approximation (Section 6.5) unless one needs to know the antiderivative. Then it might be best to take advantage of a symbolic integrator on a calculator or computer.

Definite integrals of even or odd functions over intervals of the form $[-a,a]$ can be simplified before evaluation.

EXERCISES for Section 8.S

1. Consider the definite integral $\displaystyle\int_0^{\pi/2} \sqrt{(1+\cos(\theta))^3}\sin(\theta)\,d\theta$.

(a) Use an appropriate substitution to get a simplify this integral. (b) Evaluate the integral found in (a).

2. Two of the following four indefinite integrals are elementary functions. Evaluate these two integrals.

(a) $\displaystyle\int \ln(x)\,dx$ (b) $\displaystyle\int \frac{\ln(x)}{x}\,dx$ (c) $\displaystyle\int \frac{dx}{\ln(x)}$ (d) $\displaystyle\int \frac{x}{\ln(x)}\,dx$

3. Evaluate (a) $\displaystyle\int_1^2 (1+x^3)^2\,dx$ and (b) $\displaystyle\int_1^2 (1+x^3)^2x^2\,dx$.

4. Use the Table of Integrals to evaluate the following indefinite integrals: (a) $\displaystyle\int \frac{e^x\,dx}{5e^{2x}-3}$ and (b) $\displaystyle\int \frac{dx}{\sqrt{x^2-3}}$.

5. Evaluate each indefinite integral: (a) $\displaystyle\int \frac{dx}{x^3}$, (b) $\displaystyle\int \frac{dx}{\sqrt{x+1}}$, and (c) $\displaystyle\int \frac{e^x}{1+5e^x}\,dx$.

6. Evaluate $\displaystyle\int \frac{5x^4-5x^3+10x^2-8x+4}{(x^2-1)(x-1)}\,dx$.

7. Transform $\displaystyle\int_0^3 \frac{x^3}{\sqrt{x+1}}\,dx$ into another definite integral with the following two substitutions and evaluate each new integral: (a) $u=x+1$ and (b) $u=\sqrt{x+1}$. (c) Which method was easier to apply?

8. (a) Transform the definite integral $\displaystyle\int_{-1}^4 \frac{x+2}{\sqrt{x+3}}\,dx$ into an easier definite integral by a substitution. (b) Evaluate the definite integral obtained in (a).

9. Evaluate $\displaystyle\int x^2\ln(1+x)\,dx$ (a) without consulting an integral table and (b) with the use of an integral table.

10. Verify that the factorizations into irreducible polynomials are correct.

(a) $x^3-1=(x-1)(x^2+x+1)$ (b) $x^4-1=(x-1)(x+1)(x^2+1)$ (c) $x^3+1=(x+1)(x^2-x+1)$

Express each quotient in Exercises 11 to 17 as a sum of partial fractions.

NOTE: Do not integrate.

11. $\dfrac{2x^2+3x+1}{x^3-1}$ **12.** $\dfrac{x^4+2x^2-2x+2}{x^3-1}$ **13.** $\dfrac{2x-1}{x^3+1}$

14. $\dfrac{x^4+3x^3-2x^2+3x-1}{x^4-1}$ **15.** $\dfrac{2x+5}{x^2+3x+2}$ **16.** $\dfrac{5x^3+11x^2+6x+1}{x^2+x}$

17. $\dfrac{5x^3+6x^2+8x+5}{(x^2+1)(x+1)}$

---

18. Evaluate $\displaystyle\int x\sqrt[3]{x+1}\,dx$ using (a) the substitution $u=\sqrt[3]{x+1}$ and (b) the substitution $u=x+1$.

19. The FTC can be used to evaluate one of (a) $\displaystyle\int_0^1 \sqrt[3]{x}\sqrt{x}\,dx$ and (b) $\displaystyle\int_0^1 \sqrt[3]{1-x}\sqrt{x}\,dx$,but not the other. Which one? Evaluate it.

20. Evaluate $\int \frac{x^{2/3}}{x+1}\,dx$.

21. Evaluate $\int \frac{x^3}{(x-1)^2}\,dx$ (a) using partial fractions and (b) using the substitution $u = x-1$. (c) Which approach is easier in this case?

In Exercises 22 to 25 evaluate the given definite integral.

22. $\int_0^1 (e^x+1)^3 e^x\,dx$

23. $\int_0^1 (x^4+1)^5 x^3\,dx$

24. $\int_1^e \frac{\sqrt{\ln(x)}}{x}\,dx$

25. $\int_0^{\pi/2} \frac{\cos(\theta)}{\sqrt{1+\sin(\theta)}}\,dx$

---

26. (a) Without an integral table, evaluate $\int \sin^5(\theta)\,d\theta$ and $\int \tan^6(\theta)\,d\theta$.
(b) Evaluate them with an integral table.
(c) Resolve any differences in the appearance of the antiderivatives found in (a) and (b).

27. Consider the three antiderivatives (a) $\int \sqrt{1-4\sin^2(\theta)}\,d\theta$, (b) $\int \sqrt{4-4\sin^2(\theta)}\,d\theta$, and (c) $\int \sqrt{1+\cos(\theta)}\,d\theta$. Without evaluating, which are elementary? Explain why you know they are elementary.

28. Evaluate (a) $\int \cot(3\theta)\,d\theta$ and (b) $\int \csc(5\theta)\,d\theta$.

29. Evaluate (a) $\int \sec^5(x)\tan(x)\,dx$ and (b) $\int \frac{\sin(x)}{\cos^3(x)}\,dx$.

30. Evaluate $\int \frac{x^3\,dx}{(1+x^2)^4}$ (a) by the substitution $u = 1+x^2$ and (b) by the substitution $x = \tan(\theta)$.

31. Find $\int \frac{x\,dx}{\sqrt{9x^4+16}}$ (a) without an integral table and (b) with an integral table.

32. Rewrite $\int \frac{x^2\,dx}{\sqrt{1+x}}$ using the substitutions (a) $u = \sqrt{1+x}$, (b) $y = 1+x$, and (c) $x = \tan^2(\theta)$. (d) Evaluate the easiest of these three integrals.

33. Rewrite $\int x\sqrt{(1-x^2)^5}\,dx$ using the substitutions (a) $u = x^2$, (b) $u = 1-x^2$, and (c) $x = \sin(\theta)$. (d) Evaluate the easiest of these three integrals.

34. Evaluate $\int x\sqrt{1+x}\,dx$ in three ways: (a) substituting $u = \sqrt{1+x}$, (b) substituting $x = \tan^2(\theta)$, and (c) integration by parts, with $u = x$ and $dv = \sqrt{1+x}\,dx$.

In Exercises 35 to 47, evaluate the definite integral or integrals appearing in the following exercises.

35. Exercise 21 in Section 7.1.

36. Exercise 22 in Section 7.1.

37. Exercise 23 in Section 7.1.

38. Exercise 24 in Section 7.1.

39. Exercise 25 in Section 7.1.

40. Exercise 26 in Section 7.1.

41. Exercise 27 in Section 7.1.

42. Exercise 28 in Section 7.1.

43. Exercise 30 in Section 7.1.

44. Exercise 1 in Section 7.5.

45. Exercise 2 in Section 7.5.

46. Exercise 3 in Section 7.5.

47. Exercise 4 in Section 7.5.

---

48. The region below the line $y = e$, above $y = e^x$, and to the right of the y-axis is revolved around the y-axis to form a solid $\mathscr{S}$. In Example 1 in Section 7.5 it is shown that the definite integral for the volume of $\mathscr{S}$ using disks is $\int_1^e \pi\left(\ln(y)\right)^2\,dy$ and the volume of $\mathscr{S}$ using shells is $\int_0^1 2\pi x\left(e - e^x\right)\,dx$.

(a) Which integral do you think will be easier to evaluate?
(b) Evaluate each integral.
(c) Which integral was easier to evaluate?
(d) Could you have predicted this before evaluating each integral? (Explain.)

49. The region $\mathscr{R}$ below the line $y = \frac{\pi}{2} - 1$, to the right of the y-axis, and above the curve $y = x - \sin(x)$ is revolved around the y-axis to form a solid $\mathscr{S}$. In Example 2 in Section 7.5 it is shown that the definite integral for the volume of $\mathscr{S}$ using disks cannot be evaluated in terms of elementary functions, and that the volume of $\mathscr{S}$ using shells is $\int_0^{\pi/2} 2\pi x\left(\frac{\pi}{2} - 1 - (x - \sin(x))\right)\,dx$. Find the value of the integral.

50. Evaluate each indefinite integral. (a) $\int \frac{x+1}{x^2}e^{-x}\,dx$ and (b) $\int \frac{ax-1}{ax^2}e^{ax}\,dx$ $(a \neq 0)$.

51. In Example 1 in Section 7.6 the total force on a submerged circular tank is found to be

$$\int_{-5}^{5} (0.036)(x+17)\sqrt{100-4x^2}\,dx = 0.036\int_{-5}^{5} x\sqrt{100-4x^2}\,dx + 0.036\int_{-5}^{5} 17\sqrt{100-4x^2}\,dx \text{ pounds.}$$

The value of the integral was found by seeing that the first integral had value zero because it has an odd integrand over an interval symmetric about the origin, and the second was related to the area of a quarter circle.

(a) Find the value of the first integral using the substitution $u = 100 - 4x^2$.
(b) Find the value of the second integral using the substitution $x^2 = 25\sin^2(\theta)$.
(c) Was it easier to evaluate this integral using symmetry and geometry, (as in Example 1 in Section 7.6) or by the fundamental theorem of calculus (as in this Exercise)?

52. Find $\int \frac{dx}{\sin(2x)}$ by writing $\sin(2x)$ as $2\sin(x)\cos(x)$.

53.
(a) Show that $\int_0^\infty \frac{\sin(kx)}{x}\,dx = \int_0^\infty \frac{\sin(x)}{x}\,dx$, where k is a positive constant.
(b) Show that $\int_0^\infty \frac{\sin(x)\cos(x)}{x}\,dx = \frac{1}{2}\int_0^\infty \frac{\sin(x)}{x}\,dx$.
(c) If k is negative, what is the relation between $\int_0^\infty \frac{\sin(kx)}{x}\,dx$ and $\int_0^\infty \frac{\sin(x)}{x}\,dx$?

54. Evaluate $\int_0^\infty e^{-x}\sin^2(x)\,dx$.

55. Evaluate $\int_0^\infty e^{-x}\sin(x)\,dx$. (This integral was first encountered in Example 4 in Section 7.8.)

In statistics a function $F(x)$ defined on $[0,\infty)$ is called a **probability distribution** if $F(0) = 0$, $\lim_{x\to\infty} F(x) = 1$, and F has a nonnegative derivative f. The function f is called a **probability density**. The integral $\int_0^\infty x f(x)\, dx$ is called the **expected value** or **average value** of x. Exercises 56 and 57 show that if one of the integrals $\int_0^\infty x f(x)\, dx$ and $\int_0^\infty (1 - F(x))\, dx$ is convergent, so is the other and the two are equal.

56. Assume $\int_0^\infty x f(x)\, dx$ is finite.

(a) Show that $\int_k^\infty x f(x)\, dx$ approaches zero as k approaches ∞.

(b) Use the fact that $\int_k^\infty x f(x)\, dx \geq \int_k^\infty k f(x)\, dx$, to show that $\lim_{k\to\infty} k(1 - F(k)) = 0$.

(c) Show that $\int_0^k x f(x)\, dx = k(F(k) - 1) + \int_0^k (1 - F(x))\, dx$.

(d) Use (c) to show that $\int_0^\infty x f(x)\, dx = \int_0^\infty (1 - F(x))\, dx$.

57. Assume that $\int_0^\infty (1 - F(x))\, dx$ is finite.

(a) Show that $\int_0^k x f(x)\, dx = kF(k) - \int_0^k F(x)\, dx$.

(b) Show $kF(k) - \int_0^k F(x)\, dx \leq \int_0^k (1 - F(x))\, dx$.

(c) Show that $\int_0^\infty x f(x)\, dx$ is finite.

(d) Show that $\int_0^\infty (1 - F(x))\, dx = \int_0^\infty x f(x)\, dx$.

Exercises 58 to 61 are related.

58. Show that $\int_1^\infty (\cos(x))/x^2\, dx$ is convergent.

59. Show that $\int_1^\infty (\sin(x))/x\, dx$ is convergent.

60. Show that $\int_0^\infty (\sin(x))/x\, dx$ is convergent.

61. Show that $\int_0^\infty \sin(e^x)\, dx$ is convergent.

62. In a statistics text it is asserted that for $\lambda > 0$ and n a positive integer $\int\limits_0^\infty \left(1-\left(1-e^{-\lambda t}\right)^n\right)\,dt = \frac{1}{\lambda}\sum\limits_{k=1}^n \frac{1}{k}$. Check this for (a) $n=1$ and (b) $n=2$. (c) Show that the integral is convergent for all positive integers n.

63. Let $\int\limits_{-\infty}^\infty f(x)\,dx$ be a convergent integral with value A.

(a) Express $\int\limits_{-\infty}^\infty f(x+2)\,dx$ in terms of A. (b) Express $\int\limits_{-\infty}^\infty f(2x)\,dx$ in terms of A.

64. Find the error in the following sequence of steps.

The substitution $x=y^2$ yields $dx = 2y\,dy$ and $\int\limits_0^1 \frac{1}{x}\,dx = \int\limits_0^1 \frac{2y}{y^2}\,dy = \int\limits_0^1 \frac{2}{y}\,dy = 2\int\limits_0^1 \frac{1}{y}\,dy = 2\int\limits_0^1 \frac{1}{x}\,dx$.

Hence $\int\limits_0^1 \frac{dx}{x} = 2\int\limits_0^1 \frac{dx}{x}$ from which it follows that $\int\limits_0^1 \frac{dx}{x} = 0$.

The Laplace transform of $f(t)$ is defined to be the function $F(s) = \int\limits_0^\infty e^{-st} f(t)\,dt$. Laplace transforms were introduced in Section 8.3, Exercises 51 to 59. Exercises 65 to 67 develop additional properties of Laplace transforms.

65. Let f and its derivative f' have Laplace transforms. Let P be the Laplace transform of f, and let Q be the Laplace transform of f'. Show that $Q(s) = -f(0) + sP(s)$.

66. Assume that $f(t) = 0$ for $t < 0$ and that f has a Laplace transform. Let a be a positive constant. Define $g(t)$ to be $f(t-a)$. Show that the Laplace transform of g is e^{-as} times the Laplace transform of f.

67. Let a be a positive constant, and let $g(t) = f(at)$. Let P be the Laplace transform of f, and let Q be the Laplace transform of g. Show that $Q(s) = \frac{1}{a}P\left(\frac{s}{a}\right)$.

68.
(a) Estimate $\int\limits_0^1 \frac{\sin(x)}{x}\,dx$ by approximating $\sin(x)$ by the Maclaurin polynomial $P_6(x;0)$
(b) Use the Lagrange form of the error to put an upper bound on the error in (a).

69.
(a) Estimate $\int\limits_{-1}^1 \frac{e^x}{x+2}\,dx$ by using the Taylor polynomial $P_3(x;-2)$ associated with e^x to approximate e^x.
(b) Use the Lagrange form of the error to put an upper bound on the error in (a).

70.
(a) Estimate $\int\limits_{-1}^1 \frac{e^x}{x-2}\,dx$ by using the Taylor polynomial $P_3(x;2)$ associated with e^x to approximate e^x.
(b) Use the Lagrange form of the error to put an upper bound on the error in (a).

71. Evaluate $\int \frac{\ln(x^2)}{x^2} \, dx$.

72. If a is a constant, show that $\int_{-\infty}^{\infty} e^{-(x-a)^2} \, dx = \int_{-\infty}^{\infty} e^{-x^2} \, dx = 2\int_{0}^{\infty} e^{-x^2} \, dx$.

73. If $\int_{0}^{\infty} e^{-x^2} \, dx = \frac{\sqrt{\pi}}{2}$, show that $\int_{0}^{\infty} 2^{-x^2} \, dx = \frac{\sqrt{\pi}}{2\sqrt{\ln(2)}}$.

74. When studying the normal distribution in statistics one will meet an equation that amounts to

$$\frac{\int_{-\infty}^{\infty} x \exp(-(x-\mu)^2) \, dx}{\int_{-\infty}^{\infty} \exp(-(x-\mu)^2) \, dx} = \mu,$$

where μ is a constant. Show that this equation is correct.

75. For which values of the positive constant k is $\int_{e}^{\infty} \frac{dx}{x(\ln(x))^k}$ (a) convergent? and (b) divergent?

76. The formula for the area of region OAP in Figure 6.S.2 was found in Exercise 64 in Section 6.S to be $\frac{1}{2}\cosh(t)\sinh(t) - \int_{1}^{\cosh(t)} \sqrt{x^2-1} \, dx$. Use the substitution $x = \cosh(u)$ to evaluate the definite integral.

77. The study of heat capacity of a crystal encounters this integral: $\int_{0}^{b} \frac{x^4 e^x}{(e^x-1)^2} \, dx$.

(a) Show that $\int_{0}^{b} \frac{x^4 e^x}{(e^x-1)^2} \, dx$ is convergent. (b) Is $\int_{0}^{b} \frac{x e^x}{(e^x-1)^2} \, dx$ convergent?

78. Show that $\int_{-\infty}^{\infty} \frac{dt}{(1+t^2)^{3/2}} = 2$.

79. (a) Show that $\int_{0}^{\infty} \frac{x^2}{(x^2+1)^{5/2}} \, dx$ is convergent. (b) Show that its value is $\frac{1}{3}$.

80. Let $\lambda > 0$ be a constant. In the theory of probability the equation $\int_{0}^{\infty} e^{-\lambda x} R(x) \, dx = \frac{1}{\lambda}\int_{0}^{\infty} e^{-\lambda x} R'(x) \, dx + \frac{1}{\lambda} R(0)$ occurs. Assuming the integrals converge, explain how the equation is obtained.

81. Let k be a positive constant. Justify the equation $\int_{0}^{k} \frac{1}{k} f\left(\frac{x}{k}\right) \, dx = \int_{0}^{1} f(x) \, dx$.

82. Assume that f is continuous on $[0,\infty)$ and has period one, that is, $f(x+1) = f(x)$ for all x in $[0,\infty)$.

(a) Show that $\int_{0}^{\infty} e^{-x} f(x) \, dx$ is convergent (b) Show that $\int_{0}^{\infty} e^{-x} f(x) \, dx = \frac{e}{e-1}\int_{0}^{1} e^{-x} f(x) \, dx$

83. Assume that f is continuous on $[0,\infty)$ and has period $p>0$, that is, $f(t+p)=f(t)$ for all $t\ge 0$. Let s be a positive number and assume $\int_0^\infty e^{-st}f(t)\,dt$ converges. Show that this improper integral converges and has value $\frac{1}{1-e^{-sp}}\int_0^p e^{-st}f(t)\,dt$.

84. The integral $\int_0^\infty x^{2n}e^{-kx^2}\,dx$ appears in the kinetic theory of gases. In Exercise 38 in Section 17.3 we show that $\int_0^\infty e^{-x^2}\,dx=\frac{\sqrt{\pi}}{2}$. Using this information, find the values of (a) $\int_0^\infty e^{-kx^2}\,dx$ and (b) $\int_0^\infty x^2e^{-kx^2}\,dx$.

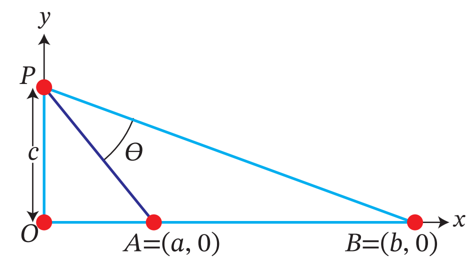


Figure 8.S.1

85. James Maxwell's "On the Geometric Mean Distance of Two Figures in a Plane", written in 1872, begins "There are several problems of great practical importance in electromagnetic measurements, in which the value of the quantity has to be calculated by taking the sum of the logarithms of the distances of a system of parallel wires from a given point." This led him to several problems, of which this is the first.

A point P is a distance c from the line that contains the line segment AB. Let O be the point on that line nearest P, as in Figure 8.S.1. Introduce a coordinate system in which O is the origin, AB lies on the x-axis, and side OP lies on the y-axis.

Let $A=(a,0)$, $B=(b,0)$, and let $f(x)$ be the distance from P to $(x,0)$.

Show that the average value of $\ln(f(x))$ for x in $[a,b]$ is $\frac{1}{2(b-a)}\left(b\ln\left(c^2+b^2\right)-a\ln\left(c^2+a^2\right)\right)-1+\frac{c\theta}{b-a}$, where θ is the measure of $\angle APB$ in radians.

86. Find the value of $\int \frac{\cos(\theta)}{(b^2+c^2\cos^2(\theta))^{1/2}}\,d\theta$.

This integral will appear again, in Exercise 21 in Section 18.S.

87. Show that $\int \sqrt{x}e^x\,dx$ is not elementary.

88. We have seen that $\int e^{x^2}\,dx$ is not elementary. Show that (a) for positive odd integers n, $\int x^ne^{x^2}\,dx$ is elementary and (b) for positive even integers n, $\int x^ne^{x^2}\,dx$ is not elementary. (c) Find nonzero values for a and b such that $\int (ax^4+bx^2)e^{x^2}\,dx$ is an elementary function.

89. We have seen that $\int e^{x^2}\,dx$ and $\int \frac{e^x}{x}\,dx$ are not elementary. Show that (a) $\int \frac{e^{x^2}}{x}\,dx$ is not elementary, (b) $\int \frac{e^{x^2}}{x^2}\,dx$ is not elementary, and (c) for positive integers n, $\int \frac{e^{x^2}}{x^n}\,dx$ is not elementary.

90. We have seen that $\int \frac{e^x}{x}\,dx$ is not elementary. Show that

(a) for positive integers n, $\int x^ne^x\,dx$ is elementary and (b) for positive integers n, $\int \frac{e^x}{x^n}\,dx$ is not elementary.

91. **Sam:** I understand what a definite integral is — the limit of sums. I accept on faith that for a continuous function this limit of sums exists. I agree that it is a handy idea, with many uses, but I don't see why I have to learn all those ways to compute it: antiderivatives, trapezoids, Simpson's method. My trusty calculator evaluates integrals to eight decimal places and a computer algebra system can often give me the exact expression for both definite and indefinite integrals.

JANE: What's your point?
SAM: I would make this text much shorter by omitting this chapter. This would allow us more time to spend on the stuff at the end.

Does Sam have a valid argument, for a change?

Exercises 92 to 98 all relate to the **bell curve** that arises in statistics. Its density function is $f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$ where μ and σ are constants (and $\sigma > 0$). Other names for a bell curve are **normal distribution** and **Gaussian distribution**. In Exercises 92 and 98 assume $\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$. This result will be established in Exercise 38 in Section 17.3.

In Exercises 92 to 97, assume $\mu = 0$.

92. Show that $\int_{-\infty}^{\infty} x^2 e^{-x^2}\,dx = \frac{1}{2}\sqrt{\pi}$.

93. Show that $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

94. Show that the graph of $y = f(x)$ has inflection points where $x = \sigma$ and $-\sigma$.

95. The **mean** of the distribution $f(x)$ is the average value of x. The average value of x is defined to be $\int_{-\infty}^{\infty} x f(x)\,dx$. Show that the mean value of x is 0.

96. The **discrepancy** from 0 is measured by σ^2. It is called the **variance**. The number σ is called the **standard deviation** of the distribution. The variance and standard deviation both measure the spread of the data. Show that

$$\int_{-\infty}^{\infty} x^2 f(x)\,dx = \sigma^2.$$

97. Assume that $\int_{-\infty}^{\infty} g(x) = 1$ and $\int_{-\infty}^{\infty} xg(x)\,dx = k$. Let $h(x) = g(x-k)$. Show that $\int_{-\infty}^{\infty} h(x)\,dx = 1$, $\int_{-\infty}^{\infty} xh(x)\,dx = k$, and $\int_{-\infty}^{\infty} (x-k)^2 h(x)\,dx = \int_{-\infty}^{\infty} x^2 g(x)\,dx$.

98. In this exercise there are no assumptions about μ. Show that (a) the graph of f is symmetric with respect to $x = \mu$, (b) $\int_{-\infty}^{\infty} f(x)\,dx = 1$, (c) $\int_{-\infty}^{\infty} xf(x)\,dx = \mu$, (d) $\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = \sigma^2$, and (e) the two inflection points of $f(x)$ occur when $x = \mu \pm \sigma$. (f) Sketch the graph of $f(x)$.

(c) shows that μ is the mean of the normal distribution.

---

99. If $f(x)$ and $g(x)$ have elementary antiderivatives, which of the following necessarily do also?
(a) $f(x)g(x)$, (b) $f(g(x))$, and (c) $f(x)+g(x)$. Justify each answer.

100. Show that (a) $e^{x^{1/2}}$ has an elementary antiderivative, (b) $e^{x^{1/3}}$ has an elementary antiderivative, and (c) for every positive integer n, $e^{x^{1/n}}$ has an elementary antiderivative.

101. Which has the larger absolute value, $\int_{0}^{\sqrt{\pi}} \sin\left(x^2\right)\,dx$ or $\int_{\sqrt{\pi}}^{\sqrt{2\pi}} \sin\left(x^2\right)\,dx$? Explain your reasoning.

102. In a letter dated May 24, 1872 Maxwell wrote: "It is strange ... that W. Weber could not correctly integrate $\int_0^{\pi} \cos(\theta)\sin(\phi)\ d\phi$ where $\tan(\theta) = \dfrac{A\sin(\phi)}{B + A\cos(\phi)}$, but that everyone should have copied such a wild result as $\dfrac{B}{\sqrt{A^2+B^2}} \cdot \dfrac{B^4 + \frac{7}{6}A^2B^2 + \frac{2}{3}A^2}{B^4 + A^2B^2 + A^4}$. Of course there are two forms of the result according as A or B is greater."
Assuming that A and B are positive, find the correct value of the integral.

103. The following calculation appears in *Electromagnetic Fields*, 2nd ed., Roald K. Wangsness, Wiley, 1986.

See also Exercise 4 in Section 12.S.

(a) Show that the substitution $\frac{\pi}{2}\cos(\theta) = \frac{1}{2}(\pi - v)$ turns $\displaystyle\int_0^{\pi} \frac{\cos^2\left(\frac{\pi}{2}\cos(\theta)\right)}{\sin(\theta)}\ d\theta$ into

$$\frac{1}{4}\left(\int_0^{2\pi} \frac{1-\cos(v)}{v}\ dv + \int_0^{2\pi} \frac{1-\cos(v)}{2\pi - v}\ dv\right). \tag{8.S.1}$$

(b) Show that introducing $w = 2\pi - v$ shows that the two integrals in (8.S.1) with respect to v are equal.

NOTE: It is still necessary to find $\displaystyle\frac{1}{2}\int_0^{2\pi} \frac{1-\cos(v)}{v}\ dv$. While this integrand does not have an elementary antiderivative, the approximate value of the definite integral can be found in many integral tables: $\displaystyle\frac{1}{2}\int_0^{2\pi} \frac{1-\cos(v)}{v}\ dv \approx 1.2188$.

In Exercises 104 and 105, verify that the derivative of the first expression is the second expression. (Assume a and p are constants with $a^2 + p^2 > 0$.)

104. $\dfrac{e^{ax}\left(a\sin(px) - p\cos(px)\right)}{a^2+p^2}$, $e^{ax}\sin(px)$

105. $\sec(x) + \ln\left(\tan\left(\frac{x}{2}\right)\right)$, $\dfrac{1}{\sin(x)\cos^2(x)}$

Calculus is Everywhere # 11

Average Speed and Class Size

There are two ways to define your average speed when jogging or driving a car. A friend could jot down your speed at regular intervals of time, say, every second. Then you would average those speeds. That average is called an average *with respect to time*. Or your friend could record your velocity at regular intervals of distance, say, every hundred feet. Their average is called an average *with respect to distance*.

How do you think they would compare? If you kept a constant speed, c, the averages would both be c. Are they always equal, even if your speed varies? Would one of the averages always tend to be larger?

CHALLENGE: Take a few minutes to think about this now. Consider some different possibilities. What happens if you stop for a period of time?

Now that you have thought about these two different average speeds, we will analyze them mathematically, with the aid of the **Cauchy-Schwarz inequality**.

There are several versions of the Cauchy-Schwarz inequality. The version we give here involves two continuous functions, f and g, defined on $[a,b]$. If $\int_a^b f(x)^2\,dx$ and $\int_a^b g(x)^2\,dx$ are small, then the absolute value of $\int_a^b f(x)g(x)\,dx$ ought to be small too. That this obvious statement is true in general is a consequence of Cauchy-Schwarz inequality.

Cauchy-Schwarz is pronounced: "koh-shee' shwartz"

Theorem C.11.1: Cauchy-Schwarz Inequality (Integral Form)

Let $f(x)$ and $g(x)$ be continuous functions on an interval $[a,b]$. Then

$$\left(\int_a^b f(x)g(x)\,dx\right)^2 \le \left(\int_a^b f(x)^2\,dx\right)\left(\int_a^b g(x)^2\,dx\right). \tag{C.11.1}$$

After showing some of its applications, we will use the quadratic formula to prove Theorem C.11.1, that is, show that (C.11.1) is true.

First we use it to see which average speed is larger, the one with respect to time or the one with respect to distance.

Call the speed at time t, $v(t)$, and the distance traveled up to time t, $s(t)$. From time a to time b the average speed with respect to time is

$$\frac{1}{b-a}\int_a^b v(t)\,dt = \frac{1}{b-a}\left(s(b)-s(a)\right).$$

The average speed with respect to distance is

$$\frac{1}{s(b)-s(a)}\int_{s(a)}^{s(b)} w(s)\,ds, \tag{C.11.2}$$

where $w(s)$ denotes the speed when the distance covered is s. The functions $v(t)$ and $w(s)$ provide two different representations of the object's speed; they are related by $w(s(t)) = v(t)$.

We know that $ds = v(t)\,dt$, so (C.11.2) becomes

$$\frac{1}{s(b)-s(a)}\int_{s(a)}^{s(b)} w(s)\,ds = \frac{1}{s(b)-s(a)}\int_a^b v(t)v(t)\,dt.$$

Substituting $s(b)-s(a)=\int_a^b v(t)\,dt$ and $b-a=\int_a^b 1\,dt$, we will show that the average with respect to time is less than or equal to the average with respect to distance, that is,

$$\frac{\int_a^b v(t)\,dt}{\int_a^b 1\,dt} \le \frac{\int_a^b v(t)^2\,dt}{\int_a^b v(t)\,dt}.$$

Or, equivalently,

$$\left(\int_a^b v(t)\,dt\right)^2 \le \left(\int_a^b 1\,dt\right)\left(\int_a^b v(t)^2\,dt\right). \tag{C.11.3}$$

Equation (C.11.3) is a special case of (C.11.1) with $f(t)=1$ and $g(t)=v(t)$.

Therefore the average with respect to time is always less than or equal to the average with respect to distance. Exercise 1 shows that the only way the two averages can be equal is that speed is constant throughout the trip. In all cases where the speed, is not constant, the average with respect to time is (strictly) less than the average with respect to distance.

Proof of Theorem C.11.1
It is easy to verify the Cauchy-Schwartz inequality if at least one of the functions $f(x)$ or $g(x)$ is the constant function 0. In the rest of the proof assume neither function is the zero function.

To verify (C.11.1) introduce $h(t)$ defined by

$$\begin{aligned} h(t) &= \int_a^b (f(x)-tg(x))^2\,dx \\ &= \int_a^b f(x)^2\,dx - 2t\int_a^b f(x)g(x)\,dx + t^2\int_a^b g(x)^2\,dx. \end{aligned} \tag{C.11.4}$$

Because the first integrand, $f(x)^2$, in (C.11.4) is never negative, $h(t)\ge 0$ for all t. Expanding (C.11.4) we can write $h(t)=r+qt+pt^2$, where

$$r=\int_a^b f(x)^2\,dx, \qquad q=-2\int_a^b f(x)g(x)\,dx, \quad \text{and} \quad p=\int_a^b g(x)^2\,dx.$$

Observe that the assumption that neither $f(x)$ nor $g(x)$ is the constant zero function means p and r are both positive. Thus, the graph of $y=h(t)$ is a parabola that never drops below the t-axis and touches it in at most one point. (If it touched the t-axis at two points, it would dip below the axis, forcing $h(t)$ to take on some negative values.) Because the quadratic equation $h(t)=0$ has at most one solution, its discriminant q^2-4pr cannot be positive. Thus, $q^2-4pr\le 0$, so $q^2\le 4pr$, from which the Cauchy-Schwarz inequality follows. ●

EXERCISES for CIE C.11

1. Show that equality holds in (C.11.1) only when $g(x)$ is a constant times $f(x)$.

2. The discrete form of the Cauchy-Schwarz inequality asserts that if $a_1, a_2, a_3, \ldots, a_n$ and $b_1, b_2, b_3, \ldots, b_n$ are real numbers, then

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \le \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2. \tag{C.11.5}$$

(a) Prove the inequality (C.11.5).
(b) In what situations are the two sides of the inequality (C.11.5) actually equal?

3. Use the inequality (C.11.5) to show that the average class size at a university as viewed by the registrar is usually smaller than the average class size as viewed by the students.

ADDITIONAL FACT: The Cauchy-Schwarz inequality also explains why the average time between buses as viewed by a dispatcher is usually shorter than the average time between buses as viewed by passengers arriving randomly at a bus stop.

4. A region R is bounded by the x-axis, the lines $x = 2$ and $x = 5$, and the curve $y = f(x)$, where f is a positive function. The area of R is A. When revolved around the x-axis it produces a solid of volume V.
(a) How large can V be? (b) How small can V be?

5. If the region $\mathscr{R}$ in the Exercise 4 is revolved around the y-axis, what can be said about the maximum and minimum values for the volume of the resulting solid? Explain.

The molecules in a gas move at various speeds. In 1859 James Maxwell developed a formula for the distribution of the speeds in a gas consisting of N molecules,

$$f(v) = 4\pi N\left(\frac{m}{2\pi kT}\right)^{3/2} v^2 e^{-mv^2/(2kT)}.$$

This means that for small values, dv, the number of molecules with speeds between v and $v + dv$ is approximately $f(v)\,dv$. In the formula k is a physical constant, T is the absolute temperature, and m is the mass of a molecule. The variable is v. Exercises 6 to 8 investigate Maxwell's model.

Exercises 6 to 8 develop the fact that the average speed of a molecule is greater than the most probable speed of a molecule.

6. Show that $\int_0^\infty f(v)\,dv = N$.

7. The average speed of the molecules is $\frac{1}{N}\int_0^\infty v f(v)\,dv$. Show that this equals $\sqrt{\frac{8kT}{\pi m}} \approx 1.5958\sqrt{\frac{kT}{m}}$.

8. The most probable speed occurs where $f(v)$ has a maximum. Show that it is $\sqrt{\frac{2kT}{m}} \approx 1.4142\sqrt{\frac{kT}{m}}$.

Chapter 9

Polar Coordinates and Plane Curves

This chapter presents further applications of the derivative and integral. Section 9.1 describes polar coordinates. Section 9.2 shows how to compute the area of a region described in polar coordinates. Section 9.3 introduces a method of describing a curve which is especially useful in the study of motion.

The speed of an object moving along a curved path is developed in Section 9.4, where we show how to express the length of a curve as a definite integral. The area of a surface of revolution is expressed as a definite integral in Section 9.5. The chapter concludes, in Section 9.6, with a discussion of how the first and second derivatives appear when finding the curvature of a curve.

9.1 Polar Coordinates

Rectangular coordinates provide one way to describe points in the plane by pairs of numbers. Other coordinate systems provide a convenient and reliable means of describing points and other geometric objects. This section introduces polar coordinates.

The rectangular coordinates x and y describe a point P in the plane as the intersection of two perpendicular lines. Polar coordinates describe P as the intersection of a circle and a ray from its center.

Figure 9.1.1

Definition: *Polar Coordinates*

Select a point in the plane and a ray emanating from it. The point is called the **pole**, and the ray the **polar axis**. (See Figure 9.1.1(a).)

The **polar coordinates** of a point P in the xy-plane is the pair of numbers (r, θ) where r is the distance from the pole to P and θ is the angle between the pole and the point P and the polar axis. (See Figure 9.1.1(b).) Positive angles are measured counterclockwise; negative angles are measured clockwise.

When we say “The storm is 10 miles northeast of the city,” we are using polar coordinates: $r = 10$ and $\theta = \pi/4$.

Observation 9.1.1: ***Properties of Polar Coordinates***

- The point (r, θ) is on the circle of radius $|r|$ whose center is the pole.
- Angles are measured relative to the polar axis. Positive angles are measured counterclockwise; negative angles are measured clockwise.
- Polar coordinates are not unique.
  (The point $(-r, \theta+\pi)$ is the same as the point (r, θ).)
- In fact, there are an infinite number of polar coordinates for any point in the plane.
  (Changing the angle by 2π does not change the point; that is, $(r,\theta) = (r,\theta+2\pi) = (r,\theta+4\pi) = \cdots = (r,\theta+2k\pi)$, for any integer k, all represent the same point.)

Algorithm: Plotting in Polar Coordinates

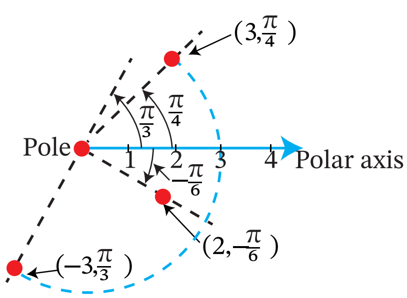
To plot the point P corresponding with polar coordinates r and θ:

- If r is positive, P is the intersection of the circle of radius r whose center is at the pole and the ray of angle θ from the pole. (See Figure 9.1.1(b).)
- If r is 0, P is the pole, no matter what θ is.
- If r is negative, P is at a distance $|r|$ from the pole on the ray directly opposite the ray of angle θ, that is, on the ray of angle $\theta+\pi$.

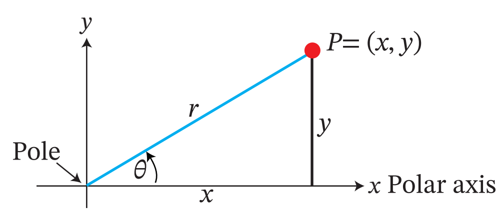
EXAMPLE 1. Plot the points (a) $\left(3, \frac{\pi}{4}\right)$, (b) $\left(2, \frac{-\pi}{6}\right)$, and (c) $\left(-3, \frac{\pi}{3}\right)$ in polar coordinates.

SOLUTION

(a) To plot $(3, \pi/4)$, go out a distance 3 on the ray of angle $\pi/4$, shown in Figure 9.1.2(a).
(b) To plot $(2, -\pi/6)$, go out a distance 2 on the ray of angle $-\pi/6$.
(c) To plot $(-3, \pi/3)$, draw the ray of angle $\pi/3$, then go a 3 units in the direction opposite from the pole.



(a)



(b)

Figure 9.1.2

It is customary to have the polar axis coincide with the positive x-axis as in Figure 9.1.2(b). This diagram also shows the relation between the rectangular coordinates (x, y) and the polar coordinates of P:

Formula 9.1.1: ***Conversions Between Polar and Rectangular Coordinates***

Polar to Rectangular	Rectangular to Polar
$x = r\cos(\theta) \quad y = r\sin(\theta)$	$r^2 = x^2 + y^2 \quad \tan(\theta) = y/x$

- These formulas hold even if r is negative.
- If r is positive, then $r = \sqrt{x^2+y^2}$.
- If $-\pi/2 < \theta < \pi/2$, then $\theta = \arctan\left(y/x\right)$.

Graphing Polar Functions: $r = f(\theta)$

Just as we may graph the set of points (x, y), where x and y satisfy an equation, we may graph the set of points (r, θ), where r and θ satisfy an equation. Although a point in the plane is specified by a unique ordered pair (x, y) in rectangular coordinates, there are many ordered pairs (r, θ) in polar coordinates that specify each point. For instance, the point whose rectangular coordinates are $(1, 1)$ has polar coordinates $(\sqrt{2}, \pi/4)$, $(\sqrt{2}, \pi/4 + 2\pi)$, $(\sqrt{2}, \pi/4 + 4\pi)$, or $(-\sqrt{2}, \pi/4 + \pi)$ and so on.

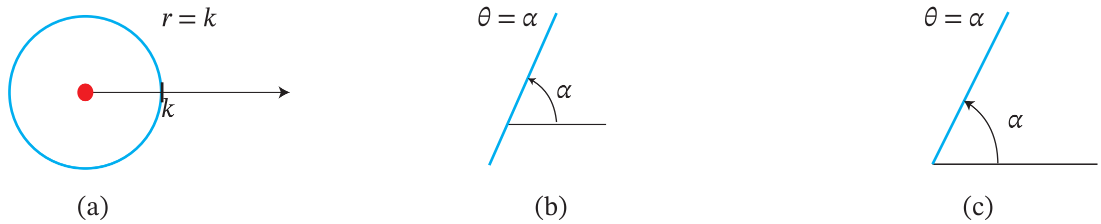


Figure 9.1.3

The simplest equation in polar coordinates is $r = k$, where k is a positive constant. Its graph is the circle of radius k, centered at the pole. (See Figure 9.1.3(a).) The graph of $\theta = \alpha$, where α is a constant, is the line through the pole with inclination α. (See Figure 9.1.3(b).) If we restrict r to be nonnegative, then $\theta = \alpha$ describes the ray (half-line) through the pole with angle α. (See Figure 9.1.3(c).)

EXAMPLE 2. Graph $r = 1 + \cos(\theta)$.

SOLUTION Since $\cos(\theta)$ has period 2π, it suffices to consider only θ in $[0, 2\pi]$. Tabulate some values:

θ	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π	$\frac{5\pi}{4}$	$\frac{3\pi}{2}$	$\frac{7\pi}{4}$	2π
$r = 1 + \cos(\theta)$	2	$1 + \frac{\sqrt{2}}{2}$ ≈ 1.7	1	$1 - \frac{\sqrt{2}}{2}$ ≈ 0.3	0	$1 - \frac{\sqrt{2}}{2}$ ≈ 0.3	1	$1 + \frac{\sqrt{2}}{2}$ ≈ 1.7	

As θ increases from 0 to π, r decreases, and as θ increases from π to 2π, r increases. The point with $\theta = 0$ is the same as the one with $\theta = 2\pi$. The graph begins to repeat itself. This heart-shaped curve, shown in Figure 9.1.4(a), is called a **cardioid**. ♦

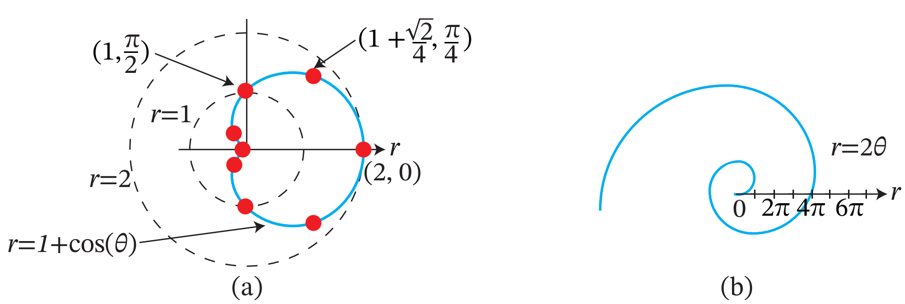


Figure 9.1.4

Spirals can be easily represented in polar coordinates: $r = 2\theta$. The next example illustrates why this is a spiral, and the simplicity with which it can be graphed.

EXAMPLE 3. Graph $r = 2\theta$ for $\theta \geq 0$.

SOLUTION Make a table:

θ	0	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	2π	$\frac{5\pi}{2}$	$\cdots$
$r = 2\theta$	0	π ≈ 3.14	2π ≈ 6.28	3π ≈ 9.43	4π ≈ 12.57	5π ≈ 15.71	$\cdots$ $\cdots$

Increasing θ by 2π does not produce the same value of r. As θ increases, r increases. The graph for $\theta \geq 0$ is a spiral, going infinitely often around the pole, as indicated in Figure 9.1.4(b). ◆

If a is a nonzero constant, the graph of $r = a\theta$ is called an **Archimedean spiral** because Archimedes was the first person to study it, finding the area within it up to any angle and also its tangent lines. The spiral with $a = 2$ is sketched in Example 3.

Polar coordinates are also convenient for describing loops arranged like the petals of a flower, as Example 4 shows.

EXAMPLE 4. Graph $r = \sin(3\theta)$.

SOLUTION The values of $\sin(3\theta)$ range from -1 to 1. For instance, when $3\theta = \pi/2$, $\sin(3\theta) = \sin(\pi/2) = 1$. That tells us that when $\theta = \pi/6$, $r = \sin(3\theta) = \sin(3(\pi/6)) = \sin(\pi/2) = 1$. This suggests that we calculate r at integer multiples of $\pi/6$, as in Table 9.1.1.

θ	0	$\frac{\pi}{18}$	$\frac{\pi}{12}$	$\frac{\pi}{9}$	$\frac{\pi}{6}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{5\pi}{6}$	π	$\frac{3\pi}{2}$	2π
3θ	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	2π	$\frac{5\pi}{2}$	3π	$\frac{9\pi}{2}$	6π
$r = \sin(3\theta)$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	0	-1	0	1	0	1	0

Table 9.1.1

The variation of r as a function of θ is shown in Figure 9.1.5(a). Because $\sin(\theta)$ has period 2π, $\sin(3\theta)$ has period $2\pi/3$.

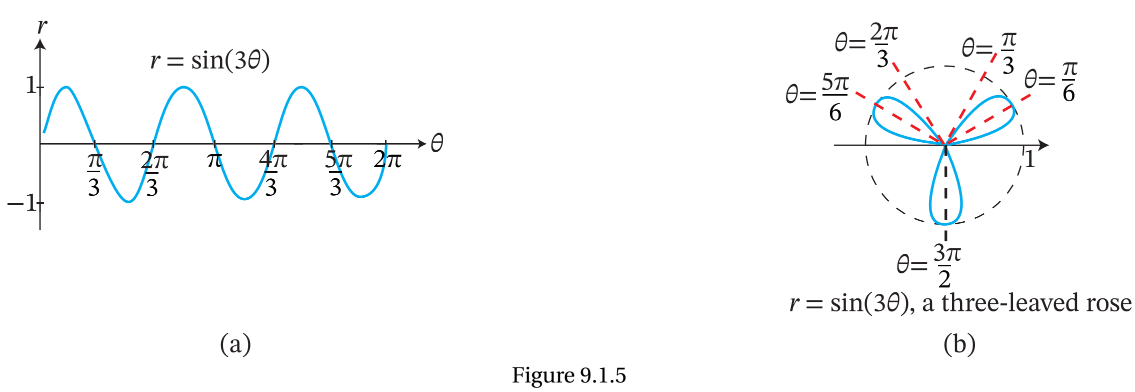


Figure 9.1.5

As θ increases from 0 to $\pi/3$, 3θ increases from 0 to π. Thus r, which is $\sin(3\theta)$, starts at 0 (for $\theta = 0$) increases to 1 (for $\theta = \pi/6$) and then decreases to 0 (for $\theta = \pi/3$). This gives one of the three loops that make up the graph of $r = \sin(3\theta)$. For θ in $[\pi/3, 2\pi/3]$, $r = \sin(3\theta)$ is negative (or 0), which is the lower loop in Figure 9.1.5(b). For θ in $[2\pi/3, \pi]$, r is again positive, and we obtain the upper left loop. Further choices of θ repeat these three loops. ◆

The graphs of $r = \sin(n\theta)$ or $r = \cos(n\theta)$ are often described as being "roses". These curves have n loops when n is an odd integer and $2n$ loops when n is an even integer. The next example illustrates the graph when n is even.

EXAMPLE 5. Graph the four-leaved rose, $r = \cos(2\theta)$.

SOLUTION To isolate one loop, find the two smallest nonnegative values of θ for which $\cos(2\theta) = 0$. They are the θ that satisfy $2\theta = \pi/2$ and $2\theta = 3\pi/2$; so $\theta = \pi/4$ and $\theta = 3\pi/4$. One leaf is described by letting θ go from $\pi/4$ to $3\pi/4$. For θ in $[\pi/4, 3\pi/4]$, 2θ is in $[\pi/2, 3\pi/2]$. Since 2θ is then a second- or third-quadrant angle, $r = \cos(2\theta)$ is negative or 0. In particular, when $\theta = \pi/2$, $\cos(2\theta)$ reaches its smallest value, -1. This loop is the bottom one in Figure 9.1.6(a). The other loops are obtained similarly. We could also sketch the graph by making a table of values.

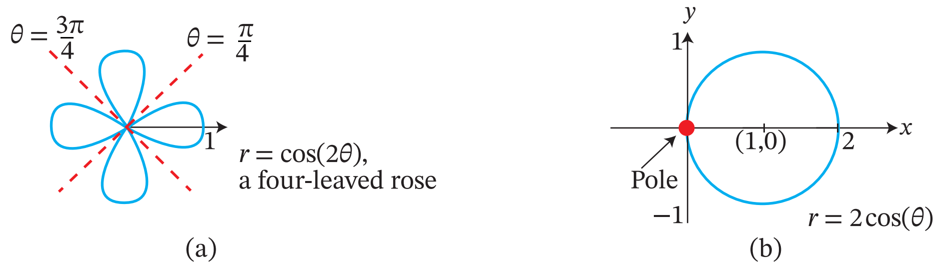


Figure 9.1.6

◆

EXAMPLE 6. Express the horizontal line $y = 2$ into polar coordinates.

SOLUTION Since $y = r\sin\theta$, the equation $y = 2$ can also be written as $r\sin\theta = 2$. Then, solving for r leads to:

$$r = \frac{2}{\sin(\theta)} = 2\csc(\theta).$$

◆

NOTE: While the polar representation of a horizontal line is more complicated than the rectangular version of this equation, it is still sometimes useful or necessary.

EXAMPLE 7. Transform the polar equation $r = 2\cos(\theta)$ into rectangular coordinates and graph it.

SOLUTION Since $r^2 = x^2 + y^2$ and $r\cos(\theta) = x$, multiply the equation $r = 2\cos\theta$ by r, obtaining $r^2 = 2r\cos(\theta)$. Hence $x^2 + y^2 = 2x$ and so $x^2 - 2x + y^2 = 0$. Then complete the square to obtain $(x-1)^2 + y^2 = 1$. The graph is a circle of radius 1 with center at $(1, 0)$. It is graphed in Figure 9.1.6(b). ◆

The step in Example 7 where we multiply by r deserves some attention. If $r = 2\cos(\theta)$, then $r^2 = 2r\cos(\theta)$. However, if $r^2 = 2r\cos(\theta)$, it does not follow that $r = 2\cos(\theta)$. We can cancel the r only when r is not 0. If $r = 0$, it is true that $r^2 = 2r\cos(\theta)$, but it not necessarily true that $r = 2\cos(\theta)$. Since $r = 0$ satisfies the equation $r^2 = 2r\cos\theta$, the pole is on the curve $r^2 = 2r\cos\theta$. Luckily, it is also on the original curve $r = 2\cos(\theta)$, since $\theta = \pi/2$ makes $r = 0$. Hence the graphs of $r^2 = 2r\cos(\theta)$ and $r = 2\cos(\theta)$ are the same.

However, as you may check, the graphs of $r = 2 + \cos(\theta)$ and $r^2 = r(2 + \cos(\theta))$ are not the same. The origin lies on the second curve, but not on the first.

The Intersection of Two Polar Curves

Finding the intersection of two curves in polar coordinates is complicated because a point has many descriptions in polar coordinates.

EXAMPLE 8. Find all intersections of the curves $r = 1 - \cos(\theta)$ and $r = \cos(\theta)$.

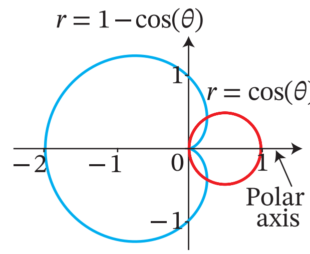


Figure 9.1.7

SOLUTION Graph the two curves. The curve $r = \cos(\theta)$ is a circle with center at $(1/2, 0)$ and radius $1/2$. The curve $r = 1 - \cos(\theta)$ is a cardioid. Both curves are shown in Figure 9.1.7. It appears that there are three points of intersection.

A point of intersection is produced when one value of θ yields the same value of r in both equations, that is, when $1 - \cos(\theta) = \cos(\theta)$, so $\cos(\theta) = 1/2$. Thus $\theta = \pm\pi/3$ or any angle differing from these by $2n\pi$, n an integer. This gives two of the three points, but it fails to give the origin. Why?

How does the origin get to be on the circle $r = \cos(\theta)$? Because when $\theta = \pi/2$, $r = 0$. How does it get to be on the cardioid $r = 1 - \cos(\theta)$? Because when $\theta = 0$, $r = 0$. The origin lies on both curves, but for different angles. This is why this intersection was not found by simply equating $1 - \cos(\theta)$ and $\cos(\theta)$. ♦

Warning: Finding Intersections of Polar Curves

When looking for the intersections of two curves, $r = f(\theta)$ and $r = g(\theta)$ in polar coordinates, it is necessary to consider the origin separately. (Recall that the origin has polar coordinates $(0, \theta)$ for any value of θ.)

Two curves may also intersect at other points not obtainable by setting $f(\theta) = g(\theta)$. This possibility is due to the fact the point (r, θ) is the same as the points $(r, \theta + 2n\pi)$ and $(-r, \theta + (2n+1)\pi)$ for any integer n. The safest procedure is to graph the two curves first, identify the intersections in the graph, and then see why the curves intersect there.

Summary

We introduced polar coordinates and showed how to graph curves given with equation $r = f(\theta)$. Some common polar curves are listed in Table 9.1.2.

Polar Equation	Curve
$r = a$, $a > 0$	circle of radius a, center at pole
$r = 1 + \cos(\theta)$	cardioid
$r = a\theta$, $a > 0$	Archimedean spiral (traced clockwise)
$r = \sin(n\theta)$, n odd	n-leafed rose (one loop symmetric about $\theta = \pi/n$)
$r = \sin(n\theta)$, n even	$2n$-leafed rose
$r = \cos(n\theta)$, n odd	n-leafed rose (one loop symmetric about $\theta = 0$)
$r = \cos(n\theta)$, n even	$2n$-leafed rose
$r = a\csc(\theta)$	the horizontal line $y = a$
$r = a\sec(\theta)$	the vertical line $x = a$
$r = a\cos(\theta)$, $a > 0$	circle of radius $a/2$ through pole and $(a, 0)$
$r = a\sin(\theta)$, $a > 0$	circle of radius $a/2$ through pole and $(a, \pi/2)$

Table 9.1.2

To find the intersection of two curves in polar coordinates, first graph them.

EXERCISES for Section 9.1

1. Plot the points with polar coordinates: (a) $\left(1, \frac{\pi}{6}\right)$, (b) $\left(2, \frac{\pi}{3}\right)$, (c) $\left(2, \frac{-\pi}{3}\right)$, (d) $\left(-2, \frac{\pi}{3}\right)$, (e) $\left(2, \frac{7\pi}{3}\right)$, and (f) $\left(0, \frac{\pi}{4}\right)$.

2. Find the rectangular coordinates of the points in Exercise 1.

3. (a) Give three pairs of polar coordinates (r, θ) with $r > 0$ for the point whose polar coordinates are $\left(3, \frac{\pi}{4}\right)$.
(b) Repeat (a) but with $r < 0$.

4. Find polar coordinates (r, θ) with $0 \le \theta < 2\pi$ and r positive, for the points whose rectangular coordinates are (a) $\left(\sqrt{2}, \sqrt{2}\right)$, (b) $\left(-1, \sqrt{3}\right)$, (c) $(-5, 0)$, (d) $\left(-\sqrt{2}, -\sqrt{2}\right)$, (e) $(0, -3)$, and (f) $(1, 1)$.

In Exercises 5 to 8 transform the equation into one in rectangular coordinates.

5. $r = \sin(\theta)$ **6.** $r = \csc(\theta)$ **7.** $r = 4\cos(\theta) + 5\sin(\theta)$ **8.** $r = \dfrac{3}{(4\cos(\theta) + 5\sin(\theta)}$

In Exercises 9 to 12 transform the equation into one in polar coordinates.

9. $x = -2$ **10.** $y = x^2$ **11.** $xy = 1$ **12.** $x^2 + y^2 = 4x$

In Exercises 13 to 22 graph the given equations.

13. $r = 1 + \sin\theta$ **14.** $r = 3 + 2\cos(\theta)$ **15.** $r = e^{-\theta/\pi}$ **16.** $r = 4^{\theta/\pi}$ $(\theta > 0)$ **17.** $r = \cos(3\theta)$
18. $r = \sin(2\theta)$ **19.** $r = 2$ **20.** $r = 3$ **21.** $r = 3\sin(\theta)$ **22.** $r = -2\cos(\theta)$

---

23. Suppose $r = \dfrac{1}{\theta}$ for $\theta > 0$.
(a) What happens to the y-coordinate of (r, θ) as $\theta \to 0^+$?
(b) What happens to the x-coordinate of (r, θ) as $\theta \to 0^+$?
(c) What happens to the y-coordinate of (r, θ) as $\theta \to \infty$?
(d) What happens to the x-coordinate of (r, θ) as $\theta \to \infty$?
(e) Use the results of (a)-(d) to help sketch the curve.

24. Suppose $r = \dfrac{1}{\sqrt{\theta}}$ for $\theta > 0$.
(a) What happens to the y-coordinate of (r, θ) as $\theta \to 0^+$?
(b) What happens to the x-coordinate of (r, θ) as $\theta \to 0^+$?
(c) What happens to the y-coordinate of (r, θ) as $\theta \to \infty$?
(d) What happens to the x-coordinate of (r, θ) as $\theta \to \infty$?
(e) Use the results of (a)-(d) to help sketch the curve.

In Exercises 25 to 30, find the intersections of the curves after drawing them.

25. $r = 1 + \cos(\theta)$ and $r = \cos(\theta) - 1$ **26.** $r = \sin(2\theta)$ and $r = 1$ **27.** $r = \sin(3\theta)$ and $r = \cos(3\theta)$
28. $r = 2\sin(2\theta)$ and $r = 1$ **29.** $r = \sin(\theta)$ and $r = \cos(2\theta)$ **30.** $r = \cos(\theta)$ and $r = \cos(2\theta)$

A curve $r = 1 + a\cos(\theta)$ (or $r = 1 + a\sin(\theta)$) is called a **limaçon**. Its shape depends on the choice of a. For $a = 1$ we have the cardioid of Example 2. Exercises 31 to 33 concern other choices of a.

31. Graph $r = 1 + 2\cos(\theta)$.

When $|a| > 1$ the graph of $r = 1 + a\cos(\theta)$ crosses itself and forms two loops.

32. Graph $r = 1 + \frac{1}{2}\cos(\theta)$.

33. Let $r = 1 + a\cos(\theta)$, where $0 \le a \le 1$.

(a) Relative to the same polar axis, graph the curves corresponding to $a = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$, and 1.

(b) For $a = \frac{1}{4}$ the graph in (a) is convex, but not for $a = 1$. Show the curve is not convex for $\frac{1}{2} < a \le 1$.

---

34. (a) Graph $r = 3 + \cos(\theta)$. (b) Find the point on the graph in (a) that has the maximum y coordinate.

35. Find the y coordinate of the highest point on the right-hand leaf of the four-leaved rose $r = \cos(2\theta)$.

36. Graph $r^2 = \cos(2\theta)$. The curve is called a **lemniscate**.

NOTE: If $\cos(2\theta) < 0$, r is not defined. If $\cos(2\theta) > 0$, there are two values of r, $r = \sqrt{\cos(2\theta)}$ and $r = -\sqrt{\cos(2\theta)}$.

The graphs of $r = \frac{1}{1 + e\cos(\theta)}$ are **conic sections** (in polar coordinates). Here, the parameter e denotes **eccentricity**, not Euler's number. When $0 < e < 1$, the graph is an ellipse; when $e = 1$ the graph is a parabola; and when $e > 1$ the graph is a hyperbola. Exercises 37 to 38 concern graphs of conic sections in polar coordinates.

37. (a) Graph $r = \frac{1}{1 + \cos(\theta)}$. (b) Find an equation in rectangular coordinates for the curve in (a).

38. (a) Graph $r = \frac{1}{1 - \frac{1}{2}\cos(\theta)}$. (b) Find an equation in rectangular coordinates for the curve in (a).

---

39. Where do the spirals $r = \theta$ and $r = 2\theta$ intersect? (Assume $r \ge 0$.)

9.2 Computing Area in Polar Coordinates

In Section 6.1 we saw how to compute the area of a region if the lengths of parallel cross sections are known. Sums based on rectangles led to the formula

$$\text{Area} = \int_a^b c(x)\,dx,$$

where $c(x)$ denoted the cross-sectional length. In polar coordinates sectors of circles, not rectangles, provide an estimate of area.

Let R be a region in the plane and P a point inside it that we take as the pole of a polar coordinate system. Assume that the distance r from P to a point on the boundary of R is known as a function $r = f(\theta)$. Also, assume that any ray from P meets the boundary of R just once, as in Figure 9.2.1(a).

The cross sections made by the rays from P are not parallel. Like the spokes in a wheel, they meet at the point P. It would be unnatural to use rectangles to estimate the area, but it is reasonable to use sectors of circles that have P as a common vertex.

In a circle of radius r a sector of central angle θ has area $\theta\ r^2/2$. (See Figure 9.2.1(b).) This formula plays the same role now as the formula for the area of a rectangle did in Section 6.1.

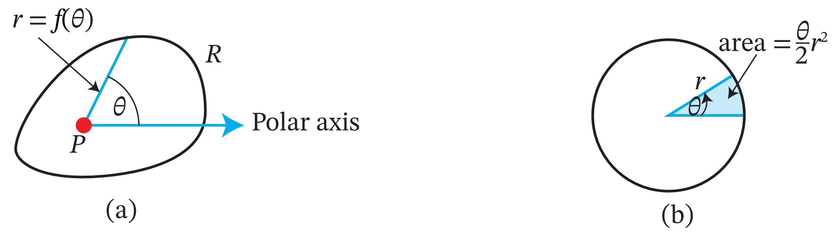


Figure 9.2.1

Area in Polar Coordinates

Assume $f(\theta) \geq 0$. Let R be the region bounded by the rays $\theta = \alpha$ and $\theta = \beta$ and by the curve $r = f(\theta)$, as shown in Figure 9.2.2(a). To obtain a **local estimate** for the area of R, consider the portion of R between the rays corresponding to the angles θ and $\theta + d\theta$, where $d\theta$ is a small positive number. (See Figure 9.2.2(b).)

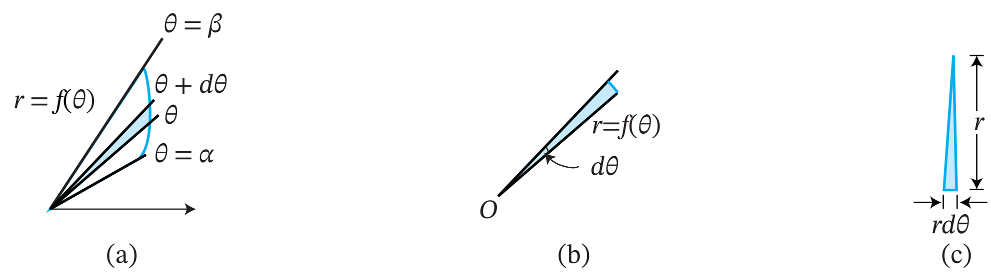


Figure 9.2.2

The area of the narrow wedge shaded in Figure 9.2.2(b) is approximately that of a sector of a circle of radius $r = f(\theta)$ and angle $d\theta$, shown in Figure 9.2.2(c). An estimate for this area is $f(\theta)^2\ d\theta/2$. Having found the local estimate of area, we conclude that the area of R is

$$\int_{\alpha}^{\beta} \frac{1}{2} f(\theta)^2\ d\theta \qquad \text{or} \qquad \int_{\alpha}^{\beta} \frac{1}{2} r^2\ d\theta. \tag{9.2.1}$$

NOTE: Equation (9.2.1) reappears in Section 15.1 (and in CIE 22, Newton's Law Implies Kepler's Three Laws, at the end of Chapter 15) in the context of motion of satellites and planets.

It may seem surprising to find $f(\theta)^2$ in the integrand. But area has the dimension of length squared. Since θ is dimensionless, because it is the length of a circular arc divided by the length of the radius, $d\theta$ is also dimensionless. Hence $f(\theta)\ d\theta$ has the dimension of length and $f(\theta)^2 d\theta/2$ has the dimension of area. For rectangular coordinates, in $f(x)\ dx$, both $f(x)$ and dx have the dimension of length, one along the y-axis, the other along the x-axis, so $f(x)\ dx$ has the dimension of area.

Observation 9.2.1: *Memory Device for Area of a Sector*

To remember the area of the sector in Figure 9.2.2(b), think of it as a triangle of height r and base $r\ d\theta$, as shown in Figure 9.2.2(c).

$$\text{Area} = \frac{1}{2} \cdot \underbrace{r}_{\text{height}} \cdot \underbrace{r d\theta}_{\text{base}} = \frac{1}{2} r^2 d\theta.$$

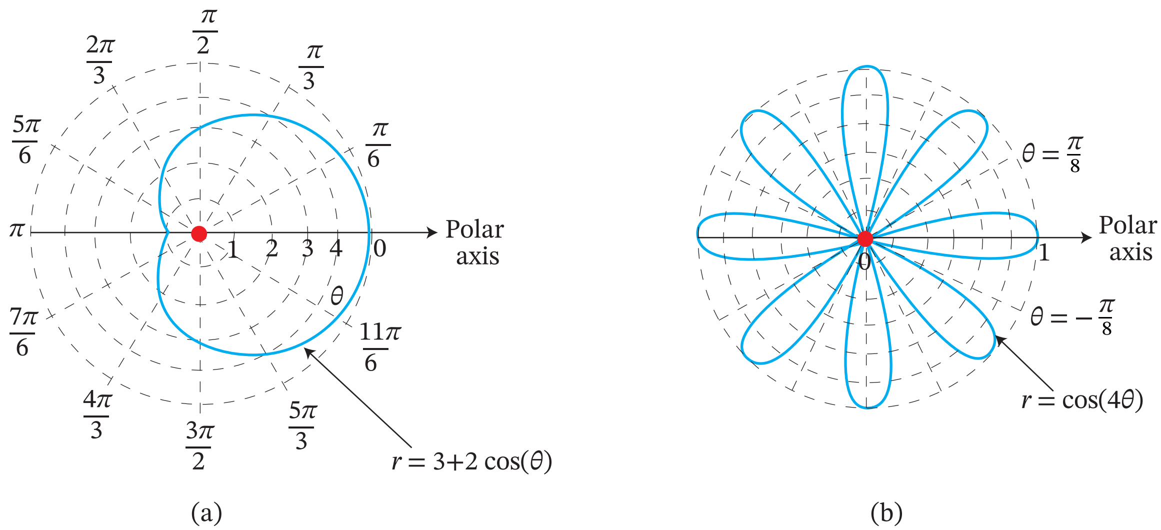


Figure 9.2.3

EXAMPLE 1. The region bounded by the curve $r = 3 + 2\cos(\theta)$ is shown in Figure 9.2.3(a). Find the region's area.

SOLUTION The curve is traced once for $0 \le \theta \le 2\pi$. By the formula just obtained, its area is

$$\begin{aligned}
\int_0^{2\pi} \frac{1}{2}(3+2\cos(\theta))^2 d\theta &= \frac{1}{2}\int_0^{2\pi} (9+12\cos(\theta)+4\cos^2(\theta))d\theta && \text{(expanding the square)}\\
&= \frac{1}{2}\int_0^{2\pi} (9+12\cos(\theta)+2(1+\cos(2\theta))\ d\theta && \text{(applying double-angle formula)}\\
&= \frac{1}{2}(9\theta+12\sin(\theta)+2\theta+\sin(2\theta))\Big|_0^{2\pi} && \text{(FTC I)}\\
&= 11\pi \approx 34.55752.
\end{aligned}$$

♦

EXAMPLE 2. Find the area of the region inside one of the eight loops of the eight-leaved rose $r = \cos(4\theta)$.

SOLUTION To graph one of the loops, start with $\theta = 0$. For that angle, $r = \cos(4\cdot 0) = \cos 0 = 1$. The point $(r,\theta) = (1,0)$ is the outer tip of a loop. As θ increases from 0 to $\pi/8$, $\cos(4\theta)$ decreases from $\cos(0) = 1$ to $\cos(\pi/2) = 0$. One of the eight loops is therefore bounded by the rays $\theta = \pi/8$ and $\theta = -\pi/8$, as shown in Figure 9.2.3(b). The area of this loop, which is bisected by the polar axis, is

$$\begin{aligned}
\int_{-\pi/8}^{\pi/8} \frac{1}{2}r^2\ d\theta = \int_{-\pi/8}^{\pi/8} \frac{1}{2}\cos^2(4\theta)\ d\theta &= \int_0^{\pi/8} \cos^2(4\theta)\ d\theta && \text{(even integrand over a symmetric interval)}\\
&= \frac{1}{2}\int_0^{\pi/8} (1+\cos(8\theta))\ d\theta && \text{(applying double-angle formula)}\\
&= \frac{1}{2}\left(\theta+\frac{\sin(8\theta)}{4}\right)\Big|_0^{\pi/8} && \text{(FTC I)}\\
&= \frac{1}{2}\left(\frac{\pi}{8}+\frac{\sin(\pi)}{8}\right) - 0 = \frac{\pi}{16} \approx 0.19635.
\end{aligned}$$

♦

Area between Two Polar Curves

Assume that $r = f(\theta)$ and $r = g(\theta)$ describe two curves in polar coordinates and that $f(\theta) \geq g(\theta) \geq 0$ for θ in $[\alpha, \beta]$. Let R be the region between them and the rays $\theta = \alpha$ and $\theta = \beta$, as shown in Figure 9.2.4(a).

The area of R is obtained by subtracting the area within the inner curve, $r = g(\theta)$, from the area within the outer curve, $r = f(\theta)$.

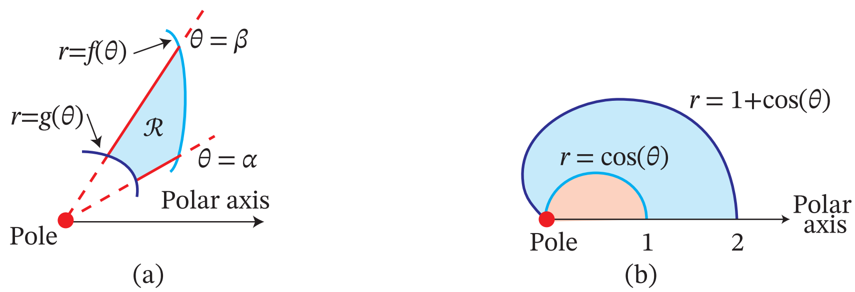


Figure 9.2.4

EXAMPLE 3. Find the area of the top half of the region inside the cardioid $r = 1 + \cos(\theta)$ and outside the circle $r = \cos(\theta)$. (See Figure 9.2.4(b).)

SOLUTION The top half of the cardioid is swept out by $r = 1 + \cos(\theta)$ as θ goes from 0 to π so its area is

$$\begin{aligned}
\frac{1}{2}\int_0^{\pi} (1+\cos(\theta))^2\, d\theta &= \frac{1}{2}\int_0^{\pi} \left(1+2\cos(\theta)+\cos^2(\theta)\right) d\theta && \text{(expanding the square)}\\
&= \frac{1}{2}\int_0^{\pi} \left(1+2\cos(\theta)+\frac{1+\cos(2\theta)}{2}\right) d\theta && \text{(applying double-angle formula)}\\
&= \frac{1}{2}\int_0^{\pi} \left(\frac{3}{2}+2\cos(\theta)+\frac{\cos(2\theta)}{2}\right) d\theta && \text{(collecting like terms)}\\
&= \frac{1}{2}\left(\frac{3\theta}{2}+2\sin(\theta)+\frac{\sin(2\theta)}{4}\right)\Bigg|_0^{\pi} && \text{(FTC I)}\\
&= \frac{3\pi}{4}
\end{aligned}$$

The top half of $r = \cos(\theta)$ is the top half of a circle centered at $(1/2, 0)$ with radius $r = 1/2$; the area of this semicircle is

$$\frac{1}{2}\pi\left(\frac{1}{2}\right)^2 = \frac{\pi}{8}.$$

Thus the area between these two curves is $3\pi/4 - \pi/8 = 5\pi/8 \approx 1.96350$. ♦

Warning

When finding the area between two polar curves, it is often the case that the can not have been written as a single definite integral because the two curves are parameterized with different intervals of θ.

For example, in Example 3, the cardioid uses $0 \leq \theta \leq \pi$ and the semicircle uses $0 \leq \theta \leq \pi/2$.

Summary

In this section we saw how to find the area within a curve $r = f(\theta)$ and the rays $\theta = \alpha$ and $\theta = \beta$. The method uses the local approximation by a narrow sector of radius r and angle $d\theta$, which has area $r^2\, d\theta/2$. This approximation leads to the formula,

$$\text{Area} = \int_\alpha^\beta \frac{1}{2} r^2\, d\theta.$$

It is prudent to remember the estimate rather than the area formula. To remember this estimate, remember that the sector resembles a triangle of height r and base $d\theta$. This, in particular, explains the factor 1/2 in the integrand.

EXERCISES for Section 9.2

In Exercises 1 to 6, draw the region enclosed by the curve and rays and then find its area.

1. $r = 2\theta$, $\alpha = 0$, $\beta = \frac{\pi}{2}$

2. $r = \sqrt{\theta}$, $\alpha = 0$, $\beta = \pi$

3. $r = \frac{1}{1+\theta}$, $\alpha = \frac{\pi}{4}$, $\beta = \frac{\pi}{2}$

4. $r = \sqrt{\sin(\theta)}$, $\alpha = 0$, $\beta = \frac{\pi}{2}$

5. $r = \tan(\theta)$, $\alpha = 0$, $\beta = \frac{\pi}{4}$

6. $r = \sec(\theta)$, $\alpha = \frac{\pi}{6}$, $\beta = \frac{\pi}{4}$

In each of Exercises 7 to 16 draw the region bounded by the curve(s) and find its area.

7. $r = 2\cos(\theta)$

8. $r = e^\theta$, $\theta = 0$, and $\theta = \pi$

9. inside $r = 3 + 3\sin(\theta)$ (cardioid) and outside $r = 3$.

10. both loops of $r = \sqrt{\cos(2\theta)}$

11. one loop of $r = \sin(3\theta)$

12. one loop of $r = \cos(2\theta)$

13. inside one loop of $r = 2\cos(2\theta)$ and outside $r = 1$

14. inside $r = 1 + \cos(\theta)$ and outside $r = \sin(\theta)$

15. inside $r = \sin(\theta)$ and outside $r = \cos(\theta)$

16. inside $r = 4 + \sin(\theta)$ and outside $r = 3 + \sin(\theta)$

---

17. Sketch the graph of $r = 4 + \cos(\theta)$. Is it a circle? How do you know?

Figure 9.2.5

18.

(a) Show that the area of the triangle in Figure 9.2.5(a) is $\int_0^\beta \frac{1}{2}\sec^2(\theta)\, d\theta$.

(b) From (a) and the fact that the area of a triangle is $\frac{1}{2}$(base)(height), show that $\tan(\beta) = \int_0^\beta \sec^2(\theta)\, d\theta$.

(c) Using the equation in (b), obtain another proof that $D(\tan(x)) = \sec^2(x)$.

19. Show that the area of the crescent between the two circular arcs, shaded cyan in Figure 9.2.5(b), is equal to the area of square $ABCD$, shaded pink.

Historical Note: Straightedge and Compass Constructions

The ideas in Exercise 19 have inspired mathematicians from the time of the ancient Greeks to try to find a method using only straightedge and compass for constructing a square whose area equals that of a circle. This was not proved impossible until the end of the nineteenth century when it was shown that π is not the root of a nonzero polynomial with integer coefficients.

20. (a) Graph $r = \frac{1}{\theta}$ for $0 < \theta \leq \frac{\pi}{2}$.

(b) Is the area of the region bounded by the curve drawn in (a) and the rays $\theta = 0$ and $\theta = \frac{\pi}{2}$ finite or infinite?

21. (a) Sketch the curve $r = \frac{1}{1+\cos(\theta)}$.

(b) What is its equation in rectangular coordinates?

(c) Find the area of the region bounded by the curve in (a) and the rays $\theta = 0$ and $\theta = \frac{3\pi}{4}$.

(d) Solve (c) using rectangular coordinates and the equation in (b).

22. Use Simpson's method to estimate the area of the region between $r = \sqrt[3]{1+\theta^2}$, $\theta = 0$, and $\theta = \frac{\pi}{2}$, correct to three decimal places.

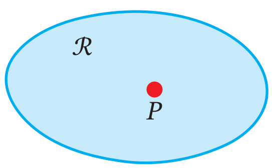


Figure 9.2.6

23. Estimate the area of the region in the first quadrant that is bounded by $r = e^{\theta}$, $r = 2\cos(\theta)$ and $\theta = 0$.

24. Figure 9.2.6 shows a point P inside a convex region $\mathscr{R}$.

(a) Assume that P cuts each chord through it into two intervals of equal length. Does each chord through P cut $\mathscr{R}$ into two regions of equal area?

(b) Assume that each chord through P cuts $\mathscr{R}$ into two regions of equal area. Must P cut each chord through P into two intervals of equal length?

25. Assume $\mathscr{R}$ is a convex region in the plane and P is a point on its boundary. Assume that every chord of $\mathscr{R}$ that has an end at P has length not more than 1.

(a) Draw several examples of such an $\mathscr{R}$. (b) Make a general conjecture about the area $\mathscr{R}$. (c) Prove your conjecture.

26. (a) Show that a line through the origin intersects the region bounded by the curve in Example 1 in a segment of length 6.

(b) A line through the center of a disk of radius 3 also intersects the disk in a segment of length 6. Does it follow that the disk and the region in Example 1 have the same areas?

27. Let P be a point inside the convex region $\mathscr{R}$. Assume that each chord through P has length 1. How small can the area of $\mathscr{R}$ be? How large?

28. Let $\mathscr{R}$ be a convex region in the plane and let P be a point in $\mathscr{R}$. If the length of each chord that passes through P is known, can you determine the area of $\mathscr{R}$ (a) if P is on the border of $\mathscr{R}$? (b) if P is in the interior of $\mathscr{R}$?

Exercises 29 to 31 are related. This set of three exercises was contributed by Rick West.

29. The graph of $r = \cos(n\theta)$ has $2n$ loops when n is even. Find the total area within them.

30. The graph of $r = \cos(n\theta)$ has n loops when n is odd. Find the total area within them.

31. Find the total area of the petals within the curve $r = \sin(n\theta)$, where n is a positive integer.

9.3 Parametric Equations

We have considered curves described in three forms: y is a function of x, x and y are related implicitly, and r is a function of θ. Sometimes a curve is described by giving x and y as functions of a third variable. We now look at this description, which arises in the study of motion. It was the basis for Uniform Sprinkler CIE (CIE 7) at the end of Chapter 5.

para meaning "together", *meter* meaning "measure".

Two Examples

EXAMPLE 1. A ball is thrown horizontally out of a window with a speed of 32 feet per second falls in a curved path. Air resistance disregarded, its position after t seconds is given by $x = 32t$, $y = -16t^2$ relative to the coordinate system in Figure 9.3.1(a). The curve is completely described, not by expressing y as a function of x, but by expressing x and y as functions of a third variable t. The third variable is called a **parameter**.

The pair of equations $x = 32t$, $y = -16t^2$ are called **parametric equations** for the curve that represents the thrown ball's path.

In this example it is easy to eliminate t and so find a direct relation between x and y:

$$t = \frac{x}{32},$$

so

$$y = -16\left(\frac{x}{32}\right)^2 = -\frac{16}{(32)^2}x^2 = -\frac{1}{64}x^2.$$

The path is part of the parabola $y = -x^2/64$. ♦

It is not always possible, as it was in Example 1, to find an explicit (and simple) formula for y as a function of x. An advantage of parametric equations is that they can provide a description of a curve that has no simple representation as $y = f(x)$. In Example 2 elimination of the parameter would lead to a complicated equation involving x and y.

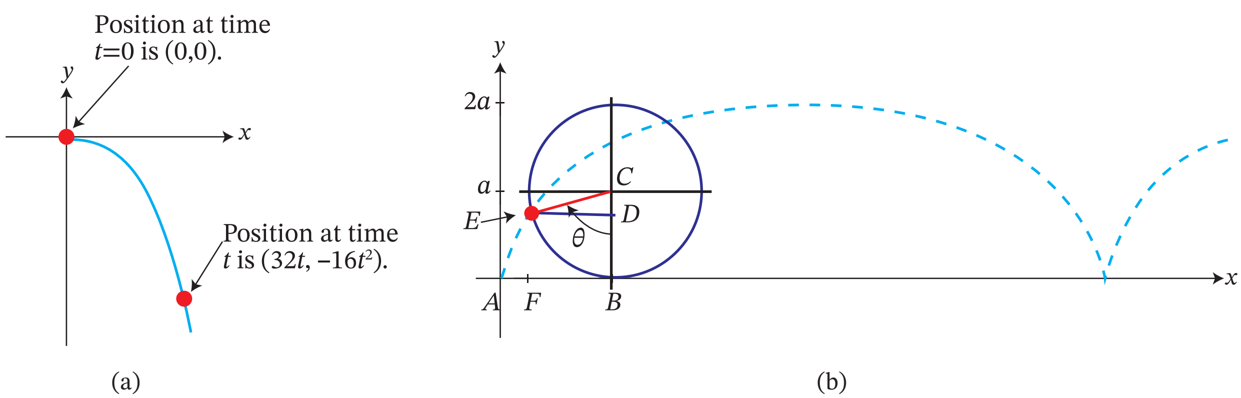


Figure 9.3.1

EXAMPLE 2. As a bicycle wheel of radius a rolls along, a tack stuck in its circumference traces out a curve called a **cycloid**, which consists of a sequence of arches, one arch for each revolution of the wheel. (See Figure 9.3.1(b).)

Find the position of the tack as a function of the angle θ through which the wheel turns.

SOLUTION Assume that the tack is initially at the bottom of the wheel. That is, at $t = 0$, the red dot in Figure 9.3.1(b) is on the ground at point $(0,0)$. At time t, after the wheel has rotated through an angle θ, the tack's position is point E.

The x coordinate of the tack, corresponding to θ, is

$$\overline{AF} = \overline{AB} - \overline{ED} = a\theta - a\sin(\theta)$$

and the y coordinate is

$$\overline{EF} = \overline{BC} - \overline{CD} = a - a\cos(\theta).$$

Thus the position of the tack as a function of the parameter θ is

$$x = a\theta - a\sin(\theta), \quad y = a - a\cos(\theta).$$

Eliminating θ leads to a complicated relation between x and y. See also Exercise 36. ♦

Observation 9.3.1: *Converting* $y = f(x)$ *to Parametric Form*

Any curve given in the form $y = f(x)$ can be described parametrically. For instance, for $y = e^x + x$ we can introduce a parameter t equal to x and write $x = t$, $y = e^t + t$. This may seem artificial, but it will be useful in the next section in order to apply results for curves expressed by means of parametric equations to curves given in the form $y = f(x)$.

Finding $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ for a Parametric Curve

How can we find the slope of a curve described parametrically as

$$x = g(t), \quad y = h(t)?$$

An often difficult, perhaps impossible, approach is to solve $x = g(t)$ for t as a function of x and substitute it into the equation $y = h(t)$, thus expressing y explicitly in terms of x and then differentiating the result to find dy/dx. Fortunately, there is an easier way. Assume that y is a differentiable function of x. Then, by the chain rule,

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt},$$

from which it follows that

Definition: *Slope of a Parametric Curve*

The slope of the graph of the parametric curve $x = x(t)$, $y = y(t)$, is given by

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}. \tag{9.3.1}$$

EXAMPLE 3. At what angle does the arch of the cycloid in Example 2 meet the x-axis at the origin?

SOLUTION The parametric equations of the cycloid, expressed in terms of the parameter θ, are

$$x = a\theta - a\sin(\theta) \quad \text{and} \quad y = a - a\cos(\theta).$$

Then $dx/d\theta = a - a\cos(\theta)$ and $dy/d\theta = a\sin(\theta$. Consequently,

$$\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta} = \frac{a\sin(\theta)}{a - a\cos(\theta)} = \frac{\sin(\theta)}{1 - \cos(\theta)}.$$

When $\theta = 0$, $(x, y) = (0,0)$ and dy/dx is not defined because $dx/d\theta = 0$. But, when θ just a little larger than 0, (x, y) is near the origin and the slope of the cycloid at $(0,0)$ can be found by looking at the limit of the slope, which is $\sin\theta/(1-\cos(\theta))$, as $\theta \to 0^+$. As this limit is indeterminate with form $0/0$, l'Hôpital's rule is an option:

$$\lim_{\theta\to 0^+} \frac{\sin(\theta)}{1-\cos(\theta)} \stackrel{\text{l'H}}{=} \lim_{\theta\to 0^+} \frac{\cos(\theta)}{\sin(\theta)} = \infty.$$

Thus the cycloid comes in vertically at the origin, as shown in Figure 9.3.1(b). ♦

We assume that in (9.3.1) dx/dt is not 0. To obtain d^2y/dx^2 just replace y in (9.3.1) by dy/dx.

Definition: *Second Derivative of a Parameterized Curve*

For a parametric curve $x = x(t)$, $y = y(t)$, the second derivative of y as a function of x is given by

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{\frac{d}{dt}\left(\frac{dy}{dx}\right)}{dx/dt}.$$

EXAMPLE 4. Find $\dfrac{d^2y}{dx^2}$ for the cycloid of Example 2.

SOLUTION In Example 3 we found $dy/dx = \sin(\theta)/(1-\cos(\theta))$. To find d^2y/dx^2, first compute

$$\begin{aligned}\frac{d}{d\theta}\left(\frac{dy}{dx}\right) &= \frac{(1-\cos(\theta))\cos(\theta) - \sin(\theta)(\sin(\theta))}{(1-\cos(\theta))^2}\\ &= \frac{\cos(\theta)-1}{(1-\cos(\theta))^2}\\ &= \frac{-1}{1-\cos(\theta)}.\end{aligned}$$

Then, since $dx/d\theta = a - a\cos(\theta)$,

$$\begin{aligned}\frac{d^2y}{dx^2} &= \frac{\frac{d}{d\theta}\left(\frac{dy}{dx}\right)}{dx/d\theta}\\ &= \frac{\frac{-1}{1-\cos(\theta)}}{a-a\cos(\theta)}\\ &= \frac{-1}{a(1-\cos(\theta))^2}.\end{aligned}$$

The denominator is zero when $\cos(\theta) = 1$, that is, when θ is an integer multiple of 2π. For all other values of θ the denominator is positive, so the quotient is negative. This agrees with Figure 9.3.1(b), which shows each arch of the cycloid is concave down except at the beginning and end of each revolution of the wheel. ♦

Summary

This section described parametric equations, where x and y are given as functions of a third variable, often time (t) or angle (θ). We also showed how to compute dy/dx and d^2y/dx^2:

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} \qquad \text{and} \qquad \frac{d^2y}{dx^2} = \frac{\frac{d}{dt}\left(\frac{dy}{dx}\right)}{dx/dt}.$$

EXERCISES for Section 9.3

For each parametric curve given in Exercises 1 to 4,

(a) Fill in the table

(b) Plot the points (x, y) obtained in (a)

(c) Graph the curve

(d) Eliminate the parameter, t, to find an equation for the curve in terms of x and y.

1. $x = 2t + 1$, $y = t - 1$,

t	-2	-1	0	1	2
x					
y					

2. $x = t + 1$, $y = t^2$.

t	-2	-1	0	1	2
x					
y					

3. $x = t^2$, $y = t^2 + t$.

t	-3	-2	-1	0	1	2	3
x							
x							
y							

4. $x = 2\cos(t)$, $y = 3\sin(t)$.

t	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π	$\frac{5\pi}{4}$	$\frac{3\pi}{2}$	$\frac{7\pi}{4}$	2π
x									
y									

In Exercises 5 to 8 express the curves parametrically with parameter t.

5. $y = \sqrt{1 + x^3}$

6. $y = \tan^{-1}(3x)$

7. $r = \cos^2(\theta)$

8. $r = 3 + \cos(\theta)$

In Exercises 9 to 14 find $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$.

9. $x = t^3 + t$, $y = t^7 + t + 1$

10. $x = \sin(3t)$, $y = \cos(4t)$

11. $x = 1 + \ln(t)$, $y = t\ln(t)$

12. $x = e^{t^2}$, $y = \tan(t)$

13. $r = \cos(3\theta)$

14. $r = 2 + 3\sin(\theta)$

In Exercises 15 to 16 find the equation of the tangent line to the curve at the point.

15. $x = t^3 + t^2$, $y = t^5 + t$; $(2, 2)$

16. $x = \dfrac{t^2 + 1}{t^3 + t^2 + 1}$, $y = \sec(3t)$; $(1, 1)$

In Exercises 17 and 18 find $\frac{d^2y}{dx^2}$.

17. $x = t^3 + t + 1$, $y = t^2 + t + 2$

18. $x = e^{3t} + \sin(2t)$, $y = e^{3t} + \cos\left(t^2\right)$

---

19. For which values of t is the curve in Exercise 17 concave up? concave down?

20. Let $x = t^3 + 1$ and $y = t^2 + t + 1$. For which values of t is the curve concave up? concave down?

21. Find the slope of the three-leaved rose, $r = \sin(3\theta)$, at the point with polar coordinates $(r,\theta) = \left(\frac{\sqrt{2}}{2}, \frac{\pi}{12}\right)$.

22. (a) Find the slope of the cardioid $r = 1 + \cos(\theta)$ at (r,θ).
(b) What happens to the slope as θ approaches π from the left?
(c) What does (b) tell us about the graph of the cardioid? (Identify this feature on the graph.)

23. Obtain parametric equations for the circle of radius a and center (h,k), using as parameter the angle θ shown in Figure 9.3.2(a).

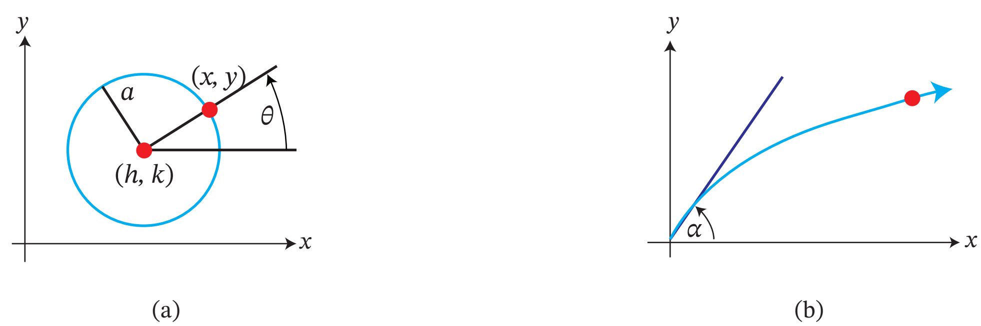


Figure 9.3.2

Exercises 24 to 26 analyze the trajectory of a ball thrown from the origin at an angle α and initial velocity v_0, as sketched in Figure 9.3.2(b). The results are used in the Uniform Sprinkler CIE (CIE 7) at the end of Chapter 5.

24. It can be shown that if time is in seconds and distance is in feet, then t seconds later the ball is at (x,y) with $x = (v_0 \cos(\alpha))\, t$, $y = (v_0 \sin(\alpha))\, t - 16t^2$.
(a) Express y as a function of x.
(b) What type of curve does the ball follow?
(c) Find the coordinates of its highest point on the curve.

25. Eventually the ball in Exercise 24 falls back to the ground.
(a) Show that the horizontal distance it travels is proportional to $\sin(2\theta)$.
(b) Use (a) to determine the angle that maximizes the horizontal distance traveled.
(c) Show that the horizontal distance traveled in (a) is the same when the ball is thrown at an angle θ or at the complementary angle $\frac{\pi}{2} - \theta$.

26. Is it possible to extend the horizontal distance traveled by throwing the ball in Exercise 24 from the top of a hill? Assume the hill has height d.

---

27. The spiral $r = e^{2\theta}$ meets the ray $\theta = \alpha$ at an infinite number of points.
(a) Graph the spiral.
(b) Find the slope of the spiral at each intersection with the ray.
(c) Show that at all the intersections the slopes are the same.

28. The spiral $r = \theta$, $\theta > 0$ meets the ray $\theta = \alpha$ at an infinite number of points (α, α), $(\alpha + 2\pi, \alpha)$, $(\alpha + 4\pi, \alpha)$, ... What happens to the angle between the spiral and the ray at the point $(\alpha + 2\pi n, \alpha)$ as $n \to \infty$?

29. Let a and b be positive numbers and a curve be given parametrically by the parametric equations $x = a\cos(t)$ and $y = b\sin(t)$.

(a) Show that the curve is the ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$.

(b) Find the area of the region bounded by the ellipse by making a substitution that expresses $4\displaystyle\int_0^a y\,dx$ in terms of an integral in which the variable is t and the range of integration is $[0, \pi/2]$.

30. For the curve given parametrically by $x = t^2 + e^t$, $y = t + e^t$,for t in $[0, 1]$:

(a) Plot the points corresponding to $t = 0$, $\dfrac{1}{2}$, and 1.
(b) Find the slope of the curve at $(1, 1)$.
(c) Find the area of the region under the curve and above the interval $[1, e + 1]$. (See Exercise 29(b).)

31. What is the slope of the cycloid in Figure 9.3.1(b) when it first has height a? See Example 1.

32. Find the volume of the solid of revolution produced when the region above the x-axis and under one arch of the cycloid $x = a\theta - a\sin(\theta)$, $y = a - a\cos(\theta)$ $(0 \le \theta \le 2\pi)$ is revolved around the x-axis.

33. Find the volume of the solid of revolution obtained by revolving the region in Exercise 32 about the y-axis.

34. Let a be a positive constant. For the curve given parametrically by the equations $x = a\cos^3(t)$, $y = a\sin^3(t)$. (a) Sketch the curve. (b) Express the slope of the curve in terms of the parameter t.

35. For a tangent line to the curve in Exercise 34 at a point P in the first quadrant, show that the length of the segment of that line intercepted by the coordinate axes is a.

36. Solve the parametric equations for the cycloid, $x = a\theta - a\sin(\theta)$, $y = a - a\cos(\theta)$, for x as a function of y. See Example 2.

37. L'Hôpital's rule asserts that if $\lim_{t\to 0} f(t) = 0$, $\lim_{t\to 0} g(t) = 0$, and $\lim_{t\to 0} \dfrac{f'(t)}{g'(t)}$ exists, then $\lim_{t\to 0} \dfrac{f(t)}{g(t)} = \lim_{t\to 0} \dfrac{f'(t)}{g'(t)}$. Interpret l'Hôpital's Rule in terms of the parameterized curve $x = g(t)$, $y = f(t)$.

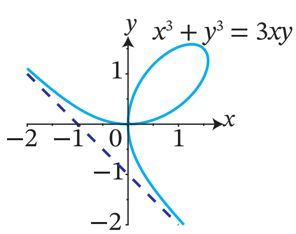


Figure 9.3.3

38. The **Folium of Descartes**, shown in Figure 9.3.3, is the graph of $x^3 + y^3 = 3xy$. It consists of a loop and two infinite pieces asymptotic to the line $x + y = -1$. Parameterize the Folium of Descartes by the slope t of the line joining the origin with (x, y).

(a) Show that $x = \dfrac{3t}{1 + t^3}$ and $y = \dfrac{3t^2}{1 + t^3}$.
(b) Find the highest point on the loop.
(c) Find the point on the loop furthest to the right.
(d) The loop is parameterized by t in $[0, \infty)$. which values of t parameterize the part in the fourth quadrant?
(e) Which values of t parameterize the part in the second quadrant?
(f) Show that the Folium of Descartes is symmetric with respect to the line $y = x$.

NOTE: An online search for "Folium of Descartes" will provide references to the history of this curve that date back to 1638. See also Exercise 33 in Section 14.S.

9.4 Arc Length and Speed on a Parameterized Curve

In Section 4.2 we studied the motion of an object moving on a line. If at time t its position is $x(t)$ then its velocity is dx/dt and its speed is $|dx/dt|$. Now we will examine the velocity and speed of an object moving along a curved path.

Arc Length and Speed in Rectangular Coordinates

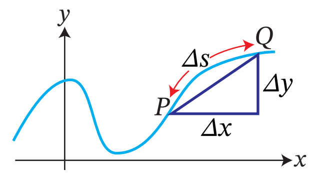


Figure 9.4.1

Suppose an object is moving on a path given parametrically by $x = g(t)$, $y = h(t)$ where g and h have continuous derivatives. If we think of t as time we can find a formula for its speed at any point on the curve.

Let $s(t)$ be the arc length covered from the initial time to time t, that is, the distance measured along the curve from its initial point to time t. In an interval of time of length Δt it travels a distance Δs along the path. We want to find

$$\lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t}.$$

During the time interval $[t, t + \Delta t]$ the object goes from P to Q on the path, covering a distance Δs, as shown in Figure 9.4.1. Its x-coordinate changes by Δx and its y-coordinate by Δy. The chord PQ has length $\sqrt{(\Delta x)^2 + (\Delta y)^2}$.

We assume then that the curve is well behaved in the sense that $\lim_{\Delta t \to 0} \Delta s/|PQ| = 1$. Then

$$\begin{aligned}
\lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \lim_{\Delta t \to 0} \left(\frac{\Delta s}{|PQ|} \frac{|PQ|}{\Delta t} \right) \qquad &= \lim_{\Delta t \to 0} \frac{\Delta s}{|PQ|} \lim_{\Delta t \to 0} \frac{|PQ|}{\Delta t} \qquad &&= 1 \cdot \lim_{\Delta t \to 0} \frac{|PQ|}{\Delta t} \\
= \lim_{\Delta t \to 0} \frac{\sqrt{(\Delta x)^2 + (\Delta y)^2}}{\Delta t} \qquad &= \lim_{\Delta t \to 0} \sqrt{\left(\frac{\Delta x}{\Delta t}\right)^2 + \left(\frac{\Delta y}{\Delta t}\right)^2} \qquad &&= \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}.
\end{aligned}$$

Observation 9.4.1: *An Important Property of Arc Length*

The rates at which x and y change determine how fast the arc length s changes, as shown in Figure 9.4.2. For a curve given parametrically as $x = x(t)$, $y = y(t)$ the rate of change of the arc length s is

$$\frac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}$$

or, in terms of differentials,

$$ds = \sqrt{(dx)^2 + (dy)^2} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt$$

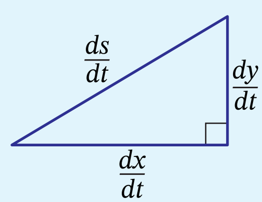


Figure 9.4.2

Now that we have a formula for ds/dt, we integrate it to get the distance along the path covered during a time interval $[a, b]$:

Definition: *Integral Definition of Arc Length (of a Plane Curve)*

The length of the parametric curve $x = x(t)$, $y = y(t)$ for $a \le t \le b$ is

$$\text{Arc length} = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt. \qquad (9.4.1)$$

If the curve is given in the form $y = f(x)$, we can use x as the parameter. A parametric representation of the curve then is $x = x$, $y = f(x)$ and (9.4.1) becomes

$$\text{Arc length} = \int_a^b \sqrt{1 + (f'(x))^2}\, dx.$$

The arc length function is, by definition, an nondecreasing function. This means ds/dt is never negative. In most applications ds/dt will never be zero either.

Historical Note: Classical Arc Length Examples

People were able to find the arc length of specific curves even before the discovery of calculus. Even today, these same curves are among the few that produce definite integrals with elementary integrands.

The first goes back to the year 1657, when the 20-year old Englishman, William Neil, found the length of an arc on the graph of $y = x^{3/2}$. His method was more complicated. (See Example 1.)

Earlier, Thomas Harriot had found the length of an arc of the spiral $r = e^{\theta}$, but his work was not widely known. (See Example 4.)

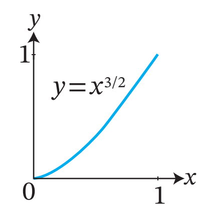


Figure 9.4.3

EXAMPLE 1. Find the arc length of the curve $y = x^{3/2}$ for x in $[0,1]$. (See Figure 9.4.3.)

SOLUTION By (9.4.1),

$$\text{Arc length} = \int_0^1 \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

Since $y = x^{3/2}$, we differentiate to find $dy/dx = (3/2)x^{1/2}$. Thus

$$\begin{aligned}
\text{Arc length} &= \int_0^1 \sqrt{1 + \left(\frac{3}{2}x^{1/2}\right)^2}\, dx = \int_0^1 \sqrt{1 + \frac{9}{4}x}\, dx \\
&= \int_1^{13/4} \sqrt{u} \cdot \frac{4}{9}\, du && \left(\text{substitution: } u = 1 + \frac{9}{4}x\right) \\
&= \frac{4}{9} \cdot \frac{2}{3} u^{3/2}\Big|_1^{13/4} \qquad = \frac{8}{27}\left(\left(\frac{13}{4}\right)^{3/2} - 1^{3/2}\right) \\
&= \frac{8}{27}\left(\frac{13^{3/2}}{8} - 1\right) \qquad = \frac{13^{3/2} - 8}{27} \approx 1.43971.
\end{aligned}$$

◆

The arc length of the curve $y = x^a$ where a is a nonzero rational number, usually cannot be computed using the fundamental theorem of calculus. The only cases in which it can be computed by the FTC are $a = 1$ (the graph of $y = x$) and $a = 1 + 1/n$ where n is an integer. Exercise 32 treats this question.

EXAMPLE 2. In Section 9.3 the parametric equations for the motion of a ball thrown horizontally with a speed of 32 feet per second (≈ 21.8 mph) were found to be $x = 32t$, $y = -16t^2$. (See Example 1 and Figure 9.3.1 in Section 9.3.) How fast is the ball moving at time t? Find the distance s that the ball travels during the first b seconds.

SOLUTION From $x = 32t$ and $y = -16t^2$ we compute $dx/dt = 32$ and $dy/dt = -32t$. Its speed at time t is

$$\text{Speed} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} = \sqrt{(32)^2 + (-32t)^2} = 32\sqrt{1 + t^2} \text{ feet per second.}$$

The distance traveled is the arc length from $t = 0$ to $t = b$. By (9.4.1),

$$\text{Arc length} = \int_0^b \sqrt{(32)^2 + (-32t)^2}\, dt = 32 \int_0^b \sqrt{1+t^2}\, dt.$$

The integral can be evaluated with an integration table or with the trigonometric substitution $x = \tan(\theta)$. Formula 47, with $p = 1$, in this book's Table of Integrals (in Appendix A) provides

$$\int \sqrt{t^2+1}\, dt = \frac{1}{2}\left(t\sqrt{1+t^2} + \ln\left(t + \sqrt{1+t^2}\right)\right),$$

so the distance traveled is

$$32 \int_0^b \sqrt{1+t^2}\, dt = 16b\sqrt{1+b^2} + 16\ln\left(b + \sqrt{1+b^2}\right).$$

◆

EXAMPLE 3. Find the length of one arch of the cycloid found in Example 2 of Section 9.3.

SOLUTION The curve can be parameterized as $x = a\theta - a\sin(\theta)$ and $y = a - a\cos(\theta)$ with θ as the parameter. Over one arch of the cycloid, θ varies from 0 to 2π.

We compute

$$\frac{dx}{d\theta} = a - a\cos(\theta) \qquad \text{and} \qquad \frac{dy}{d\theta} = a\sin(\theta).$$

The square of the speed is

$$\begin{aligned}\left(\frac{dx}{d\theta}\right)^2 + \left(\frac{dy}{d\theta}\right)^2 &= (a - a\cos(\theta))^2 + (a\sin(\theta))^2 = a^2\left((1-\cos(\theta))^2 + (\sin(\theta))^2\right)\\ &= a^2\left(1 - 2\cos(\theta) + (\cos(\theta))^2 + (\sin(\theta))^2\right) = a^2\left(2 - 2\cos(\theta)\right)\\ &= 2a^2(1-\cos(\theta)).\end{aligned}$$

Using formula (9.4.1) and the trigonometric identity $1 - \cos(\theta) = 2\sin^2(\theta/2)$, we have

$$\begin{aligned}\text{Arc length of one arch} &= \int_0^{2\pi} \sqrt{2a^2(1-\cos(\theta))}\, d\theta = a\sqrt{2}\int_0^{2\pi} \sqrt{1-\cos(\theta)}\, d\theta = a\sqrt{2}\int_0^{2\pi} \sqrt{2}\sin\left(\frac{\theta}{2}\right) d\theta\\ &= 2a\int_0^{2\pi} \sin\left(\frac{\theta}{2}\right) d\theta = 2a\left(-2\cos\left(\frac{\theta}{2}\right)\Big|_0^{2\pi}\right) = 2a\left(-2(-1) - (-2)(1)\right) = 8a.\end{aligned}$$

So, as θ varies from 0 to 2π (one revolution of the wheel), the bicycle travels a distance of $2\pi a \approx 6.28319a$ and a tack in the tread of the tire travels a distance $8a$. ◆

Arc Length and Speed in Polar Coordinates

So far in this section curves have been described in rectangular coordinates. Now we consider a curve given in polar coordinates by $r = f(\theta)$.

We will estimate the length of arc Δs corresponding to small changes $\Delta\theta$ and Δr in polar coordinates, as shown in Figure 9.4.4. The region bounded by the circular arc AB, the straight segment BC, and a part of the curve, AC, resembles a right triangle whose legs have lengths $r\Delta\theta$ and Δr. Because Δs is well approximated by its hypotenuse, $\sqrt{(r\Delta\theta)^2 + (\Delta r)^2}$, we expect

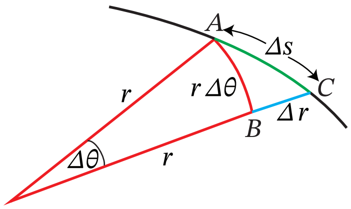


Figure 9.4.4

$$\begin{aligned}\frac{ds}{d\theta} = \lim_{\Delta\theta\to 0}\frac{\Delta s}{\Delta\theta} &= \lim_{\Delta\theta\to 0}\frac{\sqrt{(r\Delta\theta)^2+(\Delta r)^2}}{(\Delta\theta)}\\ &= \lim_{\Delta\theta\to 0}\sqrt{r^2+\left(\frac{\Delta r}{\Delta\theta}\right)^2}\\ &= \sqrt{r^2+\left(\frac{dr}{d\theta}\right)^2}.\end{aligned}$$

Observation 9.4.2: ***Arc Length in Polar Coordinates***

For a parametric curve given in polar coordinates, $r = r(\theta)$,

$$\frac{ds}{d\theta} = \sqrt{r^2+\left(\frac{dr}{d\theta}\right)^2} \quad \text{or} \quad ds = \sqrt{(r\,d\theta)^2+(dr)^2} = \sqrt{r^2+(r')^2}\,d\theta.$$

NOTE: The formulas for arc length in polar coordinates can also be obtained from the rectangular coordinate formula by using $x = r\cos(\theta)$ and $y = r\sin(\theta)$. (See Exercise 20.) However, we prefer the geometric approach because it is more direct, more intuitive, and easier to remember.

Definition: ***Arc Length of a Polar Curve*** $r = f(\theta)$

The length of the curve $r = f(\theta)$ for θ in $[\alpha,\beta]$ is

$$s = \int_\alpha^\beta ds \qquad \text{where} \qquad ds = \sqrt{r^2+(r')^2}\,d\theta = \sqrt{\left(f(\theta)\right)^2+\left(f'(\theta)\right)^2}\,d\theta.$$

EXAMPLE 4. Find the length of the spiral $r = e^{-3\theta}$ for θ in $[0,2\pi]$.

SOLUTION From

$$r' = \frac{dr}{d\theta} = -3e^{-3\theta},$$

the formula for arc length in polar coordinates gives

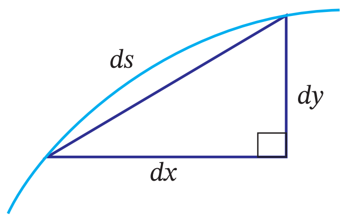
$$\begin{aligned}\text{Arc length} = \int_\alpha^\beta \sqrt{r^2+(r')^2}\,d\theta &= \int_0^{2\pi}\sqrt{\left(e^{-3\theta}\right)^2+\left(-3e^{-3\theta}\right)^2}\,d\theta\\ = \int_0^{2\pi}\sqrt{e^{-6\theta}+9e^{-6\theta}}\,d\theta &= \sqrt{10}\int_0^{2\pi}\sqrt{e^{-6\theta}}\,d\theta\\ = \sqrt{10}\int_0^{2\pi}e^{-3\theta}\,d\theta &= \sqrt{10}\left.\frac{e^{-3\theta}}{-3}\right|_0^{2\pi}\\ = \sqrt{10}\left(\frac{e^{-3\cdot 2\pi}}{-3}-\frac{e^{-3\cdot 0}}{-3}\right) &= \sqrt{10}\left(\frac{e^{-6\pi}}{-3}+\frac{1}{3}\right)\\ = \frac{\sqrt{10}}{3}\left(1-e^{-6\pi}\right) &\approx 1.05409.\end{aligned}$$

♦

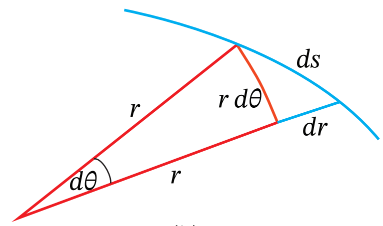
Summary

This section concerned speed along a parametric path and the length of the path. If the path is described in rectangular coordinates, then Figure 9.4.5(a) conveys the key ideas. If the curve is given in polar coordinates, Figure 9.4.5(b) is the key.

It is generally easier to recall the diagrams than the formulas for speed and arc length. Everything depends on the Pythagorean Theorem.



(a)



(b)

Figure 9.4.5

EXERCISES for Section 9.4

In Exercises 1 to 8 find the arc lengths of the curves over the intervals.

1. $y = x^{3/2}$, x in $[1,2]$

2. $y = x^{2/3}$, x in $[0,1]$

3. $y = \frac{1}{2}(e^x + e^{-x})$, x in $[0,b]$

4. $y = \frac{x^2}{2} - \frac{\ln(x)}{4}$, x in $[2,3]$

5. $x = \cos^3(t)$, $y = \sin^3(t)$, t in $[0,\pi/2]$

6. $r = e^\theta$, θ in $[0,2\pi]$

7. $r = 1 + \cos(\theta)$, θ in $[0,\pi]$

8. $r = \cos^2\left(\frac{\theta}{2}\right)$, θ in $[0,\pi]$

In each of Exercises 9 to 12 find the speed of a particle at time t, given the parametric equations for its path.

9. $x = 50t$, $y = -16t^2$

10. $x = \sec(3t)$, $y = \sin^{-1}(4t)$

11. $x = t + \cos(t)$, $y = 2t - \sin(t)$

12. $x = \csc\left(\frac{t}{2}\right)$, $y = \tan^{-1}(\sqrt{t})$

13. (a) Graph $x = t^2$, $y = t$ for $0 \le t \le 3$.
(b) Estimate its arc length from $(0,0)$ to $(9,3)$ by an inscribed polygon whose vertices have x-coordinates 0, 1, 4, and 9.
(c) Set up a definite integral for the arc length of the curve in question.
(d) Estimate the definite integral by using a partition of $[0,3]$ into three sections, each of length 1, and use the trapezoid method.
(e) Estimate the definite integral by Simpson's method with six sections.
(f) If your calculator has a program to evaluate definite integrals, use it to evaluate the definite integral in (c) to four decimal places.

14. How long is the spiral $r = e^{-3\theta}$, $\theta \ge 0$?

15. How long is the spiral $r = \frac{1}{\theta}$, $\theta \ge 2\pi$?

16. (a) Graph $y = \dfrac{1}{x^2}$ for x in $[1,2]$.

(b) Estimate the length of the arc by using an inscribed polygon whose vertices are $(1,1)$, $\left(\frac{5}{4}, \left(\frac{4}{5}\right)^2\right)$, $\left(\frac{3}{2}, \left(\frac{2}{3}\right)^2\right)$, and $\left(2, \frac{1}{4}\right)$.

(c) Set up a definite integral for the arc length of the curve.

(d) Estimate the definite integral by the trapezoid method, using four equal length sections.

(e) Estimate the definite integral by Simpson's method with four sections.

(f) If your calculator has a program to evaluate definite integrals, use it to evaluate the definite integral to four decimal places.

17. Suppose that a curve has equation $x = f(y)$ for $c \le y \le d$ in rectangular coordinates. Use a triangle whose sides have lengths dx, dy, and ds to show that the arc length of this curve is $\displaystyle\int_c^d \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy$.

18. Consider the arc length of $y = x^{2/3}$ for x in the interval $[1,8]$. See Exercise 17.

(a) Set up a definite integral for the arc length using x as the parameter.

(b) Set up a definite integral for the arc length using y as the parameter.

(c) Evaluate the easier of the two integrals found in parts (a) and (b).

19. For any time $t \ge 0$ a ball is at the point $(24t, -16t^2 + 5t + 3)$. (a) Where is ball at time $t = 0$? (b) What is its horizontal speed at time $t = 0$? (c) What is its vertical speed at time $t = 0$?

20. Earlier in this section, we obtained the formula $\dfrac{ds}{d\theta} = \sqrt{r^2 + \left(\dfrac{dr}{d\theta}\right)^2}$ geometrically.

(a) Obtain this formula by calculus, starting with $\left(\dfrac{ds}{d\theta}\right)^2 = \left(\dfrac{dx}{d\theta}\right)^2 + \left(\dfrac{dy}{d\theta}\right)^2$ and using the relations $x = r\cos(\theta)$ and $y = r\sin(\theta)$, where $r = f(\theta)$.

(b) Which derivation do you prefer? Why?

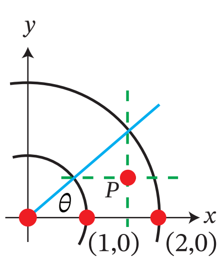


Figure 9.4.6

21. In Figure 9.4.6 the points $P = (x, y)$ lie on the horizontal line through $(\cos(\theta), \sin(\theta))$ and on the vertical line through $(2\cos(\theta), 2\sin(\theta))$ for each $0 \le \theta \le 2\pi$

(a) Sketch the curve that P sweeps out.

(b) Show that $P = (2\cos(\theta), \sin(\theta))$.

(c) Set up a definite integral for the length of the curve in (a). Do not evaluate the integral.

(d) Eliminate θ and show that P is on the ellipse $\dfrac{x^2}{4} + \dfrac{y^2}{1} = 1$.

22. (a) At time t a particle has polar coordinates $r = g(t)$, $\theta = h(t)$. How fast is it moving?

(b) Use (a) to find the speed of a particle which at time t is at the point $(r, \theta) = (e^t, 5t)$.

23. (a) How far does a bug travel from time $t = 1$ to time $t = 2$ if at time t it is at the point $(x, y) = (\cos(\pi t), \sin(\pi t))$?

(b) How fast is it moving at time t?

(c) Graph its path relative to an xy coordinate system. Where is it at time $t = 1$? At $t = 2$?

(d) Eliminate t to find a relation between x and y.

24. If $r = 1 + \cos\theta$ for θ in $[0, \pi]$ we may consider r as a function of θ or as a function of s, arc length along the curve, measured, say, from $(2,0)$. See also Exercises 11 and 12 in Section 9.S.

(a) Find the average of r with respect to θ in $[0, \pi]$.

(b) Find the average of r with respect to s.

25. Let $r = f(\theta)$ describe a curve in polar coordinates. Assume that $\frac{df}{d\theta}$ is continuous. Let θ be a function of time t and let $s(t)$ be the length of the curve corresponding to the time interval $[a, t]$.

(a) What definite integral is equal to $s(t)$? (b) What is the speed $\frac{ds}{dt}$?

26. Let a be a positive number. Find the arc length of the Archimedean spiral $r = a\theta$ for θ in $[0, 2\pi]$.

27. The function $r = f(\theta)$ describes a curve in polar coordinates, for θ in $[0, 2\pi]$. Assume r' is continuous and $f(\theta) > 0$. Prove that the average of r as a function of arc length is at least as large as $\frac{2A}{s}$, where A is the area swept out by the radius and s is the arc length of the curve. For what curves is the average equal to $\frac{2A}{s}$?

28. The equations $x = \cos(t)$, $y = 2\sin(t)$, t in $\left[0, \frac{\pi}{2}\right]$ describe a quarter of an ellipse. Draw the arc. Describe at least two different ways of estimating its length. Compare the advantages and challenges each method presents. Use the method of your choice to estimate the length of this arc.

29. When a curve is given in rectangular coordinates, its slope is $\frac{dy}{dx}$. To find the slope of the tangent line to the curve given in polar coordinates involves more work. Assume that $r = f(\theta)$. Use the relation $\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta}$, which comes from the chain rule $\left(\frac{dy}{d\theta} = \frac{dy}{dx}\frac{dx}{d\theta}\right)$.

(a) Using $y = r\sin(\theta)$ and $x = r\cos(\theta)$ find $\frac{dy}{d\theta}$ and $\frac{dx}{d\theta}$. (b) Show that the slope is $\frac{r\cos(\theta) + (dr/d\theta)\sin(\theta)}{-r\sin(\theta) + (dr/d\theta)\cos(\theta)}$.

30. Use the formula for the slope found in Exercise 29(c) to find the slope of the cardioid $r = 1 + \sin(\theta)$ at $\theta = \frac{\pi}{3}$.

31. Let $y = f(x)$ for x in $[0, 1]$ describe a curve that starts at $(0, 0)$, ends at $(1, 1)$, and lies in the square with vertices $(0, 0), (1, 0), (1, 1)$, and $(0, 1)$. Assume f has a continuous derivative.

(a) What can be said about the arc length of the curve? How small and how large can it be?
(b) Answer (a) if we assumed also that $f'(x) \geq 0$ for x in $[0, 1]$.

32. Consider the length of the curve $y = x^m$, $0 \leq x \leq 1$, where m is a rational number. Show that the fundamental theorem of calculus aids in computing this length only if $m = 1$ or if $m = 1 + \frac{1}{n}$ for some integer n.

33. If one convex polygon P_1 lies inside another polygon P_2 is the perimeter of P_1 necessarily less than the perimeter of P_2? What if P_1 is not convex?

34. One part of the cardioid $r = 1 + \sin(\theta)$ is traced as θ increases from $\frac{-\pi}{2}$ to $\frac{\pi}{2}$. Find its highest point and give its polar coordinates.

Exercises 35 and 36 form a unit. Exercise 36 provides an intuitive derivation of the result obtained in Exercise 35.

35. Figure 9.4.7(a) shows the angle, γ, between the radius and tangent line to the curve $r = f(\theta)$. Using $\gamma = \phi - \theta$ and $\tan(A - B) = \frac{\tan(A) - \tan(B)}{1 + \tan(A)\tan(B)}$, show that $\tan(\gamma) = \frac{r}{r'}$.

36. The formula $\tan(\gamma) = r/r'$ in Exercise 35 is so simple one would expect a simple geometric explanation. Use the triangle in Figure 9.4.4 that we used to obtain the formula for $\frac{ds}{d\theta}$ to show that $\tan(\gamma)$ should be $\frac{r}{r'}$.

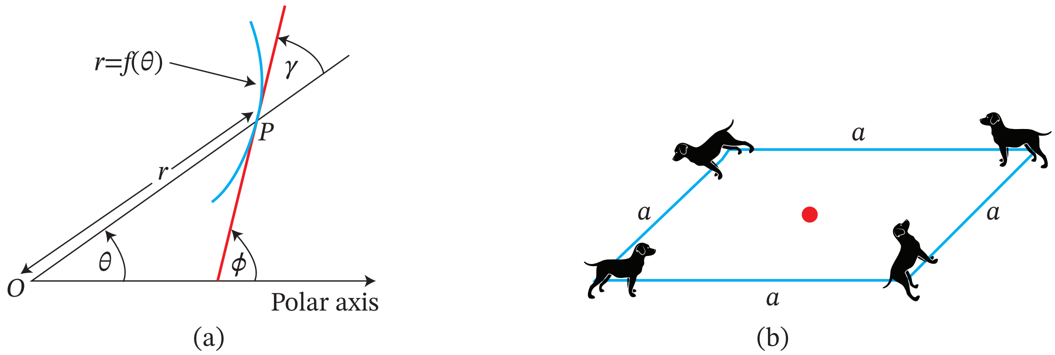


Figure 9.4.7

37. Four dogs are chasing each other counterclockwise at the same speed as shown in Figure 9.4.7(b). Initially they are at the vertices of a square of side a. As they chase each other, each running directly toward the dog in front, they approach the center of the square in spiral paths. How far does each dog travel?

(a) Find the equation of the spiral path each dog follows and use calculus to answer this question.

(b) Answer the question without using calculus.

38. This is a generalization of Exercise 37. For some integer $n \geq 3$, n dogs are chasing each other counterclockwise at the same speed. Initially they are at the vertices of a regular n-gon of side a centered at the origin. As they chase each other, each running directly toward the dog in front, they approach the origin in spiral paths.

(a) Find the equation of the spiral path each dog follows.

(b) How far does each dog run?

(c) How many times does each path circle the origin?

(d) At what point does the dog that starts on the positive x-axis return to it?

39. We assumed that a chord AB of a smooth curve is a good approximation of the arc AB when B is near to A. Show that the formula for arc length is consistent with this assumption. That is, if $y = f(x)$ has a continuous derivative, $A = (a, f(a))$, and $B = (x, f(x))$, then $\lim_{x\to a} \frac{1}{\sqrt{(x-a)^2 + (f(x)-f(a))^2}} \int_a^x \sqrt{1+f'(t)^2}\, dt = 1$.

40. This exercise outlines an approach to arc length and speed on a curve that defines arc length first, then speed.

Let $x = g(t)$, $y = h(t)$ where g and h have continuous derivatives. Let $a = t_0 < t_1 < t_2 < \cdots < t_n = b$ be a partition of $[a, b]$ into n equal sections of length $\Delta t = \frac{1}{n}(b-a)$. Let $P_i = (g(t_i), h(t_i))$, which we write as (x_i, y_i). Then the polygon $P_0P_1P_2\cdots P_n$ is inscribed in the curve. We assume that as $n \to \infty$, the length of the polygon, $\sum_{i=1}^{n} \overline{P_{i-1}P_i}$ approaches the length of the curve from $(g(a), h(a))$ to $(g(b), h(b))$.

(a) Show that the length of the polygon is $\sum_{i=1}^{n} \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}$.

(b) Show that the sum can be written as $\sum_{i=1}^{n} \sqrt{(g'(t_i^*))^2 + (h'(t_i^{**}))^2} \cdot \Delta t$ for some t_i^* and t_i^{**} in $[t_{i-1}, t_i]$.

(c) Why would you expect the limit as $n \to \infty$ of the sum found in (b) to be $\int_a^b \sqrt{(g'(t))^2 + h'(t))^2}\, dt$?

This result is typically proved in advanced courses. It is true even though t_i^* and t_i^{**} may be different.

(d) From (c) deduce that the speed is $\sqrt{(g'(t))^2 + h'(t))^2}$.

9.5 Area of a Surface of Revolution

In this section we develop a formula for the surface area of a solid of revolution as a definite integral. Along the way we will rediscover on of Archimedes' great findings in the third century B.C.: the surface area of a sphere is four times the area of a cross section through its center.

Local Estimate of Surface Area

Let $y = f(x)$ have a continuous derivative for x in some interval. Assume that $f(x) \geq 0$ on it. When the graph of f is revolved about the x-axis it sweeps out a surface, as shown in Figure 9.5.1. To develop a definite integral for this surface area, we use an informal approach.

Figure 9.5.1

A short section of the graph $y = f(x)$ is almost straight. We approximate it by a short line segment of length ds. When the segment is revolved about the x-axis it sweeps out a narrow band. (See Figures 9.5.2(a) and (b).)

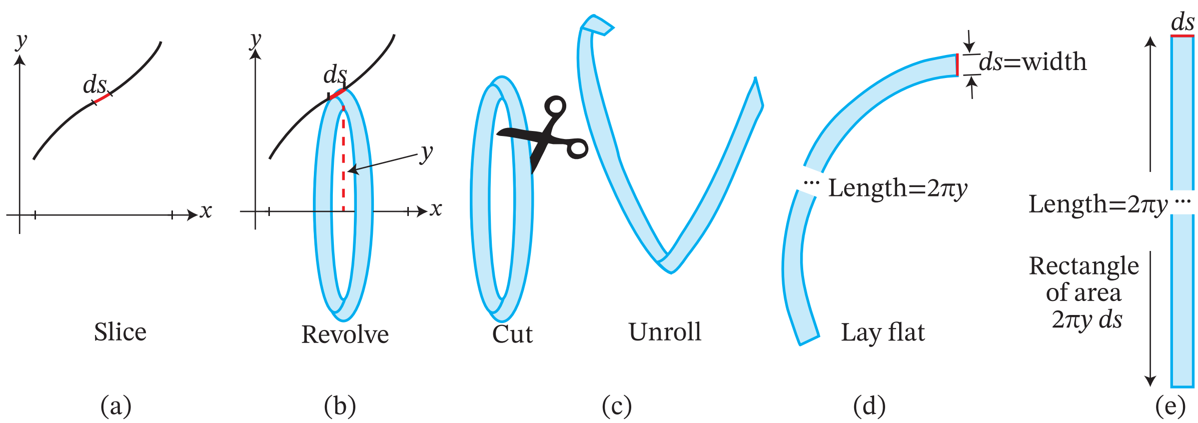


Figure 9.5.2

If we can estimate the area of the band, then we will have a local approximation of the surface area. When these are summed and when the short segments of the graph are forced to shrink towards zero, we will obtain a definite integral for the entire surface area.

To estimate the area of a narrow band, cut the band with scissors and lay it flat, as in Figures 9.5.2(c) and (d). The area of the flat band in Figure 9.5.2(d) is close to the area of a flat rectangle of length $2\pi y$ and width ds, as in

Figure 9.5.2(e). From this we are able to make a useful local approximation to the area of a surface. (The exact area of this small slice of the surface is found in Exercises 27 and 28.)

Observation 9.5.1: *Local Approximation for Area of a Surface*

The local approximation for the surface area of a short segment with length ds that is rotated around the x-axis, which is a distance y away from the segment, is

$$\text{Local Approximation of the Surface Area of One Slice} = 2\pi y\, ds.$$

Key Integral for Surface Area

From the local estimate for surface area we obtain the following integral for the total area of the curved surface:

Definition: *Area of a Surface of Revolution* ($y = f(s)$)

The surface area of the solid of revolution formed by rotating the graph of $y = f(s)$, $0 \le s \le L$, about the x-axis is, where s is arc length and L is the curve's length,

$$\text{Surface Area} = \int_0^L 2\pi y\, ds. \tag{9.5.1}$$

Recall from our previous discussions of arc length that it can be very difficult (if not impossible) to find the explicit arc length parameterization of a curve. Since s (arc length) is a clumsy parameter, for computations we will use one of the other forms for ds to change (9.5.1) into a definite integral that is more convenient to evaluate.

Say that the section of the graph $y = f(x)$ that was revolved corresponds to the interval $[a, b]$ on the x-axis, as in Figure 9.5.3. Then, assuming $y = f(x) \ge 0$ and that dy/dx is continuous for x on $[a, b]$,

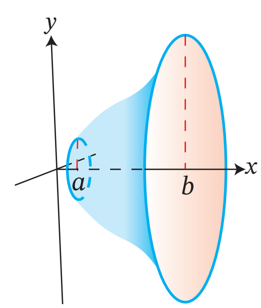


Figure 9.5.3

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$

and the surface area integral $\int_0^L 2\pi y\, ds$ provides another definition for the area of a surface of revolution.

Definition: *Area of a Surface of Revolution* ($y = f(x)$, $a \le x \le b$)

The surface area of the solid of revolution formed by rotating the graph of $y = f(x)$, $a \le x \le b$, about the x-axis is

$$\text{Surface Area} = \int_a^b 2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx. \tag{9.5.2}$$

EXAMPLE 1. Find the surface area of a sphere of radius a.

SOLUTION The circle of radius a has the equation $x^2 + y^2 = a^2$. The sphere of radius a is formed by revolving the top half of the sphere, $y = \sqrt{a^2 - x^2}$, about the x-axis. (See Figure 9.5.4.)

We want to use the fact that the surface area is given by $\int_{-a}^{a} 2\pi y\, ds$, with $y = \sqrt{a^2 - x^2}$. Then, because $dy/dx = -x/\sqrt{a^2 - x^2}$, we find that

$$\begin{aligned} ds &= \sqrt{1+\left(\frac{dy}{dx}\right)^2}\,dx & &= \sqrt{1+\left(\frac{-x}{\sqrt{a^2-x^2}}\right)^2}\,dx \\ &= \sqrt{1+\frac{x^2}{a^2-x^2}}\,dx & &= \sqrt{\frac{a^2}{a^2-x^2}}\,dx \\ &= \frac{a}{\sqrt{a^2-x^2}}\,dx. \end{aligned}$$

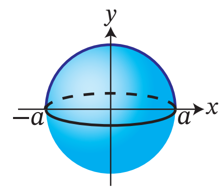


Figure 9.5.4

Thus,

$$\begin{aligned} \text{Surface Area} &= \int_{-a}^{a} 2\pi y\,ds & &= \int_{-a}^{a} 2\pi\sqrt{a^2-x^2}\,\frac{a}{\sqrt{a^2-x^2}}\,dx \\ &= \int_{-a}^{a} 2\pi a\,dx & &= 2\pi a x\Big|_{-a}^{a} \\ &= 4\pi a^2. \end{aligned}$$

♦

Historical Note: Famous Result of Archimedes

One of Archimedes' famous observations was that the surface area of a sphere is 4 times the area of its equatorial cross section. Example 1 proves this result, except that Archimedes made this discovery almost 2000 years before the invention of calculus.

If the graph is given parametrically, $x = g(t)$, $y = h(t)$, where g and h have continuous derivatives and $h(t) \geq 0$, then it is natural to express the integral $\int_0^L 2\pi y\,ds$ as an integral over an interval on the t-axis. If t varies in the interval $[a, b]$, then

$$ds = \sqrt{(dx)^2 + (dy)^2} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt$$

which leads to

Definition: *Area of a Surface of Revolution* ($x = g(t)$, $y = h(t)$, $a \leq t \leq b$)

The surface area of the solid of revolution formed by rotating the graph of the parametric curve $x = g(t)$, $y = h(t)$, $a \leq t \leq b$, about the x-axis is

$$\text{Surface Area} = \int_a^b 2\pi y\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt. \qquad (9.5.3)$$

NOTE: That (9.5.2) is the special case of (9.5.3) when the parameter is x is reassuring. In particular, this shows that these two definitions are consistent.

The formulas may seem to refer only to surfaces obtained by revolving a curve about the x-axis. In fact, they can apply to revolution about any line. The factor y in the integrand, $2\pi y\,ds$, is the distance from a point on the curve to the axis of revolution. We replace y by R (for radius) to free ourselves from coordinate systems. (We use capital R to avoid confusion with polar coordinates.) One way to write the general formula for surface area of revolution is

Definition: *Area of a Surface of Revolution (General Case)*

The surface area of the solid of revolution formed by rotating the graph of the curve about an axis is

$$\text{Surface Area } = \int_0^L 2\pi R\, ds,$$

where R denotes the distance from a point on the curve to the axis of revolution, L is the length of the curve, and the parameter s is arc length.

MEMORY AID: To remember the formula, think of a narrow circular band of width ds and radius R as having an area close to the area of the rectangle shown in Figure 9.5.5(a).

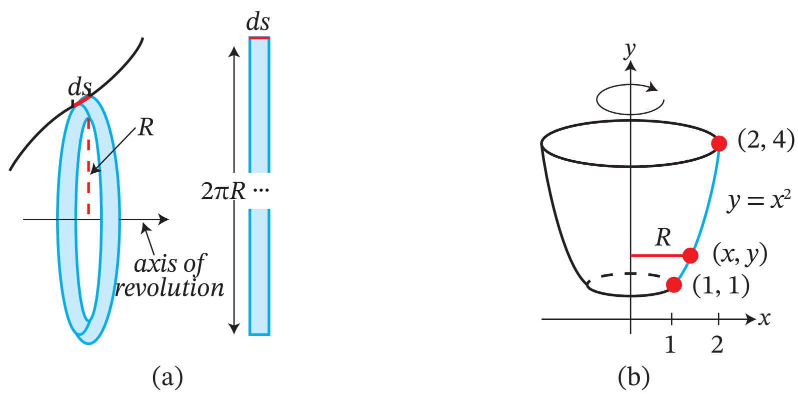


Figure 9.5.5

In practice, arc length is seldom a convenient parameter to use in computations. Instead, x, y, t or θ is used and the interval of integration describes the interval over which the parameter varies. When another parameter other than arc length is used, the limits of integration will correspond to the initial and terminal points of the curve in the chosen parameter. In particular, it is not necessary to find the length of the curve, L.

The radius, R, is found by inspection of a diagram, as the next Example illustrates.

EXAMPLE 2. Find the area of the surface obtained by revolving around the y-axis the part of the parabola $y = x^2$ that lies between $x = 1$ and $x = 2$.

SOLUTION The surface area is $\int_a^b 2\pi R\, ds$. Since the curve is described as a function of x, choose x as the parameter. From Figure 9.5.5(b), $R = x$. Because

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = \sqrt{1 + 4x^2}\, dx$$

the surface area is

$$\int_1^2 2\pi x \sqrt{1 + 4x^2}\, dx.$$

To evaluate the integral, use the substitution $u = 1 + 4x^2$, so $du = 8x\, dx$. Hence $x\, dx = du/8$. The new limits of integration are $u = 5$ and $u = 17$. Thus

$$\text{Surface Area} = \int_5^{17} 2\pi\sqrt{u}\frac{du}{8} = \frac{\pi}{4}\int_5^{17}\sqrt{u}\,du$$
$$= \frac{\pi}{4}\cdot\frac{2}{3}u^{3/2}\Big|_5^{17} = \frac{\pi}{6}(17^{3/2} - 5^{3/2}) \approx 30.84649.$$

♦

EXAMPLE 3. Find the surface area when the curve $r = \cos(\theta)$, θ in $\left[0, \frac{\pi}{2}\right]$ is (a) revolved around the x-axis and (b) revolved around the y-axis.

SOLUTION The curve shown in Figure 9.5.6(a) is the graph of $r = \cos(\theta)$, a semicircle with radius $1/2$ and center $(1/2, 0)$.

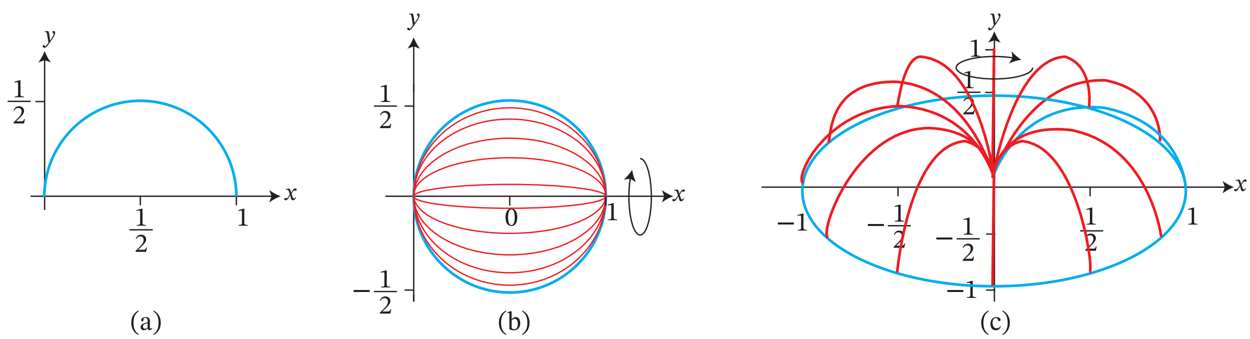


Figure 9.5.6

(a) The solid formed by revolving this semicircle about the x-axis is a sphere with radius $1/2$ (see Figure 9.5.6(b), so its surface area is $4\pi(1/2)^2 = \pi$.

The same result can be found by the ideas presented in this section. We need to find both R and $ds/d\theta$. $R = y = r\sin(\theta) = \cos(\theta)\sin(\theta)$ and using the formula for $ds/d\theta$ for a polar curve from Section 9.4:

$$\frac{ds}{d\theta} = \sqrt{r(\theta)^2 + r'(\theta)^2} = \sqrt{(\cos(\theta))^2 + (-\sin(\theta))^2} = 1.$$

Then

$$\text{Surface Area} = \int_0^{\pi/2} 2\pi R\frac{ds}{d\theta}\,d\theta \qquad = \int_0^{\pi/2} 2\pi\cos(\theta)\sin(\theta)(1)\,d\theta$$
$$= \int_0^{\pi/2} 2\pi\sin(\theta)\cos(\theta)\,d\theta = 2\pi\frac{\sin^2(\theta)}{2}\Big|_0^{\pi/2} = \pi.$$

(b) This surface is the top half of a doughnut whose hole has vanished, more like the top half of a modern bagel. See Figure 9.5.6(c). To find its surface area, $R = x = r\cos(\theta) = \cos^2(\theta)$ and, just as in (a), $ds/d\theta = 1$. Thus, remembering that $\int_0^{\pi/2}\cos^2(\theta)\,d\theta$ was evaluated, in Section 8.5, by rewriting $\cos^2(\theta)$ as $(1 + \cos(2\theta))/2$:

$$\begin{aligned}
\text{Surface Area} &= \int_0^{\pi/2} 2\pi R \frac{ds}{d\theta}\, d\theta && = \int_0^{\pi/2} 2\pi \cos^2(\theta)(1)\, d\theta \\
&= \pi \int_0^{\pi/2} 1+\cos(2\theta)\, d\theta && = \pi\left(\theta + \frac{1}{2}\sin(2\theta)\right)\Big|_0^{\pi/2} \\
&= \pi\left(\frac{\pi}{2}+\frac{1}{2}\sin(\pi) - \left(0+\frac{1}{2}\sin(0)\right)\right) = \pi\left(\frac{\pi}{2}\right) \\
&= \frac{\pi^2}{2} && \approx 4.93480.
\end{aligned}$$

Summary

Figure 9.5.7

This section developed a definite integral for the area of a surface of revolution. It rests on the use of $2\pi R\, ds$ as a local estimate of the area swept out by a short segment of length ds revolved around a line L at a distance R from the segment. (See Figure 9.5.7.) We gave an informal argument for it. Exercises 27 and 28 offer a more formal development of the same result.

EXERCISES for Section 9.5

In Exercises 1 to 4 set up a definite integral for the area of the surface using the suggested parameter. Show the radius R on a diagram. Do not evaluate the definite integrals.

1. The graph of $y = x^3$, x in the interval $[1,2]$, revolved about the x-axis with parameter x.

2. The graph of $y = x^3$, x in the interval $[1,2]$, revolved about the line $y = -1$ with parameter x.

3. The graph of $y = x^3$, x in the interval $[1,2]$, revolved about the y-axis with parameter y.

4. The graph of $y = x^3$, x in the interval $[1,2]$, revolved about the y-axis with parameter x.

---

5. Find the area of the surface formed by rotating about the x-axis that part of the curve $y = e^x$ above $[0,1]$.

6. Find the area of the surface obtained by rotating one arch of the curve $y = \sin(x)$ about the x-axis.

7. One arch of the cycloid given parametrically by $x = \theta - \sin(\theta)$, $y = 1 - \cos(\theta)$ is revolved around the x-axis. Find the area of the surface produced.

8. The curve given parametrically by $x = e^t\cos(t)$, $y = e^t\sin(t)$ $(0 \le t \le \frac{\pi}{2})$ is revolved around the x-axis. Find the area of the surface produced.

In Exercises 9 to 16 find the area of the surface formed by revolving the curve about an axis. Leave the answer as a definite integral, but indicate how it could be evaluated by the fundamental theorem of calculus.

9. $y = 2x^3$ for x in $[0,1]$, about the x-axis.

10. $y = \dfrac{1}{x}$ for x in $[1,2]$, about the x-axis.

11. $y = x^2$ for x in $[1,2]$, about the x-axis.

12. $y = x^{4/3}$ for x in $[1,8]$, about the y-axis.

13. $y = x^{2/3}$ for x in $[1,8]$, about the line $y = 1$.

14. $y = \sqrt{1-x^2}$ for x in $[-1,1]$, about the line $y = -1$.

15. $y = \frac{1}{3}x^3 + \frac{1}{4x}$ for x in $[1,2]$, about the line $y = -1$.

16. $y = \frac{1}{6}x^3 + \frac{1}{2x}$ for x in $[1,3]$, about the y-axis.

In Exercises 17 and 18 estimate the surface area obtained by revolving the arc about the given line. Find a definite integral for the surface area and then use either Simpson's method with six sections or a programmable calculator or computer to approximate the value of the integral.

17. $y = x^{1/4}$, x in $[1,3]$, about the x-axis.

18. $y = x^{1/5}$, x in $[1,3]$, about the line $y = -1$.

Exercises 19 to 22 involve finding the area of a surface obtained by revolving a curve given in polar coordinates.

19. Use a local approximation to show that the area of the surface obtained by revolving the curve $r = f(\theta)$, $\alpha \le \theta \le \beta$, around the polar axis is $\int\limits_{\alpha}^{\beta} 2\pi r \sin(\theta)\sqrt{r^2 + (r')^2}\, d\theta$.

20. Use Exercise 19 to find the surface area of a sphere of radius a.

21. Find the area of the surface formed by revolving the portion of the curve $r = 1 + \cos(\theta)$ in the first quadrant about (a) the x-axis and (b) the y-axis.

22. The curve $r = \sin(2\theta)$, θ in $\left[0, \frac{\pi}{2}\right]$, is revolved around the polar axis. Set up an integral for the surface area.

---

23. When a curve situated above the x-axis is revolved around the x-axis, the area of the resulting surface of revolution is 31. When the curve is revolved around the line $y = -2$, the surface area of this solid is 75. How long is the curve?

24. The portion of the curve $x^{2/3} + y^{2/3} = 1$ in the first quadrant is revolved around the x-axis. Find the area of the surface produced.

25. Although the fundamental theorem of calculus is of no use in computing the perimeter of an ellipse, it is useful in computing the surface area of the football-shaped surface formed when the ellipse is rotated about one of its axes.

(a) Find the area of the surface obtained when the top half of the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ (assume $a > b$), is revolved around the x-axis.

(b) Does your answer give the correct formula for the surface area of a sphere of radius a?

26. The unbounded region bounded by $y = \frac{1}{x}$ and the x-axis and situated to the right of $x = 1$ is revolved around the x-axis to produce a solid region S.

(a) Show that the volume of S is finite but its surface area is infinite.

(b) Does this mean that an infinite surface area can be painted by pouring a finite amount of paint into the solid region it bounds?

Exercises 27 and 28 obtain the formula for the area of the surface obtained by revolving a line segment about a line that does not meet it. (The area was estimated in the text.)

27. A right circular cone has slant height L and radius r, as shown in Figure 9.5.8(a). If it is cut along a line through its vertex and laid flat, it becomes a sector of a circle of radius L, as shown in Figure 9.5.8(b). By comparing Figure 9.5.8(b) to a complete disk of radius L find the area of the sector and thus the area of the cone.

28. Consider a line segment of length L in the plane that does not meet a line in the plane, called the axis. (See Figure 9.5.8(c).) When the segment is revolved around the axis, it sweeps out a curved surface. Show that its area

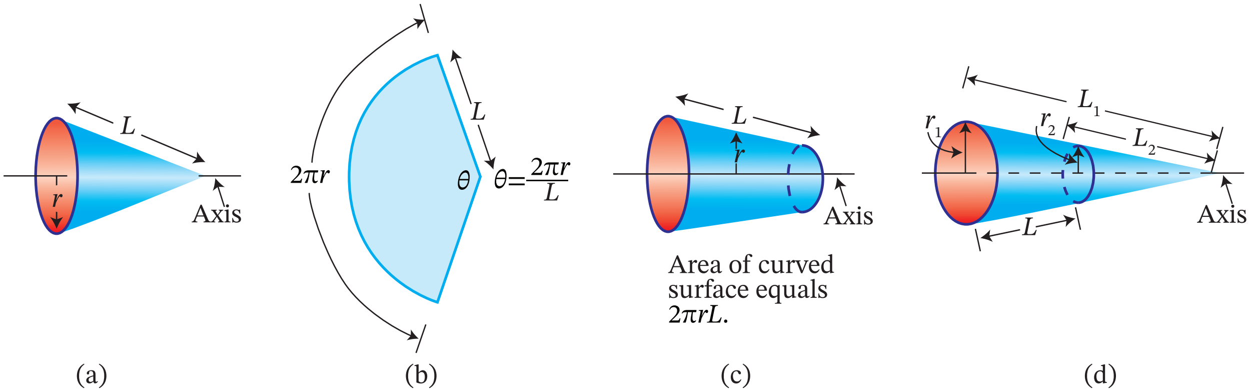


Figure 9.5.8

equals $2\pi rL$ where r is the distance from the midpoint of the line segment to the axis. The surface in Figure 9.5.2 is called a **frustum of a cone**. Follow these steps:

(a) Complete the cone by extending the frustum as shown in Figure 9.5.8(d). Label the radii and lengths as in the figure. Show that $\frac{r_1}{r_2} = \frac{L_1}{L_2}$, so $r_1 L_2 = r_2 L_1$.

(b) Show that the surface area of the frustum is $\pi r_1 L_1 - \pi r_2 L_2$.

(c) Express L_1 as $L_2 + L$ and use the result of (a) to show that $\pi r_1 L_1 - \pi r_2 L_2 = \pi r_2(L_1 - L_2) + \pi r_1 L = \pi r_2 L + \pi r_1 L$.

(d) Show that the surface area of the frustum is $2\pi rL$, where $r = \frac{r_1 + r_2}{2}$.

This justifies our use of the approximation $2\pi R\, ds$ in the definition of area of a solid of revolution.

Historical Note: Archimedes' Greatest Accomplishment

Exercise 29 was solved by Archimedes more than 2300 years ago. He considered it his greatest accomplishment. Archimedes was killed by a Roman soldier in 212 B.C. Cicero was quaestor (elected official who supervised the state treasury and conducted audits) in 75 B.C. About two centuries after Archimedes' death, Cicero wrote

> I shall call up from the dust [the ancient equivalent of a blackboard] and his measuring-rod an obscure, insignificant person belonging to the same city [Syracuse], who lived many years after, Archimedes. When I was quaestor I tracked out his grave, which was unknown to the Syracusans (as they totally denied its existence), and found it enclosed all round and covered with brambles and thickets; for I remembered certain doggerel lines inscribed, as I had heard, upon his tomb, which stated that a sphere along with a cylinder had been set up on the top of his grave. Accordingly, after taking a good look around (for there are a great quantity of graves at the Agrigentine Gate), I noticed a small column rising a little above the bushes, on which there was the figure of a sphere and a cylinder. And so I at once said to the Syracusans (I had their leading men with me) that I believed it was the very thing of which I was in search. Slaves were sent in with sickles who cleared the ground of obstacles, and when a passage to the place was opened we approached the pedestal fronting us; the epigram was traceable with about half the lines legible, as the latter portion was worn away.
>
> REFERENCE: Cicero, *Tusculan Disputations*, vol. 23, translated by J. E. King, Loef Classical Library, Harvard University, Cambridge, 1950.

29. Consider the smallest tin can that contains a given sphere.

The height and diameter of the tin can equal the diameter of the sphere.

(a) Compare the volume of the sphere with the volume of the tin can.

(b) Compare the surface area of the sphere with the total surface area of the can. SEE ALSO Exercise 31.

30. (a) Compute the area of the portion of a sphere of radius a that lies between two parallel planes at distances c and $c+h$ from the center of the sphere ($0 \le c \le c+h \le a$).

(b) The result in (a) depends only on h, not on c. What does this mean geometrically? (See Figure 9.5.9.)

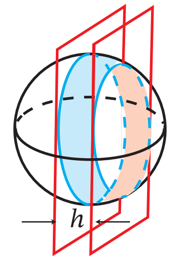


Figure 9.5.9

31. Mercator maps are discussed in CIE 12 at the end of this chapter. A fundamental characteristic of a Mercator map is that it preserves angles. A **Lambert azimuthal equal-area projection** map preserves areas, but not angles. It is made by projecting a sphere on a cylinder tangent at the equator by rays parallel to the equatorial plane and having one end on the diameter that joins the north and south poles, as shown in Figure 9.5.10.

Explain why a Lambert map preserves areas.

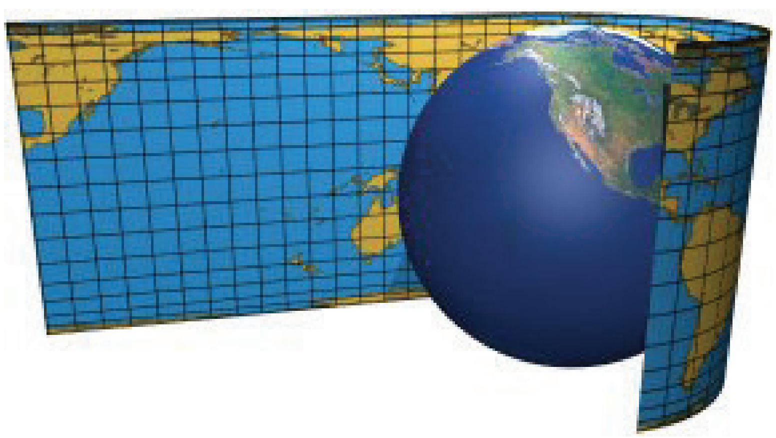


Figure 9.5.10

32. Soup is poured into a hemispherical soup bowl of radius a at a constant rate, k cubic centimeters per second. At what rate is the wetted part of the bowl increasing when the soup has depth h?

---

33. The derivative with respect to r of the volume of a sphere is its surface area: $\frac{d}{dr}\left(\frac{4}{3}\pi r^3\right) = 4\pi r^2$. Is this a coincidence?

34. For some continuous functions $f(x)$ the definite integral $\int_a^b f(x)\,dx$ depends only on the width of the interval $[a,b]$. That is, there is a function $g(x)$ such that, for all a and b, $a<b$, in some interval,

$$\int_a^b f(x)\,dx = g(b-a). \tag{9.5.4}$$

(a) Show that a constant function $f(x)$ satisfies (9.5.4).

(b) Prove that if $f(x)$ satisfies (9.5.4), then it must be constant. See also Exercise 30.

9.6 Curvature

In this section we use calculus to obtain a measure of the curviness or curvature of a curve. This concept will be generalized in Section 15.2 in the study of motion along a curved path in space.

Introduction

Imagine a bug crawling around a circle of radius one centimeter, as in Figure 9.6.1(a). As it walks a small distance, say 1 cm, it notices that its direction changes. Her direction at any point is represented by a tangent at that point. Another bug walks around a circle of radius 2 cm, as in Figure 9.6.1(b). When it goes 1 cm. his direction changes less. The first bug feels that her circle is curvier than the circle of the second bug. She is right. In fact, according to the measure of curviness that is introduced next, the smaller circle is twice as curvy as the larger circle.

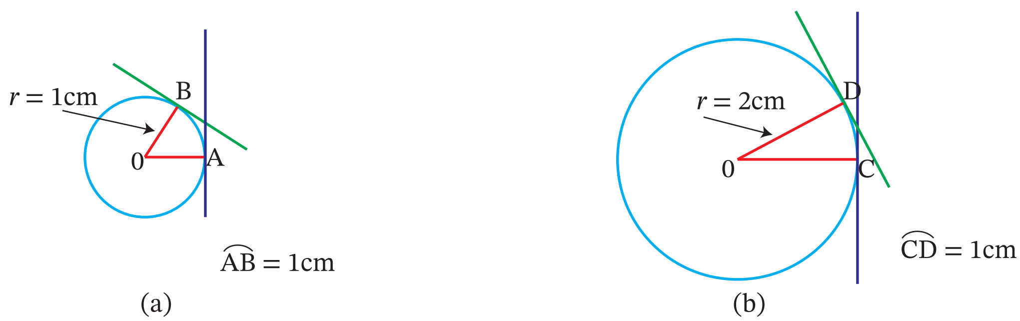


Figure 9.6.1

In this section we will provide a measure of curviness, called **curvature**, and will show how to compute curvature when a curve is given in rectangular coordinates. The introduction to Exercises 22 to 24 treats the curvature of a curve given in polar coordinates.

A straight line will have zero curvature everywhere. A circle of radius a will turn out to have curvature $1/a$ everywhere. For most other curves the curvature varies from point to point. Curvature is not defined at a corner.

Definition of Curvature

Curvature measures how rapidly the direction changes as we move a small distance along a curve. We have a way of assigning a numerical value to direction, namely, the angle of the tangent line. The rate of change of this angle with respect to arc length will be our measure of curvature.

Definition: *Curvature*

Assume that a curve is given parametrically, with s the arc length from a fixed point P_0. Let ϕ be the angle between the tangent line at P and the x-axis. The **curvature** κ at P is the absolute value of the derivative, $\frac{d\phi}{ds}$,

$$\text{Curvature} = \kappa = \left|\frac{d\phi}{ds}\right|$$

whenever the derivative exists. (See Figure 9.6.2.)

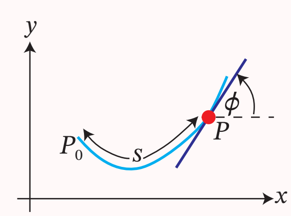


Figure 9.6.2

The Greek letter κ ("kappa") corresponds to our "k".

For a straight line the angle ϕ is constant, so its curvature is zero everywhere, as previously noted. Also, at a corner, curvature is not defined since the slope of the tangent line is discontinuous there.

The next theorem shows that curvature of a small circle is large and the curvature of a large circle is small, in agreement with the bugs' experience.

Theorem 9.6.1: Curvature of Circles

For a circle of radius a, the curvature $\left|\frac{d\phi}{ds}\right|$ is constant and equals $\frac{1}{a}$, the reciprocal of the radius.

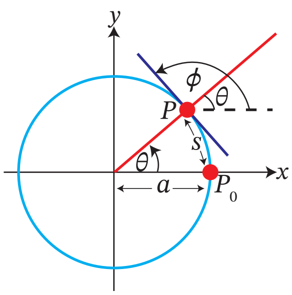


Figure 9.6.3

Proof of Theorem 9.6.1

We need to express ϕ as a function of arc length s on a circle of radius a. In Figure 9.6.3 measure arc length s counterclockwise from the point P_0 on the x-axis. Then $\phi = \pi/2 + \theta$, as Figure 9.6.3 shows. By definition of radian measure, $s = a\theta$, so that $\theta = s/a$. Then $\phi = \pi/2 + s/a$. Differentiating with respect to arc length yields:

$$\frac{d\phi}{ds} = \frac{1}{a},$$

as claimed.

Computing Curvature

When a curve is given in the form $y = f(x)$, its curvature can be expressed in terms of the first and second derivatives, dy/dx and d^2y/dx^2.

Theorem 9.6.2: Curvature of $y = f(x)$

Let arc length s be measured along the curve $y = f(x)$ from a point P_0. Assume that x increases as s increases and that y' and y'' are continuous. Then

$$\textit{Curvature} = \kappa = \frac{\left|\frac{d^2y}{dx^2}\right|}{\left(1+\left(\frac{dy}{dx}\right)^2\right)^{3/2}}.$$

Proof of Theorem 9.6.2

The chain rule, $d\phi/dx = (d\phi/ds)(ds/dx)$, implies

$$\frac{d\phi}{ds} = \frac{\frac{d\phi}{dx}}{\frac{ds}{dx}}.$$

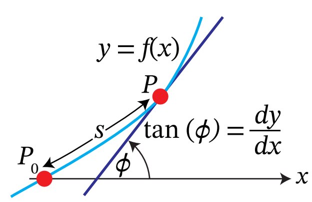


Figure 9.6.4

As was shown in Section 9.3,

$$\frac{ds}{dx} = \left(1+\left(\frac{dy}{dx}\right)^2\right)^{1/2}.$$

All that remains is to express $d\phi/dx$ in terms of dy/dx and d^2y/dx^2. In Figure 9.6.4,

$$\frac{dy}{dx} = \text{slope of tangent line to the curve} = \tan(\phi).$$

We find $d\phi/dx$ by differentiating both sides of $dy/dx = \tan(\phi)$ with respect to x. Thus

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\tan(\phi)\right) = \sec^2(\phi)\cdot\frac{d\phi}{dx} = \left(1+\tan^2(\phi)\right)\frac{d\phi}{dx} = \left(1+\left(\frac{dy}{dx}\right)^2\right)\frac{d\phi}{dx}.$$

Solving for $d\phi/dx$, we get

$$\frac{d\phi}{dx} = \frac{\frac{d^2y}{dx^2}}{1+\left(\frac{dy}{dx}\right)^2}.$$

Consequently,

$$\frac{d\phi}{ds} = \frac{\frac{d\phi}{dx}}{\frac{ds}{dx}} = \frac{\frac{d^2y}{dx^2}}{\left(1+\left(\frac{dy}{dx}\right)^2\right)\sqrt{1+\left(\frac{dy}{dx}\right)^2}} = \frac{\frac{d^2y}{dx^2}}{\left(1+\left(\frac{dy}{dx}\right)^2\right)^{3/2}},$$

and the theorem is proved.

Observation 9.6.3: *Some Geometry of Curvature*

One might have expected the curvature to depend only on the second derivative, d^2y/dx^2, since it measures the rate at which the slope changes. Theorem 9.6.2 shows that the curvature also depends on the first derivative, and exactly how the two derivatives combine to determine the curvature at any point along the graph of $y = f(x)$. See also Exercise 28.

EXAMPLE 1. Find the curvature of $y = x^2$.

SOLUTION

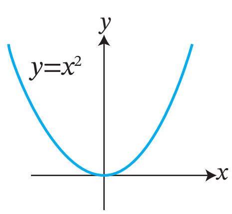


Figure 9.6.5

We have $dy/dx = 2x$ and $d^2y/dx^2 = 2$ so

$$\kappa = \frac{\left|\frac{d^2y}{dx^2}\right|}{\left(1+\left(\frac{dy}{dx}\right)^2\right)^{3/2}} = \frac{2}{\left(1+(2x)^2\right)^{3/2}}.$$

The maximum curvature occurs when $x = 0$. The curvatures at (x, x^2) and at $(-x, x^2)$ are equal. As $|x|$ increases, the curve becomes straighter and the curvature approaches 0. (See Figure 9.6.5.)

Curvature of a Parameterized Curve

Theorem 9.6.2 tells how to find the curvature if y is given as a function of x. It holds also when the curve is described parametrically, where x and y are functions of a parameter. All that is needed are appropriate expressions for the two derivatives that appear in the formula for curvature:

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} \quad\text{and}\quad \frac{d^2y}{dx^2} = \frac{\frac{d}{dt}\left(\frac{dy}{dx}\right)}{\frac{dx}{dt}}. \tag{9.6.1}$$

As explained in Section 9.3, both equations in (9.6.1) are consequences of the chain rule.

EXAMPLE 2. In Example 2 in Section 9.3 we concluded that the parametric equations for the cycloid created by a wheel of radius 1 are $x = \theta - \sin(\theta)$ and $y = 1 - \cos(\theta)$, Find the curvature of this cycloid.

SOLUTION We find dy/dx in terms of θ:

$$\frac{dx}{d\theta} = 1 - \cos(\theta) \qquad \text{and} \qquad \frac{dy}{d\theta} = \sin(\theta).$$

So

$$\frac{dy}{dx} = \frac{\sin(\theta)}{1 - \cos(\theta)}.$$

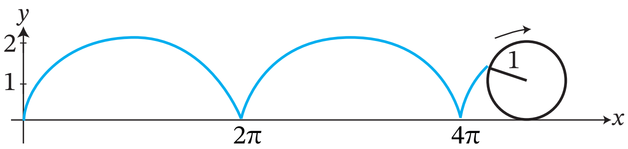


Figure 9.6.6

Similar calculations show that

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{\frac{d}{d\theta}\left(\frac{dy}{dx}\right)}{\frac{dx}{d\theta}} = \frac{\frac{d}{d\theta}\left(\frac{\sin(\theta)}{1-\cos(\theta)}\right)}{1-\cos(\theta)} = \frac{-1}{(1-\cos(\theta))^2}.$$

Thus the curvature is

$$\kappa = \frac{\left|\frac{d^2y}{dx^2}\right|}{\left(1 + \left(\frac{dy}{dx}\right)^2\right)^{3/2}} = \frac{\left|\frac{-1}{(1-\cos(\theta))^2}\right|}{\left(\frac{2}{1-\cos(\theta)}\right)^{3/2}} = \frac{1}{2^{3/2}\sqrt{1-\cos(\theta)}}.$$

Since $y = 1 - \cos(\theta)$ and $2^{3/2} = \sqrt{8}$, the curvature equals $1/\sqrt{8y}$ at all points with $y > 0$. The cycloid has a "corner" each time $y = 0$; at these points the curvature is not defined. ♦

Radius of Curvature

As Theorem 9.6.1 shows, a circle with curvature κ has radius $1/\kappa$. This suggests the definition

Definition: *Radius of Curvature*

The **radius of curvature** of a curve at a point is the reciprocal of the curvature there:

$$\text{Radius of Curvature} = \frac{1}{\text{Curvature}} = \frac{1}{\kappa}.$$

As can be checked, the radius of curvature of a circle of radius a is $1/a$.

The cycloid in Example 2 has radius of curvature at the point (x, y) equal to $\sqrt{8y}$. The higher the point on the curve, the straighter the curve. And the cycloid is nearly vertical at points near the x-axis. See Exercise 29.

Osculating Circle

The line through a point P on a curve that looks most like the curve near P is the tangent line. The circle through P that looks most like the curve near P has the same slope at P as the curve and a radius equal to the radius of curvature at P. It is called the **osculating circle**, from the Latin *osculum*, meaning kiss.

At a given point P on a curve, the osculating circle at P is defined to be that circle that passes through P, has the same slope at P as the curve does, and also has the same curvature at P.

For the parabola $y = x^2$ of Example 1, when $x = 1$, the curvature is $2/5^{3/2}$ and the radius of curvature is $5^{3/2}/2 \approx 5.5902$. Figure 9.6.7 shows the osculating circle at $(1, 1)$.

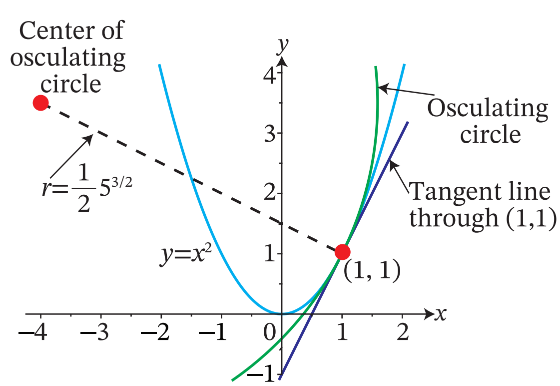


Figure 9.6.7

Notice that the osculating circle crosses the parabola as it passes through the point $(1,1)$. Although this may seem surprising, a little reflection will show why it is to be expected.

Think of driving along the parabola $y = x^2$. If you start at $(1,1)$ and drive up along the parabola, the curvature decreases. Because it is smaller than the curvature of the osculating circle at $(1,1)$, the curve would be straighter than the osculating circle at $(1,1)$ and you would be outside that circle. If you start at $(1,1)$ and move toward the origin (to the left) on the parabola, the curvature increases and is greater than that of the osculating circle at $(1,1)$. You would be driving inside the osculating circle at $(1,1)$. This shows why the osculating circle crosses the curve. The only osculating circle that does not cross the curve $y = x^2$ at its point of tangency is the one that is tangent at $(0,0)$, where the curvature is a maximum.

Summary

The curvature κ of a curve was defined as the absolute value of the rate at which the angle between the tangent line and the x-axis changes as a function of arc length: curvature equals $|d\phi/ds|$. If the curve is the graph of $y = f(x)$, then

$$\kappa = \frac{\left|\dfrac{d^2y}{dx^2}\right|}{\left(1+\left(\dfrac{dy}{dx}\right)^2\right)^{3/2}}.$$

If the curve is given in terms of a parameter t, then compute dy/dx and d^2y/dx^2 with the aid of this version of the chain rule,

$$\frac{d(\,)}{dx} = \frac{\dfrac{d(\,)}{dt}}{\dfrac{dx}{dt}}, \tag{9.6.2}$$

the both empty parentheses enclosing first y and then dy/dx, respectively.

The radius of curvature is the reciprocal of the curvature and the osculating circle is the circle that kisses the graph (by matching both slope and curvature).

EXERCISES for Section 9.6

In Exercises 1 to 6 find the curvature and radius of curvature of the curve at the point.

1. $y = x^2$ at $(1,1)$

2. $y = \cos(x)$ at $(0,1)$

3. $y = \ln(x)$ at $(e,1)$

4. $y = e^{-x}$ at $\left(1,\dfrac{1}{e}\right)$

5. $y = \tan(x)$ at $\left(\dfrac{\pi}{4},1\right)$

6. $y = \sec(2x)$ at $\left(\dfrac{\pi}{6},2\right)$

In Exercises 7 to 10 find the curvature of the curve for the value of the parameter.

7. $x = 2\cos(3t)$, $y = 2\sin(3t)$ at $t = 0$

8. $x = 1+t^2$, $y = t^3+t^4$ at $t = 2$

9. $x = e^{-t}\cos(t)$, $y = e^{-t}\sin(t)$ at $t = \dfrac{\pi}{6}$

10. $x = \cos^3(\theta)$, $y = \sin^3(\theta)$ at $\theta = \dfrac{\pi}{3}$

11.

(a) Compute the curvature and radius of curvature for the curve $y = \dfrac{e^x+e^{-x}}{2}$.

(b) Show that the radius of curvature at (x,y) is y^2.

12. Find the radius of curvature along the curve $y = \sqrt{a^2 - x^2}$, where a is a constant.

13. For what value of x is the radius of curvature of $y = e^x$ smallest?

14. For what value of x is the radius of curvature of $y = x^2$ smallest?

15. (a) Show that at a point where a curve has its tangent parallel to the x-axis its curvature is the absolute value of the second derivative. (b) Show that the curvature is never larger than the absolute value of $\frac{d^2y}{dx^2}$.

16. An engineer lays out a railroad track consisting either of circular arcs or of straight segments. After a train runs over the track, the engineer is fired because the curvature is not a continuous function. Why should the curvature be continuous?

17. Railroad curves are banked to reduce wear on the rails and flanges. The greater the radius of curvature, the less the curve must be banked. The best bank angle A satisfies $\tan(A) = v^2/(32R)$, where v is speed in feet per second and R is radius of curvature in feet. A train travels in the elliptical track $\frac{x^2}{1000^2} + \frac{y^2}{500^2} = 1$ at 60 miles per hour. Find the best angle A (a) at $(1000, 0)$ and (b) at $(0, 500)$.

CAUTION: Pay attention to the units. While x and y are measured in feet, 60 mph = 88 fps.

18. The flexure formula in the theory of beams asserts that the bending moment M required to bend a beam is proportional to the curvature, $M = c/R$, where c is a constant depending on the beam and R is the radius of curvature. A beam is bent to form the parabola $y = x^2$. Find the ratio between the bending moments (a) at $(0, 0)$ and (b) at $(2, 4)$.

Exercises 19 to 21 are related.

19. Find the radius of curvature at a point on the curve with parametric equations $x = a\cos(\theta)$, $y = b\sin(\theta)$.

20.
(a) Show, by eliminating θ, that the curve in Exercise 19 is the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.
(b) Find the radius of curvature of the ellipse (i) at $(a, 0)$ and (ii) at $(0, b)$.

21. An ellipse has major axis of length 6 and minor axis of length 4. Draw the circles that most closely approximate the ellipse at the four points that lie at the extremities of its axes. (See Exercises 19 and 20.)

In each of Exercises 22 to 24 a curve is given in polar coordinates. To find its curvature write it in rectangular coordinates with parameter θ, using $x = r\cos(\theta)$ and $y = r\sin(\theta)$.

22. Find the curvature of $r = a\cos(\theta)$.

23. Find the curvature of $r = \cos(2\theta)$.

24. Show the cardioid $r = 1 + \cos(\theta)$ has curvature $\frac{3\sqrt{2}}{4\sqrt{r}}$.

---

25. If $\frac{dy}{dx} = y^3$, express the curvature in terms of y.

26. As is shown in physics, the larger the radius of curvature of a turn, the faster a car can travel around it. The required radius of curvature is proportional to the square of the maximum speed. That says that the maximum speed around a turn is proportional to the square root of the radius of curvature. If a car moving on the path $y = x^3$ (x and y measured in miles) can go 30 miles per hour at $(1, 1)$ without sliding off, how fast can it go at $(2, 8)$?

27. Find the local extrema of the curvature of (a) $y = x + e^x$, (b) $y = e^x$, (c) $y = \sin(x)$, and (d) $y = x^3$.

28. Sam says, "I don't like the definition of curvature. It should be the rate at which the slope changes as a function of x. That is $\frac{d}{dx}\left(\frac{dy}{dx}\right)$, which is the second derivative, $\frac{d^2y}{dx^2}$." Give an example of a curve that would have constant curvature according to Sam's definition but whose changing curvature is obvious to the eye.

29. In Example 2 some of the steps were omitted in finding that the cycloid given by $x = \theta - \sin(\theta)$, $y = 1 - \cos(\theta)$ has curvature $\kappa = \frac{1}{2^{3/2}\sqrt{1-\cos(\theta)}} = \frac{1}{\sqrt{8y}}$. In this exercise you are asked to fill in the details.

(a) Verify that $\frac{dy}{dx} = \frac{\sin(\theta)}{1-\cos(\theta)}$.

(b) Verify that $\frac{d}{d\theta}\left(\frac{dy}{dx}\right) = \frac{-1}{1-\cos(\theta)}$.

(c) Verify that $\frac{d^2y}{dx^2} = \frac{-1}{(1-\cos(\theta))^2}$.

(d) Verify that $1 + \left(\frac{dy}{dx}\right)^2 = \frac{2}{1-\cos(\theta)}$.

(e) Compute the curvature, κ, in terms of θ.

(f) Express it in terms of x and y.

(g) At what points on the cycloid is the curvature largest?

(h) At what points on the cycloid is the curvature smallest?

30. Assume that g and h are functions with continuous second derivatives. and that as we move along the curve $x = g(t)$, $y = h(t)$, the arc length s from a point P_0 increases as t increases. Show that $\kappa = \frac{|\dot{x}\ddot{y} - \dot{y}\ddot{x}|}{\left(\dot{x}^2 + \dot{y}^2\right)^{3/2}}$.

NOTATION: Newton's dot notation for derivatives is sometimes used to shorten expressions involving lots of derivatives (with respect to time): $\dot{x} = dx/dt$, $\ddot{x} = d^2x/dt^2$, $\dot{y} = dy/dt$, and $\ddot{y} = d^2y/dt^2$.

31. Use the result of Exercise 30 to find the curvature of the cycloid of Example 2, which has parametric equations $x = \theta - \sin(\theta)$, $y = 1 - \cos(\theta)$

32. If a plane curve has a constant radius of curvature, R, must it be part of a circle? That the answer is "yes" is important in experiments conducted with a cyclotron. Physical assumptions imply that the path of an electron entering a uniform magnetic field at right angles to the field has constant curvature. To show that it follows that the path is part of a circle, view the curve as parameterized by the angle ϕ of the tangent.

Show that (a) $\frac{ds}{d\phi} = R$, (b) $\frac{dx}{d\phi} = R\cos(\phi)$, and (c) $\frac{dy}{d\phi} = R\sin(\phi)$. (d) Interpret (a), (b), and (c) in the special case when the curvature, κ, is constant.

CONTRIBUTED BY: G.D. Chakerian

Parts (a), (b), and (c) refer to curves in general, where the radius of curvature may not be constant. Part (d) treats the special case where R is constant. Recall that s denotes arc length.

33. In Example 2 we found the curvature of the cycloid produced by rolling a wheel with radius 1.

(a) What do you think the radius of curvature will be at the highest points on this cycloid?

(b) Show that the radius of curvature at the highest points on the cycloid is twice the rolling wheel's diameter.

(c) Since, at each moment, the wheel is rotating about its point of contact with the ground, why is the radius of curvature not just the diameter of the wheel?

34. A smooth convex curve with length L, continuous curvature, and no straight edges bounds a region.

(a) Show that the average of its curvature, as a function of arc length, is $\frac{2\pi}{L}$.

(b) Check that this is correct for a circle of radius a.

(c) Must there be a point on the curve where the curvature is $\frac{2\pi}{L}$?

9.S Chapter Summary

This chapter dealt mostly with curves described in polar coordinates and curves given parametrically, and with surfaces obtained by rotating one of these curves about an axis. Table 9.S.1 summarizes the main ideas in the chapter. Area, arc length, and surface area involve integrals while speed and curvature are based on instantaneous rates of change (that is, on derivatives).

EXERCISES for Section 9.S

1. When driving along a curved road, which is more important in avoiding car sickness, $\frac{d\phi}{ds}$ or $\frac{d\phi}{dt}$, where t is time. Explain the difference.

2. Evaluate $\int_0^{\pi/2} \cos^2(\theta)\, d\theta$ and $\int_0^{\pi} (1+\cos(\theta))^2\, d\theta$ as simply as possible.

These definite integrals are needed in the solution to Example 3 in Section 9.2.

3. A triangle ABC is inscribed in a circle, with AB a diameter of the circle.

(a) Using geometry, show that angle ACB is a right angle.

(b) Instead, using the equation of a circle in rectangular coordinates, show that AC and BC are perpendicular.

(c) Use (a) or (b) to show that the graph in the plane of $r = b\cos(\theta)$ is a circle of diameter b passing through the pole.

Exercises 4 and 5 require the use of the inequality $\int_0^{2\pi} f(\theta)g(\theta)\, d\theta \le \left(\int_0^{2\pi} f(\theta)^2\, d\theta\right)^{1/2} \left(\int_0^{2\pi} g(\theta)^2\, d\theta\right)^{1/2}$. This result is a special case of the **Cauchy-Schwarz inequality** used in a CIE 11 at the end of this chapter.

4. Let P be a point inside a region in the plane bounded by a smooth convex curve. Place the pole of a polar coordinate system at P. Let $c(\theta)$ be the length of the chord of angle θ through P. Show that $\int_0^{2\pi} c(\theta)^2\, d\theta \le 8A$, where A is the area of the region.

RECALL: A "smooth" curve is a curve that has a continuously defined tangent line.

5. Show that if $\int_0^{2\pi} c(\theta)^2\, d\theta = 8A$ then P is the midpoint of each chord through P.

---

6. A physics midterm includes the following information:

When $r = \sqrt{x^2+y^2}$ and y is a constant, recall that $\int \frac{dx}{r} = \ln(x+r)$, $\int \frac{x}{r}\, dx = r$, and $\int \frac{dx}{r^3} = \frac{x}{y^2 r}$.

Show, by differentiating, that these three equations are correct.

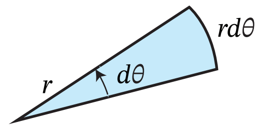
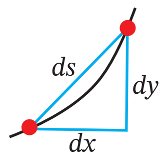
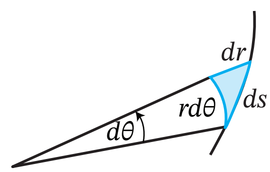
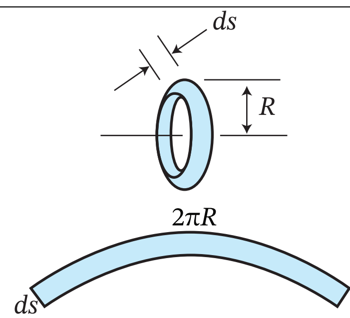
7. Let L be the line $3x+4y=1$. Consider the function that assigns to the point with polar coordinates (r,θ), r not equal to 0, the point $\left(\frac{1}{r},\theta\right)$.

(a) Plot L and at least four images of points on L.

(b) Sketch what you suspect is the image of L.

(c) Find the equation, in rectangular coordinates, of the image of L.

(d) What kind of curve is the image of L?

<table>
<tr><th>Concept</th><th>Memory Aid</th><th>Comment</th></tr>
<tr>
<td>$\text{Area} = \int_{\alpha}^{\beta} \frac{r^2}{2}\, d\theta$</td>
<td>
</td>
<td>The narrow sector resembles a triangle of base $r\ d\theta$ and height r, so
$$dA = \frac{1}{2}(r\ d\theta)(r) = \frac{1}{2}r^2\ d\theta.$$</td>
</tr>
<tr>
<td>$\text{Arc length} = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt$
$= \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$</td>
<td>
</td>
<td>A short part of the curve is almost straight, suggesting
$$(ds)^2 = (dx)^2 + (dy)^2.$$</td>
</tr>
<tr>
<td>$\text{Speed} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}$
$= \sqrt{\left(r\frac{d\theta}{dt}\right)^2 + \left(\frac{dr}{dx}\right)^2}$</td>
<td>
</td>
<td>The shaded area with two curved sides looks like a right triangle, suggesting
$$(ds)^2 = (r\ d\theta)^2 + (dr)^2.$$</td>
</tr>
<tr>
<td>Area of surface of revolution $= \int_a^b 2\pi R\, ds$</td>
<td>
</td>
<td>The surface area of a thin strip of the cylinder with radius R and width ds is
$$dS = 2\pi R\ ds.$$</td>
</tr>
<tr>
<td>$\text{Curvature} = \kappa = \left|\frac{d\phi}{ds}\right|$

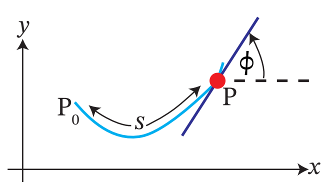
where $\tan(\phi) = \frac{dy}{dx}$
and s is arc length</td>
<td>
</td>
<td><u>Case 1:</u> $y = f(x)$
Use the chain rule to write $\left|\frac{d\phi}{ds}\right|$ as $\left|\frac{d\phi/dx}{ds/dx}\right|$, then
$$\kappa = \frac{|y''|}{\left(1+(y')^2\right)^{3/2}}.$$
<u>Case 2:</u> $x = x(t)$, $y = y(t)$
(i) Replace $\frac{dy}{dx}$ with $\frac{dy/dt}{dx/dt}$ and
(ii) replace $\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right)$ with $\frac{\frac{d}{dt}\left(\frac{dy}{dx}\right)}{dx/dt}$.</td>
</tr>
</table>

Table 9.S.1

8. Let $r = f(\theta)$ describe a convex curve surrounding the origin.

(a) Show that $\int_0^{2\pi} r\, d\theta \le$ arc length of the boundary.

(b) Show that if the equality holds in (a), the curve is a circle with center at the origin.

9. **Sam:** I have discovered an easy formula for the length of a closed curve that encloses the origin.

Jane: Well?

Sam: First of all, $\int_0^{2\pi} \sqrt{r^2 + (r')^2}\, d\theta$ is obviously greater than or equal to $\int_0^{2\pi} r\, d\theta$.

Jane: I will grant you that much, because $(r')^2$ is never negative.

Sam: Now, if a and b are not negative, $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$.

Jane: Why?

Sam: Just square both sides. So $\sqrt{r^2 + (r')^2} \le \sqrt{r^2} + \sqrt{(r')^2} = r + r'$.

Jane: Looks all right to me, so far.

Sam: Thus $\int_0^{2\pi} \sqrt{r^2 + (r')^2}\, d\theta \le \int_0^{2\pi} (r + r')\, d\theta = \int_0^{2\pi} r\, d\theta + \int_0^{2\pi} r'\, d\theta$. But $\int_0^{2\pi} r'\, d\theta$ equals $r(2\pi) - r(0)$, which is 0.

So, putting it all together, I get $\int_0^{2\pi} r\, d\theta \le \int_0^{2\pi} \sqrt{r^2 + (r')^2}\, d\theta \le \int_0^{2\pi} r\, d\theta$. So the arc length is $\int_0^{2\pi} r\, d\theta$.

Jane: That cannot be right. If it were, it would be an Exercise.

Sam: They like to keep a few things secret to surprise us on a mid-term.

Who is right, Sam or Jane? Explain.

10. (a) Let the radius of Earth be r miles. Let $f(h)$ be the fraction of Earth's surface can be seen from an object h miles above it?

(b) What is $\lim_{h\to\infty} f(h)$?

(c) What is $f(h)$ when h is 100 miles? Earth's radius is 3963 miles.

(d) What is $f(h)$ when h is 22,236 miles, the altitude of a geosynchronous satellite?

In Exercises 11 and 12, L is the length of a smooth, closed curve C and P is a point in the region R bounded by C.

See also Exercise 24 in Section 9.4.

11. (a) Let the area of R be A. Show that the average distance from P to points on the curve, averaged with respect to arc length, is greater than or equal to $\frac{2A}{L}$.

(b) Give an example when equality holds.

12. (a) Show that the average distance from P to points on the curve, averaged with respect to the polar angle, is less than or equal to $L/(2\pi)$.

(b) Give an example when equality holds.

In Exercises 13 and 14, a, b, c, m, and p are constants. Verify that the derivative of the first expression is the second expression.

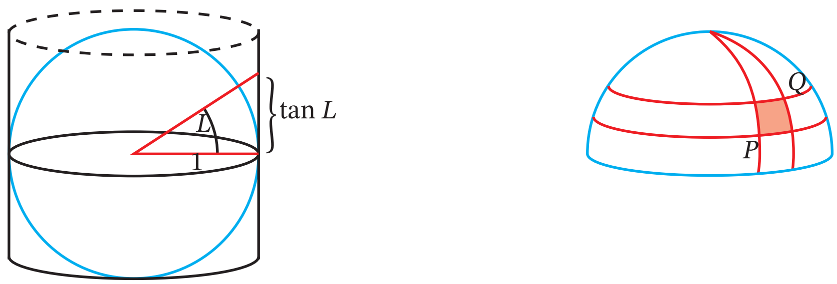
13. $\frac{1}{\sqrt{c}} \arcsin\left(\frac{cx - b}{\sqrt{b^2 + ac}}\right)$; $\sqrt{\frac{c}{a + 2bx - cx^2}}$.

14. $\frac{1}{c}\sqrt{a + 2bx + cx^2} - \frac{b}{\sqrt{c}} \ln\left(b + cx + \sqrt{c}\sqrt{a + 2bx + cx^2}\right)$; $\frac{x}{a + bx + cx^2}$

Calculus is Everywhere # 12

The Mercator Map

One way to make a map of a sphere is to wrap a paper cylinder around the sphere and project points on the sphere onto the cylinder by rays from the center of the sphere. This **central cylindrical projection** is illustrated in Figure C.12.1(a). Points at latitude L project onto points at height $\tan(L)$ from the equatorial plane.



(a) (b)

Figure C.12.1

A **meridian** is a great circle passing through the north and south poles. It corresponds to a fixed longitude. A short segment on a meridian at latitude L of length dL projects onto the cylinder in a segment of length approximately $d(\tan(L)) = \sec(L)^2\ dL$. Thus, the map magnifies short vertical segments at latitude L by the factor $\sec^2(L)$.

Points on the sphere at latitude L form a circle of radius $\cos(L)$. Its image on the cylinder is a circle of radius 1, so the projection magnifies horizontal distances at latitude L by $1/\cos(L) = \sec(L)$.

Consider the effect on a small "almost rectangular" patch bordered by two meridians and two latitude lines, shaded in Figure C.12.1(b). The vertical edges are magnified by $\sec^2(L)$, but the horizontal edges by only $\sec(L)$. The image on the cylinder will not resemble the patch, for it is stretched more vertically than horizontally.

The path a ship sailing from P to Q makes an angle with the latitude line through P. The map distorts it. The ship's captain would prefer a map without distortion, one that preserves direction. That way, to chart a voyage from A to B on the sphere corresponding to Earth's surface at a fixed compass heading, simply draw a straight line from A to B on the map to determine the compass setting.

In 1569, Gerhardus Mercator designed a map that preserves direction by making the vertical magnification the same as the horizontal distortion, $\sec(L)$.

Let y be the height on the map that represents latitude L_0. Then Δy should be approximately $\sec(L)\Delta L$. Taking the limit of $\Delta y/\Delta L$ as ΔL approaches 0, we see that $dy/dL = \sec(L)$. Thus

$$y = \int_0^{L_0} \sec(L)\ dL. \tag{C.12.1}$$

Mercator, working a century before the invention of calculus, did not have the integral or the Fundamental Theorem of Calculus. Instead, he broke the interval $[0, L_0]$ into several short sections of length ΔL, computed $(\sec(L))\Delta L$ for each, and summed to estimate y in (C.12.1).

Using calculus, we see

$$y = \int_0^{L_0} \sec(L)\ dL = \ln|\sec(L) + \tan(L)|\Big|_0^{L_0} = \ln(\sec(L_0) + \tan(L_0)) \qquad \text{for } 0 \le L_0 \le \frac{\pi}{2}.$$

In 1645, Henry Bond conjectured from numerical evidence that $\int_0^\alpha \sec(\theta)\ d\theta = \ln(\tan(\alpha/2 + \pi/4))$ but offered no proof. In 1666, Nicolaus Mercator (no relation to Gerhardus) offered the royalties on one of his inventions to the mathematician who could prove Bond's conjecture was right. Within two years James Gregory provided the proof.

Figure C.12.2

Figure C.12.2 shows a Mercator map. Though it preserves angles, it greatly distorts areas: Greenland (green) looks bigger than South America (brown) even though it is only one-eighth the size of South America. The first map we described distorts areas even more than does a Mercator map.

EXERCISES for CIE C.12

1. Draw a clear diagram showing why segments at latitude L are magnified vertically by the factor $\sec(L)$.

2. Explain why a short arc of length dL in Figure C.12.1(a) projects onto a short interval of length approximately $\sec^2(L)\, dL$.

3. On a Mercator map, what is the ratio of the distance between the lines representing latitudes 60° and 50° to the distance between the lines representing latitudes 40° and 30°?

4. What magnifying effect does a Mercator map have on areas?

5. If the distance on a Mercator map is 3 inches from latitude 0° to latitude 20° how far is it on the map (a) from 60° to 80° and (b) from 75° to 85°.

6. Bond's conjecture was first encountered in Exercise 52 in Section 8.5. Show that it is correct. That is, show that

$$\int_0^a \sec(\theta)\, d\theta = \ln\left(\tan\left(\frac{a}{2}+\frac{\pi}{4}\right)\right)$$

7. Make a Mercator map of the contiguous United States. Be sure to include lines of latitude and lines of longitude.

Historical Note: Additional Reading

1. Daners, D., The Mercator and stereographic projections, and many in between. *American Mathematics Monthly*, MAA, March 2012, 199–210.

Chapter 10

Sequences and Their Applications

In the sequence

$$0.3, 0.33, 0.333, 0.3333, \ldots$$

the more 3s we write, the closer the numbers are to $1/3$.

When estimating a definite integral $\int_a^b f(x)\,dx$ of a continuous function, for each positive integer n we divide the interval $[a,b]$ into n equal pieces of length $\Delta x = (b-a)/n$, pick a number c_i in the i^{th} interval and form the sum $E_n = \sum_{i=1}^n f(c_i)\Delta x$. We obtain a sequence of estimates,

$$E_1, \quad E_2, \quad E_3, \quad \ldots, \quad E_n, \quad \ldots.$$

As n increases the estimates approach $\int_a^b f(x)\,dx$.

In the analysis of annual percentage yield (APY) on an account at a bank, in CIE 3 at the end of Chapter 2 we encountered the sequence

$$\left(1+\frac{1}{1}\right)^1, \quad \left(1+\frac{1}{2}\right)^2, \quad \left(1+\frac{1}{3}\right)^3, \quad \ldots, \quad \left(1+\frac{1}{n}\right)^n, \quad \ldots.$$

As n increases, the numbers approach e.

What happens to the numbers

$$S_1 = 1, \quad S_2 = 1+\frac{1}{2}, \quad S_3 = 1+\frac{1}{2}+\frac{1}{3}, \quad S_4 = 1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}, \quad \ldots, \quad S_n = \sum_{k=1}^n \frac{1}{k}, \quad \ldots$$

as we add more and more reciprocals of integers? Do the sums S_1, S_2, S_n, S_4, ... get arbitrarily large or do they approach some finite number? NOTE: When we were students, neither author guessed right.

Chapters 10, 11, and 12 concern the behavior of endless sequences of numbers, which can arise in estimating a solution of an equation or in evaluating such important functions as e^x, $\sin(x)$, and $\ln(x)$. They also offer another way to evaluate indeterminate limits.

10.1 Introduction to Sequences

A **sequence** of numbers,

$$a_1, \quad a_2, \quad a_3, \quad \ldots, \quad a_n, \quad \ldots$$

is a function that assigns to each positive integer n a number a_n. The number a_n is called the n^{th} **term** of the sequence. For example, in Section 2.2 we saw the sequence

$$\left(1+\frac{1}{1}\right)^1, \quad \left(1+\frac{1}{2}\right)^2, \quad \left(1+\frac{1}{3}\right)^3, \quad \ldots, \quad \left(1+\frac{1}{n}\right)^n, \quad \ldots$$

and later showed it to be related to the number e. Its n^{th} term is

$$a_n = \left(1 + \frac{1}{n}\right)^n.$$

For example, $a_1 = (1+1)^1 = 2$, $a_2 = (1+1/2)^2 = 9/4 = 2.25$, $a_{10} = (1+1/10)^{10} \approx 2.5937$, and $a_{100} = (1+1/100)^{100} \approx 2.7048$.

NOTATION: The notation $\{a_n\}$ is an abbreviation for the sequence $a_1, a_2, \ldots, a_n, \ldots$. Read a_1 as "a sub 1" and a_n as "a sub n." ("sub" means "subscript").

If, as n gets larger, a_n approaches a number L, then L is called the **limit** of the sequence $\{a_n\}$. When the sequence $a_1, a_2, \ldots, a_n, \ldots$ has a limit L we say $\{a_n\}$ is **convergent** and write

$$\lim_{n\to\infty} a_n = L.$$

Sequences that do not converge are called **divergent** sequences. For instance, we write

$$\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e.$$

It is often convenient and informative to display terms in a sequence in a table such as Table 10.1.1.

n	a_n
1	2.0000
2	2.2500
3	2.3704
4	2.4414
5	2.4883
10	2.5937
100	2.7048
1000	2.7169
10000	2.7181

Table 10.1.1

A sequence need not begin with the term a_1. Later, sequences of the form $a_0, a_1, a_2, \ldots$ will be considered. In them, a_0 is called the **zeroth term** but we can also have sequences $a_k, a_{k+1}, a_{k+2}, \ldots$ that begin with a_k for $k > 1$. These sequences are called a **tail** of the sequence $a_1, a_2, a_3, \ldots$. Two important features of a sequence are that the terms of a sequence are defined only for integers and the sequence never ends.

The Sequence $\{r^n\}$

The next example introduces a simple but important type of sequence called a **geometric sequence**.

EXAMPLE 1. A piece of equipment depreciates in value over the years so that at the end of a year it has 80% of the value it had at the beginning of the year. (See Table 10.1.2.) What happens to its value in the long run if its value when new is \$1?

SOLUTION Let a_n be the value of the equipment at the end of the n^{th} year. Call the initial value $a_0 = 1$. At the end of year 1 the value is $a_1 = 0.8(1)$. Similarly, $a_2 = 0.8(0.8) = 0.8^2 = 0.64$ and $a_3 = 0.8(0.8^2) = 0.8^3$. After n years its value is $a_n = 0.8^n$. This question is asking about the limit of the sequence $\{0.8^n\}$.

n	$a_n = 0.8^n$
0	$0.8^0 = 1.0000$
1	$0.8^1 = 0.8000$
2	$0.8^2 = 0.6400$
3	$0.8^3 = 0.5120$
4	$0.8^4 = 0.4096$
5	$0.8^5 = 0.3277$
10	$0.8^{10} = 0.1074$
20	$0.8^{20} = 0.0115$
40	$0.8^{40} = 0.0001$

Table 10.1.2

n	0.99^n
1	0.9900
2	0.9801
3	0.9703
4	0.9606
5	0.9510
10	0.9044
20	0.8179
100	0.3660
200	0.1340

Table 10.1.3

After 5 years, the value is just under \$0.33. In another five years it is reduced to about \$0.11, and at the end of year 20, the value is roughly \$0.01. (See Table 10.1.2.) This is evidence that $\lim_{n\to\infty} 0.8^n = 0$. ♦

Even if the piece of equipment in Example 1 retained 99% of its value each year, in the long run it would still be worth less than a dime, then less than a penny, etc. The data in Table 10.1.3 indicates that 0.99^n approaches 0 as $n \to \infty$, but more slowly than 0.8^n does.

It is plausible that if $0 \le r < 1$, then $\lim_{n\to\infty} r^n = 0$. To show this we introduce a property of the real numbers that we will use often. It concerns **monotone sequences**. A sequence is **monotone** if it is nondecreasing ($a_1 \le a_2 \le a_3 \le \cdots \le a_n \le \ldots$) or nonincreasing ($a_1 \ge a_2 \ge a_3 \ge \cdots \ge a_n \ge \ldots$).

Theorem 10.1.1: Every bounded and monotone sequence converges

Let $\{a_n\}$ be a nondecreasing sequence with the property that there is a number B such that $a_n \le B$ for all n. That is, $a_1 \le a_2 \le a_3 \le a_4 \le \cdots \le a_n \le a_{n+1} \le \ldots$ and $a_n \le B$ for all n. Then $\{a_n\}$ is convergent and a_n is convergent and its limit is less than or equal to B.

Similarly, if $\{a_n\}$ is a nonincreasing sequence and there is a number B such that $a_n \ge B$ for all n, then $\{a_n\}$ is convergent and its limit is greater than or equal to B.

Figure 10.1.1 suggests the first part of Theorem 10.1.1 is plausible. The monotonicity prevents the terms from backtracking or entering a cycle. Without the bound on the terms, the sequence could continue to approach ∞. Any sequence that is both bounded and monotone has to converge to a limit.

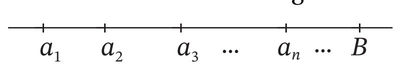


Figure 10.1.1

While the proof of Theorem 10.1.1 is deferred to a future course, the next theorem demonstrates its power.

Theorem 10.1.2

If $0 < r < 1$, then $\{r^n\}$ converges to 0.

Proof of Theorem 10.1.2

Let r be a number between 0 and 1. The sequence $r^1, r^2, r^3, \dots r^n, \dots$ is decreasing and each term is greater than 0. By Theorem 10.1.1, the sequence has a limit, L, and $L \geq 0$.

The sequence $r^2, r^3, \dots, r^{n+1}, \dots$ also approaches L. We then have

$$L = \lim_{n\to\infty} r^{n+1} = \lim_{n\to\infty} r r^n = r \lim_{n\to\infty} r^n = rL,$$

so $L = rL$, or $(1-r)L = 0$. The only way this product can be zero is for either $1-r=0$ or $L=0$. Because $0 < r < 1$, $1-r$ is not zero. Thus L has to be 0, which shows that $\lim_{n\to\infty} r^n = 0$. ●

Observation 10.1.3: *Behavior of r^n as $n \to \infty$*

The behavior of $\{r^n\}$ for other values of r is easily obtained:

1. If $r = 1$, then $r^n = 1$ for all n. So $\lim\limits_{n\to\infty} r^n = 1$.
2. If $r > 1$, then r^n gets arbitrarily large as $n \to \infty$. Hence $\{r^n\}$ is divergent.
3. If $r < -1$, then $|r|^n$ gets arbitrarily large. Thus $\lim\limits_{n\to\infty} r^n$ does not exist, so this sequence diverges.
4. If $r = -1$, then the sequence is $-1, 1, -1, 1, \dots$. which is divergent.
5. If $-1 < r < 0$, then $\lim\limits_{n\to\infty} r^n = 0$.
6. If $r = 0$, then $r^n = 0$ for all n. So $\lim\limits_{n\to\infty} r^n = 0$.

Figure 10.1.2 summarizes this information.

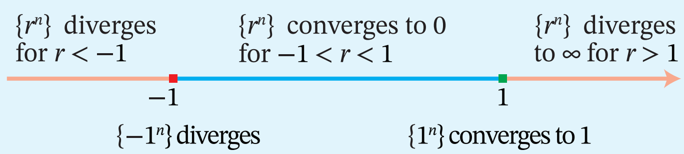


Figure 10.1.2

We prove 2. and 5.. For 2., if $r > 1$, the sequence $r, r^2, r^3, r^4, \dots, r^n, \dots$ is monotone increasing. The terms either approach a limit L or they get arbitrarily large. In the first case we would have, as before, $(1-r)L = 0$, which implies $L = 0$ (because $1-r$ is not zero). That is impossible since every term is greater than or equal to r.

To prove 5., let $-1 < r < 0$ and note that $|r^n| = |r|^n$ approaches zero as $n \to \infty$ by Theorem 10.1.2. Since the absolute value of r^n approaches 0, so does r^n.

The terms of a convergent sequence usually never equal their limit, L, but get closer to it as n increases.

If a_n becomes and remains arbitrarily large and positive as n gets larger, the sequence **diverges** and we write $\lim_{n\to\infty} a_n = \infty$. For instance, $\lim_{n\to\infty} 2^n = \infty$. Similarly, we write $\lim_{n\to\infty}(-2^n) = -\infty$. For $\lim_{n\to\infty}(-2)^n$ we can say that the sequence diverges because the values alternate between positive and negative values and $\lim_{n\to\infty} |(-2)^n| = \lim_{n\to\infty} 2^n = \infty$.

The Sequence $\left\{\dfrac{k^n}{n!}\right\}$

Example 2 introduces a type of sequence that occurs in the study of $\sin(x)$, $\cos(x)$, and e^x in Chapter 12.

EXAMPLE 2. Does the sequence defined by $a_n = \dfrac{3^n}{n!}$ converge or diverge?

SOLUTION The first terms of this sequence are in Table 10.1.4. Although a_2 is larger than a_1 and a_3 is equal to a_2, from a_4 through a_8, as Table 10.1.4 shows, the terms eventually decrease.

The numerator 3^n becomes large as $n \to \infty$, influencing a_n to grow large. The denominator $n!$ also becomes large as $n \to \infty$, influencing the quotient a_n to shrink toward 0. For $n = 1$ and $n = 2$ the first influence dominates, but then, as the table shows, the denominator $n!$ grows faster than the numerator 3^n, forcing a_n toward 0.

n	1	2	3	4	5	6	7	8
3^n	3	9	27	81	243	729	2,187	6,561
$n!$	1	2	6	24	120	720	5,040	40,320
$a_n = 3^n/n!$	3.0000	4.5000	4.500	3.3750	2.0250	1.0125	0.4339	0.1627

Table 10.1.4

To see why the denominator grows so fast that the quotient $3^n/n!$ approaches 0, consider a_{10}. This term can be expressed as the product of 10 fractions:

$$a_{10} = \frac{3^{10}}{10!} = \frac{3}{1}\cdot\frac{3}{2}\cdot\frac{3}{3}\cdot\frac{3}{4}\cdot\frac{3}{5}\cdot\frac{3}{6}\cdot\frac{3}{7}\cdot\frac{3}{8}\cdot\frac{3}{9}\cdot\frac{3}{10}.$$

The first three fractions are greater than or equal to 1, but the seven remaining fractions are all less than or equal to 3/4. Thus

$$a_{10} < \frac{3}{1}\cdot\frac{3}{2}\cdot\frac{3}{3}\cdot\left(\frac{3}{4}\right)^7.$$

Similarly,

$$a_{100} < \frac{3}{1}\cdot\frac{3}{2}\cdot\frac{3}{3}\cdot\left(\frac{3}{4}\right)^{97}.$$

More generally, for $n > 4$,

$$a_n < \frac{3}{1}\cdot\frac{3}{2}\cdot\frac{3}{3}\cdot\left(\frac{3}{4}\right)^{n-3}.$$

By Theorem 10.1.2, $\lim_{n\to\infty}(3/4)^n = 0$, from which it follows that $\lim_{n\to\infty} a_n = 0$. ◆

Reasoning like that in Example 2 shows the following fact that will be used often.

Observation 10.1.4: *Factorials Grow Faster than Powers*

For any fixed number k, $\lim\limits_{n\to\infty}\frac{k^n}{n!} = 0$. This means that the factorial grows faster than any exponential k^n.

Properties of Limits of Sequences

Limits of sequences $\{a_n\}$ behave like limits of functions $f(x)$. The most important properties are summarized in Theorem 10.1.5, without proof.

Theorem 10.1.5: Properties of Limits of Sequences

If $\lim\limits_{n\to\infty} a_n = A$ and $\lim\limits_{n\to\infty} b_n = B$ where A and B are numbers (that is, not $\pm\infty$), then

1. $\lim\limits_{n\to\infty}(a_n + b_n) = A + B$.
2. $\lim\limits_{n\to\infty}(a_n - b_n) = A - B$.
3. $\lim\limits_{n\to\infty}(a_n b_n) = AB$.
4. $\lim\limits_{n\to\infty}\left(\frac{a_n}{b_n}\right) = \frac{A}{B}$ *(provided $B \neq 0$).*
5. *If k is a constant, s $\lim\limits_{n\to\infty} k a_n = kA$. In particular, $\lim\limits_{n\to\infty}(-a_n) = -\lim\limits_{n\to\infty} a_n$.*
6. *If f is continuous on an open interval that contains A, then $\lim\limits_{n\to\infty} f(a_n) = f(A)$.*

For instance,

$$\begin{aligned}\lim_{n\to\infty}\left(\frac{3}{n}+\left(\frac{1}{2}\right)^n\right) &= 3\lim_{n\to\infty}\frac{1}{n}+\lim_{n\to\infty}\left(\frac{1}{2}\right)^n \qquad (\text{Theorem 10.1.5, Property 1})\\ &= 3\cdot 0+0\\ &= 0.\end{aligned}$$

Techniques for dealing with $\lim_{x\to\infty} f(x)$ can often help to determine the limit of a sequence because

$$\text{if } \lim_{x\to\infty} f(x)=L \quad \text{then} \quad \lim_{n\to\infty} f(n)=L.$$

EXAMPLE 3. Find $\lim_{n\to\infty}\frac{n}{2^n}$.

SOLUTION Let $f(x)=x/2^x$. By l'Hôpital's Rule (∞-over-∞ case),

$$\lim_{x\to\infty}\frac{x}{2^x}=\lim_{x\to\infty}\frac{1}{2^x\ln(2)}=0.$$

Thus

$$\lim_{n\to\infty}\frac{n}{2^n}=0.$$

◆

Warning: Converse of Theorem 10.1.5 is Not True

The converse of "if $\lim_{x\to\infty} f(x)=L$, then $\lim_{n\to\infty} f(n)=L$" is not true. For example, take $f(x)=\sin(\pi x)$. Then $\lim_{n\to\infty} f(n)=0$, but $\lim_{x\to\infty} f(x)$ does not exist.

Historical Note: k^n and Energy from the Atom

In a nuclear chain reaction, when a neutron strikes the nucleus of an atom of uranium or plutonium, neutrons split off. Call the average number of neutrons split off k. They then strike other atoms. Since each strike produces k neutrons, in the second generation there are k^2 neutrons. In the third generation there are k^3 neutrons, and so on. Each generation is born in a fraction of a second and produces energy.

If $0<k<1$, then the chain reaction dies out, since $k^n\to 0$ as $n\to\infty$. A successful chain reaction — in a nuclear reactor or an atomic bomb — requires that k be greater than 1, since then $k^n\to\infty$ as $n\to\infty$.

In September 1941, Enrico Fermi and Leo Szilard achieved $k=0.87$ with a uranium pile at Columbia University. In 1942, they obtained $k=0.918$. In the meantime, Samuel Allison at the University of Chicago, Fermi and Szilard attained $k=1.0006$, causing the neutron intensity to double every two minutes. They had achieved the first controlled, sustained, chain reaction, producing energy from the atom. Fermi let the pile run for 4.5 minutes. Had he let it go on much longer, the atomic pile, the squash court that the pile was in, the university, and part of Chicago might have disappeared.

Eugene Wigner, one of the scientists present, wrote, "We felt as, I presume, everyone feels who has done something that he knows will have very far-reaching consequences which he cannot foresee." Szilard had a different reaction: "There was a crowd there and then Fermi and I stayed there alone. I shook hands with Fermi and I said I thought this day would go down as a black day in the history of mankind."

December 2, 1942 is a historic date. Before it k was less than 1, and $\lim_{n\to\infty} k^n=0$, and after it, k was larger than 1, and $\lim_{n\to\infty} k^n=\infty$.

REFERENCE: Richard Rhodes, *The Making of the Atomic Bomb*, Simon and Schuster, New York, 1986.

Summary

We defined convergent sequences and their limits and divergent sequences, which have no limit. In particular, we have

$$\lim_{n\to\infty} r^n = 0 \quad (|r| < 1) \qquad \text{and} \qquad \lim_{n\to\infty} \frac{k^n}{n!} = 0 \quad (k \text{ any constant}).$$

A bounded monotone sequence converges, even though we may not be able to find its limit.

EXERCISES for Section 10.1

In Exercises 1 to 18 write out the first three terms of the given sequence and state whether it converges or diverges. If it converges, give its limit.

1. $\{0.999^n\}$

2. $\{1.001^n\}$

3. $\{1^n\}$

4. $\{(-0.8)^n\}$

5. $\{(-1)^n\}$

6. $\{(-1.1)^n\}$

7. $\{n!\}$

8. $\left\{\frac{10^n}{n!}\right\}$

9. $\left\{\frac{3n+5}{5n-3}\right\}$

10. $\left\{\frac{(-1)^n}{n}\right\}$

11. $\left\{\frac{\cos(n)}{n}\right\}$

12. $\left\{n\sin\left(\frac{1}{n}\right)\right\}$

13. $\left\{n\left(a^{1/n}-1\right)\right\}$

14. $\left\{\frac{n}{2^n}+\frac{3n+1}{4n+2}\right\}$

15. $\left\{\left(1+\frac{2}{n}\right)^n\right\}$

16. $\left\{\left(\frac{n-1}{n}\right)^n\right\}$

17. $\left\{\left(1+\frac{1}{n^2}\right)^n\right\}$

18. $\left\{\left(1+\frac{1}{n}\right)^{n^2}\right\}$

---

19. Assume that each year inflation decreases the value of a dollar by 2%. Let a_n be the value of a dollar after n years. (a) Find a_4. (b) Find $\lim_{n\to\infty} a_n$.

20. Let $a_n = \frac{6^n}{n!}$.

n	1	2	3	4	5	6	7	8
a_n								

(a) Fill in the empty cells in the above table:
(b) Plot the points (n, a_n). Let the n-axis be horizontal.
(c) What is the largest value of a_n? For which n?
(d) What is $\lim_{n\to\infty} a_n$?

21. What is the largest value of $\frac{(11.8)^n}{n!}$? Explain.

22. Find n such that 0.999^n is less than 0.0001
(a) by experimenting with the aid of a calculator and (b) by solving $0.999^x = 0.0001$.

23. Find the smallest n such that 1.0006^n is larger than 2
(a) by experimenting with the aid of a calculator and (b) by solving the equation $1.0006^x = 2$.

In Exercises 24 to 27 determine the limits of the sequences by identifying them as a definite integral, $\int_a^b f(x)\,dx$, for some interval $[a, b]$ and function $f(x)$. Review Section 6.2.

24. $a_n = \sum_{k=1}^{n}\left(\frac{k}{n}\right)^2\frac{1}{n}$

25. $a_n = \sum_{k=1}^{n}\left(\frac{3}{n}+\frac{k}{n^2}\right)$

26. $a_n = \sum_{k=1}^{n}\frac{1}{n}\cos\left(\frac{k\pi}{n}\right)$

27. $a_n = \sum_{k=1}^{n}\frac{n}{n^2+k^2}$

---

28. For $n \geq 1$, let $a_n = \frac{1}{n} + \frac{1}{n+1} + \cdots + \frac{1}{2n} = \sum_{k=n}^{2n} \frac{1}{k}$. For example, $a_3 = \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} = 0.95$.

(a) Compute decimal approximations to a_n for $n = 1, 2, 3, 4$, and 5.

(b) Show that $\{a_n\}$ is a monotone and bounded sequence.

(c) Show that it has a limit that is at least $\frac{1}{2}$.

29. We showed that for $-1 < r < 0$, $\lim_{n\to\infty} r^n = 0$ by considering $|r^n|$. Here is a more direct argument.

(a) Let $r = -s$, $0 < s < 1$. Show that for even n, $r^n = s^n$, and for odd n, $r^n = -(s^n)$.

(b) Show that $\{r^{2n}\}$ converges to 0.

(c) Show that $\{r^{2n-1}\}$ converges to 0.

(d) Conclude that $\lim_{n\to\infty} r^n = 0$.

30. The Binomial Theorem asserts that if n is a positive integer, then $(1+x)^n$ is equal to $1 + nx$ plus other terms that are positive if $x > 0$. Use this to show that if $r > 1$, then $\lim_{n\to\infty} r^n = \infty$.

31. Exercise 30 made use of the Binomial Theorem. This exercises shows that this was not necessary. Assume $x > 0$.

(a) Show that $(1+x)^n \geq 1 + nx$ for $n = 1$.

(b) Assume that $(1+x)^n \geq 1 + nx$ when n is 100. Show that it follows that $(1+x)^n \geq 1 + nx$ when n is 101.

(c) Explain why $(1+x)^n \geq 1 + nx$ for all positive integers n.

32. The sequence $\{a_n\}$ with $a_n = \sum_{k=n}^{2n} \frac{1}{k}$ was shown to be convergent in Exercise 28. Show that the limit of the sequences is $\ln(2)$ by expressing it as a definite integral and evaluating the integral.

33. Let $a_n = \sum_{k=2n}^{3n} \frac{1}{k}$. Does $\{a_n\}$ converge or diverge? If it converges, find its limit.

Section 3.8 developed a precise definition of $\lim_{x\to\infty} f(x) = L$ and $\lim_{x\to\infty} f(x) = \infty$. Those definitions, slightly modified, are the basis for Exercises 34 to 37.

34. Give a precise definition of $\lim_{n\to\infty} a_n = L$.

35. Use the precise definition of $\lim_{n\to\infty} a_n = L$ developed in Exercise 34 to prove that $\lim_{n\to\infty} \frac{\sin(n)}{n} = 0$.

36. Use the precise definition of $\lim_{n\to\infty} a_n = L$ developed in Exercise 34 to prove that $\lim_{n\to\infty} \frac{3}{n^2} = 0$.

37. Use the precise definition of $\lim_{n\to\infty} a_n = L$ developed in Exercise 34 to prove that $\lim_{n\to\infty} (-1)^n = 0$ is false.

---

38. Using the same approach as in the proof of Theorem 10.1.2, show that if r is greater than 1, then r^n gets arbitrarily large as n increases.

39. **Sam:** I have a better proof of Theorem 10.1.2.

Jane: You always do.

Sam: Say that r is positive and less than 1, and $r, r^2, r^3, r^4, \ldots$ approach L. Then the sequence $r^2, r^4, r^6, r^8, \ldots$ must approach L^2. Right?

Jane: Absolutely.

Sam: So L must equal L^2.

Jane: Yes, I see it.

Sam: If $L = L^2$, then L is either 0 or 1, but it can't be 1. So it's 0.

Jane: Brilliant, especially if it's correct.

Sam: I think so too.

Is Sam's argument correct? Explain.

10.2 Recursively-Defined Sequences and Fixed Points

The terms in the sequences in Section 10.1 were given by explicit formulas, $a_n = f(n)$. Often a sequence is not given explicitly. Instead, each term (after the first few) may be expressed in terms of earlier terms. For instance, the sequence of powers $a_0 = r^0 = 1$, $a_1 = r^1 = r$, $a_2 = r^2, \ldots, a_n = r^n, \ldots$ can also be described another way:

Definition: *Recursive Definition of a Geometric Sequence,* $\{r^n\}$

The first term, a_0, is 1. For $n \geq 1$, the next term is the previous term multiplied by r: $a_n = r a_{n-1}$.

In this section we will describe a technique for finding the limit of such sequences defined indirectly, if they are convergent.

Sequences Defined Recursively

A sequence given by a formula that describes the n^{th} term in terms of previous terms is said to be given **recursively**. If a_n depends only on its immediate predecessor, we would have $a_n = f(a_{n-1})$, for some function f. If a_n depends on both a_{n-1} and a_{n-2}, then there would be a function f such that $a_n = f(a_{n-1}, a_{n-2})$.

EXAMPLE 1. Let $a_0 = 1$ and $a_n = n a_{n-1}$ for $n \geq 1$. Give an explicit definition of $\{a_n\}$.

SOLUTION $a_1 = 1 \cdot a_0 = 1$, $a_2 = 2 \cdot a_1 = 2 \cdot 1$, $a_3 = 3 \cdot a_2 = 3 \cdot 2 \cdot 1$, and $a_4 = 4 \cdot a_3 = 4 \cdot 3 \cdot 2 \cdot 1$. Evidently, a_n is $n!$, n factorial, the product of all integers from 1 to n. ◆

The next example introduces the **Fibonacci sequence**, which appears often in both pure and applied mathematics,

EXAMPLE 2. (**Fibonacci sequence**) Let $b_0 = 1$, $b_1 = 1$, and $b_n = b_{n-1} + b_{n-2}$ for $n \geq 2$. Compute b_2, b_3, b_4, and b_5.

SOLUTION $b_2 = b_1 + b_0 = 1 + 1 = 2$, $b_3 = b_2 + b_1 = 2 + 1 = 3$, $b_4 = b_3 + b_2 = 3 + 2 = 5$, and $b_5 = b_4 + b_3 = 5 + 3 = 8$. The terms in the Fibonacci sequence are positive and become arbitrarily large as n increases, so it diverges to ∞. ◆

Finding the Limit of a Recursive Sequence

Assume that a sequence satisfies the relation $a_n = f(a_{n-1})$ and has a limit L. Since $a_n \to L$ as $n \to \infty$, we also have $a_{n-1} \to L$ as $n \to \infty$. If f is a continuous function, then

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} f(a_{n-1}) = f\left(\lim_{n\to\infty} a_{n-1}\right).$$

Hence, for continuous functions f

$$L = f(L), \tag{10.2.1}$$

that is, that L is a **fixed point** of f.

This reasoning is a proof of the following important fact about the limit of a recursively-defined sequence.

Theorem 10.2.1: Limit of a Recursively-Defined Sequence

If $\{a_n\}$ is a convergent recursively-defined sequence, $a_n = f(a_{n-1})$ where f is a continuous function, and if $a_n \to L$ as $n \to \infty$, then L must be a fixed point of f: $f(L) = L$.

CAUTION: *The converse is not true: not every solution of $f(L) = L$ is the limit of $\{a_n\}$.*

The next example gives a new perspective on the geometric progression $1, r, r^2, \ldots$, relating it to a continuous function.

EXAMPLE 3. Let r be in $(0,1)$ and $f(x) = rx$. If $a_1 = 1$ and $a_n = f(a_{n-1})$ for $n \geq 2$, determine if $\lim_{n\to\infty} a_n$ exists. If $\{a_n\}$ converges, find its limit.

SOLUTION Based on the recursive definition of a geometric sequence, we recognize this sequence as the geometric sequence with first term $a_1 = 1$ and recursive definition $a_n = ra_{n-1}$, $n = 2,3,4,\ldots$. The assumption that $0 < r < 1$ guarantees that the terms are monotone (decreasing) and bounded below (by 0). Therefore, by Theorem 10.1.1, this sequence converges.

Now that we know the sequence converges, and that $f(x) = rx$ is a continuous function, Theorem 10.2.1 tells us that the limit must be a fixed point of f: $L = f(L) = rL$. Since r is not 1, the only other alternative is $L = 0$. As this is the only solution of $L = f(L)$, it must be the limit of this sequence. ♦

NOTE: This is the same effectively the same argument as was used in the proof of Theorem 10.1.2 in Section 10.1. But, this reasoning based on Theorem 10.2.1 is more general.

EXAMPLE 4. Define c_n to be the ratio of successive terms in the Fibonacci sequence $\{b_n\}$: $c_n = b_n/b_{n-1}$ for all $n \geq 2$. Assuming $\{c_n\}$ converges, find its limit.

SOLUTION $c_2 = b_2/b_1 = 1/1 = 1$. For $n \geq 3$ the definition of the Fibonacci sequence can be used to obtain a recursive formula relating c_n to c_{n-1}:

$$c_n = \frac{b_n}{b_{n-1}} = \frac{b_{n-1} + b_{n-2}}{b_{n-1}} = 1 + \frac{b_{n-2}}{b_{n-1}} = 1 + \frac{1}{c_{n-1}},$$

so

$$c_n = 1 + \frac{1}{c_{n-1}} \qquad \text{for all } n \geq 3. \tag{10.2.2}$$

Thus, $c_n = f(c_{n-1})$ where $f(x) = 1 + 1/x$.

Table 10.2.1, showing the first few terms of this sequence, suggests that it s bounded ($0 < c_n \leq 2$) and probably converges. Convergence is not guaranteed because $\{c_n\}$ is neither increasing nor decreasing, so Theorem 10.2.1 does not apply.

But, since the statement of the problem tells us to assume the sequence converges and the fact that $f(x) = 1 + 1/x$ is continuous, Theorem 10.2.1 assures us that the limit must be a fixed point of $f(x) = 1 + 1/x$.

From

$$L = 1 + \frac{1}{L}.$$

it is quickly determined that Therefore,

$$L^2 - L - 1 = 0.$$

By the quadratic formula, the solutions to $L^2 - L - 1 = 0$ are

$$L = \frac{1}{2}\left(1 + \sqrt{5}\right) \qquad \text{and} \qquad L = \frac{1}{2}\left(1 - \sqrt{5}\right).$$

Since every term in the sequence is positive, the limit cannot be negative. The only positive limit is

$$L = \frac{1}{2}\left(1 + \sqrt{5}\right) \approx 1.61803.$$

Thus, $\lim_{n\to\infty} c_n = (1 + \sqrt{5})/2 \approx 1.61803$. ♦

n	b_n	c_n
1	1	–
2	1	1.000000
3	2	2.000000
4	3	1.500000
5	5	1.666667
6	8	1.600000
7	13	1.625000
8	21	1.615385
9	34	1.619048
10	55	1.617647
15	610	1.618037
25	75025	1.618034

Table 10.2.1

Definition: ***Golden Ratio***

The number $\varphi = (1+\sqrt{5})/2$ is known as the **Golden Ratio**. It appears in many situations, including art, biology, and architecture.

Historical Note: Origins of the Fibonacci Sequence

While the first description of the Fibonacci sequence can be traced back to descriptions of possible patterns of Sanskrit poetry as early as 200 BC, the sequence is named after Leonardo of Pisa, later known as Fibonacci, who introduced the Fibonacci sequence to the Western world in a problem from Chapter XII his book *Liber abaci*. This book appeared in 1202 and was revised in 1228.

> A man put a pair of rabbits in a place surrounded by a wall. How many pairs of rabbits can be produced from that pair in a year if every month each pair produces a new pair which from the second month on can produce another pair?

There are many other stories (and myths) surrounding the Golden Ratio.

REFERENCE: S. Stein, *Strength in Numbers*, John Wiley and Sons, New York, 1996 (p. 39).

A Famous Recursion

The recursion $p_{n+1} = kp_n$, where k is a constant greater than 1, describes a population of bacteria or elephants growing at a rate proportional to the amount present. If the initial population is p_1, then $p_2 = kp_1$, $p_3 = k^2p_1$, $p_4 = k^3p_1, \ldots$ and the population increases exponentially without bound. But a population cannot do that. Assume that it approaches a limiting population, which we will say is 1. As it approaches 1, the struggle to find food slows its growth. We assume that $\{p_n\}$ satisfies the **logistic equation**:

$$p_{n+1} = kp_n(1 - p_n). \tag{10.2.3}$$

The behavior of $\{p_n\}$ is surprising. For instance, if k is near 3.5699456 its behavior changes a great deal even when k is changed only a little.

Observation 10.2.2: *What is Chaos?*

A defining characteristic of **chaos** is that small changes in a parameter produces results with very different qualitative features.

EXAMPLE 5. Find the limit of the sequence given by $p_{n+1} = kp_n(1 - p_n)$ for $0 \le k \le 1$ if $0 \le p_0 < 1$.

SOLUTION For $p_0 = 0$ or 1, $p_n = 0$ for all $n \ge 1$. For $0 < p_0 < 1$, $p_1 = kp_0(1 - p_0)$ is at most $p_0(1 - p_0)$, which is less than p_0. Similarly, p_2 is less than p_1, and, in general we have $p_{n+1} < p_n$. The sequence $\{p_n\}$ decreases but stays above 0. Because the sequence is monotone and bounded, it has a limit L. A key observation is that $L \ge 0$.

Since this sequence converges, and $f(x) = kx(1 - x)$ is continuous, Theorem 11.1.3 tells us that the limit must be a fixed point of f. Solving $L = f(L) = kL(1 - L)$ for L leads us to conclude that either $L = 0$ or $L = 1 - 1/k$. But $1 - 1/k$ is either negative (if $0 < k < 1$) or 0 (if $k = 1$). So $L = 0$ is the only possible value for the limit of this sequence when $0 \le k \le 1$. ◆

Summary

A recursive sequence is one whose n^{th} term is given in terms of previous terms. If a_n depends only on its immediate predecessor, then $a_n = f(a_{n-1})$. The sequence generated by $f(x)$ starting with $a_1 = a$ has $a_2 = f(a)$, $a_3 = f(f(a))$, $a_4 = f(f(f(a)))$, ... When a recursive sequence, $a_n = f(a_{n-1})$ is known to converge, and f is a continuous function, then the limit of the sequence must be a solution of $L = f(L)$, that is, a fixed point of f.

EXERCISES for Section 10.2

In Exercises 1 to 6 give an explicit formula for a_n as a function of n.

1. $a_0 = 1$, $a_n = -a_{n-1}$ for $n \ge 1$

2. $a_0 = 2$, $a_n = 3 + a_{n-1}$ for $n \ge 1$

3. $a_0 = 3$, $a_n = \frac{1}{n} a_{n-1}$ for $n \ge 1$

4. $a_0 = 5$, $a_n = -\frac{1}{2} a_{n-1}$ for $n \ge 1$

5. $a_1 = 1$, $a_n = a_{n-1} + \frac{1}{n}$ for $n \ge 2$

6. $a_1 = 1$, $a_n = -a_{n-1} + \frac{(-1)^n}{n}$ for $n \ge 2$

In Exercises 7 to 12 describe a_n in terms of a_{n-1} and an initial term a_0.

7. $a_n = 3^n$, $n = 0, 1, 2, \ldots$

8. $a_n = \frac{5^n}{n!}$, $n = 0, 1, 2, \ldots$

9. $a_n = 3(n!)$, $n = 0, 1, 2, \ldots$

10. $a_n = 2n + 5$, $n = 0, 1, 2, \ldots$

11. $a_n = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2}$, $n = 1, 2, 3, \ldots$

12. $a_n = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{2^{n-1}}$, $n = 0, 1, 2, \ldots$

13. Define $\{b_n\}$ by $b_0 = 2$ and $b_n = \dfrac{1}{b_{n-1}}$ for $n \geq 1$.
- (a) Find $b_1, b_2, \ldots, b_5$.
- (b) Show that if $\{b_n\}$ converges, its limit is 1 or -1.
- (c) Does $\{b_n\}$ converge?
- (d) For what choices of b_0 does $\{b_n\}$ converge to 1?
- (e) For what choices of b_0 does $\{b_n\}$ converge to -1?
- (f) For what choices of b_0 does $\{b_n\}$ diverge?

14. Suppose $p_{n+1} = 2p_n(1-p_n)$.
- (a) Choose p_0 between 0 and $\frac{1}{2}$. Find values of p_n until you can guess whether $\{p_n\}$ converges or diverges.
- (b) Repeat (a) for another value of p_0 between 0 and $\frac{1}{2}$.
- (c) Repeat (a) with p_0 between $\frac{1}{2}$ and 1.
- (d) Repeat (a) for another value of p_0 between $\frac{1}{2}$ and 1.
- (e) What happens to the sequence $\{p_n\}$ if p_0 is 0 or 1?
- (f) What happens to the sequence $\{p_n\}$ if p_0 is $\frac{1}{2}$?
- (g) For what values of p_0 does $\{p_n\}$ converge? For them, to what limit?

In Exercises 15 to 18, find all values of x for which the given sequence converges.

15. $\left\{\dfrac{x^n}{n!}\right\}$ **16.** $\left\{\dfrac{x^n}{2^n}\right\}$ **17.** $\left\{\dfrac{x^n}{n^2}\right\}$ **18.** $\left\{\dfrac{x^n}{\sqrt{n}}\right\}$

---

19. Let $a_{n+2} = a_n + 2a_{n+1}$ with $a_0 = 1 = a_1$ and let $c_n = \dfrac{a_n}{a_{n-1}}$. Examine $\{c_n\}$ numerically, deciding whether it converges and, if so, what its limit might be.

20. Consider the recursion (10.2.3) with $0 < k \leq 4$. Show that if p_0 is in the interval $[0,1]$, then so is p_n for all $n \geq 0$.

21. Let $a_{n+2} = \frac{1}{4}(a_n + 3a_{n+1})$, with $a_0 = 0$ and $a_1 = 1$.
- (a) Compute enough terms of $\{a_n\}$ to guess the limit, L.
- (b) When you take limits of both sides of the recursion equation, what equation do you get for L?

22. Suppose $a_{n+2} = \dfrac{1 + a_{n+1}}{a_n}$.
- (a) Starting with $a_1 = 1$ and $a_2 = 2$, compute a_3, a_4, a_5, a_6, a_7, and a_8.
- (b) Repeat (a) with $a_1 = 3$ and $a_2 = 3$.
- (c) Repeat (a) with a_1 and a_2 of your choice.
- (d) Explain what is going on.

23. Let k and p be positive numbers and define the sequence $\{f_n\}$ as follows: given f_1, $f_{n+1} = k(f_n)^p$ for $n \geq 1$. (a) Assuming $\{f_k\}$ converges, find its limit. (b) Explain how to choose k so that it converges to 2.

24. Show that if $0 \leq k \leq 4$, $0 \leq p_0 \leq 1$, and $p_{n+1} = kp_n(1-p_n)$, then $0 \leq p_n \leq 1$ for all $n \geq 0$.

25.
- (a) Investigate the convergence or divergence of the logistic sequence $\{p_n\}$ with $k = 2$ for various values of p_0.
- (b) Make a conjecture based on (*a*).
- (c) Let $q_n = \frac{1}{2} - p_n$. Show that $q_{n+1} = -2q_n^2$.
- (d) Use (c) to discuss your conjecture in (b).

26. Define the sequence $\{a_n\}$ where $a_{n+1} = a_n - a_{n-1}$ for $n \geq 2$. For each pair of values for a_0 and a_1, find (i) the shortest interval containing all a_n and (ii) the limit (if it exists) in the following cases: (a) $a_0 = 3$, $a_1 = 4$, (b) $a_0 = 1$, $a_1 = 0$, and (c) the general case, $a_0 = a$, $a_1 = b$.

27. Figure 10.2.1 shows the graph of a decreasing continuous function f such that $f(0) = 1$ and $f(1) = 0$.

(a) Show that f has exactly one fixed point in the interval $[0,1]$. That is, show that there is one number a with $0 \leq a \leq 1$ that satisfies $f(a) = a$. (Draw the line $y = x$ on the graph of $y = f(x)$.)

(b) If $0 < x < a$, in what interval does $f(x)$ lie?

(c) If $a < x < 1$, in what interval does $f(x)$ lie?

(d) Use the graphs of $y = f(x)$ and $y = x$ to find all values of x for which $f(f(x)) > x$ and all values of x for which $f(f(x)) < x$.

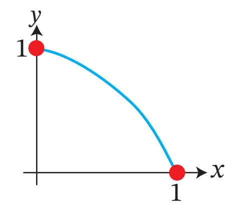


Figure 10.2.1

28. Let f be a decreasing function such that $f(0) = 1$ and $f(1) = 0$ and the graph of f is symmetric with respect to the line $y = x$. Examine the sequence x, $f(x)$, $f(f(x))$, ... for x in $[0,1]$. What can be said about its convergence?

Exercises 29 to 32 illustrate some of the characteristics that make the logistic recursion $p_{n+1} = kp_n(1 - p_n)$ so interesting. In each, create two sequences corresponding to two values of k in the range and with different values for the initial value, p_0.

29. $1 < k < 3$ **30.** $3 < k < 3.4$ **31.** $3.4 < k < 3.5$ **32.** $3.6 < k < 4$

---

33. Let k, c_1, and c_2 be positive numbers. Define the sequence $\{c_n\}$ as follows: given c_1 and c_2, $c_n = \dfrac{1 + kc_{n-1}}{c_{n-2}}$ for $n \geq 3$. Assuming it converges, find possible limits.

34. Determine the convergence or divergence of the sequence $\{a_n\}$ defined by $a_n = f(a_{n-1})$ with $f(x) = 1 - x^2$ for various inputs in $[0,1]$. Does f have a fixed point?

35. Let $f(x) = 1 - x$, $g(x) = 1 - 1.1x$, and $h(x) = 1 - 0.9x$. Let $a_0 = 0.4$. Examine what happens to the sequences determined by (a) $a_n = f(a_{n-1})$, (b) $a_n = g(b_{n-1})$, and (c) $a_n = h(a_{n-1})$.

36. Assume that f is a decreasing function for x in $[0,1]$, $f(1) = 0$, and $-1 < f'(x) < 0$.

(a) What can be said about $f(0)$?

(b) Show that f has a unique fixed point.

(c) Assume a is the fixed point of f, that is, $f(a) = a$. Show that if $1 \geq x > a$, then $f(x) < a$ and if $0 \leq x < a$, then $f(x) > a$.

(d) Let $g(x) = f(f(x))$. Determine the convergence or divergence of the sequence x, $g(x)$, $g(g(x))$, ... for x in $[0,1]$. Show that this sequence is monotone.

(e) Show that for all x in $[0,1]$ the sequence x, $f(x)$, $f(f(x))$, ..., approaches a.

37. Figure 10.2.2 shows the graph of a function for which $f(0) = 0$, $f(1) = 0$, $f''(x) \leq 0$, and $0 \leq f(x) \leq 1$.

(a) Show that f has at least one fixed point on $[0,1]$.

(b) Show that if $f'(0) < 1$, then f has only one fixed point on $[0,1]$.

(c) Show that if $f'(0) \geq 1$, it has exactly two fixed points on $[0,1]$.

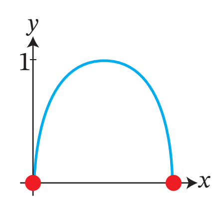


Figure 10.2.2

38. Let d_0 be a positive number and define the sequence $\{d_n\}$ by $d_n = 2d_{n-1}^2$ for $n \geq 1$. Describe the behavior of the sequence for (a) $d_0 = 0.5$, (b) $d_0 = 0.501$, and (c) $d_0 = 0.499$. If the sequence converges, give its limit. (This illustrates how sensitive a sequence can be to initial data. It is one reason predicting the future is difficult (or impossible).

39. Let $a_{n+1} = \dfrac{a_n}{1+a_n}$ for $n \geq 1$ and let a_0 be a number.

(a) If $\{a_n\}$ converges, what is its limit?

(b) Show that for $a_0 \geq 0$ or $a_0 < -1$ the sequence converges.

(c) To analyze $\{a_n\}$ for $-1 \leq a_0 < 0$ introduce the sequence $\{b_n\}$ in which $b_n = \dfrac{1}{a_n}$. Find a recursion for $\{b_n\}$.

(d) Using the recursion for $\{b_n\}$ found in (c), describe the convergence or divergence of $\{b_n\}$.

(e) WIth the aid of (d), describe the behavior of $\{a_n\}$ for $-1 \leq a_0 < 0$.

40. Let u_n be the number of ways of tiling a $2 \times n$ rectangle with 1×2 dominoes.
(a) Find u_1, u_2, and u_3. (b) Find a recursive definition for u_n.

Exercises 41 to 43 are related.

41. A path that is $1 \times n$ is to be tiled with 1×1 tiles and 1×2 tiles. Let a_n be the number of ways this can be done. (a) Obtain a recursive formula for a_n. (b) Use your answer in (a) to find a_{10}.

42. Repeat Exercise 41 with 1×1 and 1×3 tiles.

43. Repeat Exercise 41 with 1×2 and 1×3 tiles.

10.3 Bisection Method for Solving $f(x) = 0$

One way to estimate a solution (or root) of an equation $f(x) = 0$ is to zoom in on it with a graphing calculator. However, precision is limited by the resolution of the display. This section and the next describe techniques for estimating a root to as many decimal places as you need. The technique in this section is based on the fact that a continuous function that is positive at one input and negative at another has a root between them.

Bisection Method for Solving $f(x) = 0$

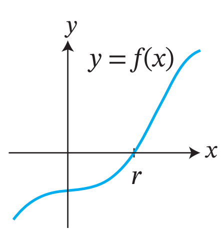


Figure 10.3.1

Let $f(x)$ be a function. A solution or **root** of the equation $f(x) = 0$ is a number r such that $f(r) = 0$. The graph of $y = f(x)$ passes through $(r, 0)$, as shown in Figure 10.3.1.

Let $f(x)$ be a continuous function defined on $[a_0, b_0]$, with $a_0 < b_0$. Assume that $f(a_0)$ and $f(b_0)$ have opposite signs. By the intermediate value theorem, $f(x)$ has at least one root in $[a_0, b_0]$.

Not knowing where in $[a_0, b_0]$ a root lies, evaluate f at the midpoint, $m_0 = (a_0 + b_0)/2$. If $f(m_0) = 0$, we have found a root and the search is over. Otherwise, the sign of $f(m_0)$ is opposite the sign of one of $f(a_0)$ and $f(b_0)$.

If $f(a_0)$ and $f(m_0)$ have opposite signs, then a root must be in the interval $[a_0, m_0]$, which is half the width of $[a_0, b_0]$. On the other hand, if $f(m_0)$ and $f(b_0)$ have opposite signs, a root lies in $[m_0, b_0]$, again half the width of $[a_0, b_0]$.

In either case we have located a root in an interval half the width of $[a_0, b_0]$. Call the shorter interval $[a_1, b_1]$. Figure 10.3.2 shows the two possibilities for $[a_1, b_1]$ when $f(a_0) > 0$ and $f(b_0) < 0$.

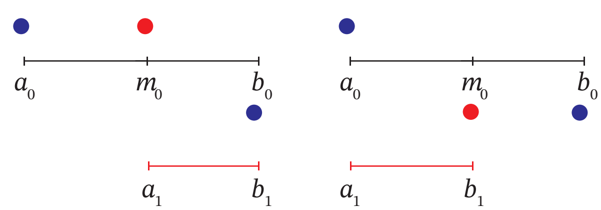


Figure 10.3.2

Repeat the process, starting at $[a_1, b_1]$. In this way we obtain a sequence of shorter and shorter intervals $[a_0, b_0]$, $[a_1, b_1]$, $[a_2, b_2]$, ..., each half as long as its predecessor. Thus the length of $[a_n, b_n]$ is $(b_0 - a_0)/2^n$.

TERMINOLOGY: The bisection method gets its name from the fact that at each step the interval is cut in half — bisected. It is a recursive algorithm.

An Illustration of the Bisection Method

When x is large and positive $f(x) = x + \sin(x) - 2$ is positive. When x is large and negative, $f(x)$ is negative. Therefore $f(x) = 0$ has at least one solution. The derivative of $f(x)$ is $1 + \cos(x)$, which is positive except at odd multiples of π, when it is zero. Thus, $f(x)$ is an increasing function, which implies that it cannot have more than one root. Let r be the unique root of $x + \sin(x) - 2 = 0$. We begin the search for r by finding an interval which contains the root.

Since $f(0) = -2$, the root must be positive. Using $\sin(x) \geq -1$ we know $f(x) = x + \sin(x) - 2 \geq x - 1 - 2 = x - 3$; thus $f(4)$ is positive. Let $a_0 = 0$ and $b_0 = 4$. The root will be found in the interval $[0, 4]$.

n	a_n	b_n	m_n	y_n	$b_n - a_n$
0	0.000000	4.000000	2.000000	0.909297	4.000000
1	0.000000	2.000000	1.000000	−0.158529	2.000000
2	1.000000	2.000000	1.500000	0.497495	1.000000
3	1.000000	1.500000	1.250000	0.198985	0.500000
4	1.000000	1.250000	1.125000	0.027268	0.250000
5	1.000000	1.125000	1.062500	−0.063925	0.125000
6	1.062500	1.125000	1.093750	−0.017895	0.062500
7	1.093750	1.125000	1.109375	0.004796	0.031250
8	1.093750	1.109375	1.101562	−0.006522	0.015625
9	1.101562	1.109375	1.105469	−0.000857	0.007812
10	1.105469	1.109375	1.107422	0.001971	0.003906
11	1.105469	1.107422	1.106445	0.000558	0.001953
12	1.105469	1.106445	1.105957	−0.000149	0.000977
13	1.105957	1.106445	1.106201	0.000204	0.000488

Table 10.3.1

The middle of the interval $[a_0, b_0] = [0, 4]$ is $m_0 = (a_0 + b_0)/2 = 2$. Evaluate $y_0 = f(m_0) = f(2) \approx 0.909297$. Because $y_0 > 0$ we now know the root is in the interval $[a_1, b_1] = [0, 2]$.

The middle of the new interval $[a_1, b_1] = [0, 2]$ is $m_1 = (a_1 + b_1)/2 = 1$. Then $y_1 = f(m_1) = f(1) \approx -0.15829$. Because $y_1 < 0$ the root is in the interval $[a_2, b_2] = [1, 2]$.

The third iteration of this process yields $m_2 = 1.5$ and $y_2 = f(1.5) \approx 0.497495$. Then $[a_3, b_3] = [1, 1.5]$. The next ten iterations are shown in Table 10.3.1.

After thirteen iterations the root is known to exist on the interval $[a_{13}, b_{13}] = [1.105957, 1.106445]$. Its midpoint, $m_{13} = 1.106201$, differs from r by at most half its width, that is, by at most 0.000244.

NOTE: Every number in $[1.105957, 1.106445]$ rounds to 1.106 (to three decimal places).

While the iterations could be continued to define sequences $\{a_n\}$ and $\{b_n\}$, in practice one stops when the length of the interval containing r is short enough to provide the needed numerical accuracy.

Observation 10.3.1: *Convergence of the Bisection Method*

The number of iterations of the bisection method needed to determine a solution of $f(x) = 0$ to a desired accuracy depends primarily on the length of the initial interval, $[a_0, b_0]$. (See Exercise 15.)

EXAMPLE 1. Use the bisection method to estimate the square root of 3 to three decimal places.

SOLUTION The square root of 3 is the positive number whose square is 3. Therefore $x^2 = 3$ or $x^2 - 3 = 0$. We are looking for the positive root of $f(x) = x^2 - 3$.

We know $\sqrt{3}$ is between 1 and 2 and f is continuous. This suggests using the bisection method with initial interval $[1,2]$. The first eleven iterations of the bisection method are displayed in Table 10.3.2. After seven iterations the approximation $\sqrt{3} \approx m_7 = 1.730469$ is accurate to two decimal places: $\sqrt{3} \approx 1.73$. After another four iterations the approximation is accurate to three decimal places: $\sqrt{3} \approx 1.732$.

n	a_n	b_n	m_n	y_n	$b_n - a_n$
0	1.000000	2.000000	1.500000	−0.750000	1.000000
1	1.500000	2.000000	1.750000	0.062500	0.500000
2	1.500000	1.750000	1.625000	−0.359375	0.250000
3	1.625000	1.750000	1.687500	−0.152344	0.125000
4	1.687500	1.750000	1.718750	−0.045898	0.062500
5	1.718750	1.750000	1.734375	0.008057	0.031250
6	1.718750	1.734375	1.726562	−0.018982	0.015625
7	1.726562	1.734375	1.730469	−0.005478	0.007812
8	1.730469	1.734375	1.732422	0.001286	0.003906
9	1.730469	1.732422	1.731445	−0.002097	0.001953
10	1.731445	1.732422	1.731934	−0.000406	0.000977
11	1.731934	1.732422	1.732178	0.000440	0.000488

Table 10.3.2

◆

Why the Bisection Method Works

The bisection method applied to $f(x)$ produces two sequences, $a_0 \le a_1 \le a_2 \le \cdots$ and $b_0 \ge b_1 \ge b_2 \ge \cdots$. If no a_n or b_n is a root of f, the sequences do not end. The sequence of left endpoints, $\{a_n\}$, is monotone increasing and the sequence of right endpoints is monotone decreasing. Because a_n is less than or equal to b_0, $\{a_n\}$ is bounded. Thus $\{a_n\}$, being bounded and monotone, has a limit, $A \le b_0$. Similarly, $\{b_n\}$ has a limit, $B \ge a_0$.

The length of the interval $[a_n, b_n]$ is $b_n - a_n = (b_0 - a_)/2^n$, so $\{b_n - a_n\}$ is a geometric sequence with ratio $1/2$. Since the ratio is between 0 and 1, $\lim_{n\to\infty}(b_n - a_n) = 0$, and we have

$$0 = \lim_{n\to\infty}(b_n - a_n) = \lim_{n\to\infty} b_n - \lim_{n\to\infty} a_n = B - A.$$

Consequently, $A = B$.

But, we still have a question to answer: Why is A a root of f?

Because f is continuous, the sequence

$$f(a_0), f(b_0), f(a_1), f(b_1), f(a_2), f(b_2), \cdots f(a_n), f(b_n), \cdots. \tag{10.3.1}$$

has a limit, $f(A)$. That one of $f(a_n)$ and $f(b_n)$ is positive implies that $f(A)$ cannot be negative. Similarly, because one of each pair of entries in (10.3.1) is negative, $f(A)$ cannot be positive. Thus, $f(A) = 0$ and A is a root of f.

Summary

In the bisection method for finding a root of f, we find two inputs a_0 and b_0 for which $f(a_0)$ and $f(b_0)$ have opposite signs. Then we evaluate f at the midpoint m_0. The function f will have opposite signs at the endpoints of one of the intervals $[a_0, m_0]$ and $[m_0, b_0]$. Call the new interval $[a_1, b_1]$, then repeat the process. Continue until the interval is short enough to give an adequate estimate of the root.

EXERCISES for Section 10.3

Exercises 8 to 27 in this section are the same as Exercises 11 to 30 in Section 10.4, except that different methods are used estimate each answer (Section 10.3 uses bisection method, Section 10.4 uses Newton's method).

In Exercises 1 and 2, use the bisection method to find a_1 and b_1.

1. $a_0 = 2$, $b_0 = 6$, $f(2) = 0.3$, $f(4) = 1.5$, $f(6) = -1.2$

2. $a_0 = 1$, $b_0 = 3$, $f(1) = -4$, $f(2) = -1.5$, $f(3) = 1$

---

3. Use the bisection method to approximate $\sqrt{2}$. Let $a_0 = 1$, $b_0 = 2$, and $f(x) = x^2 - 2$. Fill in the empty cells in the following table:

n	0	1	2	3	4	5
a_n						
b_n						

4. Estimate $\sqrt{5}$ using the bisection method.

(a) Use $f(x) = x^2 - 5$ and start with $a_0 = 2$, and $b_0 = 3$. Continue until the interval $[a_n, b_n]$ is shorter than 0.01, that is, $b_n - a_n < 0.01$.

(b) How many more steps are needed to reduce the interval by another factor of 10, that is, to $b_n - a_n < 0.001$? (This can be answered without computing every a_n and b_n.)

5. Estimate $\sqrt[3]{2}$ by the bisection method.

(a) Use $f(x) = x^3 - 2$ and start with $a_0 = 1$, and $b_0 = 2$. Continue until the interval $[a_n, b_n]$ is shorter than 0.01, that is, $b_n - a_n < 0.01$.

(b) How many more steps are needed to reduce the interval by another factor of 10, that is, to $b_n - a_n < 0.001$?

In Exercises 6 to 9 estimate the numbers to the indicated number of decimal places.

6. $\sqrt{15}$ to 3 decimal places **7.** $\sqrt{19}$ to 2 decimal places **8.** $\sqrt[3]{7}$ to 4 decimal places **9.** $\sqrt[3]{25}$ to 3 decimal places

---

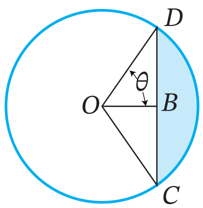


Figure 10.3.3

In Exercises 10 to 12 use the bisection method to estimate an angle θ (see Figure 10.3.3) to two decimal places. Angles are in radians. For each problem, show that there is only one answer if $0 < \theta < \frac{\pi}{2}$.

10. Find θ for which $|\widehat{CD}| = 2|\overline{OB}|$.

11. Find θ for which the areas of triangle $\triangle OCD$ and the shaded region are equal.

12. Find θ for which $|\widehat{CD}| = \frac{3}{4}|\overline{OB}|$.

Compare with Exercises 13 to 15 in Section 10.4.

Exercises 13 to 27 are the same as Exercises 16 to 30 in Section 10.4, except that the bisection method is used instead of Newton's method.

13. Let $f(x) = x^5 + x - 1$.
(a) Show that there is a root of $f(x)$ in the interval $[0, 1]$.
(b) Apply five steps of the bisection method with $a_0 = 0$ and $b_0 = 1$.
(c) Why is the root unique?

14. Let $f(x) = x^4 + x - 19$.
(a) Show that $f(2) < 0 < f(3)$ and that there is at least one root of $f(x)$ between 2 and 3? What property of f assures that there is exactly one root between 2 and 3?
(b) Using the bisection method with $[a_0, b_0] = [2, 3]$, find an interval of length no more than 0.01 containing the root.
(c) The second real root of $f(x)$ is negative. Find an interval of length 1 containing it.
(d) Repeat (b) using the interval found in (c) as the initial interval.

15. In estimating $\sqrt{3}$ with the bisection method, Sam imprudently chooses the initial interval to be $[0, 10]$.
(a) How many steps of the bisection method will Sam execute before he has an interval shorter than 0.005?
(b) Jane started with $[1, 2]$. How many steps of the bisection method will she have to execute before she has an interval shorter than 0.0005?

16. Let $f(x) = 2x^3 - x^2 - 2$.
(a) Show that there is exactly one root of $f(x) = 0$ in $[1, 2]$.
(b) Using $[a_0, b_0] = [1, 2]$ as a first interval, apply two steps of the bisection method.

17. (a) Graph $y = x$ and $y = \cos(x)$ on the same axes.
(b) Using the graph in (a), find an interval of length no more than 0.25 that contains the positive solution of the equation $x = \cos(x)$. Is there a negative solution?
(c) Using the estimate in (b) as $[a_0, b_0]$, apply the bisection method until the interval is shorter than 0.001.

18. (a) Graph $y = \cos(x)$ and $y = 2\sin(x)$ on the same axes.
(b) Without using the graph in (a), explain why the graphs intersect exactly once in $\left[0, \frac{\pi}{2}\right]$.
(c) Using $[a_0, b_0] = \left[0, \frac{\pi}{2}\right]$, apply the bisection method until the length of the interval is no more than 0.001.

19. (a) Graph $y = x\sin(x)$ for x in $[0, \pi]$.
(b) Using the first and second derivatives, show that y has a unique relative maximum in $[0, \pi]$.
(c) Show that it occurs when $x\cos(x) + \sin(x) = 0$.
(d) Use the bisection method, with $[a_0, b_0] = \left[0, \frac{\pi}{2}\right]$, to find an interval with length no more than 0.01 that contains a solution of $x\cos(x) + \sin(x) = 0$.

20. (a) Graph $y = x\cos(x)$ for x in $[0, \pi]$.
(b) Using the first and second derivatives, show that y has a unique relative maximum in $\left[0, \frac{\pi}{2}\right]$.
(c) Show that it occurs when $\cos x - x\sin(x) = 0$.
(d) Use the bisection method starting with $\left[0, \frac{\pi}{2}\right]$ to find an interval with length no more than 0.01 that contains a solution of $\cos(x) - x\sin(x) = 0$.

21. Use the bisection method to estimate the maximum value of $y = 2\sin(x) - x^2$ over the interval $\left[0, \frac{\pi}{2}\right]$.

22. Use the bisection method to estimate the maximum value of $y = x^3 + \cos(x)$ over the interval $\left[0, \frac{\pi}{2}\right]$.

23. The equation $x\tan(x) = 1$ occurs in the theory of vibrations.

(a) How many roots does it have in $\left[0, \frac{\pi}{2}\right]$?

(b) Use the bisection method to estimate each root to two decimal places.

24. The function $f(x)$ is defined to be $f(x) = \dfrac{\sin(x)}{x}$ for $x \neq 0$ and $f(0) = 1$.

(a) From a graph of $y = f(x)$, how many critical numbers does $f(x)$ have?

(b) Show that a critical number of $f(x)$ satisfies the equation $\tan x = x$.

(c) Show that $f(x)$ is an even function. Thus we need consider only positive x. (That is, if x is a critical number of $f(x)$, then so is $-x$.)

(d) Graph the functions $\tan(x)$ and x on the same axes. How often do they cross for x in $\left[0, \frac{\pi}{2}\right]$? for x in $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$? for x in $\left[\frac{3\pi}{2}, \frac{5\pi}{2}\right]$?

(e) Show that $\tan(x) - x$ is an increasing function for x in $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$. What does that tell us about the number of solutions of the equation $\tan(x) = x$ for x in $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$?

(f) Use the bisection method with $[a_0, b_0] = \left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$ to estimate its critical number in $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$ to at least two decimal places.

(g) Repeat (e) and (f) for x in $\left[\frac{3\pi}{2}, \frac{5\pi}{2}\right]$?

(h) Why is the bisection method not needed to find the critical number of $f(x)$ on the interval $\left[0, \frac{\pi}{2}\right]$?

25. How many solutions of $2x + \sin(x) = 2$ are there? Use the bisection method to evaluate them to two decimal places. Explain the steps in your solution in complete sentences.

26. How many solutions does $\sin(x) = x$ have? Explain how the bisection method could be used to estimate them.

27. Explain how to use the bisection method to estimate $\sqrt[5]{a}$.

---

28. We show that the error in the bisection method diminishes slowly. Let $[a_0, b_0]$ be the initial interval containing the root r and let $[a_1, b_1]$ be the next estimate, obtained by the bisection method.

(a) Show that $b_1 - a_1 = \frac{1}{2}(b_0 - a_0)$.

(b) Let $[a_2, b_2]$ be the next interval. Show that $b_2 - a_2 = \frac{1}{2}(b_1 - a_1) = \frac{1}{4}(b_0 - a_0)$.

(c) Explain why, in general, $b_n - a_n = \frac{1}{2}(b_{n-1} - a_{n-1}) = \frac{1}{2^n}(b_0 - a_0)$.

(d) How many steps of the bisection method are needed to obtain an interval no longer than L $(L > 0)$ containing the given root?

29. Use the bisection method to approximate the local extrema of $g(x) = 2x - (x+1)e^{-x}$ to three decimal places. What equation has to be solved? How do you know you found all extrema? See Example 3 in Section 10.4.

30. **Sam:** I have a better way than the bisection method.
Jane: What do you propose?
Sam: I trisect the interval into three equal intervals using two points.

JANE: What's so good about that?
SAM: I cut the error by a factor of 3 each step.
JANE: But you have to compute two points and evaluate the function there. That's four calculations instead of two.
SAM: But my method cuts the error so fast, it's still better, so the gain outweighs the cost.

To determine if Sam is right, answer the following question: Assume the initial interval is $[0,1]$ and estimate the cost (number of function evaluations) to reduce the length of the interval containing the root to the small number 3^{-20}. How does this compare with the cost of using the bisection method to solve this problem?

31. **SAM:** I have a better way than the bisection method.
JANE: What is it?
SAM: I break the interval into four equal intervals by three points.
JANE: Then?
SAM: I find on which of the four intervals the root must lie. I do two of the bisection steps in one step. So it must be more efficient.
JANE: That all depends. I'll think about it.

Think about it and offer your opinion.

10.4 Newton's Method for Solving $f(x) = 0$

This section presents another way to find a sequence of approximations to a solution of $f(x) = 0$. **Newton's method** uses information about f and its derivative to produce estimates that usually converge faster than the sequences obtained by the bisection method.

The Idea Behind Newton's Method

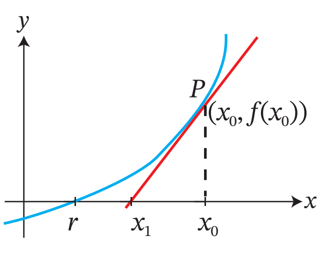


Figure 10.4.1

Figure 10.4.1 shows the graph of a function f that has a root r and a point x_0, that is an initial estimate of r. (The initial estimate comes from looking at a graph or some calculations on a calculator.)

To get a better estimate of r, find where the tangent line at $P = (x_0, f(x_0))$ crosses the x-axis. Call the new estimate x_1, as shown in Figure 10.4.1.

Repeat the process using x_1 as the estimate of the root r, and call the new estimate x_2. Continuing, we get a recursively-defined sequence $x_0, x_1, x_2, \dots, x_n, \dots$. In practice, one stops Newton's method when two successive estimates are sufficiently close together. That is, when $|x_{n+1} - x_n|$ is smaller than a given tolerance level.

The Key Formula

To obtain a formula for x_1 in terms of x_0, observe that the slope of the tangent line at P in Figure 10.4.1 is $f'(x_0)$ and also $(f(x_0) - 0)/(x_0 - x_1)$. We assume $f'(x_0)$ is not zero, so the tangent at P is not parallel to the x-axis. From $f'(x_0) = (f(x_0) - 0)/(x_0 - x_1)$ we get, since we assumed $f'(x_0) \neq 0$, $x_0 - x_1 = f(x_0)/f'(x_0)$. Solving for x_1, we find

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

The next estimate of the root is found by finding where the tangent line at $(x_1, f(x_1))$ intersects the x-axis:

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

This process can be continued and so on for $x_3, x_4, \dots$. In general, we have a recursively-defined sequence.

Definition: *Newton's Recursion*

Newton's method for solving $f(x) = 0$ with an initial guess x_0 constructs the sequence $\{x_n\}$ of approximate solutions defined recursively by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad \text{for } n = 1, 2, \ldots. \tag{10.4.1}$$

Before we examine whether the sequence defined by (10.4.1) converges to a solution of $f(x) = 0$ we look at some examples.

EXAMPLE 1. In the previous section thirteen iterations of the bisection method were needed to estimate the solution to $f(x) = x + \sin(x) - 2 = 0$ to three decimal places. Let's see how Newton's method deals with the same problem.

SOLUTION A reasonable initial estimate is $x_0 = 2$, because it cancels the -2 in $x + \sin(x) - 2$, and $\sin(2)$ is not too large. The derivative of $f(x) = x + \sin(x) - 2$ is $f'(x) = 1 + \cos(x)$. The recursion formula (10.4.1) reads

$$x_{n+1} = x_n - \frac{x_n + \sin(x_n) - 2}{1 + \cos(x_n)}.$$

The first six iterations, rounded to six decimal places, are in Table 10.4.1.

n	x_n	$f(x_n)$	$f'(x_n)$
0	2.000000	0.909297	0.583853
1	0.442592	−1.129124	1.903644
2	1.035731	−0.104034	1.509898
3	1.104632	−0.002069	1.449463
4	1.106060	−0.000001	1.448188
5	1.106060	0.000000	1.448187
6	1.106060	0.000000	1.448187

Table 10.4.1

Because, to six decimal places, $f(x_5) = 0$, all subsequent estimates will be the same as x_5. The conclusion is that $r \approx x_5 = 1.106060$ to six decimal places.

The bisection method is easier to use than Newton's method. However, Newton's method needs only five steps to obtain the approximation $x_5 \approx 1.106060$ of the root to f accurate to six decimal places while after thirteen iterations the bisection method yields an approximation, $p_{13} \approx 1.106201$, accurate to only three decimal places. ◆

EXAMPLE 2. Use Newton's method to estimate the square root of 3, the positive root of the equation $x^2 - 3 = 0$.

SOLUTION Here $f(x) = x^2 - 3$ and $f'(x) = 2x$. According to (10.4.1),

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^2 - 3}{2x_n} = \frac{1}{2}\left(x_n + \frac{3}{x_n}\right).$$

For our initial estimate, let us use $x_0=2$. Its square is 4, which is not far from 3. Then, with $n=0$:

$$x_1=\frac{1}{2}\left(x_0+\frac{3}{x_0}\right)=\frac{1}{2}\left(2+\frac{3}{2}\right)=1.75.$$

Repeat this process, using $x_1=1.75$ to obtain the next estimate:

$$x_2=\frac{1}{2}\left(x_1+\frac{3}{x_1}\right)=\frac{1}{2}\left(1.75+\frac{3}{1.75}\right)\approx 1.732142857.$$

One more step yields (to five decimal places) $x_3\approx 1.73205$, which is close to $\sqrt{3}$, whose decimal expansion begins 1.7320508. See Figure 10.4.2, which lists x_0, x_1, and the graph of $f(x)=x^2-3$. Table 10.4.2 shows the numerical values. Compare Table 10.4.2 with Table 10.3.2.

n	x_n	$f(x_n)$	$f'(x_n)$
0	2.000000	1.000000	4.000000
1	1.750000	0.062500	3.500000
2	1.732143	0.000319	3.464286
3	1.732051	0.000000	3.464102
4	1.732051	0.000000	3.464102

Table 10.4.2

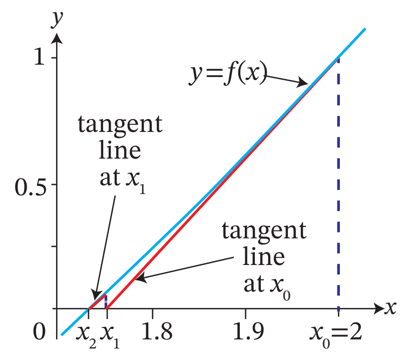


Figure 10.4.2

In fact, x_3 agrees with $\sqrt{3}$ to seven decimal places. When the same problem was solved using the bisection method in Example 1 in Section 10.3, after eleven iterations the best approximation to $\sqrt{3}$ is $p_{11}=1.732178$, which is accurate only to three decimal places. ♦

In practice we stop computing approximate solutions when either $|f(x_n)|$ or the distance between successive estimates $|x_n-x_{n-1}|$ becomes sufficiently small.

EXAMPLE 3. Use Newton's method to approximate the location of the local extrema of $g(x)=2x-(x+1)e^{-x}$.

SOLUTION This problem, which was first encountered in Exercise 29 in Section 10.3, is equivalent to asking for all roots of $f(x)=g'(x)=2+xe^{-x}$.

To find an initial guess note that $f(x)$ is positive for all positive x and $f(0)=2$. Then $f(-2)=2+(-2)e^2=2-2e^2<0$ because $e>1$. Since there is a root between 0 and -2 we choose $x_0=-1$.

The first few iterations of Newton's method with $x_0=-1$ are shown in Table 10.4.3.

n	x_n	$f(x_n)$	$f'(x_n)$	$\lvert x_n-x_{n-1}\rvert$
0	−1.000000	−0.718282	5.436564	
1	−0.867879	−0.067163	4.449017	0.132121
2	−0.852783	−0.000773	4.346941	0.015096
3	−0.852606	0.000000	4.345751	0.000177
4	−0.852606	0.000000	4.345751	0.000000

Table 10.4.3

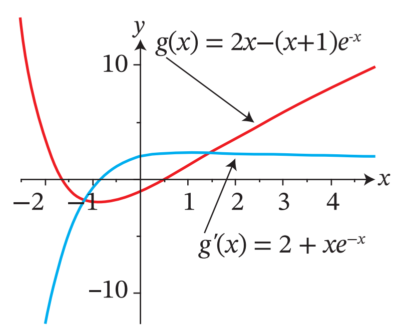


Figure 10.4.3

After four steps we stop because $f(x_3) \approx 0$. The critical number x^* of g is approximately $x_3 = -0.852606$, which is correct to all six decimal places shown.

Because $g'(x)$ is negative to the immediate left of x^* and is positive to the immediate right of x^* we conclude that x^* provides a local minimum. The graphs of g and $g' = f$ are shown in Figure 10.4.3. The only local extremum is the local minimum near $x = -0.85$. ♦

To conclude this introduction to Newton's method, we revisit Example 1 in Section 10.3 and Example 2 in this section. Both examples dealt with solving $x^2 - 3 = 0$, that is, to finding $\sqrt{3}$.

Observation 10.4.1: *How Good is Newton's Method?*

Newton's method applied to solving $f(x) = 0$ with an initial guess x_0 produces a sequence of estimates x_0, x_1, x_2, ... that hopefully approach a number r that is a root of f, that is, $f(r) = 0$. To see how quickly the x_n approach r, look at how rapidly $|x_n - r|$ approaches 0?

Newton's method (Example 2 in this section) produces a sequence that converges to the exact solution of $x^2 - 3 = 0$ much faster than the sequence obtained with the bisection method (Example 1 in Section 10.3). Now, we want to examine more closely the rate at which $|x_n - r|$ shrinks in Newton's method. The table lists x_0, x_1, x_2, and x_3 to seven decimal places for Newton's method and compares them to $\sqrt{3} \approx 1.7320508$:

Estimate	Value	Agreement with $\sqrt{3}$
x_0	2.000000000	Initial guess
x_1	$\underline{1.75}0000000$	First two digits
x_2	$\underline{1.732}142857$	First four digits
x_3	$\underline{1.7320508}10$	First eight digits

At each stage the number of correct digits tends to double. The error at one step is roughly the square of the error at the previous step and

$$|x_n - r| \le M|x_{n-1} - r|^2$$

for a constant M. In a numerical analysis course it is often shown that the constant depends on the size of the absolute values of the first and second derivatives of f.

Remarks on Newton's Method

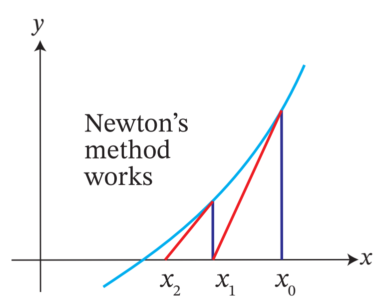


Figure 10.4.4

First, recall that the assumption that f'' exists implies f' and f are continuous.

In an interval where $f''(x)$ is positive, the graph of $y = f(x)$ is concave up and lies above its tangent lines, as shown in Figure 10.4.4. (Recall: Assuming f'' exists implies f' and f are continuous.) If x_1 is to the right of r, the sequence x_1, x_2, x_3, ... is monotone decreasing and is bounded below by r. Thus, the sequence converges to a limit $L \ge r$. To show that L is r, take limits of both sides of the Newton recursion formula (10.4.1) obtaining

$$\lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} \left(x_n - \frac{f(x_n)}{f'(x_n)} \right). \tag{10.4.2}$$

The fact that both f and f' are continuous means that both $\lim_{n\to\infty} f(x_n) = f(L)$ and $\lim_{n\to\infty} f'(x_n) = f'(L)$. If $f'(L)$ is not 0 it follows that

$$L = L - \frac{f(L)}{f'(L)}. \tag{10.4.3}$$

Thus $0 = -f(L)/f'(L)$, which implies that $f(L) = 0$, and L is a root of f.

The reasoning that obtained (10.4.3) from (10.4.2) shows, more generally, that if the sequence produced by Newton's method converges, its limit is a root of the function f.

The equation $f(x) = 0$ may not have a solution. Then the sequence produced by Newton's method does not approach a number but may vary as in Figure 10.4.5(a).

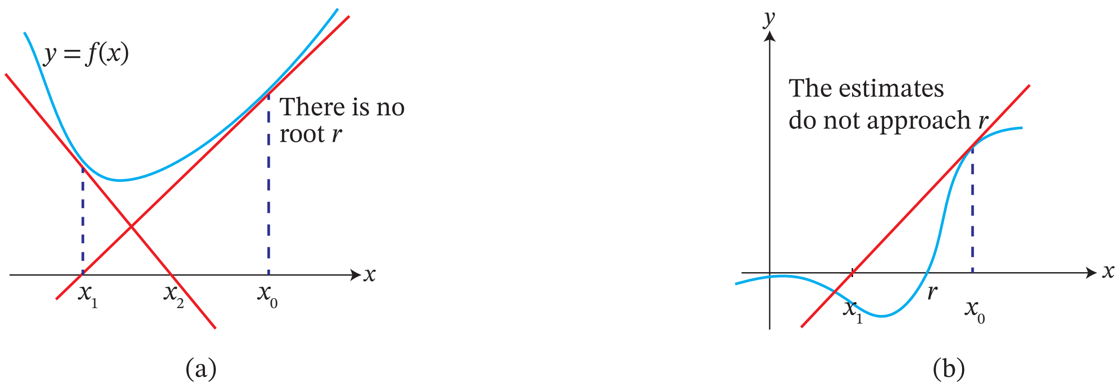


Figure 10.4.5

It is also possible that there is a root r, but the initial guess x_0 is so far from r that the sequence of estimates does not approach r. See Figure 10.4.5(b).

If x_n is a number where $f'(x_n) = 0$, then the tangent line at $(x_n, f(x_n))$ is horizontal and does not intersect the x-axis. Thus the Newton recursion, which has $f'(x_n)$ in the denominator, makes no sense.

Summary

This section developed Newton's method for estimating a root of $f(x) = 0$. We start with an estimate x_0 of the root, compute

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)},$$

then repeat, obtaining a sequence $\{x_n\}$ with

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \qquad n = 1, 2, 3, \ldots.$$

When $f'(r) \neq 0$ and f' is continuous, the iterates in Newton's method converge to r provided the initial guess is sufficiently close to r.

The Newton iterates converge quickly to the root: there is a constant M such that

$$|x_n - r| \leq M|x_{n-1} - r|^2$$

while the iterates computed by the bisection method converge more slowly:

$$|x_n - r| \leq \frac{1}{2}|x_{n-1} - r|.$$

These estimates tell us that while each iteration of the bisection method reduces the distance to the exact solution by a factor of one-half, each iteration of Newton's method is reduced by the square of the previous distance to the exact solution.

EXERCISES for Section 10.4

Exercises 11 to 30 in this section are the same as Exercises 8 to 27 in Section 10.3, except that different methods are used estimate each answer (Section 10.3 uses bisection method, Section 10.4 uses Newton's method).

In Exercises 1 and 2, use Newton's method to find x_1.

1. $x_0 = 2,\ f(2) = 0.3,\ f'(2) = 1.5$

2. $x_0 = 3,\ f(3) = 0.06,\ f'(3) = 0.3$

Definition: *Babylonian Method for Finding $\sqrt{a}$*

The **Babylonian method** for estimating $\sqrt{a}$ involves the sequence defined by starting with an estimate x_0 and then computing the next term by averaging the current term and a divided by the current term, that is: $x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right)$ for $n = 0, 1, \ldots$. That is, the next term is the average of the current term and a divided by the current term.

Exercises 3 to 5 combine to show that the Babylonian method does converge to $\sqrt{a}$ for any $a > 0$. Exercise 6 shows that the Babylonian method can also be obtained as a sequence defined by Newton's method. If x_0 is smaller than $\sqrt{a}$, then a/x_0 is larger than $\sqrt{a}$, and vice versa. So x_1 is the average of two numbers between which $\sqrt{a}$ lies. Exercises 7 and 8 provide opportunities to practice applying the Babylonian method.

3. Assume $x_n > 0$ is a typical estimate of $\sqrt{a}$ and $x_{n+1} = \frac{1}{2}\left(x + n + \frac{a}{x_n}\right)$.
   (a) Show that if $x_n < \sqrt{a}$, then its successor, x_{n+1}, is greater than $\sqrt{a}$.
   (b) Show that if $x_n > \sqrt{a}$, then its successor, x_{n+1}, is also greater than $\sqrt{a}$.

4. In light of Exercise 3, the only term in the sequence computed by the Babylonian method that is less than $\sqrt{a}$ would be the first term; all other terms will be larger than $\sqrt{a}$. Show that if $x_n > \sqrt{a}$, then its successor, x_{n+1}, is less than x_n.

5. (a) Show that a sequence of estimates of $\sqrt{a}$ obtained by the Babylonian method has a limit. Call this limit L.
   (b) Show that $L = \sqrt{a}$.

6. Let a be positive. Show that the Newton recursion formula for estimating $\sqrt{a}$ by solving $x^2 - a^2 = 0$ is given by $x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right)$.

7. Use the Babylonian method to estimate $\sqrt{15}$. Choose $x_0 = 4$ and compute x_1 and x_2 to three decimal places.

8. Use the Babylonian method to estimate $\sqrt{19}$. Choose $x_0 = 4$ and compute x_1 and x_2 to three decimal places.

---

9. Estimate $\sqrt{5}$ with Newton's method.
   (a) Use $f(x) = x^2 - 5$ and start with $x_0 = 2$. Continue until consecutive estimates differ by at most 0.01, that is, $|x_{n+1} - x_n| < 0.01$.
   (b) How many more steps are needed to obtain consecutive estimates are reduced by another factor of 10, that is, $|x_{n+1} - x_n| < 0.001$?

10. Estimate $\sqrt[3]{2}$ with Newton's method.
   (a) Use $f(x) = x^3 - 2$ and start with $x_0 = 1$. Continue until consecutive estimates differ by at most 0.01, that is, $x_{n+1} - x_n < 0.01$.

(b) How many more steps are needed to obtain consecutive estimates are reduced by another factor of 10, that is, $|x_{n+1} - x_n| < 0.001$?

11. Use Newton's method to estimate $\sqrt[3]{7}$. Choose $x_0 = 2$ and compute x_1 and x_2 to four decimal places.

12. Use Newton's method to estimate $\sqrt[3]{25}$. Choose $x_0 = 3$ and compute x_1 and x_2 to four decimal places.

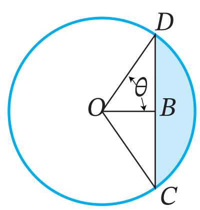


Figure 10.4.6

In Exercises 13 to 15 use Newton's method to estimate an angle θ (see Figure 10.4.6) to two decimal places. Angles are in radians. For each problem, show that there is only one answer in $0 < \theta < \frac{\pi}{2}$. Compare with the results for Exercises 10 to 12 in Section 10.3.

13. Find θ for which $|\widehat{CD}| = 2|\overline{OB}|$.

14. Find θ for which the areas of triangle $\triangle OCD$ and the shaded region are equal.

15. Find θ for which $|\widehat{CD}| = \frac{3}{4}|\overline{OB}|$.

Exercises 16 to 30 are the same as Exercises 13 to 27 in Section 10.3, except that Newton's method is used instead of the bisection method.

16. Let $f(x) = x^5 + x - 1$.
(a) Show that there is a root of $f(x) = 0$ in the interval $[0, 1]$.
(b) Why is the root unique?
(c) Using $x_0 = \frac{1}{2}$ as a initial estimate, apply Newton's method to find a first computed estimate x_1.

17. Let $f(x) = x^4 + x - 19$.
(a) Show that $f(2) < 0 < f(3)$. What property of f assures that there is exactly one root r between 2 and 3?
(b) Apply Newton's method, starting with $x_0 = 2$. Compute x_1 and x_2.
(c) Another root of $f(x)$ is negative. Find an interval of length 1 that contains it.
(d) Use the left endpoint of the interval in (c) as the initial guess for Newton's method. Compute x_1 and x_2.

18. In estimating $\sqrt{3}$ with Newton's method, Sam imprudently chooses $x_0 = 10$. What does Newton's method give for x_1, x_2, and x_3?

19. Let $f(x) = 2x^3 - x^2 - 2$.
(a) Show that there is exactly one root of $f(x) = 0$ in the interval $[1, 2]$.
(b) Using $x_0 = \frac{3}{2}$ as a first estimate, apply Newton's method to find x_2 and x_3.

20. (a) Graph $y = x$ and $y = \cos(x)$ on the same axes.
(b) Using the graph in (a), estimate the positive solution of the equation $x = \cos(x)$. Is there a negative solution?
(c) Using the estimate in (b) as x_0, apply Newton's method until consecutive estimates agree to four decimal places.

21. (a) Graph $y = \cos(x)$ and $y = 2\sin(x)$ on the same axes.
(b) Using the graph in (a), estimate the solution to $\cos(x) = 2\sin(x)$ that lies in $[0, \pi/2]$.
(c) Using the estimate in (b) as x_0, apply Newton's method until consecutive estimates agree to four decimal places.

22. (a) Graph $y = x\sin(x)$ for x in $[0, \pi]$.
(b) Using the first and second derivatives, show that it has a unique relative maximum in $[0, \pi]$.
(c) Show that it occurs when $x\cos(x) + \sin(x) = 0$.
(d) Use Newton's method with $x_0 = \frac{\pi}{2}$ to find an estimate x_1 for a root of $x\cos(x) + \sin(x) = 0$.
(e) Use Newton's method again to find x_2.

23. (a) Graph $y = x\cos(x)$ for x in $[0, \pi]$.
(b) Using the first and second derivatives, show that it has a unique relative maximum in $\left[0, \frac{\pi}{2}\right]$.
(c) Show that it occurs when $\cos(x) - x\sin(x) = 0$.
(d) Use Newton's method with $x_0 = \frac{\pi}{4}$ to find an estimate x_1 for a root of $\cos(x) - x\sin(x) = 0$.
(e) Use Newton's method again to find x_2.

24. Use Newton's method to estimate the maximum value of $y = 2\sin(x) - x^2$ over the interval $\left[0, \frac{\pi}{2}\right]$.

25. Use Newton's method to estimate the maximum value of $y = x^3 + \cos(x)$ over the interval $\left[0, \frac{\pi}{2}\right]$.

26. The equation $x\tan(x) = 1$ occurs in the theory of vibrations.
(a) How many roots does it have in the interval $\left[0, \frac{\pi}{2}\right]$.
(b) Use Newton's method to estimate the roots to two decimal places.

27. The function $f(x)$ is defined to be $f(x) = \frac{\sin(x)}{x}$ for $x \neq 0$ and $f(0) = 1$.
(a) From a graph of $y = f(x)$, how many critical numbers does $f(x)$ have?
(b) Show that a critical number of $f(x)$ satisfies the equation $\tan x = x$.
(c) Show that $f(x)$ is an even function. Thus we need consider only positive x. (That is, if x is a critical number of $f(x)$, then so is $-x$.)
(d) Graph the functions $\tan(x)$ and x on the same axes. How often do they cross for x in $\left[0, \frac{\pi}{2}\right]$? for x in $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$? for x in $\left[\frac{3\pi}{2}, \frac{5\pi}{2}\right]$?
(e) Show that $\tan(x) - x$ is an increasing function for x in $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$. What does that tell us about the number of solutions of the equation $\tan(x) = x$ for x in $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$?
(f) Use Newton's method with $x_0 = \pi$ to estimate its critical number in $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$ to at least two decimal places.
(g) Repeat (e) and (f) with $x_0 = 2\pi$ to estimate its critical number in $\left[\frac{3\pi}{2}, \frac{5\pi}{2}\right]$?
(h) Why is Newton's method not needed to find the critical number of $f(x)$ on the interval $\left[0, \frac{\pi}{2}\right]$?

28. How many solutions does $2x + \sin(x) = 2$ have? Use Newton's method to estimate them to two decimal places. Explain the steps in your solution in complete sentences.

29. How many solutions does the equation $\sin(x) = x$ have? Explain how you could use Newton's method to estimate them.

30. Explain how you could use Newton's method to obtain a formula for estimating $\sqrt[5]{a}$.

Exercises 31 and 32 show that care should be taken when applying Newton's method.

31. Let $f(x) = 2x^3 - 4x + 1$.
(a) Show that there is a root r of $f(x) = 0$ in $[0, 1]$.
(b) Take $x_0 = 1$ and apply Newton's method to obtain x_1 and x_2.
(c) Graph f, and show what is happening in the sequence of estimates.

32. Apply Newton's method, starting with $x_0 = \frac{1}{\sqrt{5}}$, to solve $f(x) = 0$ where $f(x) = x^3 - x$.
(a) Compute x_1 and x_2 exactly, not as decimal approximations.
(b) Graph f and explain why Newton's method fails.

---

33. Let $f(x) = x^2 + 1$

(a) Using Newton's method with $x_0 = 2$, compute x_1, x_2, x_3, and x_4 to two decimal places.

(b) Using the graph of f, show geometrically what is happening in (a).

(c) Using Newton's method with $x_1 = \frac{\sqrt{3}}{3}$, compute x_2 and x_3. What happens to x_n as $n \to \infty$?

(d) What happens when we use Newton's method, starting with $x_1 = 1$?

34. Assume that $f'(x) > 0$, $f''(x) < 0$ for all x, and $f(r) = 0$.

(a) Sketch a possible graph of $y = f(x)$.

(b) Describe the behavior of the sequence of Newton's estimates $x_0, x_1, \ldots, x_n, \ldots$ when you choose $x_0 > r$. Include a sketch.

(c) Describe the behavior of the sequence if you choose $x_0 < r$. Include a sketch.

35. Let $f(x) = \frac{1}{x} + 5$

(a) Graph $y = f(x)$ showing its x-intercepts.

(b) For what x_0 do the Newton iterates converge to a solution to $f(x) = 0$?

(c) For what x do the Newton iterates not converge?

36. Let $f(x) = \frac{1}{x^2} - 5$ and answer the same questions as in Exercise 35.

37. We can show that the error in Newton's method diminishes rapidly compared to the bisection method. Let x_0 be an estimate of the root r and let x_1 be the first estimate obtained by Newton's method. Assume $f'(x_0) \neq 0$.

Using the first-order Taylor polynomial with remainder centered at $a = x_0$, we may write

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(c)}{2}(x - x_0)^2 \tag{10.4.4}$$

where c is a number between x and x_0. (See Section 5.5.)

(a) In (10.4.4), replace x by r and use the definition of x_1 to show that $x_1 - r = \frac{f^{(2)}(c)}{2f'(x_0)}(r - x_0)^2$, where c is between x_0 and r.

(b) Assume that $x_0 > r$ and that $f'(x)$ and $f''(x)$ are positive for x in $[r, x_0]$. Indicate on a diagram where the numbers $x_1, x_2 \ldots$ are situated. Then use (a) to discuss how the error, $r - x_n$, behaves as n increases.

(c) According to (a), Newton's method is especially efficient when $f''(x)$ is small and $f'(x)$ is large. Using a diagram, show why this is to be expected.

38. Let p be positive.

(a) Graph $f(x) = \frac{1}{x} - p$.

(b) For what choices of x_0 will Newton's method converge to a root of f?

39. Newton's method depends on knowing f'. However, f' may be too complicated or f might be known at only a few inputs. If one makes an initial guess of a root of f, how might one find a plausible better approximation? What could be used instead of a tangent line?

10.S Chapter Summary

Infinite sequences of numbers $a_k, a_{k+1}, \ldots$ arise in many contexts. (The initial index, k, can be any integer, but is usually either 0 or 1.) For instance, sequences arise when estimating a root of an equation $f(x) = 0$. Any equation $g(x) = h(x)$ can be transformed to that form, for it is equivalent to $g(x) - h(x) = 0$.

One way to estimate a root of $f(x) = x$ is to pick an estimate, a, of a root and compute the sequence $f(a)$, $f(f(a))$, $f(f(f(a)))$, ... This is a recursively-defined sequence: $x_0 = a$, $x_{n+1} = f(x_n)$, $n = 0, 1, 2, \ldots$. If this sequence has a limit, r, then $f(r) = r$. In other words, r is a fixed point of f.

The bisection method provides estimates of a solution of $f(x) = 0$. We look for numbers a and b at which $f(x)$ has opposite signs. If f is continuous, it has a root in the interval (a, b). Let m be the midpoint of the interval. Then either m is a root of f or the sign of $f(m)$ is opposite the sign of one of $f(a)$ and $f(b)$. Repeat, using either (a, m) or (m, b) depending on which interval has ends of opposite signs (when substituted into f). Continue the process until the intervals are short enough. The midpoint of the final interval is an approximation to the root and the error estimate is half the length of the interval.

Newton's method for solving $f(x) = 0$ uses tangent line approximations to the graph of $f(x)$ to construct a sequence converging to a root. It yields the recursion $x_{n+1} = x_n - f(x_n)/f'(x_n)$, which is repeated until consecutive estimates are close enough. Convergence rates and potential problems associated with Newton's method are explored in the exercises in Section 10.4.

EXERCISES for Section 10.S

1. Let $a_0 = 0$ and $a_n = a_{n-1} + 2n - 1$ for $n \geq 1$.
- (a) Compute a few values of a_n, at least through a_5, and conjecture an explicit formula for a_n.
- (b) Show that if your formula is correct for $n = k$, then it is correct for $n = k + 1$.

2.
- (a) Graph $f(x) = \cos\left(\frac{\pi x}{2}\right)$ for x in $[0, 1]$.
- (b) Let a be the unique fixed point of f on $[0, 1]$. Estimate a by looking at the graph.
- (c) Use Newton's Method to estimate a to two decimal places.
- (d) Use the bisection method to estimate a to two decimal places.
- (e) Does the sequence $\cos\left(\frac{\pi x}{2}\right), \cos\left(\frac{\pi}{2}\cos\left(\frac{\pi x}{2}\right)\right), \ldots$ converge for every x in $[0, 1]$?

3. In Example 1 in Section 4.1 it was shown that $f(t) = (t^2 - 1)\ln\left(\frac{t}{\pi}\right)$ has one critical number on $[1, \pi]$. Use Newton's method to estimate it to three decimal digits.

4. In Example 2 in Section 4.1 it was shown that $f(x) = x^3 - 6x^2 + 15x + 3$ has exactly one real root. Use Newton's method to approximate it to three decimal places.

5. (a) Graph $y = xe^{-x^2}$. (b) Estimate the area of the region bounded by $y = xe^{-x^2}$, the line $x + y = 1$, and the x-axis.

NOTE: Use Newton's method of estimating a solution of an equation.

6. The spiral $r = \theta$ meets the circle $r = 2\sin(\theta)$ at a point other than the origin. Use Newton's method to estimate its coordinates. Give both the polar and rectangular coordinates of the point of intersection.

7. The equation $M = E - e\sin(E)$, known as **Kepler's equation**, occurs in the study of planetary motion. (Here, M is the **mean anomaly**, E is the **eccentric anomaly**, and e is the eccentricity of a planet's orbit.)
- (a) Sketch the graph of $M(E) = E - e\sin(E)$ as a function of E when $e = 0.2$.
- (b) Show that $M(E) = E - e\sin(E)$ is an increasing function of E for any $0 < e < 1$.
- (c) In view of (b), E is a function of M, $E = g(M)$. Use Newton's method to find $g(0.25)$, $g(0.5)$, and $g(1.5)$ if $e = 0.2$. Find all answers to at least three decimal digits.
- (d) Repeat (c) with $e = 0.8$.
- (e) What x_0 lead to convergent sequences? A graphing calculator or computer can be used to simplify the calculations.

Consider the problem of finding a solution to $g(x) = 0$. There are usually several ways to rewrite the equation in the form $f(x) = x$. We want to choose f so that the sequence with $a_n = f(a_{n-1})$ converges. Then $L = \lim_{n\to\infty} a_n$ is a solution to $g(x) = 0$. In Exercises 8 to 13 we develop and apply a general result, the **fixed point theorem**.

8. In this exercise we develop a version of the fixed point theorem that will explain what is happening in Exercises 9 to 11. If r is a fixed point of f, that is, a number such that $f(r) = r$, then the errors $e_n = r - a_n$ satisfy $r - e_n = f(r - e_{n-1})$.
- (a) Fill in the details to show why $r - e_n = f(r - e_{n-1})$.

(b) Replace $f(r-e_{n-1})$ with the linear approximation to f at r and derive the approximate result: $e_n \approx f'(r)e_{n-1}$ for $n \geq 0$.
(c) Show that if $e_n \approx f'(r)e_{n-1}$ for $n \geq 0$, then $e_n \approx (f'(r))^n e_0$.
(d) Explain why $e_n \to 0$ if $|f'(r)| < 1$ and $\{e_n\}$ diverges if $|f'(r)| > 1$. That is, a_n approaches r if $|f'(r)| < 1$, and $\{a_n\}$ does not approach r if $|f'(r)| > 1$.

There are several ways of turning the solving of $g(x) = x + \ln(x) = 0$ into a fixed point problem, that is, to put it in the form $f(x) = x$. Exercises 9 and 10 show some reformulations are more useful than others.

9. (a) Let $f(x) = -\ln(x)$. Verify that $g(x) = 0$ and $f(x) = x$ have the same solution.
(b) Compute $|f'(r)|$ when r is close to the solution to $g(x) = 0$. What does this imply about the sequence with $a_n = f(a_{n-1})$? (Why must r between 0 and 1?)
(c) Let $x_0 = 0.5$ and compute x_1, x_2, x_3, and x_4 using $x_n = f(x_{n-1})$. Why is it not possible to compute x_5?

10. (a) Let $f(x) = e^{-x}$. Verify that $g(x) = 0$ and $f(x) = x$ have the same solution.
(b) Compute $|f'(r)|$ when r is close to the solution to $g(x) = 0$. What does this imply about the sequence with $a_n = f(a_{n-1})$?
(c) Let $x_0 = 0.5$ and compute x_1, x_2, x_3, and x_4 using $x_n = f(x_{n-1})$. What happens as $n \to \infty$?

The function $g(x) = x^2 - 2x - 3$ has two roots, $x = 3$ and $x = -1$. Exercises 11 to 13 present three ways to use fixed point iterations to find them.

11. (a) Show that solving $g(x) = 0$ is equivalent to finding a fixed point of $f(x) = \sqrt{2x+3}$.
(b) Compute $|f'(r)|$, when r is close to either root of $g(x) = 0$. What does this imply about the sequence $a_n = f(a_{n-1})$?
(c) Let $x_0 = 0.5$ and compute x_1, x_2, x_3, and x_4 using $x_n = f(x_{n-1})$. What is $\lim_{n\to\infty} x_n$?

12. (a) Show that solving $g(x) = 0$ is equivalent to finding a fixed point of $f(x) = 3/(x-2)$.
(b) Compute $|f'(r)|$, when r is close to either root of $g(x)$. What does this imply about the sequence $a_n = f(a_{n-1})$?
(c) Let $x_0 = 0.5$ and compute x_1, x_2, x_3, and x_4 using $x_n = f(x_{n-1})$. What is $\lim_{n\to\infty} x_n$?

13. (a) Show that solving $g(x) = 0$ is equivalent to finding a fixed point of $f(x) = \frac{1}{2}(x^2 - 3)$.
(b) Compute $|f'(r)|$, when r is close to the solutions to $g(x) = 0$. What does this imply about the sequence $a_n = f(a_{n-1})$?
(c) Let $x_0 = 0.5$ and compute x_1, x_2, x_3, and x_4 using $x_n = f(x_{n-1})$. What is $\lim_{n\to\infty} x_n$?
(d) Which of the methods in the Exercises 11 to 13 is the best to find the solutions to $g(x) = 0$?

---

14. Let u_n be the number of ways of tiling a $3 \times n$ rectangle with 1×3 dominoes.
(a) Find u_1, u_2, and u_3. (b) Find a recursive definition of u. (c) Use (b) to find u_{10}.

15. A tile consists of three 1×1 squares arranged to form the letter L. Let u_n be the number of ways a $2 \times n$ rectangle can be tiled by such tiles.
(a) Find u_n for $n = 1, 2, 3, 4, 5$, and 6. (b) Find a recursion for u_n. (c) Find u_n for $n = 22$, 23, and 24.

16. Repeat Exercise 15 with u_n the number of ways a $3 \times n$ rectangle can be tiled by the L-shaped tiles.

17. Let the mass of a bacteria culture at the end of n intervals of time be C_n. If there are an adequate supply of nutrients, the mass doubles during each interval, that is, $C_{n+1} = 2C_n$. When the population is large it does not reproduce as quickly. Then, according to the Verhulst model (1848) there is a constant K such that $C_{n+1} = \dfrac{2}{1+\dfrac{C_n}{K}} C_n$.

Show that $\lim_{n\to\infty} C_n = K$.

Definition: *Binomial Distribution*

An experiment either succeeds, with probability p, or fails, with probability q, where $p+q=1$. For instance, the probability of getting a five when rolling a six on one die is $p=1/6$, so $q=5/6$. If the experiment is repeated n times we expect near pn successes. For $k=0, 1, \ldots, n$, the probability of having k successes in n experiments is

$$\frac{n!}{k!(n-k)!}p^k q^{n-k}, \tag{10.S.1}$$

called the **binomial distribution**. Exercise 18 shows how letting $n \to \infty$ in the binomial distribution leads to the **Poisson distribution**. Exercise 19 obtains an approximation of (10.S.1) when k is near pn which is related to the normal distribution.

18. Assume the n in (10.S.1) is large (there are many trials), k is a fixed positive integer, and v is the expected number of successes (which is fixed). Let $p=\frac{v}{n}$ and $q=1-p=1-\frac{v}{n}$. Then (10.S.1) becomes

$$\frac{n!}{k!(n-k)!}\left(\frac{v}{n}\right)^k\left(1-\frac{v}{n}\right)^{n-k}. \tag{10.S.2}$$

(a) Show that $\lim_{n\to\infty}\left(1-\frac{v}{n}\right)^{n-k}=e^{-v}$.

(b) Show that $\lim_{n\to\infty}\frac{n!}{(n-k)!n^k}=1$ (k is fixed).

(c) Use (a) and (b) to deduce that the limit of (10.S.2) as $n\to\infty$ is $P(k)=\frac{v^k e^{-v}}{k!}$. The function $P(k)$ describes the probability that an event occurs k times in the Poisson distribution.

19. The exercise obtains an approximation of (10.S.1) when k is "near" pn. "Near" means $\lim_{n\to\infty}(k-pn)/n=0$. We may write $k=pn+z_k$, where $\frac{z_k}{n}\to 0$ as $n\to\infty$. Note that $k\to\infty$ as $n\to\infty$.

We will use the Stirling approximation of $m!$, namely $\sqrt{2\pi m}\left(\frac{m}{e}\right)^m$, developed in Exercises 30 to 35 in Section 11.S.

(a) Show that (10.S.1) is approximated by

$$\left(\frac{n}{2\pi k(n-k)}\right)^{1/2}\left(\frac{pn}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}. \tag{10.S.3}$$

(b) Show that $\left(\frac{n}{2\pi k(n-k)}\right)^{1/2}\approx\frac{1}{\sqrt{2pq\pi n}}$.

(c) Show that the other two factors of (10.S.2) equal

$$\frac{1}{\left(1+\frac{z_k}{np}\right)^{np+z_k}}\frac{1}{\left(1-\frac{z_k}{nq}\right)^{nq-z_k}}. \tag{10.S.4}$$

(d) Using the approximation $\ln(1+x)=x-\frac{1}{2}x^2$, show that the natural logarithm (ln) of the denominator in (10.S.4) is approximately $\frac{z_k^2}{2npq}$.

(e) Using (b) and (d), show that (10.S.3) is approximately $\frac{1}{\sqrt{2\pi pqn}}e^{-(k-np)^2/(2npq)}$.

Observation 10.S.1: *Significance of e^{-x^2} in Probability Theory*

In Exercise 19(e) the function e^{-x^2} appears. This explains why that function is important in probability theory. Contrast this with Exercise 18, where the function e^{-x} appears.

---

20. The recursion $P_{n+1} = re^{-P_n/K}P_n$ was introduced by W. E. Ricker in 1954 in the study of fish populations. P_n denotes the fish population at the n^{th} time interval, while r and K are constants, with r being the maximum reproduction rate. Examine the recursion when $K = 10{,}000$, $P_0 = 5{,}000$ and (a) $r = 20$ and (b) $r = 10$.

NOTE: The unpredictable sequence $\{P_n\}$ depends dramatically on r. It was an early example of chaos.

21. SAM: I'm going to prove using the precise definition that if $0 < r < 1$, then $\lim_{n\to\infty} r^n = 0$.

JANE: I'll listen.

SAM: I want to show that there is an integer N such that $|r^n - 0| < \epsilon$ if $n > N$, in other words, $r^n < \epsilon$ if n is big enough. To find n, I take logarithms, so $n\ln(r) < \ln(\epsilon)$. Then I'll divide by $\ln(r)$.

JANE: How do you know r has a log?

SAM: Well, $r = e^{\ln(r)}$.

JANE: You mean the equation $r = e^x$ has a solution?

SAM: Sure, that's what a log is all about.

JANE: Since r is less than 1, x would be negative. May I write it as $-p$ where p is positive?

SAM: If you want to, why not?

JANE: So you're saying that r can be written as $(1/e)^p$ for some positive number p. You're assuming that no matter how small r is, there is a positive number p so that $(1/e)^p$ will equal it. Right?

SAM: Right. But why all this fuss?

JANE: To say that $\frac{1}{e^p}$ gets as small as you please is just a special case of what you're trying to prove. You're wandering in circles.

Who's right, Jane or Sam? If Sam is right, finish his proof.

22. In 1998 H. J. Brothers and J. A. Kness published this estimate of e: $f(n) = \frac{(n+2)^{n+2}}{(n+1)^{n+1}} - \frac{(n+1)^{n+1}}{n^n}$.

(a) Compute $f(n)$ for $n = 1$, 2, 3, and 4 to three decimal places.

(b) Show that $\lim_{n\to\infty} f(n) = e$.

23. Japanese tatami mats are typically twice as long as wide, and are used to cover floors in houses and temples without overlap. Such a covering is called **auspicious** if no four corners meet at a point. To accomplish this some square mats, half as large, must be used to fill gaps. This raises a question: For a positive integer n how many auspicious coverings are there of an $n \times n$ square using exactly n 1×1 squares and $(n^2 - n)/2$ tatamis? Call this number T_n. (a) Show that $T_1 = 1$ and $T_2 = 4$. (b) In 2011 it was shown that $T_n = 2^n + 4T_{n-2}$. Use this recursion to find T_n for $n = 3$, 4, 5, and 6. (c) Let $U_n = \frac{T_n}{2^{n-1}}$. Find U_1 and U_2. (d) Find a recursion for $U(n)$. (e) Show that $T(n) = n2^{n-1}$.

REFERENCE: A. Erickson, F. Ruskey, J. Woodcock, and M. Schurch, "Monomer-diamer tatami tilings of rectangular arrays," *Electronic Journal of Combinatorics* **18**(1) 2011, 109-132.

24. Define t_n by the recursion $t_n = \frac{2n}{n-1} t_{n-1}$ for $n \geq 2$, with $t_1 = 1$. Find an explicit formula for t_n. Comparing this exercise with the previous one shows that different recursions can describe the same function. (By the way, the same sequence describes the number of edges in an n-dimensional cube. However, there seems to be no relation between the cubes and tatamis.)

Chapter 11

Series

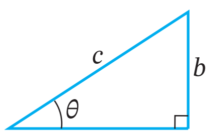


Figure 11.0.1

The ideas and techniques developed in this chapter build upon the ideas introduced in Chapter 10 and will be applied in Chapter 12.

How is $\sin(\theta)$ computed? One approach might be to draw a right triangle with one angle θ, as in Figure 11.0.1, measure the length of the opposite side b and the length of the hypotenuse c, and calculate b/c. With this method, one would be lucky to get even two decimal places correct. This cannot give the many decimal places a calculator displays for $\sin(\theta)$, even if one draws a gigantic triangle.

One way to obtain accuracy to as many decimal places as desired will be described in Chapter 12. The general idea will to use a sequence of approximating polynomials to evaluate transcendental functions like $\sin(x)$, $\arctan(x)$, e^x, and $\ln(x)$ to as many decimal places as desired. For instance, the fifth-degree Maclaurin polynomial

$$x - \frac{x^3}{6} + \frac{x^5}{120} \tag{11.0.1}$$

approximates $\sin(x)$, with an error less than 0.0002, when $|x| \le 1$ (x in radians). This means the estimate will be correct to at least three decimal places for angles less than about 57°.

RECALL: 1 radian $= 180°/\pi \approx 57.29578°$

The Maclaurin polynomial has uses other than evaluating a function. The fundamental theorem of calculus cannot be used to evaluate

$$\int_0^1 \frac{\sin(x)}{x}\, dx$$

since $\sin(x)/x$ does not have an elementary antiderivative. But, we can approximate this definite integral with

$$\int_0^1 \frac{x - \frac{x^3}{6} + \frac{x^5}{120}}{x}\, dx = \int_0^1 1 - \frac{x^2}{6} + \frac{x^4}{120}\, dx.$$

Since the integrand is now a polynomial, the fundamental theorem of calculus can be used to obtain the estimate

$$\int_0^1 \frac{\sin(x)}{x}\, dx \approx \int_0^1 1 - \frac{x^2}{6} + \frac{x^4}{120}\, dx = 1 - \frac{1}{18} + \frac{1}{600} \approx 0.94611,$$

which is accurate to three decimal places. Additional accuracy is obtained by using a higher-order Maclaurin polynomial to approximate $\sin(x)$.

11.1 Introduction to Series

Starting with a sequence $\{a_k\}$, we can form a new sequence by adding successively more terms of the original sequence: a_1, $a_1 + a_2$, $a_1 + a_2 + a_3$, $a_1 + a_2 + a_3 + a_4$, ... This new seqence of sums we call the series obtained from the original sequence. Its n^{th} term is $S_n = a_1 + a_2 + \cdots + a_n$. We generally use the index n to help distinguish the terms S_n, in the sequence of partial sums from the terms in the original sequence, a_k, where we use the index k.

Two Contrasting Examples

EXAMPLE 1. Let $a_k = \dfrac{1}{\sqrt{k}}$ for positive integers k and $S_n = \dfrac{1}{\sqrt{1}} + \dfrac{1}{\sqrt{2}} + \cdots + \dfrac{1}{\sqrt{n}}$ for positive integers n. What happens to S_n as n increases? Does it get arbitrarily large or does it stay below some fixed number?

SOLUTION Two influences operate on the sequence S_1, S_2, S_3, ... Because the number of summands increases, the sums could grow arbitrarily large. On the other hand, the summands get very small, suggesting that their sums may not grow arbitrarily large. Table 11.1.1 lists the first few values of S_n, both symbolically and numerically (to 4 decimal places).

n	$S_n = \frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n}}$	Approximation (4 decimal places)
1	$\frac{1}{\sqrt{1}}$	1.0000
2	$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}}$	1.7071
3	$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}}$	2.2845
4	$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{4}}$	2.7845
5	$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{4}} + \frac{1}{\sqrt{5}}$	3.2317

Table 11.1.1

The computations in Table 11.1.1 do not answer the question. Even if we computed S_n for $n = 1, 2, \ldots, 1,000,000$, we would not know the answer because we cannot be sure what happens when n is a billion or a google or larger. Do the sums get arbitrarily large or do they stay below some fixed number? Even the world's fastest computer cannot answer that.

However, an algebraic insight settles the question. Because

$$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n}} \tag{11.1.1}$$

has n summands and the smallest of them is $1/\sqrt{n}$, the sum (11.1.1) is at least as large as

$$\underbrace{\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}} + \cdots + \frac{1}{\sqrt{n}}}_{n \text{ summands}} = n\left(\frac{1}{\sqrt{n}}\right) = \sqrt{n}.$$

Thus $1/\sqrt{1}+1/\sqrt{2}+\cdots+1/\sqrt{n}>\sqrt{n}$. As n increases, $\sqrt{n}$ grows arbitrarily large. This implies that the sums $1/\sqrt{1}+1/\sqrt{2}+\cdots+1/\sqrt{n}$ also become arbitrarily large. They do not stay less than some fixed number. ♦

Drop a tennis ball from a height of 1 meter. It rebounds 0.6 meter. It continues to bounce, and each bounce is 60% as high as the previous bounce. (See Figure 11.1.1.) What is the total distance the ball falls?

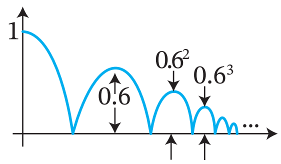


Figure 11.1.1

The first fall is 1 meter, the second 0.6 meter, the third is $(0.6)^2$ meter, and so on. In general, the n^{th} time the ball falls, it falls a distance $(0.6)^{n-1}$ meter. We know the geometric sequence converges to zero, but the question is: What happens to the sums

$$1+0.6+(0.6)^2+\cdots+(0.6)^{n-1} \quad \text{as } n\to\infty?$$

Do they become arbitrarily large, like the sums in Example 1? Do they approach a limiting value? Or, could something else happen?

EXAMPLE 2. Given the geometric progression $1, 0.6, (0.6)^2, (0.6)^3, \ldots$, form the sequence $\{S_n\}$:

$$S_1 = 1,$$
$$S_2 = 1+0.6,$$
$$S_3 = 1+0.6+(0.6)^2,$$

and, in general,

$$S_n = 1+0.6+(0.6)^2+\cdots+(0.6)^{n-1}.$$

That is, S_n is a sum of n terms of the sequence $\{a_k\}$ with $a_k = 0.6^k$ for $k = 0, 1, 2, \ldots$. Does the sequence $\{S_n\}$ converge or diverge? If it converges, what is the limit?

SOLUTION Because S_n is the sum of terms in a geometric progression whose first term is 1, has ratio 0.6, and has n terms,

$$S_n = 1+0.6+(0.6)^2+\cdots+(0.6)^{n-1} = \frac{1-(0.6)^n}{1-0.6}.$$

By Section 10.1, $\lim_{k\to\infty}(0.6)^k = 0$. Therefore

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty}\frac{1-(0.6)^n}{1-0.6} = \frac{1}{1-0.6} = 2.5.$$

♦

The content of the two examples raise a fundamental question: How can we tell whether a series $S_1, S_2, S_3, \ldots$ has a limit as $n\to\infty$? This chapter and the next address this question.

The Terminology of Series

The **series** $\{S_n\}$ with $S_1 = a_1$, $S_2 = a_1 + a_2$, $S_3 = a_1+a_2+a_3$, and, in general, $S_n = a_1+a_2+\cdots+a_n$ is denoted $\sum_{k=1}^{\infty} a_k = a_1+a_2+\cdots+a_k+\cdots$. We call $\sum_{k=1}^{\infty} a_k$ the series whose k^{th} term is a_k. The finite sum

$$S_n = a_1+a_2+\cdots+a_n = \sum_{k=1}^{n} a_k$$

is called a **partial sum** or the n^{th} **partial sum** of the series. If the sequence of partial sums of a series converges to L, then the series is said to be **convergent** and L is called the **sum** of the series. We write

$$\lim_{n\to\infty} S_n = L.$$

A series that is not convergent is called **divergent**.

The symbol

$$\sum_{k=1}^{\infty} a_k$$

is short for the sequence $S_1, S_2, \ldots, S_n, \ldots$. If the sequence of partial sums converges to a number L, we also write

$$\sum_{k=1}^{\infty} a_k = L.$$

Thus, the symbol $\sum_{k=1}^{\infty} a_k$ stands for two things: the sequence of partial sums and also, if the sequence converges, for its limit.

EMPHASIS: Even though a series involves the sum of an infinite number of terms, we never actuallly add an infinite number of terms. Instead, we look the limit of the sequence of partial sums.

In Example 1 we investigated the series $\sum_{k=1}^{\infty} 1/\sqrt{k}$. In this case the sequence of partial sums diverges, becoming arbitrarily large as n increases. We summarize this in the equation $\sum_{k=1}^{\infty} 1/\sqrt{k} = \infty$ and say the series $\sum_{k=1}^{\infty} 1/\sqrt{k}$ diverges, which means the sequence of partial sums $\{S_n\} = \left\{1/\sqrt{1} + 1/\sqrt{2} + \cdots + 1/\sqrt{n}\right\}$ diverges.

In Example 2 we investigated the series $\sum_{k=1}^{\infty} 0.6^{k-1}$, that is, the sequence of partial sums 1, $1+0.6$, $1+0.6+0.6^2$, $\ldots$, $1+0.6+0.6^2+\cdots+(0.6)^{n-1}$. The sequences converges to 2.5, and we write

$$\sum_{k=1}^{\infty} (0.6)^{k-1} = 2.5.$$

This says that the series $\sum_{k=1}^{\infty} (0.6)^{k-1}$ converges, and the limit of the sequence of partial sums is 2.5. We also say the sum of the series $\sum_{k=1}^{\infty} (0.6)^{k-1}$ is 2.5.

Just as a sequence need not start with a_1, so can a series start with any term, such as a_0 or a_7 or a_k. In these cases we might write $\sum_{j=0}^{\infty} a_j$, $\sum_{i=7}^{\infty} a_i$, or $\sum_{n=k}^{\infty} a_n$.

There is nothing special about the index for a series. The most common indices are n, k, j, and i. For instance, $\sum_{i=0}^{\infty} a_i$ and $\sum_{n=0}^{\infty} a_n$ both represent the same series, just as the definite integrals $\int_0^3 f(x)\,dx$ and $\int_0^3 f(t)\,dt$ are equal.

The n^{th}-Term Test for Divergence

Remember that a series is really a sequence of partial sums: $\{S_n\}$ where $S_n = \sum_{k=1}^{n} a_k$. The first term is $S_1 = a_1$. Then each successive partial sum can be obtained recursively from the previous partial sum: $S_2 = S_1 + a_2$, $S_3 = S_2 + a_3$, $\ldots$, $S_n = S_{n-1} + a_n$, $\ldots$. This is the basis for the following theorem.

Theorem 11.1.1: n^{th}-Term Test for Divergence

If $\lim_{n\to\infty} a_n \neq 0$, then the series $a_1 + a_2 + \cdots + a_n + \cdots$ diverges. (The same conclusion holds if $\{a_n\}$ has no limit.)

Proof of Theorem 11.1.1

Let $\{a_k\}$ be a sequence that does not converge to 0. The recursion $S_n = S_{n-1} + a_n$ implies that

$$a_n = S_n - S_{n-1} \qquad \text{for } n \geq 2.$$

Assume that the series $\{S_n\}$ converges and has limit S. Then $\lim_{n\to\infty} S_n = S$.

We also have $\lim_{n\to\infty} S_{n-1} = S$ because S_{n-1} runs through the same sums as S_n. Thus

$$\begin{aligned}
\lim_{n\to\infty} a_n &= \lim_{n\to\infty} (S_n - S_{n-1}) \\
&= \lim_{n\to\infty} S_n - \lim_{n\to\infty} S_{n-1} \\
&= S - S \\
&= 0.
\end{aligned}$$

This contradicts the hypothesis that $\{a_k\}$ does not converge to 0, so the assumption that the series $\{S_n\}$ converges must be false. Since the series does not converge, it must diverge.

There are several ways to restate this theorem. For instance it says, if the sequence of partial sums associated with the sequence $\{a_k\}$ converges, then a_k must approach 0. It also says if the a_k do not approach 0 there is no hope that the associated series converges.

However, if $a_k \to 0$ as $k \to \infty$, anything can happen. The associated series may diverge, as in Example 1, or converge as in Example 2.

Observation 11.1.2:

The only possible outcomes of the n^{th}-Term Test for Divergence are that a series diverges or that the test was inconclusive. This test can never be used to conclude a series converges.

For a series $\sum_{k=1}^{\infty} a_k$ with positive summands ($a_k > 0$) to converge, a_k must approach 0 much faster than $1/\sqrt{k}$ approaches 0 as $k \to \infty$. In the next example, the summand $1/k$ is much smaller than $1/\sqrt{k}$. Even so, it does not approach 0 fast enough for the associated series to converge. This series was given the name **harmonic series** by the Greeks because of the role $1/k$ plays in music theory. The argument that it diverges is due to the French mathematician Nicolas of Oresme, who presented it about the year 1360. (Nicole Oresme, 1323–1382, was one of the most influential philosophers of the Middle Ages.)

EXAMPLE 3. Show that the harmonic series $\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n} + \cdots$ diverges.

SOLUTION We collect the summands in longer and longer groups. Except for the first two terms, each group contains twice the number of summands as its predecessor:

$$1 + \underbrace{\frac{1}{2}}_{\text{1 term}} + \underbrace{\frac{1}{3} + \frac{1}{4}}_{\text{2 terms}} + \underbrace{\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}}_{\text{4 terms}} + \underbrace{\frac{1}{9} + \frac{1}{10} + \cdots + \frac{1}{16}}_{\text{8 terms}} + \cdots.$$

The sum of the terms in each group is at least $1/2$:

$$\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

and

$$\frac{1}{9} + \frac{1}{10} + \cdots + \frac{1}{16} > \frac{1}{16} + \frac{1}{16} + \cdots + \frac{1}{16} = \frac{8}{16} = \frac{1}{2},$$

and so on. Since the repeated addition of $1/2$'s produces sums as large as we please, the series diverges. ◆

However, the next example describes a special type of series that does converge. Recall that a sequence of the form a, ar, ar^2, $\ldots$, ar^{k-1}, $\ldots$, is called a **geometric sequence**. In Section 1.4 a short formula for the sum of the first n terms was found,

$$a + ar + ar^2 + \cdots + ar^{n-1} = a\frac{1 - r^k}{1 - r}.$$

Definition: *Geometric Series*

Let a and r be real numbers. The series $a + ar + ar^2 + \cdots + ar^{k-1} + \cdots$ is called the **geometric series** with initial term a and ratio r.

The series in Example 2 is a geometric series with initial term 1 and ratio 0.6.

Theorem 11.1.3: Convergence of a Geometric Series

If $-1 < r < 1$, *the geometric series* $a + ar + ar^2 + \cdots + ar^{k-1} + \cdots$ *converges to* $\dfrac{a}{1-r}$.

If $|r| \geq 1$, *the geometric series* $\displaystyle\sum_{k=0}^{\infty} ar^k$ *diverges.*

Proof of Theorem 11.1.3

The sum of the first n terms of the series is, since $r \neq 1$,

$$S_n = a + ar + \cdots + ar^{n-1} = a\frac{1-r^n}{1-r}.$$

By Theorem 10.1.2 in Section 10.1, since $-1 < r < 1$, $\lim_{n\to\infty} r^n = 0$ and

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \frac{a(1-r^n)}{1-r} = \frac{a(1-\lim_{n\to\infty} r^n)}{1-r} = \frac{a}{1-r}.$$

If $|r| \geq 1$, then $\lim_{n\to\infty} r^n \neq 0$ so, by Theorem 11.1.1, the sequence of partial sums cannot converge. ●

In short, when $|r| < 1$, $r^n \to 0$ so fast that $S_n = \sum_{k=1}^{n} ar^k$ does not get arbitrarily large. Instead, as $n \to \infty$, the n^{th} partial sum S_n approaches $a/(1-r)$.

EXAMPLE 4. Discuss the convergence or divergence of the following series:

(a) $\displaystyle\sum_{k=1}^{\infty} \cos(k\pi)$ (b) $\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^k}{k}$ (c) $\displaystyle\sum_{k=1}^{\infty} 3^k$ (d) $\displaystyle\sum_{k=1}^{\infty} 5(0.9)^k$

SOLUTION

(a) When k is even, $\cos(k\pi) = 1$. When k is odd, $\cos(k\pi) = -1$. The sequence $\{\cos(k\pi)\}$ is $-1, 1, -1, 1, -1, \ldots$. It does not approach 0. Thus this series diverges. To put it another way, the sequence of partial sums, $S_n = \cos(\pi) + \cos(2\pi) + \cdots + \cos(n\pi)$ does not have a limit.

(b) The series is $-1/1 + 1/2 - 1/3 + 1/4 + \cdots$. Because the k^{th} term approaches 0, there is a chance that it may converge. Also, because the negative terms shrink the sums, there is more evidence it may converge. In Section 11.5 we will show that indeed it does converge, but we do not yet have the information needed to decide whether it converges or diverges.

(c) The terms of this series are $3, 3^2, 3^3, \ldots$. Because the k^{th} term does not approach 0 the associated sequence of partial sums, with $S_n = 3 + 3^2 + \cdots + 3^n$, diverges. Therefore the series $\sum_{k=1}^{\infty} 3^k$ diverges.

(d) The series $\sum_{k=0}^{\infty} 5(0.9)^k$ is a geometric series with first term 5 and ratio 0.9. Because $|0.9|$ is less than 1, the series converges and has sum $5/(1-0.9) = 50$. ◆

An Important Example

Series in the form $\sum_{k=0}^{\infty} a^k/k!$, where a is a constant, will appear several times in Chapter 12. Example 5 treats the case when $a = 2$.

EXAMPLE 5. Does $\displaystyle\sum_{k=0}^{\infty} \frac{2^k}{k!}$ converge or diverge?

SOLUTION It was shown in Section 10.1 that $a^k/k! \to 0$ as $k \to \infty$. There is therefore the possibility that the series $\sum_{k=0}^{\infty} 2^k/k!$ converges, but we are not yet sure that it does. We must see how S_n behaves as $n \to \infty$.

RECALL: Theorem 10.1.1 in Section 10.1 tells us that an increasing sequence that is bounded above must converge.

In this example, since every summand is positive, the sequence of partial sums is an increasing sequence: $S_n = S_{n-1} + 2^n/n! > S_{n-1}$. Moreover, by the reasoning in Section 10.1, for $k \geq 2$,

$$\frac{2^k}{k!} < \frac{2}{1} \cdot \frac{2}{2} \cdot \left(\frac{2}{3}\right)^{k-2}.$$

Thus, for $n \geq 2$,

$$S_n = \sum_{k=0}^{n} a_k = a_0 + a_1 + \sum_{k=2}^{n} a_k < 1 + 2 + \sum_{k=2}^{n} 2\left(\frac{2}{3}\right)^{k-2}.$$

Note that the last term, the (finite) sum from $k = 2$ to $k = n$ is less than the (infinite) geometric series with first term 2 and ratio $2/3$, we conclude that

$$S_n < 1 + 2 + \sum_{k=2}^{n} 2\left(\frac{2}{3}\right)^{k-2} < 1 + 2 + \sum_{k=2}^{\infty} 2\left(\frac{2}{3}\right)^{k-2} = 1 + 2 + \frac{2}{1 - \frac{2}{3}} = 1 + 2 + 6 = 9.$$

Because the S_n stay below 9, the sequence of partial sums is bounded. This, combined with the earlier observation that the sequence of partial sums is increasing, allows us to conclude that sequence of partial sums must converge Consequently, $\sum_{k=1}^{\infty} 2^k/k!$ converges (to a number less that or equal to 9). ♦

The same idea can be used to prove that $\sum_{k=0}^{\infty} a^k/k!$, for any positive number a, converges. Example 5 illustrates two general principles.

First, we can disregard a finite number of terms when deciding whether a series is convergent or divergent. If a finite number of terms are deleted from a series and what is left converges, then the series converges. A partial sum, $a_1 + a_2 + \cdots + a_n$, does not influence convergence or divergence. What does matter is the tail end, $a_{n+1} + a_{n+2} + \cdots$. The sum of the series is the sum of any tail end plus the sum of the corresponding partial sum:

$$\sum_{k=1}^{\infty} a_k = \underbrace{\sum_{k=1}^{m} a_k}_{\text{front end}} + \underbrace{\sum_{k=m+1}^{\infty} a_k}_{\text{tail end}}.$$

Second, suppose that $\sum_{k=1}^{\infty} p_k$ is a series with positive terms and there is a number B such that every partial sum $S_1 = p_1$, $S_2 = p_1 + p_2, \ldots, S_n = p_1 + p_2 + \cdots + p_n$, is less than or equal to B. These partial sums then have a limit L less than or equal to B. Thus $\sum_{k=1}^{\infty} p_k$ is convergent and its sum is less than or equal to B. This will be useful in establishing the convergence of a series of nonnegative terms, though it does not yield the sum of the series.

Summary

Given a sequence, $\{a_k\}$, we can obtain from it another sequence formed by its partial sums: $\{S_n\}$ where $S_n = \sum_{k=1}^{n} a_k$. The sequence formed that way is called a series. There is no general way to decide whether the sequence of partial sums has a limit as $n \to \infty$. However, if it does converge, then a_n must approach 0. It is important to remember that the converse of this fact is not true: $\lim_{n\to\infty} a_n = 0$ does not imply anything about the convergence of the series $\sum_{k=1}^{\infty} a_k$. But, if $\lim_{n\to\infty} a_n$ is not zero, then the series $\sum_{k=1}^{\infty} a_k$ must diverge.

The series $\sum_{k=1}^{\infty} 1/\sqrt{k}$ diverges. So does the harmonic series, $\sum_{k=1}^{\infty} 1/k$.

A geometric series $\sum_{k=0}^{\infty} ar^k$ converges if and only if $|r| < 1$; in these cases the sum of the series is $a/(1-r)$.

EXERCISES for Section 11.1

1. Define series, using no mathematical symbols.

2. Is every series a sequence? Is every sequence a series? Explain your answers.

3. This exercise concerns the geometric series $\sum_{k=1}^{\infty} 5\left(\frac{-1}{2}\right)^k$.

(a) Express the fourth term of the series as a decimal.
(b) Express the fourth partial sum of the series as a decimal.
(c) Find $\lim_{k\to\infty} 5\left(\frac{-1}{2}\right)^k$.
(d) Find the limit as $n \to \infty$ of the n^{th} partial sum of the series.
(e) Does the series converge? If so, what is its sum?

4. This exercise concerns the geometric series $\sum_{k=1}^{\infty} 3\left(\frac{1}{10}\right)^k$.

(a) Express the third term of the series as a decimal.
(b) Express the third partial sum of the series as a decimal.
(c) Find $\lim_{k\to\infty} 3\left(\frac{1}{10}\right)^k$.
(d) Find the limit as $n \to \infty$ of the n^{th} partial sum of the series.
(e) Does the series converge? If so, what is its sum?

5. (a) What is the total distance traveled, both up and down, by the ball in the opening paragraph of this section?
(b) What is the total time the ball spends falling if it falls $16t^2$ feet in t seconds?

6. Example 1 shows that $\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}}$ diverges and Example 3 that $\sum_{k=1}^{\infty} \frac{1}{k}$ diverges.

(a) Does the divergence of $\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}}$ imply the divergence of $\sum_{k=1}^{\infty} \frac{1}{k}$?
(b) Does the divergence of $\sum_{k=1}^{\infty} \frac{1}{k}$ imply the divergence of $\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}}$?

7. Does the series $\sum_{k=1}^{\infty} \frac{\sin^2(k)}{2^k}$ converge or diverge?

8. (a) Check that $\frac{1}{1\cdot 2} = \frac{1}{1} - \frac{1}{2}$, $\frac{1}{2\cdot 3} = \frac{1}{2} - \frac{1}{3}$, and $\frac{1}{3\cdot 4} = \frac{1}{3} - \frac{1}{4}$.
(b) Does the pattern in (a) continue?
(c) Show that $\sum_{k=1}^{\infty} \frac{1}{n(n+1)}$ converges.
(d) Find the sum of the series in (c).

9. Using the result in Exercise 8, show that $\sum_{k=1}^{\infty} \frac{1}{k^2}$ converges.

In Section 11.2 this sum is shown to be $\frac{\pi^2}{6}$.

In Exercises 10 to 17 determine whether the geometric series converges. If it does, find its sum.

10. $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \left(\frac{1}{2}\right)^{k-1} + \cdots$

11. $1 - \frac{1}{3} + \frac{1}{9} - \frac{1}{27} + \cdots + \left(\frac{-1}{3}\right)^{k-1} + \cdots$

12. $\frac{-3}{2} + \frac{3}{4} - \frac{3}{8} + \cdots + \frac{3}{(-2)^k} + \cdots$

13. $\sum_{k=1}^{\infty} 10^{-k}$

14. $\sum_{k=1}^{\infty} 10^k$

15. $\sum_{k=1}^{\infty} 5(0.99)^k$

16. $\sum_{k=1}^{\infty} 7(-1.01)^k$

17. $\sum_{k=1}^{\infty} 4\left(\frac{2}{3}\right)^k$

Exercises 18 to 21 concern a general series $\sum_{k=1}^{\infty} a_k$ and the sequence of its partial sums $\{S_n\}$. In each problem, select all of the statements that are true. These problems are based on suggestions by James T. Vance Jr.

18. Suppose that $a_n \to 0$ as $n \to \infty$.
(a) The series converges.
(b) The series diverges.
(c) There is not enough information to decide whether the series diverges or converges.
(d) More information is needed to determine the sum of the series.
(e) $S_n \to 0$ as $n \to \infty$.
(f) $\sum_{k=1}^{\infty} a_k = 0$.

19. Suppose that $a_k \to 6$ as $k \to \infty$.
(a) The series converges.
(b) The series diverges.
(c) There is not enough information to decide whether the series diverges or converges.
(d) More information is needed to determine the sum of the series.
(e) $S_n \to 0$ as $n \to \infty$.
(f) $\sum_{k=1}^{\infty} a_k = 6$.

20. Suppose that $S_n \to 3$ as $n \to \infty$.
(a) The series converges.
(b) The series diverges.
(c) There is not enough information to decide whether the series diverges or converges.
(d) More information is needed to determine the sum of the series.
(e) The sum of the series is 3.
(f) $\sum_{k=1}^{\infty} a_k = 3$.
(g) $\lim_{k\to\infty} a_k = 0$.

21. Suppose the sequence of terms $\{a_n\}$ produces a sequence of partial sums $\{S_n\}$ where $S_n = \frac{n}{n+1}$.
(a) The series converges.
(b) The series diverges.
(c) The sequence of terms converges.
(d) The sequence of terms diverges.
(e) $\sum_{k=1}^{\infty} a_k = 1$.
(f) $a_k = \frac{1}{k(k+1)}$.
(g) The series is a geometric series.

In Exercises 22 to 27 determine whether the series converges or diverges. If a series converges find its sum.

22. $-5+5-5+5-\cdots+(-1)^k 5+\cdots$

23. $\sum_{k=1}^{\infty} \left(1+\frac{1}{k}\right)^{-k}$

24. $\sum_{k=1}^{\infty} \frac{2}{k}$

25. $\sum_{k=1}^{\infty} \frac{k}{2k+1}$

26. $\sum_{k=1}^{\infty} 6\left(\frac{4}{5}\right)^k$

27. $\sum_{k=1}^{\infty} 100\left(\frac{-8}{9}\right)^k$

Exercises 28 to 30 are related. In each problem a geometric series is used to convert a repeating fraction to an equivalent fraction.

28. The repeating decimal $3.171717\ldots$, where the 17's continue forever, can be viewed as 3 plus a geometric series: $3+\frac{17}{100}+\frac{17}{100^2}+\frac{17}{100^3}+\cdots$. Using the formula for the sum of a geometric series, write the decimal as a fraction.

29. Write the repeating decimal $0.3333\cdots$ as a fraction.

30. Write the repeating decimal $4.1256256256\ldots$ (with 256 repeating) as a fraction.

---

31. This is a quote from an economics text: "The present value of the land, if a new crop is planted at time t, $2t$, $3t$, etc., is $P = g(t)e^{-rt} + g(t)e^{-2rt} + g(t)e^{-3rt} + \cdots$. By the formula for the sum of a geometric series, $P = \frac{g(t)e^{-rt}}{1-e^{-rt}}$." Check that the missing step, which simplified the formula for P, is correct.

32. A patient takes A grams of a medicine every 6 hours. The amount of each dose active in the body t hours later is Ae^{-kt} grams, where k is a positive constant and time is measured in hours. See also Exercise 108 in Section 5.S.

(a) Show that immediately after taking the medicine for the n^{th} time, the amount active in the body is

$$S_n = A + Ae^{-6k} + Ae^{-12k} + \cdots + Ae^{-6(n-1)k}.$$

(b) If, as $n \to \infty$, $S_n \to \infty$, the patient would be in danger. Does $S_n \to \infty$? If not, what is $\lim_{n\to\infty} S_n$?

33. Does the series $\sum_{k=1}^{\infty} \frac{1}{\sqrt[3]{k}}$ converge or diverge? Explain.

34. Each sequence $\{a_k\}$ produces another sequence $\{S_n\}$, where $S_n = a_1 + a_2 + \cdots + a_n$. This raises a question. Given a sequence $\{b_n\}$, is there a sequence $\{a_k\}$ such that $b_n = a_1 + a_2 + \cdots + a_n$? Explain your answer.

35. Oresme, around the year 1360, summed the series $\sum_{k=1}^{\infty} \frac{k}{2^k}$ by drawing the endless staircase shown in Figure 11.1.2, in which each stair after the first has width 1 and is half as high as the stair immediately to its left.

(a) By looking at the staircase in two ways, show that

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \cdots.$$

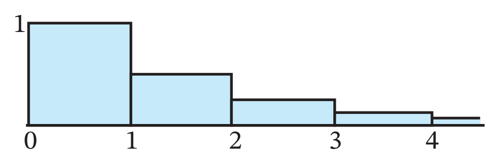


Figure 11.1.2

(b) Use (a) to sum $\sum_{k=1}^{\infty} \frac{k}{2^k}$.

(c) Use the same idea to find $\sum_{k=1}^{\infty} kp^k$, when $0 < p < 1$.

The formula for the sum of a finite geometric progression, $\sum_{k=0}^{n-1} ar^k = a\frac{1-r^n}{1-r}$ for $r \neq 1$, implies that $\frac{1}{1-t}$ can be written as $1 + t + t^2 + t^3 + \cdots t^{n-1} + \frac{t^n}{1-t}$. Replacing t with $-t$, we obtain an alternate representation for $\frac{1}{1+t}$:

$$\frac{1}{1+t} = 1 - t + t^2 - t^3 + \cdots (-1)^{n-1} t^{n-1} + \frac{(-1)^n t^n}{1+t} \qquad (\text{for } x \neq -1). \tag{11.1.2}$$

With the aid of (11.1.2) and integration, Exercises 36 and 37 obtain series that converge to $\ln(1+x)$ and $\arctan(x)$.

36. (a) Integrate both sides of (11.1.2) over the interval from 0 to x, $x > 0$, to show that

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + \frac{(-1)^{n-1}x^n}{n} + (-1)^n \int_0^x \frac{t^n}{1+t}\, dt.$$

(b) Use the fact that $\frac{1}{1+t} \leq 1$ for $t \geq 0$ to show that for $0 \leq x \leq 1$, $\int_0^x \frac{t^n}{1+t}\, dt$ approaches 0 as n increases.

(c) Show that for x in $[0,1]$, $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + \frac{(-1)^{k-1}x^k}{k} + \cdots$.

37. Replacing t in (11.1.2) by t^2 shows that $\frac{1}{1+t^2} = 1 - t^2 + t^4 - t^6 + \cdots + (-1)^{n-1} t^{2n-2} + \frac{(-1)^n t^{2n}}{1+t^2}$ for all t.

(a) For $0 \le x \le 1$, integrate both sides of the above equation over $[0, x]$ to get

$$\arctan(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + (-1)^{n-1} \frac{x^{2n-1}}{2n-1} + (-1)^n \int_0^x \frac{t^{2n}}{1+t^2}\, dt. \tag{11.1.3}$$

(b) Show that for fixed x, $0 \le x \le 1$, the integral in (11.1.3) approaches 0 as $n \to \infty$.

(c) Use the polynomial in (a) with $n = 5$ to estimate $\arctan(1)$.

(d) Use the result in (c) to estimate π.

(e) Show that for x in $[0, 1]$, $\arctan(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + (-1)^k \frac{x^{2k+1}}{2k+1} + \cdots$.

Exercise 38 provides background for the discussion in the historical note below about Newton's calculation of the area under a hyperbola to more than fifty decimal places. See also Exercises 29 and 30 in Section 6.S.

38. Let c be a positive constant.

(a) Show that the area under the curve $y = \frac{1}{1+x}$ above the interval $[0, c]$ is $-\sum_{k=1}^{\infty} \frac{(-c)^k}{k}$.

(b) Show that the area under the curve $y = \frac{1}{1+x}$ above the interval $[-c, 0]$ is $\sum_{k=1}^{\infty} \frac{c^k}{k}$.

Historical Note: Isaac Newton, Area, Logarithms, and Series

The area under the curve $y = \frac{1}{1+x}$ above the interval $[0, c]$, when c is positive, is approximated by $c - c^2/2 + \cdots \pm c^n/n$. When c is negative, the area above the interval $[c, 0]$ is approximated by $-c + c^2/2 - c^3/3 + \cdots + c^n/n$.

When he was a student, Isaac Newton calculated the area under the curve $y = \frac{1}{1+x}$ and above the intervals $[-0.1, 0]$ $(c = -0.1)$ and $[0, 0.1]$ $(c = 0.1)$ to 53 decimal places. (See Figure 11.1.3; provided courtesy of the Cambridge University Library.)

The first area equals

$$(0.1) - \frac{(0.1)^2}{2} + \frac{(0.1)^3}{3} - \frac{(0.1)^4}{4} + \frac{(0.1)^5}{5} - \cdots.$$

The area above $[0.9, 1]$ is

$$(0.1) + \frac{(0.1)^2}{2} + \frac{(0.1)^3}{3} + \frac{(0.1)^4}{4} + \frac{(0.1)^5}{5} + \cdots.$$

The manuscript shows Newton's orderly calculations, done with a quill pen, not with a computer. Note that he found - and corrected - an error in the value of $\frac{1}{23}(0.1)^{23}$.

We recognize that $\int_0^c \frac{dx}{1+x}$ is not only related to the area under $y = \frac{1}{1+x}$ above $[0, c]$, but also to $\ln(1+c)$. This second point of view was not available at his time.

See Exercises 29 and 30 in Section 6.S and Exercise 38 in this section.

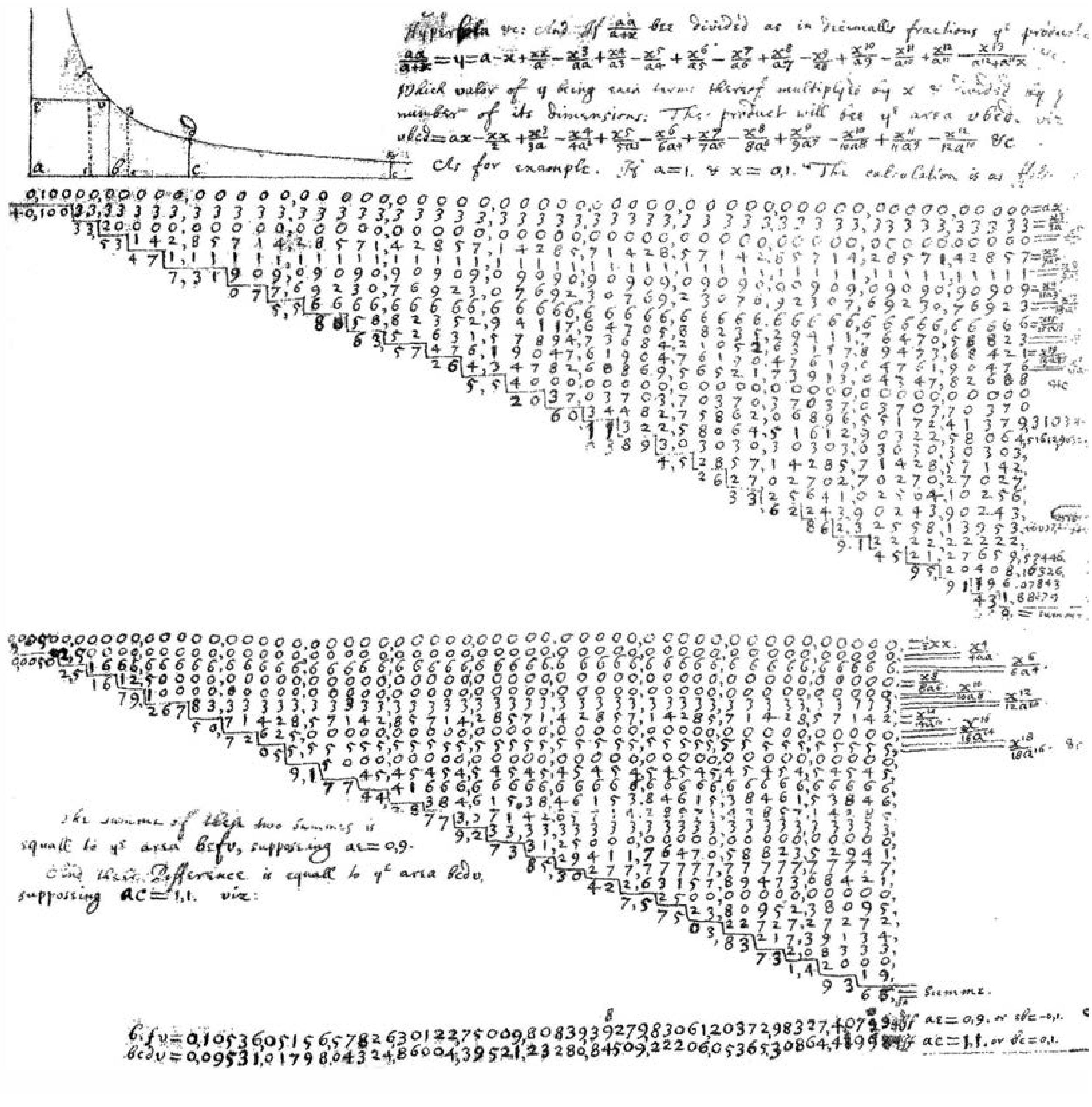


Figure 11.1.3

11.2 The Integral Test

In this section we use integrals of the form $\int_a^\infty f(x)\,dx$ to establish convergence or divergence of series whose terms are positive and decreasing. We also analyze the error when we use a partial sum to estimate the sum of such a series.

The Integral Test

Let $f(x)$ be a decreasing positive function. We obtain a sequence from $f(x)$ by defining a_n to be $f(n)$. For instance, the sequence $1/1, 1/2, 1/3, \ldots, 1/n, \ldots$ is obtained from the function $f(x) = 1/x$. It turns out that the convergence (or divergence) of the series $\sum_{k=1}^\infty a_k$ is closely connected with the convergence (or divergence) of the improper integral $\int_1^\infty f(x)\,dx$. The connection is described in the following theorem:

Theorem 11.2.1: Integral Test

Let $f(x)$ be a continuous decreasing function such that $f(x) > 0$ for $x \geq 1$. Let $a_n = f(n)$ for each positive integer n. Then

(a) *If $\int_1^\infty f(x)\,dx$ is convergent, then so is the series $\sum_{k=1}^\infty a_k$.*

(b) *If $\int_1^\infty f(x)\,dx$ is divergent, then so is the series $\sum_{k=1}^\infty a_k$.*

REMINDER: *It is important to remember that when the Integral Test tells us that a series converges, it does not say the series converges to the value of the improper integral.*

Proof of Theorem 11.2.1

The two parts of Figure 11.2.1 are the keys to the proof of the two parts of Theorem 11.2.1.

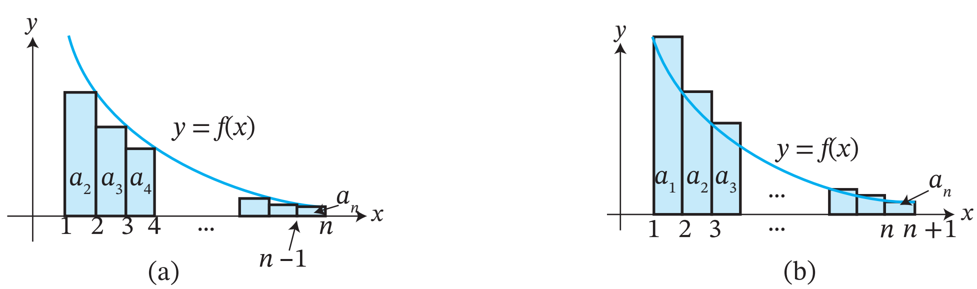


Figure 11.2.1

In Figure 11.2.1(a) the rectangles lie below the curve $y = f(x)$. Each rectangle has width 1. The shaded region has area $f(2)+f(3)+\cdots+f(n) = a_2+a_3+\cdots+a_n$. This staircase area lies completely under the curve, which translates into the inequality $a_2 + a_3 + \cdots + a_n < \int_1^n f(x)\,dx$, and therefore

$$a_1 + a_2 + a_3 + \cdots + a_n < a_1 + \int_1^n f(x)\,dx. \tag{11.2.1}$$

If $\int_1^\infty f(x)\,dx$ is convergent with value S then $a_1 + a_2 + \cdots + a_n < a_1 + S$. Since the partial sums of the series $\sum_{k=1}^\infty a_k$ are all bounded by $a_1 + S$, the series $\sum_{k=1}^\infty a_k$ converges and its sum is less than or equal to $a_1 + S$.

In a similar fashion, the shaded region in Figure 11.2.1(b) has area $f(1)+f(2)+\cdots+f(n)=a_1+a_2+\cdots+a_n$. As each of these rectangles extends above the curve,

$$a_1+a_2+\cdots+a_n > \int_1^{n+1} f(x)\,dx. \tag{11.2.2}$$

If follows that if $\int_1^\infty f(x)\,dx$ diverges, then so must the series $\sum_{k=1}^\infty a_k$.

EXAMPLE 1. Use the integral test to determine the convergence or divergence of

(a) $\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{k}+\cdots=\sum_{k=1}^\infty \frac{1}{k}$

(b) $\frac{1}{1^{1.01}}+\frac{1}{2^{1.01}}+\cdots+\frac{1}{k^{1.01}}+\cdots=\sum_{k=1}^\infty \frac{1}{k^{1.01}}$

Even though the graphs of $y=\frac{1}{x}$ and $y=\frac{1}{x^{1.01}}$ are very similar, $\int \frac{dx}{x}$ and $\int \frac{dx}{x^{1.01}}$ behave very differently.

SOLUTION

(a) This is the harmonic series, which was shown to diverge in Example 3 in Section 11.1. To apply the integral test to it let $f(x)=1/x$, a decreasing positive function for $x>0$, so that $a_k=f(k)=1/k$. Then

$$\begin{aligned}
\int_1^\infty \frac{dx}{x} &= \lim_{b\to\infty}\int_1^b \frac{dx}{x} && (\text{ evaluating improper integral })\\
&= \lim_{b\to\infty} (\ln(b)-\ln(1)) && (\text{ FTC I })\\
&= \infty
\end{aligned}$$

By Theorem 11.1.3(b), since $\int_1^\infty dx/x$ is divergent, so is the series $\sum_{i=1}^\infty 1/n$.

(b) Let $f(x)=1/x^{1.01}$, which is a decreasing positive function. Then $a_k=f(k)=1/k^{1.01}$. Now

$$\begin{aligned}
\int_1^\infty \frac{dx}{x^{1.01}} &= \lim_{b\to\infty}\int_1^b \frac{dx}{x^{1.01}} && (\text{ evaluating improper integral })\\
&= \lim_{b\to\infty} \left.\frac{x^{-1.01+1}}{-1.01+1}\right|_1^b && (\text{ FTC I })\\
&= \lim_{b\to\infty} \left.\frac{x^{-0.01}}{-0.01}\right|_1^b\\
&= \lim_{b\to\infty}\left(\frac{b^{-0.01}}{-0.01}-\frac{1^{-0.01}}{-0.01}\right)\\
&= 0-(-100)\\
&= 100.
\end{aligned}$$

By Theorem 11.1.3(a), since $\int_1^\infty dx/x^{1.01}$ is convergent, so is $\sum_{k=1}^\infty 1/k^{0.01}$.
And, while the Integral Test does not provide the actual value of the convergent infinite series, (11.2.1) in its proof does tell us that $\sum_1^\infty 1/k^{1.01}$ is less than $a_1+100=101$. ◆

The argument in Example 1 extends to a family of series known as p-series.

Definition: *p-series*

For a positive number p, the series $\sum_{k=1}^\infty \frac{1}{k^p}$ is called a ***p*-series**.

For example, when $p=1$ we obtain the harmonic series $\sum_{k=1}^\infty 1/k$ and for $p=1.01$, the series $\sum_{k=1}^\infty 1/k^{1.01}$. An argument similar to those in Example 1 establishes the following theorem.

Theorem 11.2.2

If $0 < p \le 1$, the p-series $\sum_{k=1}^{\infty} 1/k^p$ diverges. If $p > 1$, the p-series $\sum_{k=1}^{\infty} 1/k^p$ converges.

There is a p-series for each positive number p. If p is negative, $\lim_{k\to\infty} 1/k^p = \infty$, and the series diverges because its terms do not approach 0 as the index increases without bound. For example, if $p = -3$,

$$\sum_{k=1}^{\infty} \frac{1}{k^p} = \sum_{k=1}^{\infty} k^3,$$

and $k^3 \to \infty$.

Controlling the Error

Partial sum = front end;
Error = tail end

When we use a front end of a series (a partial sum) to estimate the sum of the whole series there will be an error, namely, the sum of the corresponding tail end. For the sum of a front end to be a good estimate of the sum of the whole series, we must be sure that the sum of the tail end is small. Otherwise we would be like the carpenter who measures a board as 5 feet long with an error of perhaps as much as 5 feet. We wish to be sure that the sum of the tail end is small.

Let S_n be the sum of the first n terms of a convergent series $\sum_{k=1}^{\infty} a_k$ whose sum is S. Define R_n to be the difference between the infinite series, S, and its n-term partial sum, S_n:

$$R_n = S - S_n = a_{n+1} + a_{n+2} + a_{n+3} + \cdots$$

The difference R_n is called the **remainder** or **error** in using the sum of the first n terms to approximate the sum of the series. That is,

$$\underbrace{a_1 + a_2 + \cdots + a_n}_{\text{partial sum } S_n} + \underbrace{a_{n+1} + a_{n+2} + \cdots}_{\text{tail end } R_n} = \underbrace{a_1 + a_2 + \cdots + a_n + a_{n+1} + a_{n+2} + \cdots}_{\text{sum of series } S},$$

or more briefly,

$$S_n + R_n = S.$$

For a series whose terms are positive and decreasing, we can use an improper integral to estimate the error by comparing a staircase of rectangles with the area under a curve.

Recall that $f(x)$ is a continuous decreasing positive function. The error in using $S_n = f(1) + f(2) + \cdots + f(n) = \sum_{k=1}^{n} f(k)$ to approximate $\sum_{k=1}^{\infty} f(k)$ is $\sum_{k=n+1}^{\infty} f(k)$. It is the area of the endless staircase of rectangles shown in Figure 11.2.2(a). Comparing the rectangles with the region under the curve $y = f(x)$ and above the unbounded interval $[n+1, \infty)$, we conclude that

$$R_n = a_{n+1} + a_{n+2} + \cdots = f(n+1) + f(n+2) + \cdots > \int_{n+1}^{\infty} f(x)\, dx,$$

which gives a lower estimate of the error.

The staircase in Figure 11.2.2(b), which lies below the graph of $y = f(x)$ and above the unbounded interval $[n, \infty)$, gives an upper estimate of the error. It shows that

$$R_n = a_{n+1} + a_{n+2} + \cdots = f(n+1) + f(n+2) + \cdots < \int_{n}^{\infty} f(x)\, dx.$$

Combining these lower and upper bounds for R_n proves the following theorem.

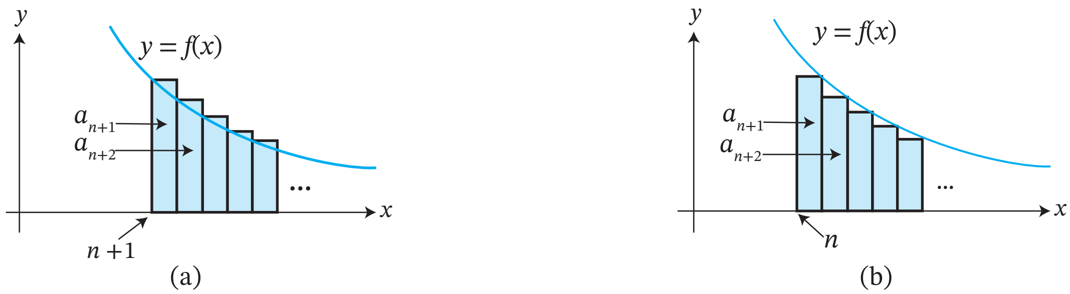


Figure 11.2.2

Theorem 11.2.3: Error Bound for Partial Sums

Let $f(x)$ be a continuous decreasing positive function such that $\int_1^\infty f(x)\,dx$ is convergent. Then the error R_n in using the partial sum $f(1)+f(2)+\cdots+f(n)$ to estimate the series $\sum_{k=1}^\infty f(k)$ satisfies the inequality

$$\int_{n+1}^\infty f(x)\,dx < R_n < \int_n^\infty f(x)\,dx. \tag{11.2.3}$$

EXAMPLE 2. The first five terms of the series $\frac{1}{1^2}+\frac{1}{2^2}+\cdots+\frac{1}{n^2}+\cdots$ are used to estimate the sum of the series.

(a) Put upper and lower bounds on the error in using just those terms.

(b) Use the bounds in (a) to estimate $\sum_{k=1}^\infty \frac{1}{k^2}$.

SOLUTION The series with terms $a_k = 1/k^2$ is the p-series with $p=2$. Since $p>1$, it converges. Also, the function $f(x) = 1/x^2$ is continuous, decreasing, and positive for $x \ge 1$.

(a) By Theorem 11.2.3, specifically (11.2.3),

$$\int_6^\infty \frac{dx}{x^2} < R_5 < \int_5^\infty \frac{dx}{x^2}.$$

Then, evaluating these two (convergent) improper integrals:

$$\int_5^\infty \frac{dx}{x^2} = \left.\frac{-1}{x}\right|_5^\infty = 0 - \left(\frac{-1}{5}\right) = \frac{1}{5}$$

and

$$\int_6^\infty \frac{dx}{x^2} = \frac{1}{6},$$

we arrive at the following lower and upper bounds on the error between the partial sum with five terms and the full infinite series:

$$\frac{1}{6} < R_5 < \frac{1}{5}.$$

(b) The sum of the first five terms of the series is

$$S_5 = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} \approx 1.463611.$$

GOOD HABIT: Keep more digits than needed until all calculations have been completed. Then, round down lower bounds and round up upper bounds.

Since the sum of the remaining terms (the tail end) is between 1/6 and 1/5, the sum of the series is between 1.463611 + 0.166666 and 1.463611 + 0.2, hence between 1.6302 and 1.6636. ♦

In the 18th century Euler, using complex numbers, proved that $\sum_{k=1}^{\infty} 1/k^2 = \pi^2/6 \approx 1.644934068$.

Estimating a Partial Sum S_n

We continue to assume that there is a continuous, positive, and decreasing function $f(x)$ such that $f(k) = a_k$.

As we can use an (improper) integral to estimate the sum of the tail, we can use a (definite) integral to estimate a partial sum $S_n = a_1 + a_2 + \cdots + a_n$.

In Theorem 11.2.1, we obtained (11.2.1) and (11.2.2), which give

Theorem 11.2.4: Bounds for a Partial Sum of $\sum_{k=1}^{\infty} a_k$

Let $f(x)$ be a continuous decreasing positive function. Then the partial sum $S_n = a_1 + a_2 + \cdots + a_n$ of the series $\sum_{k=1}^{\infty} a_k$ with terms $a_k = f(k)$ for $k = 1, 2, 3, \ldots$, satisfies the following upper and lower bounds:

$$\int_1^{n+1} f(x)\, dx < S_n = a_1 + a_2 + \cdots + a_n < a_1 + \int_1^n f(x)\, dx. \tag{11.2.4}$$

The estimates in (11.2.4) are valid whether the series $\sum_{k=1}^{\infty} a_k$ converges or diverges.

If we can evaluate $\int_1^{n+1} f(x)\, dx$ and $\int_1^n f(x)\, dx$ by the fundamental theorem of calculus, we may use (11.2.4) to put upper and lower bounds on $S_n = \sum_{k=1}^{n} a_k$.

EXAMPLE 3. Use (11.2.4) to estimate the sum of the first million terms of the harmonic series.

SOLUTION By (11.2.4)

$$\int_1^{1,000,001} \frac{dx}{x} < \sum_{k=1}^{1,000,000} \frac{1}{k} < 1 + \int_1^{1,000,000} \frac{dx}{x}.$$

hence, after evaluating these convergent improper integrals yields bounds on the value of the sum of the first one million terms of the harmonic series:

$$\ln(1,000,001) < \sum_{k=1}^{1,000,000} \frac{1}{k} < 1 + \ln(1,000,000).$$

Evaluating the logarithms with a calculator, these bounds

$$13.8155 < \sum_{i=1}^{1,000,000} \frac{1}{i} < 14.8156.$$

♦

Summary

We developed a test for convergence or divergence for series whose terms a_k are of the form $f(k)$ for a continuous, positive, decreasing function $f(x)$. The series converges if $\int_1^\infty f(x)\,dx$ converges and diverges if $\int_1^\infty f(x)\,dx$ diverges.

We also used integrals to estimate a partial sum S_n and the error in using it to estimate the sum of the series.

MEMORY AID: Rather than memorizing formulas, draw the appropriate staircase diagrams. (See Figure 11.2.2.)

We assumed $f(x)$ is decreasing for $x \geq 1$. Theorem 11.2.1 holds if we just assume that $f(x)$ is decreasing from some point on, that is, there is some number a such that $f(x)$ is decreasing for $x \geq a$. The argument uses similar staircase diagrams.

EXERCISES for Section 11.2

Use the integral test in Exercises 1 to 8 to determine whether the series diverges or converges.

1. $\sum_{k=1}^{\infty} \frac{1}{k^{1.1}}$ **2.** $\sum_{k=1}^{\infty} \frac{1}{k^{0.9}}$ **3.** $\sum_{k=1}^{\infty} \frac{k}{k^2+1}$ **4.** $\sum_{k=1}^{\infty} \frac{1}{k^2+1}$

5. $\sum_{k=1}^{\infty} \frac{1}{k\ln(k)}$ **6.** $\sum_{k=1}^{\infty} \frac{1}{k+1{,}000}$ **7.** $\sum_{k=1}^{\infty} \frac{\ln(k)}{k}$ **8.** $\sum_{k=1}^{\infty} \frac{k^3}{e^k}$

Use Theorem 11.2.2 in Exercises 9 to 12 to determine whether the series diverges or converges.

9. $\sum_{k=1}^{\infty} \frac{1}{\sqrt[3]{k}}$ **10.** $\sum_{k=1}^{\infty} \frac{1}{k^3}$ **11.** $\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}}$ **12.** $\sum_{k=1}^{\infty} \frac{1}{k^{0.999}}$

13. (a) Prove that if $p > 1$, the p-series converges. (b) Give upper and lower bounds for the sum.

14. (a) If S_{100} is used to estimate $\sum_{k=1}^{\infty} \frac{1}{k^2}$, what could be said about the error R_{100}?
(b) How large should n be to be sure that R_n is less than 0.0001?

15. (a) If S_{1000} is used to estimate $\sum_{k=1}^{\infty} \frac{1}{k^3}$, what could be said about the error R_{1000}?
(b) How large should n be to be sure that R_n is less than 0.0001?

16. (a) How many terms of $\sum_{k=1}^{\infty} \frac{1}{k^4}$ should be used to be sure that the error in using their sum to estimate the sum of the series is less than 0.0001? (b) Estimate $\sum_{k=1}^{\infty} \frac{1}{k^4}$ to three decimal places.

17. Repeat Exercise 16 for $\sum_{k=1}^{\infty} \frac{1}{k^5}$.

In Exercises 18 to 21, (a) compute the sum of the first four terms of the series to four decimal places, (b) give upper and lower bound on the error R_4, and (c) combine (a) and (b) to estimate the sum..

18. $\sum_{k=1}^{\infty} \frac{1}{k^3}$ **19.** $\sum_{k=1}^{\infty} \frac{1}{k^4}$ **20.** $\sum_{k=1}^{\infty} \frac{1}{k^2+1}$ **21.** $\sum_{k=1}^{\infty} \frac{1}{k^2+k}$

22. Prove that if $p \le 1$, the p-series diverges.

23. What does the integral test say about the geometric series $\sum_{k=1}^{\infty} p^k$ when $0 < p < 1$?

24. Let $f(x)$ be a positive continuous function that is decreasing for $x \ge a$. Let $a_k = f(k)$. Show, in your own words with appropriate diagrams and exposition, why $\int_a^{\infty} f(x)\,dx$ and $\sum_{k=1}^{\infty} a_k$ both converge or both diverge.

25. Show that $\sum_{k=1}^{\infty} k^{10}e^{-k}$ converges. See also Exercise 24.

26. Show that for $n \ge 2$, $2\sqrt{n+1} - 2 < \sum_{k=1}^{n} \frac{1}{\sqrt{k}} < 2\sqrt{n} - 1$.

27. (a) By comparing the sum with integrals, show that $\ln\left(\frac{201}{100}\right) < \frac{1}{100} + \frac{1}{101} + \frac{1}{102} + \cdots + \frac{1}{200} < \ln\left(\frac{200}{99}\right)$.

(b) Find $\lim_{n\to\infty}\left(\frac{1}{n} + \frac{1}{n+1} + \cdots + \frac{1}{2n}\right)$.

28. In Example 1 we showed that the p-series for $p = 1$ diverges but the p-series for $p = 1.01$ converges. This occurs though the terms of the two series resemble each other closely. (For instance, $1/7^{1.01} \approx 0.140104$, $1/7^1 \approx 0.142857$.) What happens to the ratio $\frac{1/k^{1.01}}{1/k}$ as $k \to \infty$?

Exercises 29 and 30 extend our knowledge about series, that is, sums of an infinite number of numbers, to products of an infinite number of numbers.

29. Let $\{a_n\}$ be a sequence of positive numbers. Denote the product $(1+a_1)(1+a_2)\cdots(1+a_n)$ by $\prod_{k=1}^{n}(1+a_k)$.

(a) Show that $\sum_{k=1}^{n} a_k \le \prod_{k=1}^{n}(1+a_k)$. (b) Show that if $\lim_{k\to\infty}\prod_{k=1}^{n}(1+a_k)$ exists, then $\sum_{k=1}^{\infty} a_k$ is convergent.

30. (a) Show that $1 + a_k \le e^{a_k}$. (b) Show that if $\sum_{k=1}^{\infty} a_k$ is convergent, then $\lim_{n\to\infty}\prod_{k=1}^{n}(1+a_k)$ exists.

---

31. Assume that there is only a finite number of primes, $p_1, p_2, \ldots, p_m$.

(a) Show that $\frac{1}{1-\frac{1}{p_k}} = 1 + \frac{1}{p_k} + \frac{1}{p_k^2} + \frac{1}{p_k^3} + \cdots$.

(b) Then show that $\frac{1}{1-\frac{1}{p_1}}\frac{1}{1-\frac{1}{p_2}}\cdots\frac{1}{1-\frac{1}{p_m}} = \sum_{k=1}^{\infty}\frac{1}{k}$.

(c) From (b) obtain a contradiction. This proves that there are an infinite number of prime numbers.

11.3 The Comparison Tests

So far in this chapter three tests for the convergence (or divergence) of a series have been presented. The first, the n^{th}-term test for divergence, asserts that if the n^{th} term of a series does not approach 0, the series diverges. The second was for geometric series. The third, the integral test, applies to certain series of positive terms. In this section two further tests are developed, the comparison and limit comparison tests. We continue to focus our attention on series that have only positive terms.

Comparison Tests

The first test is similar to the comparison test for improper integrals in Section 7.8.

Theorem 11.3.1: Comparison Tests for Convergence and Divergence

(a) *If* $0 \le p_k \le c_k$ *for each positive integer* k *and* $\sum_{k=1}^{\infty} c_k$ *converges, then the series* $\sum_{k=1}^{\infty} p_k$ *also converges.*

(b) *If* $0 \le d_k \le p_k$ *for each positive integer* k *and* $\sum_{k=1}^{\infty} d_k$ *diverges, then the series* $\sum_{k=1}^{\infty} p_k$ *also diverges.*

REMINDER: *The convergence of the comparison series,* $\sum_{k=1}^{\infty} c_k$ *or* $\sum_{k=1}^{\infty} c_k$*, must be shown separately.*

Proof of Theorem 11.3.1

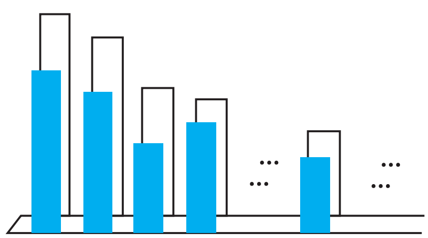
(a) Let the sum of the series $c_1 + c_2 + \cdots$ be C. Let S_n denote the partial sum $p_1 + p_2 + \cdots + p_n$. Then

$$S_n = p_1 + p_2 + \cdots + p_n \le c_1 + c_2 + \cdots + c_n \le C.$$

Since the p_n's are nonnegative, the sequence of partial sums is nondecreasing: $S_1 \le S_2 \le \cdots \le S_n \le \cdots$.
Because each S_n is less than or equal to C, Theorem 10.1.1 of Section 10.1 assures us that the bounded and monotone sequence $\{S_n\}$ converges to L (where L must be less than or equal to C). So $p_1 + p_2 + \cdots$ converges and its sum is less than or equal to the sum $c_1 + c_2 + \cdots$.

(b) The divergence test follows from the convergence test. If the series $p_1 + p_2 + \cdots$ converged, so would the series $d_1 + d_2 + \cdots$, which is assumed to diverge.

Figure 11.3.1 presents the two comparison tests in Theorem 11.3.1 as endless staircases.



a. If the unshaded staircase has finite area, so does the shaded lower staircase.

b. If the shaded staircase has infinite area, so does the unshaded staircase.

Figure 11.3.1

To apply the comparison test to a series of positive terms we have to compare it to a series whose convergence or divergence we already know. What series can we use for comparison? We know the p-series converges for $p > 1$ and diverges for $p \le 1$ and that a geometric series $\sum_{k=1}^{\infty} r^k$ with positive terms converges for $0 \le r < 1$ but diverges for $r \ge 1$. When we multiply a series by a nonzero constant, we do not affect its convergence or divergence.

EXAMPLE 1. Does the series $\displaystyle\sum_{k=1}^{\infty} \frac{k+1}{k+2}\frac{1}{k^2} = \frac{2}{3}\cdot\frac{1}{1^2} + \frac{3}{4}\cdot\frac{1}{2^2} + \frac{4}{5}\cdot\frac{1}{3^2} + \cdots$ converge or diverge?

SOLUTION The series resembles the p-series with $p = 2$: $\sum_{k=1}^{\infty} 1/k^2$, which was shown by the integral test to be convergent. The difference is that each term is multiplied by an additional factor, $(k+1)/(k+2)$. For $k = 1$, $(k+1)/(k+2) = 2/3$, for $k = 2$ it is $3/4$, and as k increases, $(k+1)/(k+2)$ increases towards 1. So, the extra factor in these terms of this series is always between $1/2$ and 1. (It is also true that they are between $2/3$ and 1, but using $1/2$ will be easier for reasons seen in Example 2.) Based on these initial thoughts, we are thinking the series in this example might converge.

Building on these ideas with the goal of using the comparison test to show convergence, Theorem 11.3.1 (a), define $p_k = (k+1)/((k+2)k^2)$ and $c_k = 1/k^2$. (There is no need for d_k since we are only using only Theorem 11.3.1 (a).)

Then, as we just noted, the p_k are all positive, every p_k is smaller than the corresponding c_k, and the comparison series $\sum_{k=1}^{\infty} c_k$ converges. By the comparison test for convergence, the series $\sum_{k=1}^{\infty}(k+1)/(k+2)\cdot 1/k^2$ converges. ♦

RECALL: The comparison test, like the integral test, does not reveal the sum of the series.

EXAMPLE 2. Does the series $\sum_{k=1}^{\infty} \frac{k+1}{k+2}\cdot\frac{1}{k} = \frac{2}{3}\cdot\frac{1}{1}+\frac{3}{4}\cdot\frac{1}{2}+\cdots+\frac{k+1}{k+2}\cdot\frac{1}{k}+\cdots$ converge or diverge?

SOLUTION The only difference from the first example is the change from $1/k^2$ to $1/k$. But, this is a big difference. The p-series with $p=1$ (that is, the harmonic series) diverges. This suggests that the series in this example probably diverges. We hope to be able to use part (b) of Theorem 11.3.1 to confirm this initial thought.

Unfortunately, the terms in this series are less than the terms of the harmonic series $\sum_{k=1}^{\infty} 1/k$. Therefore, the divergence portion of the comparison test, part (b) of Theorem 11.3.1, does not apply when compared with the harmonic series. However, the observation that $(k+1)/(k+2) \geq 1/2$ for all $k \geq 1$suggests the following approach.

To confirm that the series with terms $p_k = (k+1)/(k+2)\cdot 1/k$ diverges we apply part (b) of Theorem 11.3.1 with a comparison series with terms $d_k = 1/(2k)$. Each p_k is positive, each p_k is larger than the corresponding d_k, and the comparison series diverges (because it is a multiple of the harmonic series). Then, as expected, by part (b) of the comparison test, the series $\sum_{k=1}^{\infty}(k+1)/(k+2)\cdot 1/k$ diverges. ♦

Limit Comparison Tests

A variation of Theorem 11.3.1, called the **limit comparison test**, produces a much quicker approach to use when determining the convergence or divergence of a series. After giving its proof, we will revisit Example 2.

Theorem 11.3.2: Limit Comparison Tests for Convergence and Divergence

Let $\sum_{k=1}^{\infty} p_k$ be a series of positive terms.

(a) *Let $\sum_{k=1}^{\infty} c_k$ be a convergent series of positive terms. If $\lim_{k\to\infty}\frac{p_k}{c_k}$ exists, then $\sum_{k=1}^{\infty} p_k$ also converges.*

(b) *Let $\sum_{k=1}^{\infty} d_k$ be a divergent series of positive terms. If either (i) $\lim_{k\to\infty}\frac{p_k}{d_k}$ exists and is not 0 or (ii) if the limit does not exist because it is infinite, then $\sum_{k=1}^{\infty} p_k$ also diverges.*

RECALL: *A limit "is infinite" when the limit does not exist because the ratios are unbounded.*

(c) *Let $\sum_{k=1}^{\infty} d_k$ be a divergent series of positive terms. If $\lim_{k\to\infty}\frac{p_k}{d_k}$ exists and is 0, then the limit comparison test (with comparison series $\sum_{k=1}^{\infty} d_k$) is inconclusive.*

TERMINOLOGY: Sometimes, to distinguish between Theorems 11.3.1 and 11.3.2, Theorem 11.3.1 is referred to as the **termwise comparison test**.

Proof of Theorem 11.3.2

We prove only parts (a) and (c) of the limit comparison test. See Exercise 31 for the proofs of part (b).

(a) Since as $k\to\infty$, $p_k/c_k \to a$, there is an integer N such that, for all $n \geq N$, p_k/c_k remains less than, say, $a+1$. Thus

$$p_k < (a+1)c_k \qquad \text{for } n \geq N.$$

The series

$$\sum_{k=N}^{\infty}(a+1)c_k = (a+1)c_N + (a+1)c_{N+1} + \cdots + (a+1)c_k + \cdots,$$

being $a+1$ times the tail of a convergent series, is convergent. Note that this series starts with the index $k=N$; it is the tail of the full series. By the comparison test,

$$p_N + p_{N+1} + \cdots + p_k + \cdots$$

is convergent. To extend this to the full series, add the first $N-1$ terms. While these $N-1$ additional terms greatly influence the sum of the series, they have no impact on the convergence of the series. Hence $\sum_{k=1}^{\infty} p_k = p_1 + p_2 + \cdots + p_k + \cdots$ converges.

(c) To prove part (c), all that is needed is to produce two examples that satisfy all of the hypotheses for this part of the theorem but have different outcomes: one converges and one diverges.

The first example is the series in Example 2, with $p_k = (k+1)/(k+2)\cdot 1/k$ and $d_k = 1/\sqrt{k}$. The comparison series is a p-series with $p = 1/2$, so it diverges. Also,

$$\lim_{k\to\infty}\frac{p_k}{d_k} = \lim_{k\to\infty}\frac{\frac{k+1}{k+2}\frac{1}{k}}{\frac{1}{\sqrt{k}}} = \lim_{k\to\infty}\frac{k+1}{k+2}\frac{1}{\sqrt{k}} = 1\cdot 0 = 0.$$

And, as we learned in Example 2, the series $\sum_{k=1}^{\infty}(k+1)/(k+2)\cdot 1/k$ diverges.

For the second example, consider the series with $p_k = 1/k^2$ and the comparison series with $d_k = 1/k$. This comparison series is the harmonic series, which diverges. And, $p_k/d_k = (1/k^2)/(1/k) = 1/k$, which approaches 0 as $k\to\infty$. But, the series $\sum_{k=1}^{\infty} 1/k^2$ is a p-series with $p=2$ which converges.

Because there are two series (and divergent comparison series) which satisfy $\lim_{k\to\infty} p_k/d_k = 0$, this combination of facts cannot classify the original series, $\sum_{k=1}^{\infty} p_k$ as either convergent or divergent — the test is inconclusive.

●

The next example shows how much more convenient the limit comparison test can be, compared with the comparison test. Compare the solution in Example 3 with that in Example 2.

EXAMPLE 3. Does the series $\displaystyle\sum_{k=1}^{\infty}\frac{k+1}{k+2}\cdot\frac{1}{k} = \frac{2}{3}\cdot\frac{1}{1} + \frac{3}{4}\cdot\frac{1}{2} + \cdots + \frac{k+1}{k+2}\cdot\frac{1}{k} + \cdots$ converge or diverge?

SOLUTION The key observation is the same as in Example 2: the terms of this series are very to the terms in the harmonic series. To take advantage of this observation in the context of the limit comparison test, we define $p_k = (k+1)/(k+2)\cdot 1/k$ and $d_k = 1/k$ and examine the ratio between corresponding terms:

$$\lim_{k\to\infty}\frac{p_k}{c_k} = \lim_{k\to\infty}\frac{\frac{k+1}{k+2}\cdot\frac{1}{k}}{\frac{1}{k}} = \lim_{k\to\infty}\frac{k+1}{k+2} = 1.$$

Since the above limit exists and is not 0, and the harmonic series diverges, the limit comparison test shows that $\displaystyle\sum_{k=1}^{\infty}(k+1)/(k+2)\cdot 1/k$ diverges. ◆

EXAMPLE 4. Does $\displaystyle\sum_{k=1}^{\infty} \frac{\left(1+\frac{1}{k}\right)^k \left(1+\left(\frac{-1}{2}\right)^k\right)}{2^k}$ converge or diverge?

SOLUTION As $k \to \infty$, $(1+1/k)^k \to e$ and $1+(-1/2)^k \to 1$. The series resembles the convergent geometric series with first term $1/2$ and ratio also $1/2$, that is: $\sum_{k=1}^{\infty} 1/2^k = 1/2+1/4+\cdots+1/2^k+\cdots$. (Geometric series were discussed in Section 2.2.) Using the limit comparison theorem, we compute

$$\lim_{k\to\infty} \frac{\dfrac{\left(1+\frac{1}{k}\right)^k \left(1+\left(\frac{-1}{2}\right)^k\right)}{2^k}}{\dfrac{1}{2^k}} = \lim_{k\to\infty} \left(1+\frac{1}{k}\right)^k \left(1+\left(\frac{-1}{2}\right)^k\right) = e\cdot 1 = e.$$

Since, as shown previously, $\sum_{k=1}^{\infty} 2^{-k}$ is a convergent geometric series, the given series must also converge. ◆

The final example shows that the comparison test is not completely replaced by the limit comparison test.

EXAMPLE 5. Does $\displaystyle\sum_{k=1}^{\infty} k^3 3^{-k}$ converge or diverge?

SOLUTION The typical term $k^3 3^{-k}$ is dominated by the exponential factor, $1/3^k$. We suspect that the series $\sum_{k=1}^{\infty} k^3 3^{-k}$ might converge. We try the limit comparison test, with $p_k = k^3 3^{-k} = k^3/3^k$ and $c_k = 3^{-k} = 1/3^k$, obtaining

$$\lim_{k\to\infty} \frac{p_k}{c_k} = \lim_{k\to\infty} \frac{\frac{k^3}{3^k}}{\frac{1}{3^k}} = \lim_{k\to\infty} k^3 = \infty.$$

Since the above limit is not finite, the test gives no information. We will try something else.

Because k^3 approaches ∞ much more slowly than 3^k, we still suspect that $\sum_{k=1}^{\infty} k^3/3^k$ converges. L'Hôpital's rule can be used to show that k^3 approaches ∞ more slowly than any exponential b^k with $b > 1$. (For example, for large k, k^3 is less than $(1.5)^k$.) Thus, for large k

$$\frac{k^3}{3^k} < \frac{(1.5)^k}{3^k} = (0.5)^k.$$

The geometric series $\sum_{k=1}^{\infty}(0.5)^k$ converges. Since $k^3/3^k < (0.5)^k$ for all but a finite number of values of k, the comparison test tells us that $\sum_{k=1}^{\infty} k^3/3^k$ converges. ◆

Observation 11.3.3: *Differences Between the Comparison and Limit Comparison Tests*

Unlike Theorem 11.3.1 where there is one part leading to a conclusion that the series converges and one part that leads to a conclusion that the series diverges, Theorem 11.3.2 has an additional part whose conclusion is that the limit comparison test is inconclusive (at least for that comparison series).

Other tests for the convergence or divergence of an infinite series that we will learn about in upcoming sections also have the possibility that the test is inconclusive. When a series test is inconclusive, it will be necessary to look for a different comparison series or a different test that does lead to a definite conclusion about the convergence or divergence of the series being analyzed.

Summary

We developed two new tests for convergence or divergence of a series with positive terms, $\sum_{k=1}^{\infty} p_k$.

The (termwise) comparison test (Theorem 11.3.1) has two parts: If every p_k is less than the corresponding term of a convergent series of positive terms, then $\sum_{k=1}^{\infty} p_k$ converges. And, if every p_k is larger than the corresponding term of a divergent series of positive terms, then $\sum_{k=1}^{\infty} p_k$ diverges.

The limit comparison test (Theorem 11.3.2) is based on the same ideas as the (termwise) comparison test, but the limit comparison test is often much easier to apply. All three parts of Theorem 11.3.2 involve the same limit of the ratio of p_k to the corresponding term of a series of positive terms whose convergence or divergence is known. The basic idea is that if this limit exists and is positive then the unknown series, $\sum_{k=1}^{\infty} p_k$, and the comparison series either both converge or both diverge. One special case must be remembered: When the comparison limit is zero and the comparison series diverges, no conclusion about the divergence of $\sum_{k=1}^{\infty} p_k$ can be made. (Another test must be used.)

Learning to master the use of the different tests for the convergence or divergence of a series takes practice.

EXERCISES for Section 11.3

Use the comparison test in Exercises 1 to 4 to determine whether the series converges or diverges.

1. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^2+3}$ **2.** $\displaystyle\sum_{k=1}^{\infty} \frac{k+2}{(k+1)\sqrt{k}}$ **3.** $\displaystyle\sum_{k=1}^{\infty} \frac{\sin^2(k)}{k^2}$ **4.** $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k2^k}$

Use the limit comparison test in Exercises 5 to 8 to determine whether the series converges or diverges.

5. $\displaystyle\sum_{k=1}^{\infty} \frac{5k+1}{(k+2)k^2}$ **6.** $\displaystyle\sum_{k=1}^{\infty} \frac{2^k+k}{3^k}$ **7.** $\displaystyle\sum_{k=1}^{\infty} \frac{k+1}{(5k+2)\sqrt{k}}$ **8.** $\displaystyle\sum_{k=1}^{\infty} \frac{(1+1/k)^k}{k^2}$

In Exercises 9 to 28 use any test to determine whether the series converges or diverges.

9. $\displaystyle\sum_{k=1}^{\infty} \frac{k^2k}{3^k}$ **10.** $\displaystyle\sum_{k=1}^{\infty} \frac{2^k}{k^2}$ **11.** $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^k}$ **12.** $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k!}$

13. $\displaystyle\sum_{k=1}^{\infty} \frac{4k+1}{(2k+3)k^2}$ **14.** $\displaystyle\sum_{k=1}^{\infty} \frac{k^2(2^k+1)}{3^k+1}$ **15.** $\displaystyle\sum_{k=1}^{\infty} \frac{1+\cos(k)}{k^2}$ **16.** $\displaystyle\sum_{k=1}^{\infty} \frac{\ln(k)}{k}$

17. $\displaystyle\sum_{k=1}^{\infty} \frac{\ln(k)}{k^2}$ **18.** $\displaystyle\sum_{k=1}^{\infty} \frac{5^k}{k^k}$ **19.** $\displaystyle\sum_{k=1}^{\infty} \frac{2^k}{k!}$ **20.** $\displaystyle\sum_{k=2}^{\infty} \frac{1}{\sqrt{k}\ln(k)}$

21. $\displaystyle\sum_{k=1}^{\infty} \frac{e^{2k}}{\pi^k}$ **22.** $\displaystyle\sum_{k=1}^{\infty} \frac{k^2e^k}{\pi^k}$ **23.** $\displaystyle\sum_{k=1}^{\infty} \frac{3k+1}{2k+10}$ **24.** $\displaystyle\sum_{k=1}^{\infty} \frac{4}{2k^2-k}$

25. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{\ln(k+1)}$ **26.** $\displaystyle\sum_{k=1}^{\infty} \frac{1}{\sin(1/k)}$ **27.** $\displaystyle\sum_{k=1}^{\infty} \left(\frac{k+1}{k+3}\right)^k$ **28.** $\displaystyle\sum_{k=1}^{\infty} \left(\frac{k}{2k-1}\right)^k$

29. For what values of the positive number x does the series $\displaystyle\sum_{k=1}^{\infty} \frac{x^k}{k2^k}$ converge? diverge?

30. For what values of the positive exponent m does the series $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^m\ln(k+1)}$ converge? diverge?

31. Prove part (b) of the limit comparison test for convergence and divergence.

32. For what constants p does $\sum_{k=1}^{\infty} k^p e^{-k}$ converge?

33. (a) Show that $\sum_{k=1}^{\infty} \frac{1}{1+2^k}$ converges. (b) Show that the sum in (a) is between 0.64 and 0.77.

34. (a) Show that $\sum_{k=n+1}^{\infty} \frac{1}{k!}$ is less than the sum of the geometric series with first term and $\frac{1}{(n+1)!}$ and ratio $\frac{1}{n+2}$.

(b) Use (a) with $n=4$ to show that $1+1+\frac{1}{2!}+\frac{1}{3!}+\frac{1}{4!} < \sum_{k=0}^{\infty} \frac{1}{k!} < 1+1+\frac{1}{2!}+\frac{1}{3!}+\frac{1}{4!}+\frac{1}{5!}\cdot\frac{1}{1-1/6}$.

(c) From (b) deduce that $2.71 < \sum_{k=0}^{\infty} \frac{1}{k!} < 2.72$.

(d) Find n so that $\sum_{k=n+1}^{\infty} \frac{1}{k!} < 0.0005$.

(e) Use (d) to estimate $\sum_{k=0}^{\infty} \frac{1}{k!}$ to three decimal places.

In Exercises 35 to 40, assume that $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$ are series with positive terms. What, if anything, can we conclude about the convergence or divergence of $\sum_{k=1}^{\infty} a_k$

35. if $\sum_{k=1}^{\infty} b_k$ is divergent and $\lim_{k\to\infty} \frac{a_k}{b_k} = 0$?

36. if $\sum_{k=1}^{\infty} b_k$ is convergent and $\lim_{k\to\infty} \frac{a_k}{b_k} = \infty$?

37. if $\sum_{k=1}^{\infty} b_k$ is convergent and $3b_k \le a_k \le 5b_k$?

38. if $\sum_{k=1}^{\infty} b_k$ is divergent and $3b_k \le a_k \le 5b_k$?

39. if $\sum_{k=1}^{\infty} b_k$ is convergent and $a_k < b_k^2$?

40. if $\sum_{k=1}^{\infty} b_k$ is divergent, $b_k \to 0$ as $k \to \infty$, and $a_k < b_k^2$?

---

41. Find a number B such that $\sum_{k=1}^{\infty} \frac{\ln(k)}{k^2} < B$.

42. Find a number B such that $\sum_{k=1}^{\infty} \frac{k+2}{k+1}\cdot\frac{1}{n^3} < B$.

43. Estimate $\sum_{k=1}^{\infty} \frac{1}{k2^k}$ to three decimal places.

44. Prove the following result, which is used in the statistical theory of stochastic processes:

> Let $\{a_n\}$ and $\{c_n\}$ be sequences of nonnegative numbers such that $\lim_{n\to\infty} c_n = 0$ and $\sum_{k=1}^{\infty} a_k c_k$ converges. Then $\sum_{k=1}^{\infty} a_k c_k^2$ converges.

45. Let $\sum_{k=1}^{\infty} a_k$ be a convergent series with only positive terms. Must $\sum_{k=1}^{\infty} (a_k)^2$ also converge?

46. Let $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$ be convergent series with only positive terms. Must $\sum_{k=1}^{\infty} a_k b_k$ converge?

47. Let $p_1, p_2, \ldots$ be a sequence of nonnegative numbers such that $\sum_{k=1}^{\infty} p_k = 1$. Also let $x_1, x_2, \ldots$ be a sequence of nonnegative numbers. Assuming that $\sum_{k=1}^{\infty} p_k x_k^2$ converges, show that $\sum_{k=1}^{\infty} p_i x_i$ cnoverges.

48. **SAM:** I don't like Exercise 47. The hint solves it and I learn nothing.
JANE: That idea of adding 1 to x_k^2 came from out of the blue, a rabbit out of the hat.
SAM: I have a natural way: There are two types of x_k, $x_k > 1$ and $0 \leq x_k \leq 1$. In the first case $x_k < x_k^2$ and in the second $0 \leq p_k x_k \leq p_k$. That does it.
JANE: That's too fast.

Explain Sam's logic to Jane.

11.4 The Ratio and Root Tests

The next test is suggested by the test for the convergence of a geometric series. In a geometric series the ratio between consecutive terms is constant. The ratio test concerns series when this ratio is essentially constant.

The Ratio Test

Theorem 11.4.1: Ratio Test

Let $\sum_{k=1}^{\infty} p_k = p_1 + p_2 + \cdots + p_k + \cdots$ be a series of positive terms. Assume $\lim\limits_{k\to\infty} \dfrac{p_{k+1}}{p_k}$ exists and call it r.

Note that r is zero or a positive number, it cannot be negative.

(a) *If $r < 1$, the series $\sum_{k=1}^{\infty} p_k$ converges.*
(b) *If $r > 1$ or $r = \infty$, the series $\sum_{k=1}^{\infty} p_k$ diverges.*
(c) *If $r = 1$ or r does not exist, no conclusion can be drawn (the series $\sum_{k=1}^{\infty} p_k$ may converge or it may diverge).*

Proof of Theorem 11.4.1

The idea behind the Ratio Test is to compare the series to a geometric series.

(a) Assume $r = \lim_{k\to\infty} p_{k+1}/p_k < 1$. Select a number s such that $r < s < 1$. Then there is an integer N such that for all $k \geq N$, $p_{k+1}/p_k < s$ and therefore $p_{k+1} < sp_k$. Using this, we see that

$$\begin{aligned} p_{N+1} &< sp_N \\ p_{N+2} &< sp_{N+1} < s(sp_N) = s^2 p_N \\ p_{N+3} &< sp_{N+2} < s(s^2 p_N) = s^3 p_N, \end{aligned}$$

and so on.

Thus the terms of the tail end of the series $p_N + p_{N+1} + p_{N+2} + \cdots$ are less than the corresponding terms of the geometric series $p_N + sp_N + s^2 p_N + \cdots$ (except for the first term, p_N, which equals the first term of the geometric series). Since $s < 1$, the latter series converges. By the comparison test, $p_N + p_{N+1} + p_{N+2} + \cdots$ converges. Adding in the front end,

$$p_1 + p_2 + \cdots + p_{N-1},$$

we still have a convergent series.

(b) If $r > 1$ or is infinite, then for all k from some point on p_{k+1} is larger than p_k. Thus $\lim_{k\to\infty} p_k$ cannot be 0. By the n^{th}-term test for divergence the series diverges.

(c) When $r = 1$ or r does not exist, anything can happen. The series may diverge or it may converge, as Exercise 23 illustrates. We need other tests to determine whether the series diverges or converges.

Observation 11.4.2: *When to Apply the Ratio Test*

Because the ratio test is based on geometric series, it is often the first test to try if the formula for the terms involves powers of a number or factorials.

The next two examples illustrate the use of ratio test.

EXAMPLE 1. Show that the series $p + 2p^2 + 3p^3 + \cdots + kp^k + \cdots$ converges for $0 < p < 1$.

SOLUTION Let a_k denote the k^{th} term of the series. Then $a_k = kp^k$ and $a_{k+1} = (k+1)p^{k+1}$. The ratio between consecutive terms is

$$\frac{a_{k+1}}{a_k} = \frac{(k+1)p^{k+1}}{k\,p^k} = \frac{k+1}{k}p.$$

Thus

$$\lim_{k\to\infty} \frac{a_{k+1}}{a_k} = p < 1,$$

and the series converges. ♦

The sum of the series in Example 1 is found in Exercise 35 in Section 11.1.

Observation 11.4.3: *Most Series Tests Do Not Provide the Sum of Convergent Series*

The two comparison tests in Section 11.3, the ratio test, and root test can be used to determine the convergence or divergence of a series. None of these tests provide a way to find the sum of the series. (The integral test does provide a way to estimate the sum of an infinite series whose convergence was previously determined (by the integral test).

In Example 2 the terms of the series depend on the variable x, actually, k^{th} powers of x: x^k. This does not change the basic approach to determining convergence or divergence, except that instead of getting just one answer, the series may converge for some values of x and diverge for others.

EXAMPLE 2. Determine the positive values of x for which the series

$$\frac{1}{0!} + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^k}{k!} \cdots$$

converges and for which values of x it diverges. Each choice of x determines a specific series with constant terms.

SOLUTION If we start the series with the index $k = 0$, then the k^{th} term, a_k, is $x^k/k!$. Thus

$$a_{k+1} = \frac{x^{k+1}}{(k+1)!}$$

and therefore

$$\frac{a_{k+1}}{a_k} = \frac{\frac{x^{k+1}}{(k+1)!}}{\frac{x^k}{k!}} = x\frac{k!}{(k+1)!} = \frac{x}{k+1}.$$

Since x is fixed,

$$\lim_{k\to\infty} \frac{a_{k+1}}{a_k} = \lim_{k\to\infty} \frac{x}{k+1} = 0.$$

By the ratio test, the series converges for all positive x. ♦

In the next section we show the series converges for all negative values of x, too. This type of series will be studied is one of the main topics of Chapter 12.

The next example uses the ratio test to establish divergence.

EXAMPLE 3. Show that the series $\frac{2}{1}+\frac{2^2}{2}+\cdots+\frac{2^k}{k}+\cdots$ diverges.

SOLUTION Here $a_k = 2^k/k$ and

$$\frac{a_{k+1}}{a_k} = \frac{\frac{2^{k+1}}{k+1}}{\frac{2^k}{k}} = \frac{2^{k+1}}{k+1}\frac{k}{2^k} = \frac{2k}{k+1},$$

so $r = \lim_{k\to\infty} a_{k+1}/a_k = \lim_{k\to\infty} 2k/(k+1) = 2$, which is larger than 1. This, by the ratio test (with $r = 2$), this series diverges. ♦

Observation 11.4.4: ***Five Alternate Solutions of Example 3***

The ratio test is not the only series test that could be used to show $\sum_{k=1}^{\infty} 2^k/k$ diverges. Here are some alternate ways to approach this problem, including at least one that does not work. Of course, in each case the final conclusion must be the same: the series diverges.

1. The fact that $r = 2$ in the series test suggests trying to use a comparison test with the geometric series $\sum_{k=1}^{\infty} 2^k$. Direct comparison with $p_k = 2^k/k$ and $d_k = 2^k$ does not work because $2^k/k < 2^k$.
2. But, for the limit comparision test,

$$\lim_{k\to\infty}\frac{p_k}{d_k} = \lim_{k\to\infty}\frac{\frac{2^k}{k}}{2^k} = \lim_{k\to\infty}\frac{1}{k} = 0.$$

   The fact that this limit is 0 means the limit comparision test is inconclusive.
3. Another idea is to do a comparison with the harmonic series, $d_k = 1/k$. Because $2^k > 1$ for all positive integers k, $2^k/k > 1/k$. By the comparison test, since the harmonic series diverges, so does $\sum_{k=1}^{\infty} 2^k/k$.
4. The limit comparison test with the harmonic series is another idea that looks appealing. This means using Theorem 11.3.2 with $p_k = 2^k/k$ and $d_k = 1/k$. Then

$$\lim_{k\to\infty}\frac{p_k}{d_k} = \lim_{k\to\infty}\frac{\frac{2^k}{k}}{\frac{1}{k}} = \lim_{k\to\infty} 2^k = \infty.$$

   This is one of the cases under which the limit comparison test allows us to conclude that the original series diverges.
5. The last approach might just be the easiest and fastest way to approach this problem. Since the powers 2^k grow faster than k, $\lim_{k\to\infty} 2^k/k = \infty$. This means this series' terms become arbitrarily large. So, by the n^{th}-term test, $\sum_{k=1}^{\infty} 2^k/k$ diverges.

ADVICE: We hope you are beginning to understand that there is no single series test that will work for all problems. And, sometimes more than one test can be used. While gaining experience in analyzing series, remember to think about other ways that could be used to determine the convergence or divergence of a given series.

The Root Test

The next test, the root test, is closely related to the ratio test. Its proof is outlined in Exercises 24 and 25.

Theorem 11.4.5: Root Test

Let $\sum_{k=1}^{\infty} p_k$ be a series of positive terms. Assume $\lim_{k\to\infty} \sqrt[k]{p_k}$ exists and call it r. Then

(a) *If $r < 1$, the series converges.*
(b) *If $r > 1$ or $r = \infty$, the series diverges.*
(c) *If $r = 1$ or r does not exist, the ratio test is inconclusive (the series may converge or may diverge).*

EXAMPLE 4. Use the root test to determine whether $\sum_{k=1}^{\infty} \frac{3^k}{k^{k/2}}$ converges or diverges.

SOLUTION We have

$$r = \lim_{k\to\infty} \sqrt[k]{\frac{3^k}{k^{k/2}}} = \lim_{k\to\infty} \frac{3}{\sqrt{k}} = 0.$$

By the root test, the series converges. ♦

Observation 11.4.6: *When to Apply the Root Test*

The root test is generally most useful when the k^{th} term contains only k^{th} powers, such as k^k or 3^k. It is not useful if $k!$ is present.

Summary

This section introduced two more tests for convergence or divergence of a series $\sum_{k=1}^{\infty} p_k$ with positive terms. Both tests are motivated by geometric series. The ratio test involves $r = \lim_{k\to\infty} p_{k+1}/p_k$ and the root test involves $r = \lim_{k\to\infty} \sqrt[k]{p_k}$.

The ratio test is useful when the terms involve powers and factorials.

The root test is useful when there are no factorials but there are powers involving the summation index.

EXERCISES for Section 11.4

In Exercises 1 to 6 apply the ratio test to decide whether the series converges or diverges. If the test gives no information, use another test to decide.

1. $\sum_{k=1}^{\infty} \frac{k^2}{3^k}$

2. $\sum_{k=1}^{\infty} \frac{(k+1)^2}{k2^k}$

3. $\sum_{k=1}^{\infty} \frac{k\ln(k)}{3^k}$

4. $\sum_{k=1}^{\infty} \frac{k!}{3^k}$

5. $\sum_{k=1}^{\infty} \frac{(2k+1)(2^k+1)}{3^k+1}$

6. $\sum_{k=1}^{\infty} \frac{k!}{k^k}$

In Exercises 7 to 10 use the root test to determine whether the series converges or diverges.

7. $\sum_{k=1}^{\infty} \frac{k^k}{3^{k^2}}$

8. $\sum_{k=1}^{\infty} \frac{e^k}{((2k+1)/k)^k}$

9. $\sum_{k=1}^{\infty} \frac{(1+1/k)^{2k}}{\left(2+1/\sqrt{k}\right)^k}$

10. $\sum_{k=1}^{\infty} \frac{(1+1/k)^k(2k+1)^k}{(3k+1)^k}$

The series in Exercises 11 to 16 converge. Find a number B that is larger than the sum.

11. $\sum_{k=1}^{\infty} \frac{k^2}{2^k}$ **12.** $\sum_{k=1}^{\infty} \frac{k}{3^k}$ **13.** $\sum_{k=1}^{\infty} \frac{1}{k^3}$ **14.** $\sum_{k=1}^{\infty} \frac{\sin^2(k)}{k^2}$ **15.** $\sum_{k=1}^{\infty} \frac{\ln(k)}{k^2}$ **16.** $\sum_{k=1}^{\infty} \frac{(1+2//k)^k}{1.1^k}$

The series in Exercises 17 to 20 diverge. Find a number m such that the m^{th} partial sum of the series exceeds 1,000.

17. $\sum_{k=1}^{\infty} \frac{\ln(k)}{k}$ **18.** $\sum_{k=1}^{\infty} \frac{k}{k^2+1}$ **19.** $\sum_{k=1}^{\infty} (1.01)^k$ **20.** $\sum_{k=1}^{\infty} \frac{(k+2)^2}{k+1} \cdot \frac{1}{\sqrt{k}}$

---

21. Use the result of Example 2 to show that, for $x > 0$, $\lim_{k\to\infty} \frac{x^k}{k!} = 0$. This was established directly in Section 11.1.

22. Solve Example 3 using the root test.

23. This exercise shows that the ratio test gives no information if $r = \lim_{k\to\infty} \frac{p_{k+1}}{p_k} = 1$.

(a) Show that for $p_k = \frac{1}{k}$, $\sum_{k=1}^{\infty} p_k$ diverges and $r = \lim_{k\to\infty} \frac{p_{k+1}}{p_k} = 1$.

(b) Show that for $p_k = \frac{1}{k^2}$, $\sum_{k=1}^{\infty} p_k$ converges and $r = \lim_{k\to\infty} \frac{p_{k+1}}{p_k} = 1$.

24. This exercise shows that the root test gives no information if $r = \lim_{k\to\infty} \sqrt[k]{p_k} = 1$.

(a) Show that for $p_k = \frac{1}{k}$, $\sum_{k=1}^{\infty} p_k$ diverges and $r = \lim_{k\to\infty} \sqrt[k]{p_k} = 1$.

(b) Show that for $p_k = \frac{1}{k^2}$, $\sum_{k=1}^{\infty} p_k$ converges and $r = \lim_{k\to\infty} \sqrt[k]{p_k} = 1$.

25. This exercise outlines a proof of the root test, Theorem 11.4.5.

(a) Assume that $r = \lim_{k\to\infty} \sqrt[k]{p_k} < 1$. Pick s with $r < s < 1$, and then pick N such that $\sqrt[k]{p_k} < s$ for all $k > N$. Show that $p_k < s^k$ for $k > N$ and compare a tail end of $\sum_{k=1}^{\infty} p_k$ to a geometric series.

(b) Assume that $r = \lim_{k\to\infty} \sqrt[k]{p_k} > 1$. Pick s with $1 < s < r$, and then pick N such that $\sqrt[k]{p_k} > s$ for all $k > N$. Show that $p_k > s^k$ for all $k > N$. From this conclude that $\sum_{k=1}^{\infty} p_k$ diverges.

In Exercises 26 to 28 a, b, and c are constants. Verify the following derivative formulas.

26. $\frac{d}{dx} a^2 x \sin(ax) = \sin(ax) + ax\cos(ax)$

27. $\frac{d}{dx} \ln\left|ax^2+bx+c\right| = \frac{2ax+b}{ax^2+bx+c}$

28. $\frac{d}{dx}\left(x\arctan(ax) - \frac{1}{2a}\ln\left(1+a^2x^2\right)\right) = \arctan(ax)$

In Exercises 29 to 34 a, b, c, and n are constants and n is a positive integer. Use integration techniques to obtain each of the following reduction formulas.

29. $\int x^n \sin(ax)\,dx = -\frac{1}{a}\cos(ax) + \frac{n}{a}\int x^{n-1}\cos(ax)\,dx$

30. $\int x^n \cos(ax)\,dx = \frac{1}{a}\cos(ax) - \frac{n}{a}\int x^{n-1}\sin(ax)\,dx$

31. $\int \frac{dx}{x^2\sqrt{ax+b}} = \frac{-\sqrt{ax+b}}{bx} - \frac{a}{2b}\int \frac{dx}{x\sqrt{ax+b}}$

32. $\int \frac{dx}{(ax^2+c)^{n+1}} = \frac{1}{2nc}\frac{x}{(ax^2+c)^n} + \frac{2n-3}{2nc}\int \frac{dx}{(ax^2+c)^n}$

33. $\int \frac{dx}{(ax^2+bx+c)^{n+1}} = \frac{2ax+b}{n(4ac-b^2)(ax^2+bx+c)^n} + \frac{2(2n-1)a}{n(4ac-b^2)}\int \frac{dx}{(ax^2+bx+c)^n}$

34. $\int (\ln(ax))^2\,dx = x^2\left((\ln(ax))^2 - 2\ln(ax) + 2\right)$

11.5 The Tests for Series with Both Positive and Negative Terms

The tests for convergence or divergence in Sections 11.2 to 11.4 apply only to series whose terms are positive. This section examines series that have both positive and negative terms. Two tests for their convergence are presented. The alternating series test applies to series whose terms alternate in sign (+, -, +, -, ...), or vice versa, and decrease in absolute value. In the absolute convergence test the signs may appear in any pattern.

Alternating Series

Definition: *Alternating Series*

If $p_1, p_2, \ldots, p_n, \ldots$ is a sequence of positive numbers, then the series

$$\sum_{k=1}^{\infty}(-1)^{k+1}p_k = p_1 - p_2 + p_3 - p_4 + \cdots + (-1)^{k+1}p_k + \cdots$$

and

$$\sum_{k=1}^{\infty}(-1)^{k}p_k = -p_1 + p_2 - p_3 + p_4 - \cdots + (-1)^{k}p_k + \cdots,$$

in which the terms alternate in sign, are called **alternating series**.

Two examples of alternating series are

$$\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{2k-1} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots + (-1)^{k+1}\frac{1}{2k-1} + \cdots$$

and

$$\sum_{k=0}^{\infty}(-1)^{k} = 1 - 1 + 1 - 1 + \cdots + (-1)^{k} + \cdots.$$

By the n^{th}-term test, the second series diverges. The following theorem can be used to show that the first series converges.

Theorem 11.5.1: Alternating Series Test

Consider the alternating series $\sum_{k=1}^{\infty}(-1)^{k+1}p_k = p_1 - p_2 + p_3 - \cdots + (-1)^{k+1}p_k + \cdots$. *If the terms come from a sequence of positive numbers* $\{p_1, p_2, \ldots, p_k, \ldots\}$ *that (a) is decreasing* ($p_{k+1} < p_k$) *and (b) approaches zero as k increases* ($\lim_{k\to\infty} p_k = 0$), *then the alternating series* $\sum_{k=1}^{\infty}(-1)^k p_k$ *converges.*

Under the same conditions on the terms p_k, $\sum_{k=1}^{\infty}(-1)^k p_k = -p_1 + p_2 - p_3 + \cdots + (-1)^k p_k + \cdots$ *also converges.*

Proof of a Special Case of the Alternating Series Test (Theorem 11.5.1)
We prove the theorem for the **alternating harmonic series**:

$$\sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + (-1)^{k+1}\frac{1}{k} + \cdots.$$

The argument can be generalized to a general alternating series that satisfies the hypotheses of Theorem 11.5.1. (See Exercise 21.)

Recall that the sequence of partial sums has terms $S_n = \sum_{k=1}^{n}(-1)^{k+1}/k$. To show this sequence converges, look separately at the terms with even and odd indices.

The plan is to show the sequence $\{S_2, S_4, \ldots, S_{2n}, \ldots\}$ converges by showing that it is both monotone and bounded. Next, it will be much easier to show that that the sequence $\{S_1, S_3, \ldots, S_{2n+1}, \ldots\}$ converges to the same limit. This sounds like a lot, but it is not as difficult as it sounds.

To show the sequence of partial sums with an even number of summands is monotone increasing, group the summands in pairs:

$$\begin{aligned}
S_2 &= \left(1 - \frac{1}{2}\right) & & \\
S_4 &= \left(1 - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) & &= S_2 + \left(\frac{1}{3} - \frac{1}{4}\right) \\
S_6 &= \left(1 - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \left(\frac{1}{5} - \frac{1}{6}\right) & &= S_4 + \left(\frac{1}{5} - \frac{1}{6}\right)
\end{aligned}$$

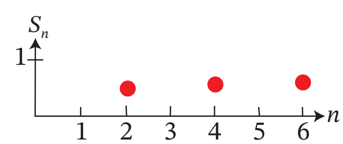


Figure 11.5.1

Since $1/3$ is larger than $1/4$, the difference $1/3 - 1/4$ is positive. Therefore, S_4, which equals $S_2 + (1/3 - 1/4)$, is larger than S_2. Similarly, $S_6 > S_4$. In general, for any positive integer n,

$$S_{2n+2} = S_{2n} + \frac{1}{2n+1} - \frac{1}{2n+2} = S_{2n} + \frac{1}{(2n+1)(2n+2)} > S_{2n}.$$

Since $1/((2n+1)(2n+2))$ is positive, $S_{2n+2} > S_{2n}$. This proves the sequence of even partial sums, $\{S_{2n}\}$ is an increasing sequence. (See Figure 11.5.1.)

We next show that each S_{2n} is less than the first term of the series, 1. To do this we will use induction to show $S_{2n} \le 1 - 1/(2n)$ for all positive integers n, and then observe that each of these bounds is less than 1. First, for $n = 1$, $S_2 = 1 - 1/2 = 1 - 1/(2\cdot 1)$. Next, for the inductive step, suppose $S_{2n} \le 1 - 1/(2n)$ for some positive integer n. Then,

$$\begin{aligned}
S_{2n+2} &= S_{2n} + (-1)^{(2n+1)+1}\frac{1}{2n+1} + (-1)^{(2n+2)+1}\frac{1}{2n+2} && (\text{ partial sum for alternating harmonic series }) \\
&= S_{2n} + \frac{1}{2n+1} - \frac{1}{2n+2} && (\, 2n+2 \text{ is always even; } 2n+3 \text{ is always odd }) \\
&\le S_{2n-2} - \frac{1}{2n} + \frac{1}{2n+1} - \frac{1}{2n+2} && (\text{ inductive hypothesis }) \\
&= S_{2n} - \frac{1}{2n+2} && (\text{ partial sum for alternating harmonic series }).
\end{aligned}$$

Therefore, in general, $S_{2n} < 1 - 1/(2n)$ for all n. And, as $1/(2n)$ is always positive, every $S_{2n} < 1$.

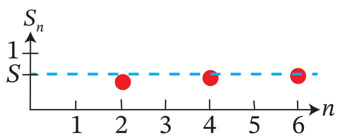


Figure 11.5.2

The sequence of partial sums with an even index $\{S_2, S_4, S_6, \ldots\}$ is therefore increasing and is bounded above (by 1). By Theorem 10.1.1 of Section 10.1, $\lim_{n\to\infty} S_{2n}$ exists. Call the limit S, and note that S is both positive and less than or equal to 1. (See Figure 11.5.2.)

To show that the sequence of partial sums with an odd index $\{S_1, S_3, S_5, \ldots\}$ also converges to S, observe that

$$\begin{aligned} S_3 &= 1 - \frac{1}{2} + \frac{1}{3} &&= S_2 + \frac{1}{3}, \\ S_5 &= 1 - \frac{1}{3} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} &&= S_4 + \frac{1}{5}, \end{aligned}$$

and, in general,

$$\begin{aligned} S_{2n+1} &= S_{2n} + (-1)^{(2n+1)+1}\frac{1}{2n+1}. && (\text{ partial sum for alternating harmonic series }) \\ &= S_{2n} + \frac{1}{2n+1} && (\, 2n+1 \text{ is always odd }) \\ &\to S + 0 && (\text{ as } n \to \infty\,) \\ &= S. \end{aligned}$$

Since the sequence of partial sum with an even index and the sequence of partial sums with an odd index have the same limit, S, it follows that $\lim_{k\to\infty} S_k = S$. Thus the alternating harmonic series

$$\sum_{k=1}^{\infty} (-1)^k \frac{1}{k} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots$$

converges. NOTE: In Exercise 29 in Section 12.4 it will be shown that its sum is ln(2).

A similar argument applies to any alternating series whose k^{th} term approaches 0 and whose terms decrease in absolute value.

An alternating series, such as the alternating harmonic series, whose terms decrease in absolute value as k increases is sometimes referred to as a **decreasing alternating series**. The alternating series test says that a decreasing alternating series whose k^{th} term approaches zero (as $k \to \infty$) converges.

EXAMPLE 1. Estimate the sum S of the alternating harmonic series.

SOLUTION The first five partial sums are

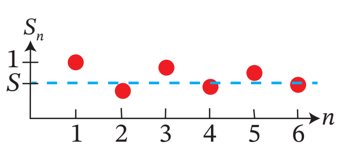


Figure 11.5.3

$$\begin{aligned} S_1 &= 1 && = 1.00000 \\ S_2 &= 1 - \frac{1}{2} && = 0.50000 \\ S_3 &= 1 - \frac{1}{2} + \frac{1}{3} \approx 0.5 + 0.33333 && = 0.83333 \\ S_4 &= S_3 - \frac{1}{4} \approx 0.83333 - 0.25 && = 0.58333 \\ S_5 &= S_4 + \frac{1}{5} \approx 0.58333 + 0.2 && = 0.78333 \end{aligned}$$

Figure 11.5.3 is a graph of S_n as a function of n. This graph confirms that the even partial sums, S_2, S_4, S_6, $\ldots$, approach S from below: $S > S_4 \approx 0.58333$. It also suggests that the odd partial sums, S_1, S_3, S_5, $\ldots$, approach S from above: $S < S_5 \approx 0.78333$. From the first five partial sums we know the sum of the alternating series, S, is between S_4

and S_5: $0.58333 < S < 0.78333$. More generally, $S_{2n} < S < S_{2n+1}$, with the estimate becoming increasing accurate as n increases. ♦

As Figure 11.5.3 suggests, a partial sum of any series satisfying the hypotheses of the alternating series test differs from the sum of the series by less than the absolute value of the first omitted term. That is, if S_n is the sum of the first n terms of the series and S is the sum of the series, then the error $R_n = S - S_n$ has absolute value at most the absolute value of the first omitted term. Moreover, S is between S_n and S_{n+1} for every n. To be precise, if n is even, then $S_n < S < S_{n+1}$ and, if n is odd, then $S_{n+1} < S < S_n$.

EXAMPLE 2. Does the series $\frac{3}{1!} - \frac{3^2}{2!} + \frac{3^3}{3!} - \frac{3^4}{4!} + \frac{3^5}{5!} - \cdots + (-1)^{k+1}\frac{3^k}{k!} + \cdots$ converge or diverge?

SOLUTION This is an alternating series. By Example 5 of Section 11.1, its k^{th} term approaches 0. Let us see whether the absolute values of the terms decrease in size, term-by-term. The first few absolute values are

$$\frac{3}{1!} = 3, \qquad \frac{3^2}{2!} = \frac{9}{2} = 4.5, \qquad \frac{3^3}{3!} = \frac{27}{6} = 4.5, \quad \text{and} \quad \frac{3^4}{4!} = \frac{81}{24} = 3.375.$$

At first, they increase. However, the fourth term is less than the third. Let us show that the rest of the terms decrease in size. For instance,

$$\frac{3^5}{5!} = \frac{3}{4}\frac{3^4}{4!} < \frac{3^4}{4!},$$

and, for $k \geq 3$,

$$\frac{3^{k+1}}{(k+1)!} = \frac{3}{k+1}\frac{3^k}{k!} < \frac{3^k}{k!}.$$

By the alternating series test, the tail end that begins

$$\frac{3^3}{3!} - \frac{3^4}{4!} + \frac{3^5}{5!} - \frac{3^6}{6!} - \cdots$$

converges. Call its sum S. If the front end

$$\frac{3}{1!} - \frac{3^2}{2!}$$

is added, we obtain the original series, which therefore converges and has sum $3/(1!) - 3^2/(2!) + S$. ♦

Observation 11.5.2:

As Example 2 illustrates, the alternating series test works as long as the k^{th} term approaches 0 and the terms decrease in size from some point on.

Does any alternating series whose k^{th} term approaches 0 converge? This is not the case, as shown by

$$\frac{2}{1} - \frac{1}{1} + \frac{2}{2} - \frac{1}{2} + \frac{2}{3} - \frac{1}{4} + \cdots, \tag{11.5.1}$$

whose terms alternate $+2/k$ and $-1/k$. Let S_n be the sum of the first n terms in (11.5.1). Then

$$\begin{aligned}
S_2 &= \frac{2}{1} - \frac{1}{1} &&= \frac{1}{1},\\
S_4 &= \left(\frac{2}{1} - \frac{1}{1}\right) + \left(\frac{2}{2} - \frac{1}{2}\right) &&= \frac{1}{1} + \frac{1}{2},\\
S_6 &= \left(\frac{2}{1} - \frac{1}{1}\right) + \left(\frac{2}{2} - \frac{1}{2}\right) + \left(\frac{2}{3} - \frac{1}{3}\right) &&= \frac{1}{1} + \frac{1}{2} + \frac{1}{3},
\end{aligned}$$

and, in general,

$$S_{2n} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

Notice that the partial sums with an even number of summands are the partial sums of the harmonic series and, even though the terms become closer and closer to zero, their sum diverges. Since S_{2n} gets arbitrarily large as $n \to \infty$, the series (11.5.1) diverges.

Absolute Convergence

If a series $a_1 + a_2 + \cdots + a_n \cdots$, terms that may be positive, negative, or zero it is reasonable to expect it to behave at least as nicely as the corresponding series with nonnegative terms $|a_1| + |a_2| + \cdots + |a_n| + \cdots$. This is true because making all the terms positive gives the series more chances to diverge. This is similar to the improper integrals in Section 7.8, where it was shown that if $\int_a^\infty |f(x)|\,dx$ converges then so does $\int_a^\infty f(x)\,dx$. The next theorem (and its proof) is similar to the absolute convergence test for improper integrals in Section 7.8.

Theorem 11.5.3: Absolute Convergence Test

If the series $\sum_{k=1}^{\infty} |a_k|$ *converges, then so does* $\sum_{k=1}^{\infty} a_k$*. In fact, if* $\sum_{k=1}^{\infty} |a_k| = S$*, then* $\sum_{k=1}^{\infty} a_k$ *is between* $-S$ *and* S*.*

Proof of the Absolute Convergence Test (Theorem 11.5.3

We introduce two series, one for the positive terms and one for the negative terms in $\sum_{k=1}^{\infty} a_k$. Let

$$b_k = \begin{cases} a_k & \text{if } a_k \text{ is positive} \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad c_k = \begin{cases} a_k & \text{if } a_k \text{ is negative} \\ 0 & \text{otherwise} \end{cases}.$$

Then $a_k = b_k + c_k$. To establish the convergence of $\sum_{k=1}^{\infty} a_k$ we show that $\sum_{k=1}^{\infty} b_k$ and $\sum_{k=1}^{\infty} c_k$ converge. Because b_k is nonnegative and $b_k \le |a_k|$, the series of positive terms, $\sum_{k=1}^{\infty} b_k$, converges by the comparison test; its sum is a number P in $[0, S]$.

Since c_k is nonpositive and $c_k \ge -|a_k|$, the series of negative terms, $\sum_{k=1}^{\infty} c_k$, converges; its sum is a number N in $[-S, 0]$.

Thus $\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty} (b_k + c_k)$ converges to $P + N$, which is between $-S$ and S. ●

EXAMPLE 3. For what values of x does the series $\dfrac{\cos(x)}{1^2} + \dfrac{\cos(2x)}{2^2} + \dfrac{\cos(3x)}{3^2} + \cdots + \dfrac{\cos(kx)}{k^2} + \cdots$ converge? diverge?

SOLUTION Depending on the value of x, the numbers $\cos(kx)$ may be positive, negative, or zero, in an irregular manner. However, for all k, $|\cos(kx)| \le 1$.

The positive-term series

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{k^2}$$

is the p-series with $p = 2$, which was shown to converge by the integral test in Section 11.2.

Since $|\cos(kx)/k^2| \le 1/k^2$, the series

$$\frac{|\cos(x)|}{1^2} + \frac{|\cos(2x)|}{2^2} + \frac{|\cos(3x)|}{3^2} + \cdots + \frac{|\cos(kx)|}{k^2} + \cdots$$

converges by the comparison test for any x. And, finally, Theorem 11.5.3 tells us that the original series converges for all x. ♦

NOTE: In Section 12.7 it is shown that the series in Example 3 sums to $(3x^2 - 6\pi x + 2\pi^2)/12$ for $0 \le x \le 2\pi$.

Observation 11.5.4: *The Converse of Theorem 11.5.3 is False*

If $\sum_{k=0}^{\infty} a_k$ converges, then $\sum_{k=0}^{\infty} |a_k|$ may converge or diverge. A counterexample to the converse of Theorem 11.5.3 is the alternating harmonic series, $1/1 - 1/2 + 1/3 - \cdots$. It converges, as was shown by the alternating series test (Theorem 11.5.1). But when terms are replaced by their absolute values, we obtain the series $1/1 + 1/2 + 1/3 + \cdots$, which diverges (it is the harmonic series).

The following definitions describe two different kinds of convergence.

Definition: *Absolute and Conditional Convergence*

1. A series $\sum_{k=1}^{\infty} a_k = a_1 + a_2 + \cdots$ is said to **converge absolutely** if the series $\sum_{k=1}^{\infty} |a_k| = |a_1| + |a_2| + \cdots$ converges.
2. A series $\sum_{k=1}^{\infty} a_k = a_1 + a_2 + \cdots$ is said to **converge conditionally** if it converges but does not converge absolutely, that is, if $\sum_{k=1}^{\infty} a_k$ converges and $\sum_{k=1}^{\infty} |a_k|$ diverges.

Theorem 11.5.3 can then be stated as

"if a series converges absolutely, then it converges."

Observation 11.5.5: *The Alternating Harmonic Series* $1 - 1/2 + 1/3 - 1/4 + \cdots$ *is Conditionally Convergent*

In a conditionally convergent series the terms do not approach 0 fast enough for the series to converge if all the terms were made positive. However, the negative terms help prevent the sum of the positive terms from growing arbitrarily large.

Rearrangements

The sum of a finite number of numbers does not depend on the order in which they are added. A series that converges absolutely is similar: however its terms are rearranged, the new series converges and to the same sum.

It may be a surprise to learn that this is not true for the alternating harmonic series

$$\frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots. \tag{11.5.2}$$

To show this, rearrange the terms so that two positive terms alternate with one negative term:

$$\frac{1}{1} + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \cdots. \tag{11.5.3}$$

We emphasize that (11.5.2) and (11.5.3) involve the same summands, just in different orders. It turns out that the sum in (11.5.3) is $3\ln(2)/2$, while the sum of (11.5.2) is $\ln(2)$. See also Exercise 28 in Section 12.4.

Warning: Rearranging Terms in a Conditionally Convergent Series is Dangerous

The fact that rearranging the terms in the alternating harmonic series changes the sum of the series is not limited to this one example. In fact, every conditionally convergent series can be made to sum to any number that one chooses by a suitable rearrangement of its terms.

Summary

Earlier in this chapter we described ways to test for the convergence or divergence of series whose terms are all positive. This section described tests for series that may have positive and negative terms.

The alternating series test says: if (i) the signs of the terms alternate and (ii) the absolute value of the terms decrease and approach 0, then the series converges.

The absolute convergence test involves an alternating series and the corresponding series when all the terms are made positive; if the series with positive terms converges, then the alternating series also converges.

It is important to understand these tests: when they apply, when they do not apply, and what they say — both formally and informally.

EXERCISES for Section 11.5

In Exercises 1 to 8 determine which series converge and which diverge. Explain your reasoning.

1. $\frac{1}{2}-\frac{2}{3}+\frac{3}{4}-\frac{4}{5}+\cdots+(-1)^{k+1}\frac{k}{k+1}+\cdots$

2. $-\frac{1}{1+\frac{1}{2}}+\frac{1}{1+\frac{1}{4}}-\frac{1}{1+\frac{1}{8}}+\cdots+(-1)^k\frac{1}{1+\frac{1}{2^k}}+\cdots$

3. $\frac{1}{\sqrt{1}}-\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{3}}-\frac{1}{\sqrt{4}}+\cdots+(-1)^{k+1}\frac{1}{\sqrt{k}}+\cdots$

4. $\frac{5}{1!}-\frac{5^2}{2!}+\frac{5^3}{3!}-\frac{5^4}{4!}+\cdots+(-1)^{k+1}\frac{5^k}{k!}+\cdots$

5. $\frac{3}{\sqrt{1}}-\frac{2}{\sqrt{1}}+\frac{3}{\sqrt{2}}-\frac{2}{\sqrt{2}}+\frac{3}{\sqrt{3}}-\frac{2}{\sqrt{3}}+\cdots$

6. $\sqrt{1}-\sqrt{2}+\sqrt{3}-\sqrt{4}+\cdots+(-1)^{k+1}\sqrt{k}+\cdots$

7. $\frac{1}{3}-\frac{2}{5}+\frac{3}{7}-\frac{4}{9}+\frac{5}{11}-\cdots+(-1)^{k+1}\frac{k}{2k+1}+\cdots$

8. $\frac{1}{1^2}-\frac{1}{2^2}+\frac{1}{3^2}-\frac{1}{4^2}+\cdots+(-1)^{k+1}\frac{1}{k^2}+\cdots$

---

9. Answer the following questions for the alternating harmonic series $\sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k}$. (a) Compute S_5 and S_6 to five decimal places. (b) Is the estimate S_5 smaller or larger than the sum of the series? (c) Use (a) and (b) to find two numbers between which the sum of the series must lie.

10. For the series $\sum_{k=1}^{\infty}(-1)^{k+1}\frac{2^{-k}}{k}$. (a) Estimate its sum of the series using S_6. (b) Estimate the error R_6.

11. Does the series $\frac{2}{1}-\frac{3}{2}+\frac{4}{3}-\frac{5}{4}+\cdots+(-1)^{k+1}\frac{n+1}{n}+\cdots$ converge or diverge? Why?

12. The series $\sum_{k=1}^{\infty}(-1)^{k+1}2^{-k}$ is both a geometric series and a decreasing alternating series whose k^{th} term approaches 0. (a) Compute S_6 to three decimal places. (b) Using the fact that the series is a decreasing alternating series, put a bound on R_6. (c) Using the fact that the series is a geometric series, compute R_6 exactly.

In Exercises 13 to 20 determine whether the series diverges, converges absolutely, or converges conditionally. Explain your answers. NOTE: In Exercise 20, two +'s alternate with two −'s.

13. $\sum_{k=1}^{\infty}\frac{(-1)^k}{\sqrt[3]{k^2}}$

14. $\sum_{k=1}^{\infty}\ln\left(\frac{1}{k}\right)$

15. $\sum_{k=1}^{\infty}\frac{(-1)^k}{k\ln(k)}$

16. $\sum_{k=1}^{\infty}\frac{\sin(k)}{k^{1.01}}$

17. $\sum_{k=1}^{\infty}\left(1-\cos\left(\frac{\pi}{k}\right)\right)$

18. $\sum_{k=1}^{\infty}(-1)^k\cos\left(\frac{\pi}{k^2}\right)$

19. $\sum_{k=1}^{\infty}\frac{(-2)^k}{k!}$

20. $\frac{1}{1^2}+\frac{1}{2^2}-\frac{1}{3^2}-\frac{1}{4^2}+\frac{1}{5^2}+\frac{1}{6^2}-\cdots$

---

21. The alternating series test was proved only for the alternating harmonic series. Prove it in general.

22. Add the alternating harmonic series to half of itself:

$$\begin{array}{cccccccccccccccc}
1 & -\frac{1}{2} & +\frac{1}{3} & -\frac{1}{4} & +\frac{1}{5} & -\frac{1}{6} & +\frac{1}{7} & -\frac{1}{8} & +\frac{1}{9} & -\frac{1}{10} & -\frac{1}{11} & +\frac{1}{12} & +\cdots & = & S \\
 & \frac{1}{2} & & -\frac{1}{4} & & +\frac{1}{6} & & -\frac{1}{8} & & +\frac{1}{10} & & -\frac{1}{12} & +\cdots & = & \frac{1}{2}S \\
\hline
1 & & +\frac{1}{3} & -\frac{1}{2} & +\frac{1}{5} & & +\frac{1}{7} & -\frac{1}{4} & +\frac{1}{9} & & -\frac{1}{11} & & +\cdots & = & \frac{3}{2}S
\end{array}$$

Rearranging the last line produces the alternating harmonic series, whose sum is S. Thus $S = \frac{3}{2}S$, from which it follows that $S = 0$. What is wrong with this argument?

23. (a) How many terms of $\sum_{k=1}^{\infty} \frac{\sin(k)}{k^2}$ are needed to be sure the error in using their sum to estimate the sum of the series is less than 0.005? Explain. (b) Estimate $\sum_{k=1}^{\infty} \frac{\sin(k)}{k^2}$ to two decimal places.

24. Estimate $\sum_{k=0}^{\infty} (-1)^k \frac{1}{k!} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots$ to two decimal places. Explain your reasoning.

A sequence that has both positive and negative terms can be divided into two sequences, the sequence formed by the negative terms in order and the sequence of positive terms in order. For example the terms of the alternating harmonic series $\sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k}$ can be separated into the sequence of negative terms $\frac{-1}{2}, \frac{-1}{4}, \frac{-1}{6}, \ldots$ and the sequence of positive terms $\frac{1}{1}, \frac{1}{3}, \frac{1}{5}, \ldots$ Exercises 25 to 28 concern the series formed by summing the terms in these two sequences.

25. If $\sum_{k=1}^{\infty} a_k$ is absolutely convergent, what, if anything, can be said about the convergence of the series associated with its sequences of negative and positive terms?

26. If $\sum_{k=1}^{\infty} a_k$ is conditionally convergent, what, if anything, can be said about the convergence of the series associated with its sequences of negative and positive terms?

27. If the series associated with the negative sequence and the series associated with the positive sequence both converge, what, if anything, can be said about the convergence of the series associated with the full sequence?

28. Let $\sum_{k=1}^{\infty} a_k$ be a series whose terms are both positive and negative. Assume the series associated with the sequence of negative terms converges to -7 and the series associated with the sequence of positive terms converges to 11. What, if anything, can be said about the series $\sum_{k=1}^{\infty} a_k$?

---

29. Assume that $p_1 > p_2 > p_3 > \cdots > p_k > \cdots$ is a sequence with $\lim_{k\to\infty} p_k = 0$ and $p_1 - p_2 + p_3 - p_4 + \cdots + (-1)^{k+1} p_k + \cdots$ converges. Form a new sequence $q_1, q_2, q_3, \ldots$ by grouping the p_k's in pairs, $q_1 = p_1 - p_2$, $q_2 = p_3 - p_4$, $q_3 = p_5 - p_6$, and so on. (a) Show that all the q_k are positive. (b) Does $\sum_{k=1}^{\infty} q_k$ converge?

30. Could a conditionally convergent series have only a finite number of negative terms? Explain the reasoning to support your answer.

31. A series satisfies the assumptions of the alternating series test and converges conditionally to 8. Let N_k be the sum of the first k negative terms and P_k the sum of the first k positive terms. What, if anything, can be said about

(a) $\lim_{k\to\infty} P_k$? (b) $\lim_{k\to\infty} N_k$? (c) $\lim_{k\to\infty} (P_k + N_k)$? (d) $\lim_{k\to\infty} (P_k - N_k)$? (e) $\lim_{k\to\infty} \frac{P_k}{N_k}$?

32. **SAM:** I have a neat proof that absolute convergence implies convergence. First of all, $a_n = a_n + |a_n| - |a_n|$.
JANE: True, but why do that?
SAM: Don't interrupt. Just wait. Now $a_n + |a_n|$ is 0 if a_n is negative and it's $2|a_n|$ if a_n is positive. Right?
JANE: If you say so.
SAM: Just think.
JANE: Yes, I agree.
SAM: So $0 \le a_n + |a_n| \le 2|a_n|$. Right? So $\sum_{n=1}^{\infty} (a_n + |a_n|)$ converges.
JANE: Yes.
SAM: You can fill in the rest, yes?
JANE: Oh, neat.
SAM: Yeh, mathematicians really like this proof.

Is the proof correct? Explain your answer. What are the advantages and disadvantages of each proof?

33. In the proof of the absolute convergence theorem, why does $\sum_{k=1}^{\infty} c_k$ converge and have a sum greater than or equal to $-S$?

34. The absolute convergence test asserts that $\sum_{k=1}^{\infty} a_k$ is between $-S$ and S. Why is that?

11.6 Applications of the Absolute Convergence Test

This section describes two tests for convergence that are formed by combining the absolute convergence test with the limit comparison test and with the ratio test. These new tests are shortcuts that allow for the analysis of many series in one step instead of having to apply two separate tests in succession. These tests will be used repeatedly in Chapter 12.

Absolute Limit Comparison Test

Combining the limit comparison test for positive series with the absolute convergence test gives

Theorem 11.6.1: Absolute Limit Comparison Test

Let $\sum_{k=1}^{\infty} a_k$ be a series whose terms may be negative or positive. Let $\sum_{k=1}^{\infty} c_k$ be a convergent series of positive terms. If $\lim_{k\to\infty} |\frac{a_k}{c_k}|$ exists, then $\sum_{k=1}^{\infty} a_k$ is absolutely convergent, and hence convergent.

Proof of Theorem 11.6.1

Because c_k is positive, $|a_k/c_k| = |a_k|/c_k$. The limit comparison test tells us that $\sum_{k=1}^{\infty} |a_k|$ converges. Then the absolute convergence test assures us that $\sum_{k=1}^{\infty} a_k$ converges.

EXAMPLE 1. Show that $\frac{3}{2}\left(\frac{1}{2}\right)-\frac{5}{2}\left(\frac{1}{2}\right)^2+\frac{7}{3}\left(\frac{1}{2}\right)^3-\cdots+(-1)^{k+1}\frac{2k+1}{k}\left(\frac{1}{2}\right)^k+\cdots$ converges.

SOLUTION Consider the associated series with positive terms:

$$\frac{3}{2}\left(\frac{1}{2}\right)+\frac{5}{2}\left(\frac{1}{2}\right)^2+\frac{7}{3}\left(\frac{1}{2}\right)^3+\cdots+\frac{2k+1}{k}\left(\frac{1}{2}\right)^k+\cdots.$$

Because $(2k+1)/k \to 2$ as $k\to\infty$, the limit comparison test can be used to compare the second series to the convergent geometric series $\sum_{k=1}^{\infty}(1/2)^k$. We have

$$\lim_{k\to\infty}\frac{\frac{2k+1}{k}\left(\frac{1}{2}\right)^k}{\left(\frac{1}{2}\right)^k}=2.$$

Thus the series with positive terms, $\sum_{k=1}^{\infty}(2k+1)/(k2^k)$, converges. Consequently, the original series, with both positive and negative terms, converges absolutely. Hence it also converges. ♦

Absolute Ratio Test

The ratio test of Section 11.4 also has an analog that applies to series with both negative and positive terms.

Theorem 11.6.2: Absolute Ratio Test

Let $\sum\limits_{k=1}^{\infty} a_k$ *be a series such that* $\lim\limits_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right|=r$.

(a) *If* $r<1$, *then* $\sum\limits_{k=1}^{\infty} a_k$ *converges.*

(b) *If either* (i) $r>1$ *or* (ii) $\lim\limits_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right|=\infty$, *then* $\sum\limits_{k=1}^{\infty} a_k$ *diverges.*

(c) *If* $r=1$, *then the absolute ratio test gives no information.*

Observation 11.6.3: *The Absolute Ratio Test Avoids Having to Work with Minus Signs*

Theorem 11.6.2 establishes the convergence of the series in Example 1. Let $a_k=(-1)^{k+1}(2k+1)/(k2^k)$. Then

$$\left|\frac{a_{k+1}}{a_k}\right|=\left|\frac{(-1)^{k+2}\frac{(2k+3)}{(k+1)2^{k+1}}}{(-1)^{k+1}\frac{(2k+1)}{k2^k}}\right|=\frac{2k+3}{2k+1}\cdot\frac{k}{k+1}\cdot\frac{1}{2},$$

which approaches $r=\frac{1}{2}$ as $k\to\infty$. Thus $\sum\limits_{k=1}^{\infty} a_k$ converges absolutely.

Proof of Absolute Ratio Test (Theorem 11.6.2)

(a) If $r<1$, the ratio test shows that $\sum_{k=1}^{\infty}|a_k|$ converges. Since $\sum_{k=1}^{\infty}|a_k|$ converges, then $\sum_{k=1}^{\infty}a_k$ also converges.
(b) If $\lim_{k\to\infty}|a_{k+1}/a_k|=\infty$, the ratio $|a_{k+1}|/|a_k|$ gets arbitrarily large as $k\to\infty$. From some point on, the positive numbers $|a_k|$ increase. By the n^{th}-term test for divergence, $\sum_{k=1}^{\infty}a_k$ is divergent. (For $r>1$, see Exercise 17.)
(c) In the case where $r=1$ there is nothing to prove. The examples used to prove Theorem 11.4.1(c) in Section 11.4 are easily modified for this setting.

●

Summary

The two tests introduced in this section provide tests for absolute convergence of a sequence $\sum_{k=1}^{\infty} a_k$.

The absolute limit comparison test is applicable when one can identify a convergent series $\sum_{k=1}^{\infty} c_k$ with positive terms for which $\lim_{k\to\infty} \frac{|a_k|}{c_k}$ exists. In that case the original series $\sum_{k=1}^{\infty} a_k$ converges absolutely.

The absolute ratio test applies when $\lim_{n\to\infty} |a_{k+1}/a_k| = r$. If $r < 1$, $\sum_{k=1}^{\infty} a_k$ converges absolutely. For $r > 1$, $\sum_{k=1}^{\infty} a_k$ diverges.

EXERCISES for Section 11.6

In Exercises 1 to 14 (a) if the absolute limit comparison test applies, what does it say about the convergence or divergence of the series? (b) if the absolute ratio test applies, what does it say about the convergence or divergence of the series?

1. $\sum_{k=1}^{\infty} (-1)^k \frac{3k+2}{3k-1}\left(\frac{2}{3}\right)^k$

2. $\sum_{k=1}^{\infty} (-1)^k \frac{k^2+k}{k^2+2k}\left(\frac{5}{7}\right)^k$

3. $\sum_{k=1}^{\infty} \frac{3k+1}{(4k+2)(-2)^k}$

4. $\sum_{k=1}^{\infty} \frac{k^2+k}{(3k^2+5)(-3)^k}$

5. $\sum_{k=1}^{\infty} (-1)^k \frac{\ln(k+2)}{2^k}$

6. $\sum_{k=1}^{\infty} (-1)^k \frac{k\ln(k+1)}{3^k}$

7. $\sum_{k=1}^{\infty} (-1)^k \frac{\sin(1/k)}{2^k}$

8. $\sum_{k=1}^{\infty} (-1)^k \frac{\tan^2(1/k)}{2^k}$

9. $\sum_{k=1}^{\infty} \sin^2(1/k)$

10. $\sum_{k=1}^{\infty} \frac{3k+2}{k}$

11. $\sum_{k=1}^{\infty} \frac{(-1.01)^k}{k!}$

12. $\sum_{k=1}^{\infty} \frac{3k+2}{3k-1}\left(\frac{2}{3}\right)^k$

13. $\sum_{k=1}^{\infty} \frac{(-9)^k}{10^k+k}$

14. $\sum_{k=1}^{\infty} \frac{(-\pi)^{2k+1}}{(2k+1)!}$

---

15. If $\sum_{k=1}^{\infty} c_k$ is a divergent series of positive terms and $\lim_{k\to\infty} \frac{|a_k|}{c_k} = 3$, must $\sum_{k=1}^{\infty} a_k$ also be divergent?

16. Theorem 11.5.1 does not mention divergence. If $\sum_{k=1}^{\infty} c_k$ is a divergent series of positive terms and $\lim_{k\to\infty} \frac{|a_k|}{c_k}$ exists, must $\sum_{k=1}^{\infty} a_k$ also be divergent?

17. This exercise treats the case (i) in the second part of the absolute ratio test, when $r > 1$.

(a) Show that if $r = \lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right| > 1$, then $|a_k| \to \infty$ as $k \to \infty$.

(b) From (a) deduce that a_k does not approach 0 as $k \to \infty$.

18. If $\sum_{k=1}^{\infty} a_k$ converges absolutely, determine the convergence or divergence of $\sum_{k=1}^{\infty} \sin(a_k)$.

19. If $\sum_{k=1}^{\infty} a_k$ converges and $a_k > 0$ for all k, determine the convergence or divergence of

(a) $\sum_{k=1}^{\infty} \sin(a_k)$ and (b) $\sum_{k=1}^{\infty} \cos(a_k)$.

20. Let $P(x)$ and $Q(x)$ be polynomials of degree at least one. Assume that for $n \geq 1$, $Q(n) \neq 0$. What relation must there be between the degrees of $P(x)$ and $Q(x)$

(a) if $\lim_{k\to\infty} \frac{P(k)}{Q(k)} = 0$, (b) if $\sum_{k=1}^{\infty} \frac{P(k)}{Q(k)}$ converges absolutely, and (c) if $\sum_{k=1}^{\infty} (-1)^k \frac{P(k)}{Q(k)}$ converges absolutely.

21. (a) Show $\sum_{k=1}^{\infty} \frac{2^k}{k!}$ converges and (b) estimate its sum to two decimal places.

22. Assume that $a_1, a_2, a_3, \ldots$ is a sequence of positive terms such that $\sum_{k=1}^{\infty} a_k$ converges. Why does every rearrangement of the terms produce a convergent sequence with the same sum?

23. Show that there is a rearrangement of the terms of the alternating harmonic series that sums to $\sqrt{2}$.

(a) Exhibit the first twelve summands in the new series. (b) Why do the sums converge to $\sqrt{2}$?

24. This exercise implies that in the process in Exercise 23 one never encounters a sum that is an integer. It also implies that the only partial sum of the harmonic series that is an integer is the one with only one term, 1.

(a) In a finite sequence of consecutive even integers each one is divisible by a power of 2. Show that only one of them is divisible by the highest power of 2 that divides any of them. For instance, in the sequence 18, 20, 22, 24, 26, 28 the highest power of 2 that divides any of them is 2^3, and it divides only 24.

(b) Show that for $n > 1$ the partial sum of the first n terms of the harmonic series is not an integer. (Assume it is an integer, m. Let 2^c be the largest power of 2 that divides the denominators. To obtain a contradiction, multiply both sides of the equation $\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = m$ by the product of 2^{c-1} and all the odd denominators.)

(c) Show why the process in Exercise 23 never produces a sum that is an integer.

11.S Chapter Summary

This chapter concerned sequences formed by adding a finite number of terms from another sequence: $S_n = a_1 + a_2 + \cdots + a_n$. There were two key questions:

(i) Does the sequence of partial sums converge, that is, doeslim$_{n\to\infty} S_n$ exist?

(ii) If the sequence of partial sums converges, what is its value?

If the sequence of partial sums converges, its limit is denoted $\sum_{k=1}^{\infty} a_k$ — though we never add an infinite number of summands.

Some of the tests for convergence or divergence of an infinite series apply only to series whose terms all have one sign (typically positive): the integral test, the comparison tests, and the ratio and root tests. For series whose terms a_k may be both positive and negative, the key is to understand absolute convergence: if $\sum_{k=1}^{\infty} |a_k|$ converges, then so must $\sum_{k=1}^{\infty} a_k$, and if $\sum_{k=1}^{\infty} |a_k| = L$, then $-L \leq \sum_{k=1}^{\infty} a_k \leq L$.

If the series has terms that alternate in sign, $a_1 - a_2 + a_3 - a_4 + \cdots$, and $a_k \to 0$ monotonically, then $\sum_{k=1}^{\infty} a_k$ converges. This is the alternating series test.

The integral and alternating series tests and the formula for the sum of a geometric series provide ways to estimate the error in using a partial sum S_n to approximate the sum.

EXERCISES for Section 11.S

1. Explain in your own words.

(a) Why the comparison test for convergence works.

(b) Why the ratio test for convergence works.

(c) Why the alternating series test works.

(d) Why the absolute convergence test works.

2. How many terms of the series $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^2}$ need to be used to estimate its sum to three decimal place accuracy?

3. For what type of series does the test given below imply convergence? (a) alternating series test, (b) integral test, (c) comparison test, (d) absolute convergence test, and (e) absolute ratio test.

4. Assume that $|a_k| \leq \frac{1}{2^k}$ for $k \geq 1$.

(a) Must $\sum_{k=1}^{\infty} |a_k|$ converge? If so, what can be said about its sum?

(b) Must $\sum_{k=1}^{\infty} a_k$ converge? If so, what can be said about its sum?

Sometimes convergence or divergence of a series can be determined by more than one test. In Exercises 5 to 10 determine the convergence or divergence of the series by as many tests as can be applied.

5. $\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}$

6. $\sum_{i=1}^{\infty} \frac{(-1)^i}{3^i}$

7. $\sum_{k=1}^{\infty} \frac{\sqrt{k}}{k^2+1}$

8. $\sum_{k=1}^{\infty} \frac{\sqrt{k}}{k^2-2}$

9. $\sum_{i=1}^{\infty} \left(\frac{3+1/i}{2+1/i}\right)^i$

10. $\sum_{n=1}^{\infty} \left(\frac{2}{3+1/n}\right)^n$

---

11. After using the (termwise) comparison test to determine a series converges, how might you go about estimating the error when a partial sum is used to estimate the sum of the series? Illustrate this with an example.

12. What does it mean when we say an infinite series (a) converges, (b) converges conditionally, and (c) converses absolutely?

13. What tests could be used to test a series for convergence if it is known that $\lim_{k\to\infty} \frac{a_{k+1}}{a_k} = \frac{-1}{3}$? Explain.

---

14. Assume that $\lim_{k\to\infty} a_k = 2$. For which values of s does $\sum_{k=1}^{\infty} a_k s^k$ converge?

15. For which values of p does $\sum_{k=1}^{\infty} \frac{1}{k^p}$ converge?

16. If $\lim_{k\to\infty} \frac{a_{k+1}}{a_k = 1}$, what can we conclude about the series $\sum_{k\to\infty} a_k$?

17. For which values of q does $\sum_{k=1}^{\infty} (-1)^k k^q$ (a) converge? (b) converge absolutely?

18. If $\sum_{k=0}^{\infty} a_k$ is convergent, which of the following must be true:

(a) $\lim_{k\to\infty} a_k = 0$, (b) $\lim_{k\to\infty} (a_k + a_{k+1}) = 0$, (c) $\lim_{n\to\infty} \sum_{k=n}^{2n} a_k = 0$, and (d) $\lim_{n\to\infty} \sum_{k=n}^{\infty} a_k = 0$.

19. Let $\sum_{k=0}^{\infty} a_k$ be a conditionally convergent series. It contains a subsequence of nonnegative terms and a subsequence of negative terms. (a) Could both subsequences be convergent? (b) Could exactly one of them be convergent? (c) Could neither be convergent?

20. (a) Graph $y = \dfrac{\cos(x)}{\sqrt{x}}$ and $y = \cos(x^2)$ for $x > 0$. (b) Show that $\int_0^\infty \dfrac{\cos(x)}{\sqrt{x}}\,dx$ is convergent. (Its value is $\sqrt{\pi/2}$.)
(c) Use the results in (b) to show that $\int_0^\infty \cos(x^2)\,dx$ is convergent and equals $\sqrt{\dfrac{\pi}{8}}$.

In Exercises 21 to 28 show that the integrals are convergent. Their values are found by more advanced techniques, including the use of complex numbers.

21. $\int_0^\infty e^{-nx}\sqrt{x}\,dx = \dfrac{1}{2n}\sqrt{\dfrac{\pi}{n}},\ n > 0.$

22. $\int_0^\infty e^{-ax}\cos(mx)\,dx = \dfrac{a}{a^2+m^2},\ a > 0.$

23. $\int_0^\infty e^{-ax}\sin(mx)\,dx = \dfrac{m}{a^2+m^2},\ a > 0.$

24. $\int_0^\infty \sin(x^2)\,dx = \dfrac{1}{2}\sqrt{\dfrac{\pi}{2}}.$

25. $\int_0^\infty \dfrac{\cos(mx)}{1+x^2}\,dx = \dfrac{\pi}{2}e^{-m},\ m > 0.$

26. $\int_0^\infty \dfrac{\sin^2(x)}{x^2}\,dx = \dfrac{\pi}{2}.$

27. $\int_0^\infty \dfrac{\ln(x)}{1+x}\,dx = \dfrac{-\pi^2}{12}.$

28. $\int_0^\infty \dfrac{\ln(x)}{1-x}\,dx = \dfrac{-\pi^2}{6}.$

---

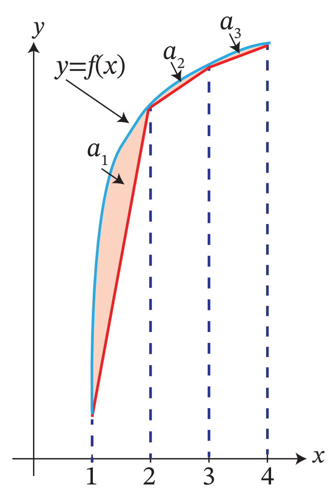
29. Give an example of a convergent series of positive terms $\{a_k\}$ such that $\lim_{k\to\infty} \dfrac{a_{k+1}}{a_k}$ does not exist and is not ∞.

Exercises 30 to 35 provide a formula for estimating $n!$.

30. Let f be such that for $x \geq 1$, $f(x) \geq 0$, $f'(x) > 0$, and $f''(x) < 0$. Let a_n be the area of the region below the graph of $y = f(x)$ and above the line segment that joins $(n, f(n))$ with $(n+1, f(n+1))$, $n = 1, 2, 3, \ldots$.

(a) Draw a large version of Figure 11.S.1. The regions of areas a_1, a_2, a_3, and a_4 should be clear and not too narrow.

(b) Using Figure 11.S.1, show that the series $a_1 + a_2 + a_3 + \cdots$ converges and has a sum no larger than the area of the triangle with vertices $(1, f(1))$, $(2, f(2))$, $(1, f(2))$.



$a_1 + a_2 + a_3 \cdots$ converges

Figure 11.S.1

31. Let $y = \ln(x)$.

(a) Using Exercise 30, show that as $n \to \infty$,

$$\int_1^n \ln(x)\,dx - \left(\frac{\ln(1)+\ln(2)}{2} + \frac{\ln(2)+\ln(3)}{2} + \cdots + \frac{\ln(n-1)+\ln(n)}{2}\right)$$

has a limit, C.

(b) Show that (a) is equivalent to $\lim_{n\to\infty} \left(n\ln(n) - n + 1 - \ln(n!) + \ln(\sqrt{n})\right) = C$.

32. From Exercise 31(b), deduce that there is a constant k such that $\lim_{n\to\infty} \dfrac{n!}{k(n/e)^n\sqrt{n}} = 1$.

Exercises 33 and 34 are related. Review Example 8 of Section 8.3 first.

33. Let $I_n = \int_0^{\pi/2} \sin^n(\theta)\, d\theta$, where n is is a nonnegative integer.

(a) Evaluate I_0 and I_1.

(b) Show that $I_{2n} = \frac{2n-1}{2n} \cdot \frac{2n-3}{2n-2} \cdot \dots \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{\pi}{2}$ and $I_{2n+1} = \frac{2n}{2n+1} \cdot \frac{2n-2}{2n-1} \cdot \dots \cdot \frac{4}{5} \cdot \frac{2}{3}$.

(c) Show that $\frac{I_7}{I_6} = \frac{6}{7} \cdot \frac{6}{5} \cdot \frac{4}{5} \cdot \frac{4}{3} \cdot \frac{2}{3} \cdot \frac{2}{1} \cdot \frac{2}{\pi}$.

(d) Show that $\frac{I_{2n+1}}{I_{2n}} = \frac{2n}{2n+1} \cdot \frac{2n}{2n-1} \cdot \frac{2n-2}{2n-1} \cdot \dots \cdot \frac{2}{3} \cdot \frac{2}{1} \cdot \frac{2}{\pi}$.

(e) Show that $\frac{2n}{2n+1} I_{2n} < \frac{2n}{2n+1} I_{2n-1} = I_{2n+1} < I_{2n}$, and thus $\lim_{n\to\infty} \frac{I_{2n+1}}{I_{2n}} = 1$.

(f) From (d) and (e), deduce that $\lim_{n\to\infty} \frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6\cdot \dots}{1\cdot 3\cdot 3\cdot 5\cdot 5\cdot 7\dots} = \frac{\pi}{2}$. This is **Wallis's formula**, which is usually written as $\frac{2\cdot 2}{1\cdot 3}\frac{4\cdot 4}{3\cdot 5}\frac{6\cdot 6}{5\cdot 7}\cdots = \frac{\pi}{2}$.

34. (a) Show that $2\cdot 4\cdot 6\cdot 8\cdot \dots \cdot 2n = 2^n n!$.

(b) Show that $1\cdot 3\cdot 5\cdot 7\cdot \dots \cdot (2n-1) = \frac{(2n)!}{2^n n!}$.

(c) From Exercise 33 deduce that $\lim_{n\to\infty} \frac{(n!)^2 4^n}{(2n)!\sqrt{2n+1}} = \sqrt{\frac{\pi}{2}}$.

35. (a) Using Exercise 34(c), show that k in Exercise 32 equals $\sqrt{2\pi}$. Thus a good estimate of $n!$ is provided by $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$, which is known as **Stirling's formula**.

(b) Using the factorial key on a calculator, compute $(20)!$. Then compute $\frac{\sqrt{2\pi n}}{n!}\left(\frac{n}{e}\right)^n$ for $n = 20$.

(c) What is the ratio between 20! and the estimate in (b)?

---

36. Let $\{a_k\}$ and $\{b_k\}$ be sequences of positive terms. Assume that for all k $\frac{a_{k+1}}{a_k} \le \frac{b_{k+1}}{b_k}$.

(a) Prove that if $\sum_{k=1}^{\infty} b_k$ converges, so does $\sum_{k=1}^{\infty} a_k$.

(b) Use the result in (a) to prove that if $\lim_{k\to\infty} \frac{a_{k+1}}{a_k} = r < 1$, then $\sum_{k=1}^{\infty} a_k$ converges.

37. An energy problem leads to the need to evaluate the definite integral $\int_0^{\pi/2} \frac{\sin x}{e^x - 1}\, dx$. The integrand is not defined at $x = 0$. Is the improper integral convergent or divergent? NOTE: Do not try to evaluate the integral.

Calculus is Everywhere # 13

$$E = mc^2$$

The equation $E = mc^2$ relates the energy associated with an object to its mass and the speed of light. Where does it come from? Newton's second law of motion says that force is the rate at which the momentum of an object changes. The momentum of an object of mass m and velocity v is mv. Denoting the force by F, we have

$$F = \frac{d}{dt}(mv). \tag{C.13.1}$$

If the mass is constant, this reduces to the familiar "force equals mass times acceleration." But what if the mass m is not constant?

According to Einstein's Special Theory of Relativity, announced in 1905, the mass of an object changes with its velocity so that

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{C.13.2}$$

Here m_0 is the mass at rest, v is the velocity, and c is the velocity of light. If v is not zero, m is larger than m_0. When v is small compared to the velocity of light m is only slightly larger than m_0. However, as v approaches the velocity of light, the mass becomes arbitrarily large.

NOTE: For a satellite circling the Earth at 17,000 miles per hour, v/c is less than $1/2500 = 0.0004$.

An object, initially at rest, moves in a straight line. If its velocity at time t is $v(t)$, then the displacement is $x(t) = \int_0^t v(s)\,ds$. Assuming the object is initially at rest, so $v(0) = 0$, the work done by a force F in moving the object during the time interval $[0, T]$ is

$$\begin{aligned}
\int_0^T F(t)v(t)\,dt &= \int_0^T (mv)'v\,dt && \text{(by (C.13.1))}\\
&= (mv)v\Big|_0^T - \int_0^T mv(v')\,dt && \text{(integration by parts)}\\
&= m(v(T))^2 - \int_0^T \frac{m_0 v v'}{\sqrt{1-\frac{v^2}{c^2}}}\,dt && \text{(using (C.13.2))}\\
&= m(v(T))^2 - \left(-c^2 m_0\sqrt{1-\frac{v^2}{c^2}}\right)\Bigg|_0^T && \text{(FTC I)}\\
&= m(v(T))^2 - \left(-c^2 m_0\sqrt{1-\frac{(v(T))^2}{c^2}} + c^2 m_0\sqrt{1-\frac{0^2}{c^2}}\right) && \text{(substitution: } u = 1 - v^2/c^2\text{)}\\
&= m(v(T))^2 + c^2 m_0\sqrt{1-\frac{(v(T))^2}{c^2}} - m_0c^2 && \text{(simplification)}\\
&= m(v(T))^2 + mc^2\left(1-\frac{(v(T))^2}{c^2}\right) - m_0c^2 && \text{(using (C.13.2))}\\
&= m(v(T))^2 + mc^2 - m(v(T))^2 - m_0c^2 && \text{(more simplification)}\\
&= mc^2 - m_0c^2.
\end{aligned} \tag{C.13.3}$$

We interpret (C.13.3) as saying that the total energy associated with the object increases from m_0c^2 to mc^2. The energy of the object at rest is then m_0c^2, called its rest energy.

That is the mathematics behind the equation $E = mc^2$. It suggests that mass may be turned into energy, as Einstein predicted. In a nuclear reactor some of the mass of the uranium is turned into energy in the fission process. Also, the mass of the sun decreases as it emits radiant energy.

What about the equation that states kinetic energy is half the product of the mass and the square of the velocity? Exercise 2 uses (C.13.3) when v is small (compared to c) to show the total increase in energy is approximated by the familiar kinetic energy

$$mc^2 - m_0c^2 \approx \frac{1}{2}m_0v^2. \tag{C.13.4}$$

EXERCISES for CIE C.13

1. Provide a brief explanation for each step in the derivation of (C.13.3).

2. (a) Use a linear approximation to show that when x is near 0, $f(x) = (1-x)^{-1/2}$ is approximately $1 + x/2$.

(b) Use (a) to show that (C.13.4) holds when v is small when compared to c.

Chapter 12

Applications of Series

The last chapter developed tests for determining the convergence or divergence of an infinite series. Section 12.1 introduces a special type of infinite series, called Taylor series, that are used to approximate functions such as e^x, $\sin(x)$, and $\ln(1+x)$. Taylor series are applied in Section 12.2 to evaluate integrals and to find limits in the indeterminate form zero-over-zero. Sections 12.3 and 12.4 develop the theory and methods for working with general power series.

The second half of the chapter begins by reviewing complex numbers in Section 12.5, then Section 12.6 show a connection between trigonometric and exponential functions.

The final application of infinite series is Fourier series. The presentation of Fourier series in Section 12.7 shows how these infinite sums of trigonometric functions can be particularly useful when approximating a periodic function.

12.1 Taylor Series

Section 5.5 introduced the n^{th}-order Taylor polynomial of a function f centered at a as the polynomial P_n that agreed with f and its first n derivatives at $x = a$:

$$P_n(x;a) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

The Taylor polynomials $P_0(x;a)$, $P_1(x;a)$, ..., $P_n(x;a)$, ... can be viewed as the sequence of partial sums of the infinite series

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

The series is called the **Taylor series at a for the function f**. When $a = 0$, the series is also called the **Maclaurin series for f**. A partial sum of a Taylor series is a **Taylor polynomial** and a partial sum of a Maclaurin series is a **Maclaurin polynomial**.

EXAMPLE 1. Show that the Maclaurin series associated with e^x has a sum equal to e^x.

SOLUTION By Section 5.5 the series is $\sum_{k=0}^{\infty} x^k/k!$. We want to show that it converges to e^x. The absolute ratio test shows that the series converges for all x, but it does not tell us that its limit is e^x. Also by Section 5.5, the difference between $f(x)$ and its Maclaurin polynomial up through the power x^n has the form

$$\frac{f^{(n+1)}(c_n)}{(n+1)!}x^{n+1} \tag{12.1.1}$$

for some number c_n between 0 and x. Therefore, if $f(x) = e^x$, we have $f^{(n+1)}(x) = e^x$. Thus $f^{(n+1)}(c_n) = e^{c_n}$ and the error (12.1.1) equals

$$\frac{e^{c_n}x^{n+1}}{(n+1)!}.$$

For $x > 0$, we know $c_n < x$, which implies $c^{c_n} < e^x$. For $x < 0$, $c_n < 0$, and we have $e^{c_n} < 1$. In either case e^{c_n} is less than a fixed number, which we call M. That is, $e^{c_n} < M$ for all n. Because x is fixed, we see that

$$\lim_{n\to\infty} \frac{\left|e^{c_n}x^{n+1}\right|}{(n+1)!} \le M\frac{|x|^{n+1}}{(n+1)!}. \tag{12.1.2}$$

It was shown in Section 11.1 that $\lim_{n\to\infty} k^n/n!$ is 0 for any k. Thus (12.1.2) approaches 0 as $n \to \infty$, which means that the sum of the series is e^x. ◆

As a result of Example 1, we have

Definition: *Maclaurin Series for* e^x

$$\text{For all } x, \quad e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^k}{k!} + \cdots = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

One consequence of this definition is that e^x can be estimated using only addition, multiplication, and division. When $x = 1$, we have a series representation of e,

$$e = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} + \cdots.$$

Euler used this formula to evaluate e to 23 decimal places.

In Example 1 the Maclaurin series associated with e^x represents the function in the sense that its sum is the exponential function. However, there are functions for which the associated Maclaurin series does not represent the function, as Exercise 30 illustrates.

EXAMPLE 2. Use the Maclaurin series in Example 1 to estimate $\sqrt{e} = e^{1/2}$ with an error of at most 0.001.

SOLUTION The error in using the front end $\sum_{k=0}^{n}(1/2)^k/k!$ has the form

$$e^{c_n}\frac{\left(\frac{1}{2}\right)^{n+1}}{(n+1)!}$$

where c_n is between 0 and 1/2. Then $e^{c_n} < e^{1/2}$. Now, $e^{1/2} = \sqrt{e} < \sqrt{4} = 2$. Therefore we want to find n large enough so that

$$\frac{2\left(\frac{1}{2}\right)^{n+1}}{(n+1)!} < 0.001.$$

To find n, we experiment by making a table of values with four decimal place accuracy. Table 12.1.1 stops at $n = 4$ with an error less than 0.001. Rounded to five decimal places, the estimate for $\sqrt{e}$ is

$$1 + \frac{1}{2} + \frac{\left(\frac{1}{2}\right)^2}{2!} + \frac{\left(\frac{1}{2}\right)^3}{3!} + \frac{\left(\frac{1}{2}\right)^4}{4!} \approx 1.64843,$$

which is close to the 1.6487 that a calculator shows. ◆

n	1	2	3	4
$\dfrac{2}{(n+1)!}\left(\dfrac{1}{2}\right)^{n+1}$	0.2500	0.0417	0.0026	0.0005

Table 12.1.1

In Section 5.5 we found the Maclaurin polynomial for $\sin(x)$. Using it, we conclude that the Maclaurin series for $\sin(x)$ is

$$\sum_{k=0}^{\infty}(-1)^k\frac{x^{2k+1}}{(2k+1)!}=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots+(-1)^k\frac{x^{2k+1}}{(2k+1)!}+\cdots.$$

The next example shows that its sum is $\sin(x)$.

EXAMPLE 3. Show that $\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)!}x^{2k+1}=\sin(x)$.

SOLUTION To demonstrate that the series converges to $\sin(x)$ we show that the difference between $\sin(x)$ and $\sum_{k=0}^{n}(-1)^k x^{2k+1}/(2k+1)!$ approaches 0 as $n\to\infty$.

To do this we use Lagrange's formula, which involves the higher derivatives of $\sin(x)$, which are $\pm\sin(x)$ and $\pm\cos(x)$. In any case, if $f(x)=\sin(x)$, $\left|f^{(n)}(x)\right|\le 1$. Thus we have

$$\left|\frac{f^{(n+1)}(c_n)x^{n+1}}{(n+1)!}\right|\le\frac{|x|^{n+1}}{(n+1)!}.$$

Because $|x|^{n+1}/(n+1)!$ approaches 0 as $n\to\infty$ no matter what x is, the difference between the Maclaurin polynomials and $\sin(x)$ approaches 0 as the degree of the polynomial increases. We conclude that the Maclaurin series converges to $\sin(x)$ for all x. ♦

Therefore

Definition: *Maclaurin Series for* $\sin(x)$

For all x, $\quad \sin(x)=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots+(-1)^k\frac{x^{2k+1}}{(2k+1)!}+\cdots=\sum_{k=0}^{\infty}(-1)^k\frac{x^{2k+1}}{(2k+1)!}.$

In a similar manner, we have

Definition: *Maclaurin Series for* $\cos(x)$

For all x, $\quad \cos(x)=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\cdots+(-1)^k\frac{x^{2k}}{(2k)!}+\cdots=\sum_{k=0}^{\infty}\frac{(-1)^k x^{2k}}{(2k)!}.$

Taylor Series in Powers of $x-a$

Just as there are Taylor polynomials around 0, there are Taylor polynomials around a. The Taylor series around a for $f(x)$ has powers of $x-a$ instead of powers of x $(=x-0)$,

$$\sum_{k=0}^{\infty}\frac{f^{(k)}(a)}{k!}(x-a)^k.$$

The series may or may not converge and if it converges, it may not converge to $f(x)$.

EXAMPLE 4. Find the Taylor series for $\frac{1}{x}$ in powers of $x-1$.

SOLUTION Here $f(x) = 1/x$ and, as can be checked,

$$f^{(n)}(x) = (-1)^n n! x^{-(n+1)} \qquad \text{for } n \geq 1.$$

Table 12.1.2 shows a few of the higher derivatives evaluated at 1.

n	1	2	3	4	5
$f^{(n)}(x)$	$\frac{-1}{x^2}$	$\frac{2}{x^3}$	$\frac{-3\cdot 2}{x^4}$	$\frac{4\cdot 3\cdot 2}{x^5}$	$\frac{-5\cdot 4\cdot 3\cdot 2}{x^6}$
$f^{(n)}(1)$	-1	2	$-3!$	$4!$	$-5!$

Table 12.1.2

In general, $f^{(n)}(1) = (-1)^n n!$ and the term containing $(x-1)^n$ in the Taylor series around 1 is

$$\frac{(-1)^n n!(x-1)^n}{n!} = (-1)^n (x-1)^n.$$

The series begins

$$1 - (x-1) + (x-1)^2 - (x-1)^3 + \cdots.$$

By the n^{th} term test, the series does not converge if $|x-1| > 1$, that is, if $x > 2$ or $x < 0$.

If $x = 0$, the series becomes $\sum_{k=0}^{\infty} (-1)^k (-1)^k = \sum_{k=0}^{\infty} 1$, which, by the n^{th} term test does not converge. When $x = 2$ it is $\sum_{k=0}^{\infty} (-1)^k$, which also does not converge. For x in $(0,2)$ we use the absolute ratio test, examining

$$\lim_{n\to\infty} \left| \frac{(-1)^{n+1}(x-1)^{n+1}}{(-1)^n (x-1)^n} \right| = \lim_{n\to\infty} |x-1| = |x-1|.$$

Thus, if $|x-1| < 1$ the series converges. But does it converge to $f(x) = 1/x$?

The Lagrange formula for the remainder is

$$\frac{f^{(n+1)}(c_n)(x-1)^{n+1}}{(n+1)!} = \frac{(-1)^{n+1}(n+1)!}{c_n^{n+2}} \frac{(x-1)^{n+1}}{(n+1)!} = \frac{(-1)^{n+1}(x-1)^{n+1}}{c_n^{n+2}},$$

where c_n is between 1 and x. We want to show that

$$\left| \frac{(-1)^{n+1}(x-1)^{n+1}}{c_n^{n+2}} \right| = \frac{1}{|c_n|} \left| \frac{x-1}{c_n} \right|^{n+1} \to 0 \qquad \text{as } n \to \infty.$$

However, if x is near 0, $|x-1|$ is near 1 and c_n may be near 0, for we know only that c_n is between x and 1. Perhaps the ratio $|(x-1)/c_n|$ is a large number.

But when x is in $(1,2)$ we have $c_n > 1$ while $|x-1| < 1$, so

$$0 < \left| \frac{x-1}{c_n} \right| < |x-1| < 1.$$

Thus the remainder approaches 0 as $n \to \infty$. So we see that for x in $(1,2)$, $1 - (x-1) + (x-1)^2 - (1-x)^3 - \cdots = 1/x$. The Lagrange formula justifies the same conclusion for x in $(-1/2, 1)$, but not for x in $(0, 1/2]$, as Exercise 28 shows.

Because $1-(x-1)+(x-1)^2-(x-1)^3-\cdots$ is a geometric series with first term 1 and ratio $r = -(x-1)$ it converges to

$$\frac{1}{1-r} = \frac{1}{1-(-(x-1))} = \frac{1}{1+x-1} = \frac{1}{x}.$$

This argument covers all x in $(0,2)$ at once. ♦

The General Binomial Theorem

If r is 0 or a positive integer, $(1+x)^r$, when multiplied out, is a polynomial of degree r. Therefore its derivatives from order $r+1$ on are the zero function. That implies its Maclaurin series has only a finite number of nonzero terms, the one of highest degree being x^r. The formula

$$(1+x)^r = \sum_{k=0}^{r} \frac{r!}{k!(r-k)!} x^k = \sum_{k=0}^{r} \frac{r(r-1)\cdots(r-(k-1))}{1\cdot 2\cdots k} x^k$$

is known as the **binomial formula**. (See Exercises 23 to 26 in Section 5.5.)

MEMORY AID: One way to remember the binomial formula is to observe that the coefficient of x^k has k factors in both the numerator and denominator. The k factors in the numerator start from r and decrease by 1. The k factors in the denominator start from 1 and increase by 1.

Example 5 generalizes the binomial theorem to arbitrary exponents r.

EXAMPLE 5. Find the Maclaurin series for $f(x) = (1+x)^r$, when r is not zero or a positive integer and determine for which x it converges.

SOLUTION Table 12.1.3 will help in computing $f^{(k)}(0)$:

k	$f^{(k)}(x)$	$f^{(k)}(0)$
0	$(1+x)^r$	1
1	$r(1+x)^{r-1}$	r
2	$r(r-1)(1+x)^{r-2}$	$r(r-1)$
3	$r(r-1)(r-2)(1+x)^{r-3}$	$r(r-1)(r-2)$
$\vdots$	$\vdots$	$\vdots$
k	$r(r-1)\cdots(r-k+1)(1+x)^{r-k}$	$r(r-1)(r-2)\cdots(r-k+1)$

Table 12.1.3

Consequently the Maclaurin series associated with $(1+x)^r$ is

$$1 + rx + \frac{r(r-1)}{1\cdot 2}x^2 + \frac{r(r-1)(r-2)}{1\cdot 2\cdot 3}x^3 + \cdots. \tag{12.1.3}$$

The series has an infinite number of nonzero terms if r is not a positive integer or zero.

For $x=0$, the series has only one nonzero term and hence converges. For $x \neq 0$ use the absolute ratio test. Let a_k be the term in the series that contains x^k. Then

$$a_k = \frac{r(r-1)(r-2)\cdots(r-k+1)}{1\cdot 2\cdot 3\cdots k} x^k \qquad \text{and} \qquad a_{k+1} = \frac{r(r-1)(r-2)\cdots(r-k+1)(r-k)}{1\cdot 2\cdot 3\cdots k(k+1)} x^{k+1}.$$

Thus

$$\left|\frac{a_{k+1}}{a_k}\right| = \left| \frac{\frac{r(r-1)(r-2)\cdots(r-k+1)(r-k)}{1\cdot 2\cdot 3\cdots k(k+1)} x^{k+1}}{\frac{r(r-1)(r-2)\cdots(r-k+1)}{1\cdot 2\cdot 3\cdots k} x^k} \right| = \left| \frac{r-k}{k+1} x \right|$$

and

$$\lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right| = |x|.$$

By the absolute ratio test, when r is neither zero nor a positive integer, the Maclaurin series for $(1+x)^r$, (12.1.3), has radius of convergence $R=1$. ♦

Observation 12.1.1: *Endpoint Convergence of Binomial Series*

Example 5 does not address the convergence of a binomial series at the endpoints of its interval of convergence. It turns out that the actual interval of convergence depends on the value of r. If r is zero or a positive integer, the binomial series is a polynomial and the interval of convergence is $(-\infty,\infty)$. When r is neither 0 nor a positive integer, the interval of convergence is $(-1,1)$ if $r \leq -1$, $(-1,1]$ if $-1 < r < 0$, and $[-1,1]$ if $r > 0$.

CAUTION: While Example 5 proves that the Maclaurin series for $(1+x)^r$ converges for $|x| < 1$, this does not mean that it converges to $(1+x)^r$.

Let us check the case $r = -1$. The binomial series becomes

$$1 + (-1)x + \frac{(-1)(-2)}{1\cdot 2}x^2 + \frac{(-1)(-2)(-3)}{1\cdot 2\cdot 3}x^3 + \cdots$$

or

$$1 - x + x^2 - x^3 + \cdots,$$

a geometric series with first term 1 and ratio $-x$. It therefore converges for $|x| < 1$, but not for either $x = -1$ or for $x = 1$. Moreover, it represents the function $(1+x)^r = (1+x)^{-1}$.

The series (12.1.3) is called the **binomial expansion** of $(1+x)^r$. That $(1+x)^r$ is represented by this series is known as the **general binomial theorem** or, simply, the **binomial theorem**. Exercises 33 to 36 in Section 12.4 prove this result.

Summary

The Taylor series associated with a function is the series whose partial sums are its n^{th}-order Taylor polynomials. It represents the original function only for inputs such that the remainder of the n^{th}-order Taylor polynomial approaches zero as $n \to \infty$. The Lagrange form of the remainder, Theorem 5.5.2 from Section 5.5, helps to show when the remainder converges to zero, though, as Example 3 illustrates, sometimes it may not be strong enough to do that.

Table 12.1.4 summarizes the series representations developed in this section.

EXERCISES for Section 12.1

1. State, without using any mathematical symbols, the formula for the terms of a Taylor series of a function around a number that may not be zero and Lagrange's formula for the remainder.

2. State, without using any mathematical symbols, the formula for the terms of a Maclaurin series of a function and Lagrange's formula for the remainder.

In Exercises 3 to 10 for the given function compute the Maclaurin series associated with it and write out its first five nonzero terms.

3. $\frac{1}{1+x}$ **4.** $\frac{1}{1-x}$ **5.** $\ln(1+x)$ **6.** $\ln(1-x)$

7. $\sin(x)$ **8.** $\cos(x)$ **9.** e^{-x} **10.** $\sqrt{1+x}$

---

11. Let $f(x) = e^x$. Show that $\lim_{n\to\infty} R_n(x;0) = 0$ for a negative number x. This completes the proof that the exponential function is represented by its Maclaurin series for all x (see Example 2).

Function	Series	Interval of Convergence
e^x	$\sum_{k=0}^{\infty} \frac{x^k}{k!}$	$-\infty < x < \infty$
$\sin(x)$	$\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}$	$-\infty < x < \infty$
$\cos(x)$	$\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!}$	$-\infty < x < \infty$
$\frac{1}{x}$	$\sum_{k=0}^{\infty} (-1)^k (x-1)^k = \sum_{k=0}^{\infty} (1-x)^k$	$0 < x < 2$
$(1+x)^r$	$\sum_{k=0}^{\infty} \frac{r(r-1)(r-2)\cdots(r-k+1)}{1\cdot 2\cdot 3\cdots k} x^k$	$-1 < x < 1$, if $r \le -1$ $-1 < x \le 1$, if $-1 < r < 0$ $-1 \le x \le 1$, if $r \ge 0$ (and not a positive integer) all x, if $r = 0$ or a positive integer:

Table 12.1.4

12. Show that the Maclaurin series associated with $\sin(x)$ represents $\sin(x)$ for all x.

13. Show that the Maclaurin series associated with e^{-x} represents e^{-x} for all x.

14. (a) Why will there be no terms of even degree in the Maclaurin series for $\arctan(x)$? (That is, all terms of the form x^{2k} have coefficient zero.)
(b) Obtain the first two nonzero terms of the Maclaurin series for $\arctan(x)$.

15. (a) Use the Lagrange formula to show that the Maclaurin series for $\frac{1}{1+x}$ represents $\frac{1}{1+x}$ for $-\frac{1}{2} < x < 1$.
(b) Use the fact that it is a geometric series to show that the representation holds for $-1 < x < 1$.

16. Show that the Taylor series in powers of $x-a$ for e^x represents e^x for all x.

17. Show that the Taylor series in powers of $x-a$ for $\cos(x)$ represents $\cos(x)$ for all x.

18. (a) Write the first four terms of the binomial expansion of $(1+x)^{-2} = \frac{1}{(1+x)^2}$. (b) What is the coefficient of x^n?

19. Write the first four terms of the binomial expansion of $(1+x)^{1/2} = \sqrt{1+x}$.

20. What is the n^{th} term in the Maclaurin series for $(1-x)^r$? (Use the binomial expansion of $(1+x)^r$.)

21. Suppose we use the Maclaurin series for e^x to find e^{100}.
(a) What are the first four terms?
(b) Does the series converge to e^{100}?
(c) If your answer to (b) is "yes" how many terms are needed to estimate e^{100} with an error less than 0.005?
(d) What term in the series is largest?

22. (a) Use the Maclaurin series for e^x to estimate $\sqrt[3]{e}$ to three decimal places.
(b) Compare the answer in (a) to the value of $\sqrt[3]{e}$ given by a calculator.

23. Find the Maclaurin series for $\dfrac{1}{(1-x)^2}$.

24. This problem examines ways to estimate the error in using a partial sum of $\sum_{k=0}^{\infty} \dfrac{(-1)^k}{k!}$ to estimate e^{-1}.

(a) Use the Lagrange formula to estimate the error when the e^{-1} is estimated by a partial sum that ends with the term $\dfrac{(-1)^m}{m!}$.
(b) Estimate the error by noticing the series is alternating and the terms decrease in absolute value
(c) Estimate the error by comparing $\sum_{k=m+1}^{\infty} \left| \dfrac{(-1)^k}{k!} \right|$ to a geometric series that is easy to sum.
(d) Which of the methods provides the smallest estimate of the error?

25. (a) Use the Taylor series around $\dfrac{\pi}{4}$ to estimate $\cos(50°)$ to two decimal places. Use $\pi \approx 3.1416$ and $\sqrt{2} \approx 1.4142$.
(b) Check the result with a calculator.

26. Let f have derivatives of all orders for all x. Assume that $\left|f^{(n)}(x)\right| \le n$ for all n. Show why $f(x)$ is represented by its Maclaurin series for all x.

27. (a) From the Maclaurin series for $\cos(x)$ and a trigonometric identity obtain the Maclaurin series for $\sin^2(x)$.
(b) From (a) and another trigonometric identity obtain the Maclaurin series for $\cos^2(x)$.

28. Explain why it is not possible to use the Lagrange formula to show that the Taylor series in powers of $(x-1)$ for $\dfrac{1}{x}$ converges to $\dfrac{1}{x}$ for $0 < x < \dfrac{1}{2}$. See Exercise 15(a).

Exercises 29 and 30 present a nonzero function whose Maclaurin series has the value 0 for all x and therefore does not represent the function. The function is so flat at the origin that all its derivatives are zero there.

29. Show that $\lim_{x\to 0} \dfrac{e^{1/x^2}}{x^n} = 0$ for all positive n:

(a) Why does it suffice to consider only $x > 0$?
(b) Let $v = \dfrac{1}{x^2}$ and translate the limit to $\lim_{v\to\infty} v^{n/2} e^{-v}$.
(c) The limit in (b) is similar to a limit treated in Section 5.6. Show that it equals 0.
(d) Show that $\lim_{n\to\infty} p(x) \dfrac{e^{-1/x^2}}{x^n} = 0$ for any polynomial $p(x)$.

30. Let $f(x) = e^{-1/x^2}$ if $x \ne 0$ and $f(0) = 0$.

(a) Show f is continuous at 0.
(b) Show f is differentiable at 0.
(c) Show that $f'(0) = 0$.
(d) Show that $f''(0) = 0$.
(e) Explain why $f^{(n)}(0) = 0$ for all $n \ge 0$.
(f) What is the Maclaurin series for f?
(g) Why does the example use e^{-1/x^2} instead of $e^{-1/x}$?

12.2 Two Applications of Taylor Series

This section provides examples showing how the partial sums of a Taylor series can be used to evaluate limits and estimate definite integrals.

Using a Taylor Series to Find a Limit

The next example uses a Maclaurin series to evaluate the limit of a quotient that is an indeterminate form.

EXAMPLE 1. Find $\lim_{x\to 0} \frac{\sin(x^2)}{\sqrt{1+3x^2}-1}$.

SOLUTION Replacing x by x^2 in the Maclaurin series for $\sin(x)$ produces a series whose value is $\sin(x^2)$:

$$\sin\left(x^2\right) = x^2 - \frac{\left(x^2\right)^3}{6} + \cdots.$$

Likewise, replacing x by $3x^2$ in the Maclaurin series for $\sqrt{1+x}$ yields a series whose value is $\sqrt{1+3x^2}$:

$$\sqrt{1+3x^2} = 1 + \frac{\left(3x^2\right)}{2} - \frac{\left(3x^2\right)^2}{8} + \cdots.$$

With these expansions we have

$$\begin{aligned}
\lim_{x\to 0} \frac{\sin(x^2)}{\sqrt{1+3x^2}-1} &= \lim_{x\to 0} \frac{x^2 - \frac{1}{6}\left(x^2\right)^3 + \cdots}{1 + \frac{1}{2}\left(3x^2\right) - \frac{1}{8}\left(3x^2\right)^2 + \cdots - 1} && \text{(Maclaurin series of both numerator and denominator)} \\
&= \lim_{x\to 0} \frac{x^2 - \frac{1}{6}x^6 + \cdots}{\frac{3}{2}x^2 - \frac{9}{8}x^6 + \cdots} && \text{(expand terms in numerator and denominator)} \\
&= \lim_{x\to 0} \frac{1 - \frac{1}{6}x^4 + \cdots}{\frac{3}{2} - \frac{9}{8}x^4 + \cdots} && \text{(simplify quotient)} \\
&= \frac{1}{3/2} = \frac{2}{3} && \text{(evaluate limit and simplify)}
\end{aligned}$$

◆

In Example 1 we needed only enough terms of the series to know the smallest power of x that appeared in the numerator and in the denominator. Keeping one additional term shows how the subsequent terms are smaller, and so do not contribute to the final result after taking limits. The next example also illustrates this.

EXAMPLE 2. Find $\lim_{x\to 0} \frac{\sqrt{1+x}-\sqrt{1+2x}}{\sqrt{1+2x}-\sqrt{1+4x}}$.

SOLUTION This limit could be found using l'Hôpital's rule. However, it is faster to use Taylor series.

Recall that the binomial theorem asserts that for $|u| < 1$

$$(1+u)^r = 1 + ru + \frac{r(r-1)}{2}u^2 + \cdots.$$

Evaluating this with $r = 1/2$ for $u = x$, $u = 2x$ (twice), and $u = 4x$, allows the straightforward evaluation of the limit: Thus the limit is

$$\begin{aligned}
&\lim_{x\to 0} \frac{\sqrt{1+x}-\sqrt{1+2x}}{\sqrt{1+2x}-\sqrt{1+4x}} \\
&= \lim_{x\to 0} \frac{\left(1+\frac{1}{2}x-\frac{1}{8}(x)^2+\cdots\right)-\left(1+\frac{1}{2}(2x)-\frac{1}{8}(2x)^2+\cdots\right)}{\left(1+\frac{1}{2}(2x)-\frac{1}{8}(2x)^2+\cdots\right)-\left(1+\frac{1}{2}(4x)-\frac{1}{8}(4x)^2+\cdots\right)} && \text{(substitute series expansion (4 times))} \\
&= \lim_{x\to 0} \frac{-\frac{1}{2}x-\frac{3}{8}x^2+\cdots}{-x+\frac{3}{2}x^2+\cdots} && \text{(simplify numerator and denominator)} \\
&= \lim_{x\to 0} \frac{-\frac{1}{2}-\frac{3}{8}x+\cdots}{-1+\frac{3}{2}x+\cdots} && \text{(cancel common factor of } x \text{)} \\
&= \frac{-1/2}{-1} = \frac{1}{2} && \text{(evaluate limit and simplify)}
\end{aligned}$$

♦

Observation 12.2.1:

While the limits in Examples 1 and 2 could also have been evaluated using l'Hôpital's rule, the use of series also provides more information about the behavior of the expression near the limit point.

Using a Taylor Series to Estimate an Integral

The **error function**, $\mathrm{erf}(x) = \int_0^x e^{-t^2}\,dt$, is important in statistics, probability, and partial differential equations. The integrand is an example of the famous bell curve. Since e^{-t^2} does not have an elementary antiderivative, the integral cannot be evaluated by the fundamental theorem of calculus.

Some definitions of $\mathrm{erf}(x)$ include a leading coefficient of $1/\sqrt{2\pi}$.

The next example shows how to estimate $\int_a^b f(x)\,dx$ when $f(x)$ is represented by a Taylor series. It also shows how Taylor polynomials can be used to obtain bounds on an unknown exact value.

EXAMPLE 3. Use the Maclaurin series for e^x to obtain lower and upper estimates of $\mathrm{erf}(1) = \displaystyle\int_0^1 e^{-t^2}\,dt$.

SOLUTION The first step is to obtain a series representing the integrand. In the Maclaurin series for e^x,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots,$$

replace x with $-t^2$ to obtain

$$e^{-t^2} = 1 - t^2 + \frac{t^4}{2!} - \frac{t^6}{3!} + \cdots. \tag{12.2.1}$$

For $-\infty \le t \le \infty$, (12.2.1) is a convergent alternating series. A partial sum that ends with a negative term is smaller than e^{-t^2} and a partial sum that ends with a positive term is larger than e^{-t^2}. For example, for all values of t,

$$1 - t^2 + \frac{t^4}{2!} - \frac{t^6}{3!} < e^{-t^2} < 1 - t^2 + \frac{t^4}{2!} - \frac{t^6}{3!} + \frac{t^8}{4!}.$$

Integrating these inequalities over the interval $0 \le t \le 1$ yields

$$\int_0^1 \left(1 - t^2 + \frac{t^4}{2!} - \frac{t^6}{3!}\right) dx < \int_0^1 e^{-t^2}\, dx < \int_0^1 \left(1 - t^2 + \frac{t^4}{2!} - \frac{t^6}{3!} + \frac{t^8}{4!}\right) dt,$$

so

$$1 - \frac{1}{3} + \frac{1}{5\cdot 2!} - \frac{1}{7\cdot 3!} < \int_0^1 e^{-x^2}\, dx < 1 - \frac{1}{3} + \frac{1}{5\cdot 2!} - \frac{1}{7\cdot 3!} + \frac{1}{9\cdot 4!}.$$

Thus

$$0.7428 < \int_0^1 e^{-x^2}\, dx < 0.7475.$$

Using four and five terms in the Taylor series representation of the integrand shows that $0.7428 < \text{erf}(1) < 0.7475$. In fact, from tables, $\text{erf}(1) = 0.74682$ to five decimal places. ◆

Summary

Taylor series were used to evaluate limits and estimate definite integrals. Usually only the first few terms are needed, and, in many cases, error estimates are also readily available.

EXERCISES for Section 12.2

In Exercises 1 to 4, use Taylor series to find the limits.

1. $\lim_{x\to 0} \dfrac{\sqrt{1+x}-1}{\sqrt{1+3x}-1}$

2. $\lim_{x\to 0} \dfrac{\sin(4x)}{\sqrt{1+3x}-1}$

3. $\lim_{x\to 0} \dfrac{e^{x^2}-1}{\sin\left(x^2\right)}$

4. $\lim_{x\to 0} \dfrac{\cos(x)-1+\frac{x^2}{2}}{\sin\left(x^4\right)}$

In Exercises 5 to 11, find the limits (a) by using a Taylor series and (b) by using l'Hôpital's rule.

5. $\lim_{x\to 0} \dfrac{\cos(x)e^{2x^2}-1}{x\sin(x)}$

6. $\lim_{x\to 0} \dfrac{\sqrt{1+3x}(e^x-1)x}{1-\cos(2x)}$

7. $\lim_{x\to 0} \dfrac{\cos(x)-\sqrt{1+x}}{\cos(2x)-\sqrt[3]{1+2x}}$

8. $\lim_{x\to 0} \dfrac{\ln(1+3x)}{\sin(2x)}$

9. $\lim_{x\to 0} \dfrac{e^{x^2}-1}{e^{3x^2}-1}$

10. $\lim_{x\to 0} \dfrac{\left(\sin(x^2)+e^{x^3}-1\right)\sqrt[3]{5+x}}{\sqrt{1+5x^2}-1}$

11. $\lim_{x\to 4} \dfrac{(8-2x)e^{x^2}}{\sqrt[3]{4-x}}$

12.
(a) Write the first four terms of the binomial series for $(1+x)^{-2}$.
(b) What is the n^{th} term?

13.
(a) Find the limit in Example 2 by l'Hôpital's rule.
(b) Find the limit in Example 1 by l'Hôpital's rule.

14.
(a) Show $\displaystyle\int_0^1 \frac{e^x-1}{x}\,dx$ is finite, even though the integrand is not defined at 0.
(b) Show that $1+\frac{1}{2\cdot 2!}+\frac{1}{3\cdot 3!}+\frac{1}{4\cdot 4!}+\frac{1}{5\cdot 5!}$ is an estimate of the integral.
(c) The error in using the sum in (b) is $\frac{1}{6\cdot 6!}+\frac{1}{7\cdot 7!}+\frac{1}{8\cdot 8!}+\frac{1}{9\cdot 9!}+\cdots$.
Show that this is less than $\frac{1}{6\cdot 6!}\left(1+\left(\frac{1}{7}\right)+\left(\frac{1}{7}\right)^2+\left(\frac{1}{7}\right)^3+\cdots\right)$.
(d) From (c) deduce that the error is less than 0.00027.

15.
(a) Show that for x in $[0,2]$ $x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots+\frac{x^n}{n!}\le e^x-1\le x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots+\frac{x^n}{n!}+\frac{e^2x^{n+1}}{(n+1)!}$.
(b) Use (a) to find $\displaystyle\int_0^2 \frac{e^x-1}{x}\,dx$ to three decimal places.

16. Find $\displaystyle\int_0^1 \frac{1-\cos(x)}{x}\,dx$ to three decimal places, using an approach like that in Exercise 15.

17. Estimate $\displaystyle\int_0^\infty e^{-5x^2}\,dx$.
(a) Find b so that $\displaystyle\int_b^\infty e^{-5x^2}\,dx<0.0005$.
(b) Let b be the number found in (a). Estimate $\displaystyle\int_0^b e^{-5x^2}\,dx$ with an error of less than 0.0005.
(c) Combine (a) and (b) to get a three decimal place estimate of $\displaystyle\int_0^\infty e^{-5x^2}\,dx$.

18. Estimate $\displaystyle\int_0^\infty \frac{\cos\left(\frac{x^6}{100}\right)-1}{x^6}\,dx$.
(a) Find b so that $\left|\displaystyle\int_b^\infty \frac{\cos\left(\frac{x^6}{100}\right)-1}{x^6}\,dx\right|<0.005$.
(b) Let b be the number you found in (a). Estimate $\displaystyle\int_0^b \frac{\cos\left(\frac{x^6}{100}\right)-1}{x^6}\,dx$, with an error less that 0.005.
(c) Combine (a) and (b) to get a two decimal place estimate for $\displaystyle\int_0^\infty \frac{\cos\left(\frac{x^6}{100}\right)-1}{x^6}\,dx$.

19. Evaluate $\displaystyle\int_0^1 \frac{dx}{1+x^2}$ by
(a) the fundamental theorem of calculus (approximate π to 3 decimal places),
(b) Simpson's method (six sections),
(c) trapezoid method (six sections),
(d) using the first six nonzero terms of the series $1-x^2+x^4-\cdots$ for $\frac{1}{1+x^2}$.

The binomial theorem does not apply to $(a+b)^r$ if neither a nor b is 1. In such situations write $(a+b)^r$ as $a^r\left(1+\frac{b}{a}\right)^r$ or $b^r\left(1+\frac{a}{b}\right)^r$, to which the binomial theorem applies. This idea is used in Exercises 20 to 22.

20. If $\left|\frac{a}{b}\right| < 1$, use the binomial theorem to expand $(a+b)^r$ as a sum of terms of the form ca^pb^q.

21. If $\left|\frac{b}{a}\right| < 1$, use the binomial theorem to expand $(a+b)^r$ as a sum of terms of the form ca^pb^q.

22. Write the first four terms of the series for $(8+x)^{1/3}$ if (a) $x > 8$ and (b) $x < 8$. See Exercises 20 and 21.

---

23. Sam and Jane continue to work together to learn calculus. Here is another of their exchanges.

Sam: I was playing with the binomial theorem.

Jane: Is that possible?

Sam: I looked at $(3+5)^{1/3}$, which I know is 2. But I can write it as $5^{1/3}\left(1+\frac{3}{5}\right)^{1/3}$ and get

$$5^{1/3}\left(1+\frac{1}{3}\frac{3}{5}+\frac{1}{2!}\frac{1}{3}\left(\frac{-2}{3}\right)\left(\frac{3}{5}\right)^2+\cdots\right)$$

so

$$2 = 5^{1/3}+\frac{1}{3}5^{-2/3}(3)-\frac{1}{9}5^{-5/3}3^2+\cdots.$$

Jane: That's a fancy way to estimate 2.

Sam: But I can write $(3+5)^{1/3}$ as $3^{1/3}\left(1+\frac{5}{3}\right)^{1/3}$ and get

$$2 = 3^{1/3}+\frac{1}{3}3^{-2/3}(5)-\frac{1}{9}3^{-5/3}5^2+\cdots$$

Jane: Another nutty way to estimate 2.

Sam: My point is that they can't both be right.

Can they both be right? Explain your answer.

24. Repeat Exercise 19 for $\int_0^1 \frac{dx}{1+x^3}$.

25. In R. P. Feynman, *Lectures on Physics*, Addison-Wesley, Reading, MA 1963, this appears in Section 15.8 of Volume 1:

> An approximate formula to express the increase of mass, for the case when the velocity is small, can be found by expanding $m_0/\sqrt{1-v^2/c^2} = m_0(1-v^2/c^2)^{-1/2}$ in a power series, using the binomial theorem. We get
>
> $$m_0\left(1-\frac{v^2}{c^2}\right)^{-1/2} = m_0\left(1+\frac{1}{2}\frac{v^2}{c^2}+\frac{3}{8}\frac{v^4}{c^4}+\cdots\right).$$
>
> We see clearly from the formula that the series converges rapidly when v is small and the terms after the first two or three are negligible.

Check the expansion and justify the equation.

26. A fluid mechanics text has the following in a discussion of flow through a nozzle:

> The pressure p equals $\left(1+\frac{\gamma-1}{2}M^2\right)^{\gamma/(1-\gamma)}$. By the binomial theorem and the fact that $v^2 = M^2\gamma RT$:
>
> $$p = 1 - \frac{1}{2}\frac{v^2}{RT} + \frac{\gamma(2\gamma-1)}{8}M^4 + \cdots.$$

Fill in the steps that confirm the quoted three-term expansion of the pressure, p.

27. (a) The ellipse $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ has the parameterization $x = a\cos(t)$, $y = b\sin(t)$ for $0 \le t \le 2\pi$. Show that the arc length of one quadrant of the ellipse is $b\int_0^{\pi/2}\sqrt{1-\left(1-\left(\frac{a}{b}\right)^2\right)\sin(t)^2}\,dt$. (The integrand does not have an elementary antiderivative.)

(b) If $a < b$, the integral has the form $b\int_0^{\pi/2}\sqrt{1-k^2\sin(t)^2}\,dt$, where $0 < k < 1$. Use the binomial theorem and the formula for $\int_0^{\pi/2}\sin^n(\theta)\,d\theta$ to obtain the first four terms in the series expansion of $b\int_0^{\pi/2}\sqrt{1-k^2\sin(t)^2}\,dt$ in powers of k^2.

12.3 Power Series and Their Intervals of Convergence

Our use of Taylor polynomials to approximate a function led us to consider series

$$\sum_{k=0}^{\infty} b_k(x-a)^k = b_0 + b_1(x-a) + b_2(x-a)^2 + \cdots + b_k(x-a)^k + \cdots,$$

called a **power series** in $x-a$. If $a = 0$, we obtain a series in powers of x, that is, a Maclaurin series:

$$\sum_{k=0}^{\infty} b_k x^k = b_0 + b_1 x + b_2 x^2 + \cdots + b_k x^k + \cdots.$$

We will now look at some properties of power series and see that they behave much like polynomials.

The Interval of Convergence of a Power Series in x: $\sum_{k=0}^{\infty} b_k x^k$

The power series $b_0 + b_1 x + b_2 x^2 + \cdots$ converges when $x = 0$. It may or may not converge for other choices of x. However, as Theorem 12.3.1 will show, if it converges at c, it converges at any number x whose absolute value is less than $|c|$, that is, throughout the interval $(-|c|, |c|)$.

Theorem 12.3.1

Let c be a nonzero number. Assume that $\sum_{k=0}^{\infty} b_k c^k$ converges. Then, if $|x| < |c|$, the series converges absolutely.

MAIN TAKEAWAY: *The x's for which the series converges form an interval with 0 at its midpoint.*

The proof of Theorem 12.3.1 is presented at the end of this section.

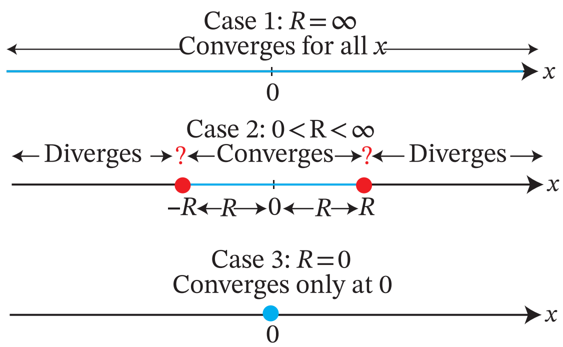


Figure 12.3.1

The series $\sum_{k=0}^{\infty} b_k x^k$ clearly converges when $x = 0$, because it consists of just one nonzero term, b_0. If it converges for 3 or at -3, for example, then according to Theorem 12.3.1, it converges at any number x in $(-3, 3)$.

There are three possibilities. If there are arbitrarily large r's such that the series converges for x in $(-r, r)$, then it converges for all x. If there is an upper bound on the numbers r such that the series converges for x in $(-r, r)$, then it is shown in advanced courses that there is then a smallest upper bound; call it R. Note that while R is generally positive, it could be zero. These observations can be summarized in the following three cases (see also Figure 12.3.1):

Case 1: $b_0 + b_1 x + b_2 x^2 + \cdots$ converges for all x.

Case 2: there is a positive number R such that $b_0 + b_1 x + b_2 x^2 + \cdots$ converges for all x such that $|x| < R$ and diverges for $|x| > R$.

NOTE: Convergence or divergence at $x = R$ and $x = -R$ is not mentioned.

Case 3: $b_0 + b_1 x + b_2 x^2 + \cdots$ converges only at $x = 0$.

The number R is called the **radius of convergence** of the series. If the series converges for all x, $R = \infty$. For the geometric series $1 + x + x^2 + \cdots + x^k + \cdots$, $R = 1$, since the series converges when $|x| < 1$ and diverges when $|x| > 1$. It also diverges when $x = 1$ and $x = -1$. A power series with radius of convergence R may or may not converge at R and at $-R$. We have

Theorem 12.3.2: Radius of Convergence of $\sum\limits_{k=0}^{\infty} b_k x^k$

Let R be the radius of convergence for $\sum\limits_{k=0}^{\infty} b_k x^k$. There are three possibilities:

Case 1: *If $R = 0$, the series converges only for $x = 0$.*

Case 2: *If R is positive, the series converges for $|x| < R$ and diverges for $|x| > R$.*
(Convergence at the endpoints, $x = R$ and $x = -R$ must be checked separately.)

Case 3: *If R is ∞, the series converges for all x.*

EXAMPLE 1. Find the radius of convergence for $x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \dfrac{x^4}{4} + \cdots + (-1)^{k+1}\dfrac{x^k}{k} + \cdots$.

SOLUTION We use the absolute ratio test with $a_k = (-1)^{k+1} x^k / k$. The absolute value of the ratio of successive terms is

$$\left|\frac{a_{k+1}}{a_k}\right| = \left|\frac{\dfrac{(-1)^{k+2} x^{k+1}}{k+1}}{\dfrac{(-1)^{k+1} x^k}{k}}\right| = \frac{k}{k+1}|x|.$$

As $k \to \infty$, $k/(k+1) \to 1$. Thus,

$$r = \lim_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right| = \lim_{k\to\infty}\frac{k}{k+1}|x| = |x|.$$

Consequently, by the absolute ratio test, if $|x| < 1$ the series converges. If $|x| > 1$, it diverges. Thus, the radius of convergence is $R = 1$.

Convergence at the two endpoints, $x = 1$ and $x = -1$, must be checked separately.

At the right-hand endpoint, $x = 1$, the series is

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots.$$

This is immediately recognized as the alternating harmonic series which, by the alternating series test, converges (conditionally).

At the left-hand endpoint, $x = -1$, the series reduces to

$$-1 - \frac{1}{2} - \frac{1}{3} - \frac{1}{4} - \cdots - \frac{1}{k} - \cdots,$$

which, being the negative of the harmonic series, diverges.

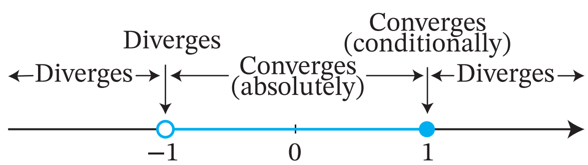


Figure 12.3.2

The series converges only for $-1 < x \leq 1$, as shown in Figure 12.3.2. ♦

Earlier we saw that $\sum_{k=0}^{\infty} x^k/k!$ has radius of convergence $R = \infty$, which is consistent with the fact that the series converges for $-\infty < x < \infty$. The next example illustrates the opposite extreme, $R = 0$.

EXAMPLE 2. Find the radius of convergence of $\displaystyle\sum_{k=1}^{\infty} k^k x^k = 1^1 x + 2^2 x^2 + 3^3 x^3 + \cdots + k^k x^k + \cdots$.

SOLUTION The series converges for $x = 0$, where the value is also 0.

If $x \neq 0$, consider the k^{th} term $k^k x^k$, which can be written as $(kx)^k$. As $k \to \infty$, $|kx| \to \infty$. By the n^{th} term test, this series diverges. Consequently the series converges only when $x = 0$. The radius of convergence is $R = 0$. ♦

REMEMBER: Every power series converges for at least one value of x.

The Radius of Convergence of a Power Series in $x - a$: $\sum_{k=0}^{\infty} b_k (x-a)^k$

As a power series in x has an associated radius of convergence, so does a power series in $x - a$. To see this for

$$\sum_{k=0}^{\infty} b_k (x-a)^k = b_0 + b_1(x-a) + b_2(x-a)^2 + \cdots \tag{12.3.1}$$

let $u = x - a$. Then it becomes a power series in u:

$$\sum_{k=0}^{\infty} b_k u^k = b_0 + b_1 u + b_2 u^2 + \cdots. \tag{12.3.2}$$

Both series have the same radius of convergence, say R. (Recall that R may be zero, positive, or infinite.) The series in (12.3.2) converges for $|u| < R$ and diverges for $|u| > R$. Consequently, the series in (12.3.1) converges for $|x-a| < R$ and diverges for $|x - a| > R$.

Figure 12.3.3 shows $\sum_{k=0}^{\infty} b_k (x-a)^k$ converges in an interval $(a - R, a + R)$, whose midpoint is a. The question marks in Figure 12.3.3 indicate that the series may converge or may diverge at the ends of the interval, $a - R$ and $a + R$. These cases must be looked at separately.

These results are summarized as a corollary to Theorem 12.3.2:

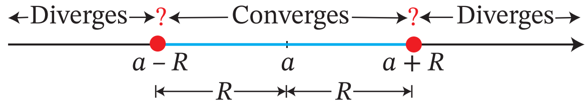


Figure 12.3.3

Corollary 12.3.3: Radius of Convergence of $\sum_{k=0}^{\infty} b_k (x-a)^k$

Let R be the radius of convergence for $\sum_{k=0}^{\infty} b_k (x-a)^k$.

Case 1: *If $R = 0$, the series converges only for $x = a$.*

Case 2: *If R is positive (and finite), the series converges for $|x-a| < R$ and diverges for $|x-a| > R$. (Convergence at the endpoints, $x = a + R$ and $x = a + -R$ must be checked separately.)*

Case 3: *If $R = \infty$, the series converges for all x.*

EXAMPLE 3. Find all x for which $\sum_{k=1}^{\infty} (-1)^{k-1} \frac{(x-1)^k}{k} = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} + \cdots$ converges.

SOLUTION This is Example 1 with x replaced by $x-1$. Thus $x-1$ plays the role that x played in that series. Recall that the series in Example 1 was found to have a radius of convergence, $R = 1$, and to converge for $-1 < x \leq 1$.

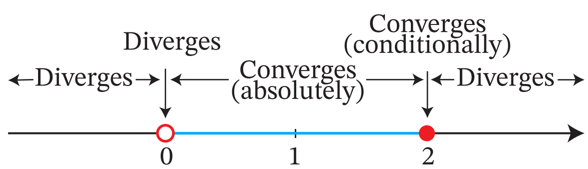


Figure 12.3.4

First, the series in $x-1$ has the same radius of convergence; it is still $R = 1$. But, this series converges for $-1 < x-1 \leq 1$, that is, for $0 < x \leq 2$, and diverges for other values of x. The set of values where the series converges is $(0,2]$, as shown in Figure 12.3.4. ♦

Proof of Theorem 12.3.1

Now that we have some experience finding the radius of convergence of a power series, we conclude by providing a proof of Theorem 12.3.1. This proof is based on the comparison test and the absolute convergence test that were introduced in Chapter 11.

Proof of Theorem 12.3.1

By assumption, $\sum_{k=0}^{\infty} b_k c^k$ converges, so its k^{th} term, $b_k c^k$, approaches 0 as $k \to \infty$. Thus there is an integer N such that for $k \geq N$, $|b_k c^k| \leq 1$. From here on we suppose $k \geq N$. From

$$b_k x^k = b_k c^k \left(\frac{x}{c}\right)^k$$

we have

$$\left|b_k x^k\right| = \left|b_k c^k\right| \left|\frac{x}{c}\right|^k.$$

Consequently, for $k \geq N$,

$$\left|b_k x^k\right| \leq \left|\frac{x}{c}\right|^k \qquad \text{(since } |b_k c^k| \leq 1 \text{ for } k \geq N\text{).}$$

Because $\sum_{k=0}^{\infty} |x/c|^k$ is a geometric series with ratio $|x/c| < 1$ it converges.

Since $|b_k x^k| \leq |x/c|^k$ for $k \geq N$, $\sum_{k=N}^{\infty} \left|b_k x^k\right|$ converges by the comparison test. Thus $\sum_{k=N}^{\infty} b_k x^k$ converges (in fact, absolutely). Putting in the first N terms, $\sum_{k=0}^{N-1} b_k x^k$, we conclude that $\sum_{k=0}^{\infty} b_k x^k$ converges absolutely for all $|x| < |c|$. •

Observation 12.3.4: *Why R is Called a "Radius" of Convergence*

You may well wonder why R is called the radius of convergence, when no circles seem to be involved. Be patient! There is a good, and understandable, explanation but you will have to wait until Sections 12.5 and 12.6, which use complex numbers.

Summary

Motivated by Taylor series, we investigated power series in x, $\sum_{k=0}^{\infty} b_k x^k$, and, more generally, power series in $x-a$: $\sum_{k=0}^{\infty} b_k (x-a)^k$. For each series there is a radius of convergence R.

If the series converges for all x, $R = \infty$. If $\sum_{k=0}^{\infty} b_k x^k$ has radius of convergence R, then it converges (absolutely) for all x in $(-R, R)$ and diverges for all x such that $|x| > R$.

Similarly, if $\sum_{k=0}^{\infty} b_k (x-a)^k$ has radius of convergence R, it converges for all x in $(a-R, a+R)$ and diverges for $|x-a| > R$.

Convergence or divergence at the endpoints of the interval of convergence must be checked separately.

EXERCISES for Section 12.3

In Exercises 1 to 12 draw diagrams like Figure 12.3.4 showing where the series converge and diverge. Check each endpoint separately and explain your work.

1. $\displaystyle\sum_{k=1}^{\infty} \frac{x^k}{k^2}$

2. $\displaystyle\sum_{k=1}^{\infty} \frac{x^k}{\sqrt{k}}$

3. $\displaystyle\sum_{k=0}^{\infty} \frac{x^k}{3^k}$

4. $\displaystyle\sum_{k=1}^{\infty} k^2 e^{-k} x^k$

5. $\displaystyle\sum_{k=0}^{\infty} \frac{2k^2+1}{k^2-5} x^k$

6. $\displaystyle\sum_{k=1}^{\infty} \frac{x^k}{k}$

7. $\displaystyle\sum_{k=0}^{\infty} \frac{x^k}{(2k)!}$

8. $\displaystyle\sum_{k=0}^{\infty} \frac{2^k x^k}{k!}$

9. $\displaystyle\sum_{k=0}^{\infty} \frac{x^k}{(2k+1)!}$

10. $\displaystyle\sum_{k=0}^{\infty} k! x^k$

11. $\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^{k+1} x^k}{k}$

12. $\displaystyle\sum_{k=1}^{\infty} \frac{2^k x^k}{n}$

In Exercises 13 to 24 draw diagrams showing where the series converge and diverge.

13. $\displaystyle\sum_{k=0}^{\infty} \frac{(x-2)^k}{k!}$

14. $\displaystyle\sum_{k=0}^{\infty} \frac{(x-1)^k}{k3^k}$

15. $\displaystyle\sum_{k=0}^{\infty} \frac{(x-1)^k}{k+3}$

16. $\displaystyle\sum_{k=0}^{\infty} \frac{(x-4)^k}{2k+1}$

17. $\displaystyle\sum_{k=0}^{\infty} \frac{k(x-2)^k}{2k+3}$

18. $\displaystyle\sum_{k=0}^{\infty} \frac{(x-5)^k}{k\ln(k)}$

19. $\displaystyle\sum_{k=0}^{\infty} \frac{(x+3)^k}{5^k}$

20. $\displaystyle\sum_{k=0}^{\infty} k(x+1)^k$

21. $\displaystyle\sum_{k=0}^{\infty} \frac{(x-5)^k}{k^2}$

22. $\displaystyle\sum_{k=0}^{\infty} (-1)^k \frac{(x+4)^k}{k+2}$

23. $\displaystyle\sum_{k=0}^{\infty} k!(x-1)^k$

24. $\displaystyle\sum_{k=0}^{\infty} \frac{k^2+1}{k^3+1} (x+2)^k$

In Exercises 25 to 30 write the first five nonzero terms of the binomial expansion of the expression.

25. $(1+x)^{1/2}$ **26.** $(1+x)^{1/3}$ **27.** $(1+x)^{3/2}$ **28.** $(1+x)^{-2}$ **29.** $(1+x)^{-3}$ **30.** $(1+x)^{-4}$

---

31. If $\sum_{k=0}^{\infty} b_k 6^k$ converges, what can be said about the convergence of the following series.

(a) $\sum_{k=0}^{\infty} b_k(-6)^k$? (b) $\sum_{k=0}^{\infty} b_k 5^k$? (c) $\sum_{k=0}^{\infty} b_k(-5)^k$?

32. If $\sum_{k=0}^{\infty} b_k x^k$ converges whenever x is positive, must it converge whenever x is negative?

33. Assume that $\sum_{k=0}^{\infty} b_k x^k$ converges for $x=9$ and diverges when $x=-12$. What, if anything, can be said about

(a) convergence when $x=7$?
(b) absolute convergence when $x=-7$?
(c) absolute convergence when $x=9$?
(d) convergence when $x=-9$?
(e) divergence when $x=10$?
(f) divergence when $x=-15$?
(g) divergence when $x=15$?
(h) divergence when $x=30$?

34. Assume that $\sum_{k=0}^{\infty} b_k x^k$ converges for $x=-5$ and diverges when $x=8$. What, if anything, can be said about

(a) convergence when $x=4$?
(b) absolute convergence when $x=4$?
(c) convergence when $x=7$?
(d) absolute convergence when $x=-5$?
(e) convergence when $x=-9$?
(f) divergence when $x=-18$?

35. (a) If $\sum_{k=0}^{\infty} b_k x^k$ diverges when $x=3$, at which other values of x must it diverge?
(b) If $\sum_{k=0}^{\infty} c_k(x+5)^k$ diverges when $x=-3$, at which other values of x must it diverge?

36. If $\sum_{k=0}^{\infty} b_k(x-3)^k$ converges for $x=7$, at what other values of x must the series necessarily converge?

37. Find the radius of convergence of $\sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!}$.

38. Find the radius of convergence for the Maclaurin series for each expression.
(a) e^x (b) $\sin(x)$ (c) $\cos(x)$ (d) $\ln(1+x)$ (e) $\arctan(x)$ (f) $(1+x)^{1/3}$ (g) $(1+2x)^{3/5}$

39. If $\sum_{k=0}^{\infty} b_k x^k$ has a radius of convergence 3 and $\sum_{k=0}^{\infty} c_k x^k$ has a radius of convergence 5, what can be said about the radius of convergence of $\sum_{k=0}^{\infty} (b_k+c_k)x^k$?

40. The integral $S=\int_0^1 \sqrt{1+x^3}\,dx$ cannot be evaluated by the fundamental theorem of calculus.
(a) Use the first four nonzero terms of the Maclaurin series for $\sqrt{1+x^3}$ to estimate S.
(b) Evaluate S to three decimal places by Simpson's method.

41. (a) Write the first four terms of the Maclaurin series for $f(x)=(1+x)^{-3}$.
(b) Find a formula for the n^{th} term in the series.
(c) Replace x by $-x$ to obtain the first four nonzero terms in the Maclaurin series for $(1-x)^{-3}$.

12.4 Manipulating Power Series

Where they converge, power series behave like polynomials. We can differentiate or integrate them term-by-term. We can add, subtract, multiply, and divide them, whether the power series is in x or $x-a$. Most of the discussion and examples in this section will involve power series in x, but remember that the same ideas and results also apply to power series in $x-a$.

While the proofs of these facts will be deferred to more advanced courses, it is very reasonable to understand and to become fluent with these ideas and methods now.

Differentiating a Power Series

In Section 3.3 we showed that we can differentiate the sum of a finite number of functions by adding their derivatives. Theorem 12.4.1 generalizes this to power series in x.

Theorem 12.4.1: Differentiating a Power Series

Assume $R>0$ and that $\sum_{k=0}^{\infty} b_k x^k = b_0 + b_1 x + b_2 x^2 + \cdots$ converges to $f(x)$ for all $|x|<R$. Then, for every $|x|<R$,

(a) *f is differentiable,*

(b) *$\sum_{k=1}^{\infty} k b_k x^{k-1}$ converges to $f'(x)$, and*

(c) *the derivative of $f(x)$ can be written as* $\quad f'(x) = \sum_{k=1}^{\infty} k b_k x^{k-1} = b_1 + 2b_2x^2 + 3b_3x^3 + \cdots.$

This theorem is important because it says, if $f(x)$ is represented by a power series, then $f'(x)$ is represented by the power series obtained by differentiating that power series term-by-term. This is very different from the earlier property of differentiation that says the derivative of the sum of a finite number of functions is the sum of their derivatives.

Because f is differentiable it is continuous. Thus the limit as x approaches 0 of $\sum_{k=0}^{\infty} b_k x^k$ is b_0, the value of the series when $x=0$. This was used without justification in Example 2 in Section 12.2. Theorem 12.4.1 tells us that that manipulation of power series is justified.

EXAMPLE 1. Obtain a power series for the function $\dfrac{1}{(1-x)^2}$ from the power series for $\dfrac{1}{1-x}$.

SOLUTION The first observation is that $1/(1-x)^2$ is the square of $1/(1-x)$, and that $1/(1-x)$ has a simple series representation as a geometric series. But, multiplying series term-by-term can be tedious, as we will experience later in this section. A more useful observation is that the derivative of $1/(1-x)$ is $1/(1-x)^2$.

From the formula for the sum of a geometric series, we know that

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots \qquad \text{for } |x|<1.$$

According to Theorem 12.4.1, differentiating both sides of this equation produces a valid equation, namely

$$\begin{aligned}
\frac{1}{(1-x)^2} &= \frac{d}{dx}\left((1-x)^{-1}\right) && (\text{ differentiation, chain rule })\\
&= \frac{d}{dx}\left(1 + x + x^2 + x^3 + \cdots + x^k + \cdots\right) && (\text{ geometric series } (R=1)\text{ })\\
&= 0 + 1 + 2x + 3x^2 + \cdots + kx^{k-1} + \cdots && (\text{ differentiation, term-by-term }).
\end{aligned}$$

This can also be answered using summation notation. The geometric series is $1/(1-x) = \sum_{k=0}^{\infty} x^k$. When we differentiate both sides, we obtain $1/(1-x)^2 = \sum_{k=1}^{\infty} kx^{k-1}$. Figure 12.4.1 shows that $y = 1+2x+3x^2$ (in red) approximates $y = 1/(1-x)^2$ (in cyan), particularly for x near 0.

In either case, the resulting series has the same radius of convergence as the original series, namely $R = 1$. But, Theorem 12.4.1 does not say anything about convergence of the series for $1/(1-x)^2$ at the endpoints of the interval of convergence. These still have be checked individually.

When $x = 1$ the series is $\sum_{k=1}^{\infty} k$, which diverges because the terms do not approach 0. This is not surprising, because the derivative and the original function are not defined when $x = 1$. When $x = -1$, $1/(1-x)^2 = 1/4$, and the derivative is well-defined. But, when the series for the derivative is evaluated at $x = -1$ we get the series $\sum_{k=0}^{\infty} (-1)^{k-1} k$. As when $x = 1$, its terms do not converge to zero and the series diverges.

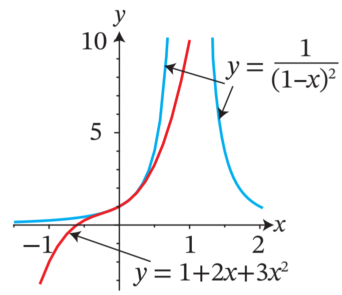


Figure 12.4.1

The interval of convergence for $\sum_{k=1}^{\infty} kx^{k-1}$, the power series for $1/(1-x)^2$ is $-1 < x < 1$. ♦

When $f(x)$ has a power series representation $b_0 + b_1x + b_2x^2 + \cdots$, Theorem 12.4.2 gives meaning to its coefficients.

Theorem 12.4.2: Formula for Coefficients in $\sum_{k=0}^{\infty} b_k x^k$

Suppose the function $f(x)$ is represented by the power series with a positive radius of convergence, $R > 0$:

$$f(x) = \sum_{k=0}^{\infty} b_k x^k = b_0 + b_1 x + b_2 x^2 + \cdots + b_k x^k + \cdots \qquad \text{for } |x| < R.$$

Then, the coefficient of x^k is given by $b_k = \dfrac{f^{(k)}(0)}{k!}$ for $k = 1, 2, 3, \ldots$.

The proof of Theorem 12.4.2 is practically the same as the derivation of the formulas for the coefficients of Taylor polynomials in Section 5.5. It consists of repeated differentiation of $f(x)$ and evaluation of the higher derivatives of $f(x)$ at $x = 0$.

Theorem 12.4.2 also tells us that there can be at most one series of the form $\sum_{k=0}^{\infty} b_k x^k$ that represents $f(x)$, for the coefficients b_k are completely determined by $f(x)$ and its derivatives. It must be the Maclaurin series we obtained in Section 12.1. For instance, the series $1+2x+3x^2+4x^3+\cdots$, which we obtained by differentiating another series, must be the Maclaurin series associated with $1/(1-x)^2$. To put it another way, if two power series represent the same function throughout an interval they are identical term-by-term.

Integrating a Power Series

We can also integrate a power series term-by-term.

Theorem 12.4.3: Integrating a Power Series

Assume that

$$f(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_k x^k + \cdots = \sum_{k=0}^{\infty} b_k x^k \qquad \text{for } |x| < R$$

where R is positive. Then the term-by-term antiderivative of the power series for $f(x)$,

$$b_0 x + b_1 \frac{x^2}{2} + b_2 \frac{x^3}{3} + \cdots + b_k \frac{x^{k+1}}{k+1} + \cdots = \sum k = 0^{\infty} b_k x^{k+1},$$

has radius of convergence R and is an antiderivative of $f(x)$ for $|x| < R$.

The next example shows the power of Theorem 12.4.3.

EXAMPLE 2. Integrate the power series for $\dfrac{1}{1+x}$ to obtain the power series for $\ln(1+x)$.

SOLUTION In the geometric series $1/(1-x) = 1 + x + x^2 + \cdots$ for $|x| < 1$ replace x by $-x$ and obtain, for all $|x| < 1$,

$$\begin{aligned}\frac{1}{1+x} &= 1 - x + x^2 - x^3 + x^4 - \cdots \\ &= \sum_{k=0}^{\infty} (-1)^k x^k\end{aligned}$$

Both $\ln(1+x)$ and the term-by-term antiderivative of the Maclaurin series for $1/(1+x)$, namely, $x - x^2/2 + x^3/3 - x^4/4 + x^5/5 - \cdots$ are antiderivatives of $1/(1+x)$. Therefore they differ by a constant. To determine this constant, evaluate both expressions when $x = 0$: $\ln(1+0) = \ln(1) = 0$ and $\sum_{k=1}^{\infty}(-1)^{k+1}\frac{0^k}{k} = 0$. Because these are equal, the constant is 0. Thus, for $|x| < 1$,

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \cdots = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^k}{k}$$

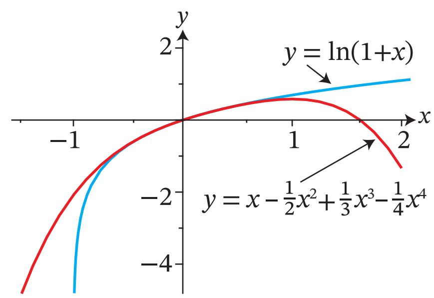


Figure 12.4.2

The convergence for $|x| < 1$ is guaranteed by Theorem 12.4.3. But, convergence at the endpoints must be checked separately.

When $x = 1$, the series is the alternating harmonic series, which converges conditionally. And, when $x = -1$, the series is the negative of the harmonic series, which diverges. The actual interval of convergence for the Maclaurin series for $\ln(1+x)$ is $-1 < x \leq 1$.

Figure 12.4.2 shows the graph of $y = \ln(1+x)$ (in cyan) and the graph of the sum of the first four nonzero terms of its Maclaurin series (in red). ♦

The Algebra of Power Series

We may also add, subtract, multiply, and divide power series like polynomials, as Theorem 12.4.4 asserts.

Theorem 12.4.4: The Algebra of Power Series

Assume that

$$f(x) = \sum_{k=0}^{\infty} b_k x^k = b_0 + b_1 x + b_2 x^2 + \cdots \quad \textit{for } |x| < R_1$$

and

$$g(x) = \sum_{k=0}^{\infty} c_k x^k = c_0 + c_1 x + c_2 x^2 + \cdots \quad \textit{for } |x| < R_2.$$

Then,

$$\begin{aligned}
f(x) + g(x) &= \sum_{k=0}^{\infty} (b_k + c_k) x^k = (b_0 + c_0) + (b_1 + c_1)x + (b_2 + c_2)x^2 + \cdots \\
f(x) - g(x) &= \sum_{k=0}^{\infty} (b_k - c_k) x^k = (b_0 - c_0) + (b_1 - c_1)x + (b_2 - c_2)x^2 + \cdots \\
f(x)g(x) &= \sum_{k=0}^{\infty} p_k x^k \qquad \textit{where } p_k = \sum_{i=0}^{k} b_i c_{k-i} \\
&= b_0 c_0 + (b_0 c_1 + b_1 c_0)x + (b_0 c_2 + b_1 c_1 + b_2 c_0)x^2 + \cdots
\end{aligned}$$

$$\frac{f(x)}{g(x)} = \sum_{k=0}^{\infty} q_k x^k \qquad \textit{where } c_0 \neq 0,\ q_0 = \frac{b_0}{c_0} \textit{ and } q_k = \frac{1}{c_0}\left(b_k - \sum_{i=0}^{k-1} q_i c_{k-i}\right)$$
$$= \frac{b_0}{c_0} + \frac{b_1 - \frac{b_0 c_1}{c_0}}{c_0} x + \cdots$$

NOTES:

1. *The radius of convergence of each new series is the smaller of the two radii of the given series, $R = \min(R_1, R_2)$, except that the radius of convergence for a quotient must also be limited by the requirement that the interval of convergence cannot include any point where $g(x) = 0$.*
2. *The assumption that $c_0 \neq 0$ is equivalent to assuming $g(0) \neq 0$.*

Observation 12.4.5: *Multiplying and Dividing Power Series in Practice*

The formulas for multiplying and dividing power series given in Theorem 12.4.4 are recursive. In general, it is easier to look for another way to perform the required operation.

For a product, two power series are multiplied the same way polynomials are multiplied — term-by-term, starting with the constant terms and working systematically to include all terms for each power.

For a quotient, long division is often the best approach.

EXAMPLE 3. Find the first four nonzero terms of the Maclaurin series for $\frac{e^x}{1-x}$.

SOLUTION There are at least three approaches to this problem. The direct approach is to use Theorem 12.4.2. This requires finding the first three derivatives of $e^x/(1-x)$ evaluated at $x = 0$. A second way is to divide the power series for e^x by $1-x$. The third is to multiply the power series for e^x by the power series for $1/(1-x)$.

We choose the third approach. The power series for e^x is $1 + x + x^2/2! + x^3/3! + \cdots$ with radius of convergence ∞ and the power series for $1/(1-x)$ is $1 + x + x^2 + x^3 + \cdots$ with radius of convergence is 1,

$$\begin{aligned} e^x \cdot \frac{1}{1-x} &= \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\right)\left(1 + x + x^2 + x^3 + \cdots\right) \\ &= (1\cdot 1) + (1\cdot 1 + 1\cdot 1)x + \left(1\cdot 1 + 1\cdot 1 + \frac{1}{2!}\cdot 1\right)x^2 + \left(1\cdot 1 + 1\cdot 1 + \frac{1}{2!}\cdot 1 + \frac{1}{3!}\cdot 1\right)x^3 + \cdots \\ &= 1 + 2x + \frac{5}{2}x^2 + \frac{8}{3}x^3 + \cdots. \end{aligned}$$

According to Theorem 12.4.2, the series has radius of convergence $R = 1$, the smaller of 1 and ∞. ♦

EXAMPLE 4. Find the first four nonzero terms of the Maclaurin series for $\frac{e^x}{\cos(x)}$.

SOLUTION We use Theorem 12.4.4. The Maclaurin series for $e^x/\cos(x)$ is the quotient of the Maclaurin series for e^x and $\cos(x)$. Long division shows that

$$\begin{aligned} \frac{e^x}{\cos(x)} &= \frac{1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots}{1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots} \\ &= 1 + x + x^2 + \frac{2x^3}{3} + \cdots \end{aligned}$$

♦

Observation 12.4.6: ***Utility of Using the Algebra of Power Series***

Example 4 could have been answered using Theorem 12.4.2, but this quickly becomes complicated.

Moreover, from the work done above, the radius of convergence of the quotient is $\pi/2$ even though the power series for e^x and $\cos(x)$ both have infinite radius of convergence, Why? (Because the denominator is zero at odd multiples of $\pi/2$.) QUESTION: Why is this not a contraction of Theorem 12.4.4?

Power Series in a

The theorems and methods of this section were stated for power series in $x = x - 0$. Analogous theorems hold for power series in $x - a$. They may be differentiated and integrated term by term inside the interval in which they converge. For instance, Theorem 12.4.2 generalizes to

Corollary 12.4.7: Formula for b_k in $\sum_{k=0}^{\infty} b_k(x-a)^k$

Let R be positive and suppose that $f(x)$ equals $\sum_{k=0}^{\infty} b_k(x-a)^k$ for $|x-a| < R$. Then $b_k = \frac{f^{(k)}(a)}{k!}$.

The proof of Corollary 12.4.7 is similar to the one of Theorem 12.4.2. It still involves the repeated differentiation of $f(x)$, but now the higher derivatives of $f(x)$ must be evaluated at $x = a$.

Convergence at Endpoints

Observation 12.4.8: ***Convergence at the Endpoints of Interval of Convergence***

1. The theorems in this section include information on the radius of convergence but the convergence at the endpoints is never mentioned. Each endpoint must be checked separately in every case.
2. When performing algebra or calculus operations on series, the resulting series may converge at both endpoints, at neither endpoint, or at only one of the two endpoints.

In Example 1 it was shown that the series for $1/(1-x)^2$ does not converge at either end of the interval. Its interval of convergence is $(-1, 1)$.

In Example 2 the power series for $\ln(1+x)$ was found to be

$$x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}x^k}{k} \tag{12.4.1}$$

for $|x| < 1$. When $x = 1$ the series is $\sum_{k=1}^{\infty}(-1)^{k+1}/k$, the alternating harmonic series, which converges to $\ln(2)$, as Exercise 29 shows. When $x = -1$ the series is $\sum_{k=1}^{\infty} -1/k$ which diverges because it is the negative of the harmonic series. The interval of convergence for (12.4.1) is $(-1, 1]$.

And, in Example 3, even though we have only the first four nonzero terms of the series, it is clear that each coefficient is positive and larger than the previous one. As a result, when evaluated at either $x = 1$ or $x = -1$, the terms do not approach zero. As a result, the interval of convergence is $-1 < x < 1$.

These examples confirm the last part of Observation 12.4.8: some series converge at both endpoints, some at neither, and others at just one of the two endpoints.

Summary

We showed how to operate with power series to obtain new power series by differentiation, by integration, or by an algebraic operation, such as multiplying or dividing two series. For instance, from the geometric series for $1/(1+x)$, we can obtain the series for $\ln(1+x)$ by integration, or the series for $-1/(1+x)^2$ by differentiation.

In many cases the radius of convergence for a derived power series can be determined from the radius of convergence of the original series. But, do not forget that convergence at the endpoints must be checked separately before knowing the interval of convergence.

EXERCISES for Section 12.4

1. Differentiate the Maclaurin series for $\sin(x)$ to obtain the Maclaurin series for $\cos(x)$.

2. Differentiate the Maclaurin series for e^x to show that $D(e^x) = e^x$.

3. Prove Theorem 12.4.2 by carrying out the necessary differentiations.

4. Since $e^x e^y = e^{x+y}$, the product of the Maclaurin series for e^x and e^y should be the Maclaurin series for e^{x+y}. Check that this is the case for terms up to degree three in the series for e^{x+y}.

In Exercises 5 to 10, use power series to determine

5. $\lim_{x\to 0} \dfrac{(1-\cos(x))^3}{x^6}$

6. $\lim_{x\to 0} \dfrac{\sin(3x)}{\sin(2x)}$

7. $\lim_{x\to 0} \dfrac{\sin^2(x^3)e^x}{(1-\cos(x^2))^3}$

8. $\lim_{x\to 0} \left(\dfrac{1}{\sin(x)} - \dfrac{1}{\ln(1+x)}\right)$

9. $\lim_{x\to 0} \dfrac{(e^x-1)^2(\cos(3x))^2}{\sin(x^2)}$

10. $\lim_{x\to 0} \dfrac{\sin(x)(1-\cos(x))}{e^{x^3}-1}$

In Exercises 11 and 12, (a) use algebraic operations to obtain the first three nonzero terms in the Maclaurin series and (b) state the radius of convergence.

11. $e^x \sin(x)$

12. $\dfrac{x}{\cos(x)}$

---

13. Answer these questions. See Exercise 22 for the interval of convergence of the power series for $\arctan(x)$.

(a) Show that, for $|t| < 1$, $\dfrac{1}{1+t^2} = 1 - t^2 + t^4 - t^6 + \cdots$.

(b) Use Theorem 12.4.3 to show that, for $|x| < 1$, $\arctan(x) = x - \dfrac{x^3}{3} + \dfrac{x^5}{5} - \dfrac{x^7}{7} + \cdots$.

(c) Give the formula for the k^{th} term of the series in (b).

(d) How many terms of the series in (b) are needed to approximate $\arctan\left(\dfrac{1}{2}\right)$ to three decimal places?

(e) Use the formula in (b) to estimate $\arctan\left(\dfrac{1}{2}\right)$ to three decimal places.

14.

(a) Using Theorem 12.4.3, show that for $|x| < 1$, $\displaystyle\int_0^x \frac{dt}{1+t^3} = x - \frac{x^4}{4} + \frac{x^7}{7} - \frac{x^{10}}{10} + \cdots$.

(b) Use (a) to express $\displaystyle\int_0^{0.7} \frac{dt}{1+t^3}$ as a series whose terms are numbers.

(c) How many terms are needed to estimate $\displaystyle\int_0^{0.7} \frac{dt}{1+t^3}$ to three decimal places?

(d) Use (b) to evaluate $\displaystyle\int_0^{0.7} \frac{dt}{1+t^3}$ to three decimal places.

(e) Describe how to evaluate $\displaystyle\int_0^{0.7} \frac{dt}{1+t^3}$ using the fundamental theorem of calculus. Do not evaluate!

(f) Use a computer algebra system to find the exact value of $\displaystyle\int_0^{0.7} \frac{dt}{1+t^3}$.

15. Estimate $\displaystyle\int_0^{1/2} \sqrt{x}e^{-x}\,dx$ to four decimal places.

16. Let $f(x) = \sum_{k=0}^{\infty} k^2 x^k$. (a) What is the domain of f? (b) Find $f^{(100)}(0)$.

17. Let $f(x) = \arctan(x)$. Using the Maclaurin series for $\arctan(x)$, find (a) $f^{(100)}(0)$ and (b) $f^{(101)}(0)$.

18. Find the first four nonzero terms of the Maclaurin series for $\dfrac{e^x}{1-x^2}$ in two different ways: (a) by division of series and (b) by using the formula for the coefficient of the k^{th} term, $b_k = \dfrac{1}{k!}f^{(k)}(0)$.

19. Find the first four nonzero terms of the Maclaurin series for $\dfrac{1-\cos(x)}{1-x^2}$ in two different ways: (a) by division of series and (b) by multiplication of series.

20. Find the first three nonzero terms of the Maclaurin series for $\tan(x)$ in two different ways: (a) by dividing the series for $\sin(x)$ by the series for $\cos(x)$ and (b) by using the formula for the coefficient of the k^{th} term, $b_k = \dfrac{1}{k!}f^{(k)}(0)$.

21. (a) Give a numerical series whose sum is $\displaystyle\int_0^1 \sqrt{x}\sin(x)\,dx$., (b) How many terms are needed to approximate the integral to four decimal places?, and (c) Use (a) to evaluate the integral to four decimal places..

22. The Taylor series for $\arctan(x)$ is $\sum_{k=0}^{\infty}(-1)^k \dfrac{x^{2k+1}}{2k+1}$. Its interval of convergence is $[-1,1]$ but Theorem 12.4.3 tells us only that it converges to $\arctan(x)$ at least on the open interval $(-1,1)$.

(a) Use the Lagrange form for the remainder to show that, when $x = 1$, the series sums to $\arctan(1)$.

(b) Repeat (a), using $x = -1$.

(c) Because $\arctan(1) = \dfrac{\pi}{4}$, the Maclaurin series for $\arctan(1)$ provides a way to approximate $\dfrac{\pi}{4}$ and thus π. Do this by using the first five nonzero terms to estimate π.

(d) Estimate the error in the approximation to π found in (c).

(e) How many terms are needed to obtain an approximate value of π accurate (i) to 2 decimal places? (ii) to 4 decimal places? and (iii) to 12 decimal places?

23. (a) From the Maclaurin series for $\cos(x)$, obtain the Maclaurin series for $\cos(2x)$.

(b) Using the identity $\sin^2(x) = \dfrac{1}{2}(1-\cos(2x))$, obtain the Maclaurin series for $\dfrac{\sin^2(x)}{x^2}$.

(c) Estimate $\displaystyle\int_0^1 \left(\frac{\sin(x)}{x}\right)^2 dx$ using its first three nonzero terms.

(d) Find a bound on the error in the estimate in (c).

24. Let $\sum_{k=0}^{\infty} b_k x^k$ and $\sum_{k=0}^{\infty} c_k x^k$ converge for $|x| < 1$. If they converge to the same limit for each x in $(-1,1)$ must $b_k = c_k$ for every $k = 0,1,2,\dots$?

25. This exercise outlines a way to compute logarithms of numbers larger than 1.

(a) Show that every number $y > 1$ can be written in the form $\frac{1+x}{1-x}$ for some x in $(0,1)$.

(b) When $y = 3$, find x.

(c) Show that if $y = \frac{1+x}{1-x}$, then $\ln(y) = 2\left(x + \frac{x^3}{3} + \cdots + \frac{x^{2n+1}}{2n+1} + \dots\right)$.

(d) Use (b) and (c) to estimate ln(3) to two decimal places.

(e) Is the error in (d) less than the first omitted term?

26. Sam has an idea:

SAM: I have a more direct way of estimating $\ln(y)$ for $y > 1$. I just use the identity $\ln(y) = -\ln\left(\frac{1}{y}\right)$. Because $\frac{1}{y}$ is in $(0,1)$ I can write it as $1-x$, and x is still in (0,1). In short,

$$\ln(y) = -\ln\left(\frac{1}{y}\right) = -\ln(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \dots.$$

It's even an easier formula. And it's better because it doesn't have that coefficient 2 in front.

(a) Is Sam's formula correct?

(b) Use Sam's method to estimate ln(3) to two decimal places.

(c) Which approach is better, Sam's method or the one in Exercise 25?

NOTE: Here, "better" means requires fewer terms for the same level of accuracy.

27. Use the method of Exercise 25 to estimate ln(5) to two decimal places. Include a description of your procedure.

28. Here are six ways to estimate ln(2).

(i) The series for $\ln(1+x)$ when $x = 1$.

(ii) The series for $\ln(1+x)$ when $x = \frac{-1}{2}$.

(iii) The series for $\ln\left(\frac{1+x}{1-x}\right)$ when $x = \frac{1}{3}$.

(iv) The root of $e^x = 2$.

(v) The sum $\frac{1}{n} + \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n}$

(vi) Simpson's method applied to $\int_1^2 \frac{dx}{x}$.

Compute ln(2) using (i)–(vi). Rank them in order from most efficient to least efficient. Explain your rankings.

29. In the discussion of endpoints for the Maclaurin series for $\ln(1+x)$, we showed that the series converges for $x = 1$, but we did not show that its sum is ln(2). To show that it is ln(2), integrate over [0,1] both sides of

$$\frac{1+(-x)^{n+1}}{1+x} = 1 - x + x^2 - x^3 + \cdots + (-1)^n x^n.$$

30. (a) For which numbers x does $\sum_{k=0}^{\infty} k^2 x^k$ converge?

(b) Differentiate the Maclaurin series for $\frac{1}{1-x}$ to obtain the one for $\frac{1}{(1-x)^2}$.

(c) Differentiate the Maclaurin series for $\frac{x}{(1-x)^2}$ to sum the power series in (a).

(d) Use (c) to evaluate the series in (a) when $x = \frac{1}{3}$.

31. This exercise uses power series to give a new perspective on l'Hôpital's rule. Assume that f and g have power series in an open interval containing $x = 0$: $f(x) = \sum_{k=0}^{\infty} b_k x^k$ and $g(x) = \sum_{k=0}^{\infty} c_k x^k$. Assume that $f(0) = 0$, $g(0) = 0$, and $g'(0) \neq 0$. Explain why $\lim_{x\to 0} \frac{f(x)}{g(x)} = \lim_{x\to 0} \frac{f'(x)}{g'(x)}$.

32. In R. P. Feynman, *Lectures on Physics*, Addison-Wesley, Reading, MA, 1963, appears this remark:

> Thus the average velocity is
> $$\langle E\rangle = \frac{\hbar\omega(0 + x + 2x^2 + 3x^3 + \cdots)}{1 + x + x^2 + \cdots}.$$
> Now the two sums which appear here we shall leave for the reader to play with and have some fun with. When we are all finished summing and substituting for x in the sum, we should get — if we make no mistakes in the sum —
> $$\langle E\rangle = \frac{\hbar\omega}{e^{\hbar\omega/kT} - 1}.$$
> This, then, was the first quantum-mechanical formula ever known, or ever discussed, and it was the beautiful culmination of decades of puzzlement.

Do the summing and substituting, given that $x = e^{-\hbar\omega/kT}$.

Exercises 33 to 36 outline a proof that the Maclaurin series for $(1+x)^r$ converges to $(1+x)^r$ for $|x| < 1$. This justifies the assertion that $(1+x)^r = \sum_{k=0}^{\infty} \binom{n}{k} x^k$ for $|x| < 1$. The notation $\binom{n}{k}$ stands for $\frac{n!}{k!(n-k)!}$.

33. Show that $k\binom{r}{k} + (k+1)\binom{r}{k+1} = r\binom{r}{k}$.

34. Let $f(x) = \sum_{k=0}^{\infty} \binom{r}{k} x^k$. (a) Find the interval of convergence. (b) Show that $(1+x)f'(x) = rf(x)$.

35. Using the result from Exercise 34, show that the derivative of $\frac{f(x)}{(1+x)^r}$ is 0.

36. Show that $\frac{f(x)}{(1+x)^r} = 1$, which implies that $\sum_{k=0}^{\infty} \binom{r}{k} x^k = (1+x)^r$. What is the interval of convergence?

---

37. Newton obtained the Maclaurin series for $\arcsin(x)$ by using the binomial series for $\sqrt{1-x^2}$ as follows:

Define the angle $\theta = \arcsin(x) = \angle QOR$ where $Q = (x, y)$ is a point on the unit circle $x^2 + y^2 = 1$, as shown in Figure 12.4.3.

Then $\frac{\theta}{2} = \text{Area } OQR = \text{Area } OPQR - \text{Area } OPQ = \int_0^x \sqrt{1-t^2}\,dt - \frac{1}{2}x\sqrt{1-x^2}$.

(a) Use this setup to obtain Newton's result $\theta = x + \frac{1}{6}x^3 + \frac{3}{40}x^5 + \frac{5}{112}x^7 + \cdots$.

(b) Use $\theta = \arcsin(x) = \int_0^x \frac{dt}{\sqrt{1-t^2}}$ to derive Newton's result (in (d)).

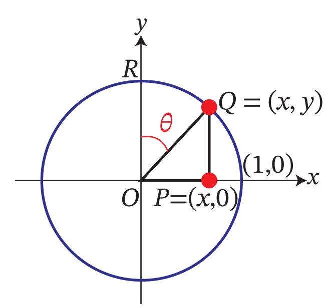


Figure 12.4.3

Historical Note: How Some Calculators Find e^x

The power series in x for e^x is $1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^k}{k!} + \cdots$. When this series is evaluated with $x = 10$, this gives

$$e^{10} = 1 + 10 + \frac{10^2}{2!} + \frac{10^3}{3!} + \cdots + \frac{10^k}{k!} + \cdots$$
$$\approx 1 + 10 + 50 + 333.33333 + \cdots.$$

Although the terms eventually become very small, the first few terms are large. (The fifth term, $10^4/4!$, is about 417. and the largest term, $10^{10}/10!$, is more than 2500.) When x is large, using the series for e^x would not be a good way of calculating e^x.

Some calculators use another method, using pre-computed values of e^x that are permanently stored in memory. In particular,

$$e^1 \approx 2.718281828459 \qquad e^{0.1} \approx 1.1051709181 \qquad e^{10} \approx 22,026.46579$$
$$e^{0.01} \approx 1.0100501671 \qquad e^{100} \approx 2.6881171 \times 10^{43} \qquad e^{0.001} \approx 1.0010005002.$$

To find $e^{315.425}$ the calculator uses identities $e^{x+y} = e^x e^y$ and $(e^x)^y = e^{xy}$ and computes

$$e^{315.425} \approx \left(e^{100}\right)^3 \left(e^{10}\right)^1 \left(e^1\right)^5 \left(e^{0.1}\right)^4 \left(e^{0.01}\right)^2 \left(e^{0.001}\right)^5 \approx 9.71263198 \times 10^{136},$$

which is accurate to six decimal places.

12.5 Complex Numbers

The number line of real numbers coincides with the x-axis of the xy rectangular coordinate system. With its addition, subtraction, multiplication, and division, it is a small part of a number system that occupies the plane, and which obeys the usual rules of arithmetic. This section describes that system, known as the complex numbers.

The Complex Numbers

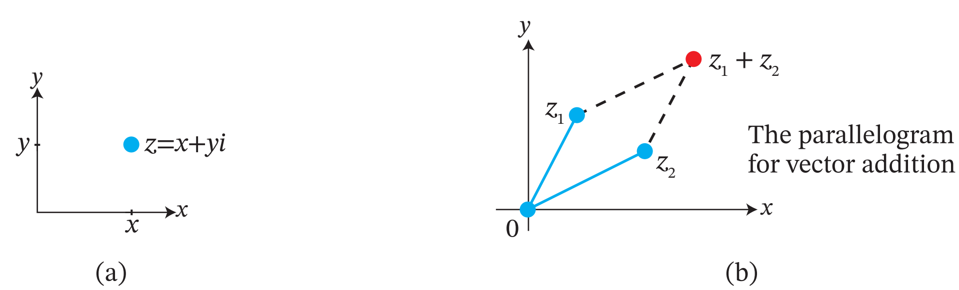


Figure 12.5.1

A **complex number** z is an expression of the form $x + yi$ or $x + iy$, where x and y are real numbers and i is a symbol with the property that $i^2 = -1$. We identify $x + yi$ with the point (x, y) in the xy-plane, as in Figure 12.5.1(a). Every point in the xy-plane may therefore be thought of as a complex number.

Because $(-i)(-i) = i^2 = -1$, both i and $-i$ are square roots of -1. The symbol $\sqrt{-1}$ denotes i rather than $-i$. A complex number that lies on the y-axis is called **imaginary**. A complex number z is the sum of a real number and an imaginary number, $z = x + yi$. The number x is called the **real part** of z, and y is called the **imaginary part**. We write $\mathrm{Re}(z) = x$ and $\mathrm{Im}(z) = y$. Real numbers are on the x-axis, imaginary numbers are on the y-axis.

To add or multiply two complex numbers, follow the usual rules of arithmetic of real numbers, with one additional rule: Whenever you see i^2, replace it by -1. For example, $i^7 = (i^6)i = (i^2)^3 i = (-1)^3 i = -i$.

Observation 12.5.1: ***Geometry of Addition of Complex Numbers***

The sum of the complex numbers z_1 and z_2 is the fourth vertex (opposite O) in the parallelogram determined by the origin O and the points z_1 and z_2, as shown in Figure 12.5.1(b).

For instance, to add the complex numbers $3+2i$ and $-4+5i$, collect the real terms and the imaginary terms:

$$(3+2i)+(-4+5i) = (3-4)+(2i+5i) = -1+7i.$$

(See Figure 12.5.2(a).) Addition does not make use of the defining property of i: $i^2 = -1$. However, multiplication does, as Example 1 will demonstrate.

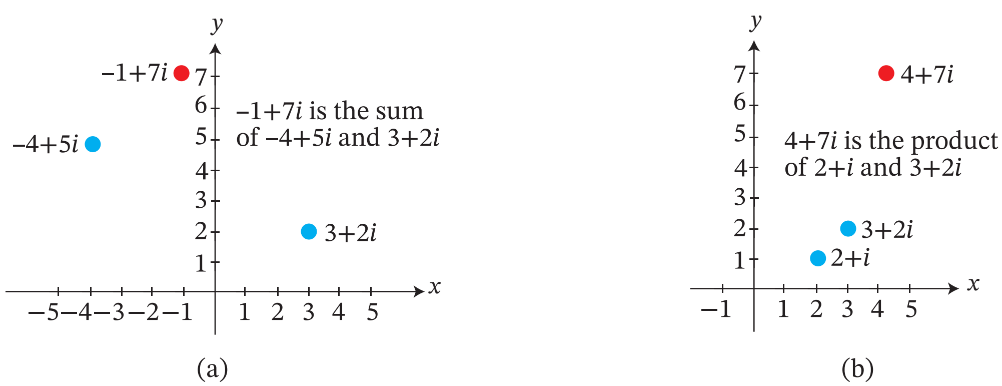


Figure 12.5.2

EXAMPLE 1. Compute the product $(2+i)(3+2i)$.

SOLUTION We multiply complex numbers just as we would multiply binomials. We have

$$(2+i)(3+2i) = 2\cdot 3 + 2\cdot 2i + i\cdot 3 + i\cdot 2i = 6+4i+3i+2i^2 = 6+4i+3i-2 = 4+7i.$$

Figure 12.5.2(b) shows the complex numbers $2+i$, $3+2i$, and their product $4+7i$. ♦

Subtraction of complex numbers is straightforward. For instance,

$$(3+2i)-(4-i) = (3-4)+(2i-(-i)) = -1+3i.$$

Division of complex numbers requires **rationalizing the denominator**. This involves the conjugate of a complex number. The **conjugate** of the complex number $z = x + yi$ is $\overline{z} = x - yi$. For example,

$$\begin{aligned} z\overline{z} &= (x+yi)(x-yi) = x^2+y^2 \\ z+\overline{z} &= (x+yi)+(x-yi) = 2x \\ z-\overline{z} &= (x+yi)-(x-yi) = 2yi. \end{aligned}$$

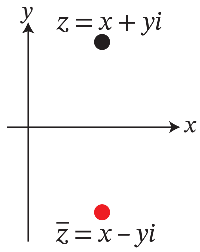


Figure 12.5.3

Thus, $z\bar{z}$ and $z+\bar{z}$ are real, and $z-\bar{z}$ is purely imaginary. Figure 12.5.3 shows that $\bar{z}$ is the image of z reflected across the x-axis. To rationalize the denominator of the quotient of two complex numbers means to find an equivalent fraction with a real-valued denominator. If the fraction is w/z, the denominator can be rationalized by multiplying by $\bar{z}/\bar{z}$.

CLARIFICATION: "Rationalizing" the denominator does not make the denominator a rational number. It involves multiplying by 1, in the special form $\bar{z}/\bar{z}$. Doing so ensures the denominator is real-valued, but not necessarily rational.

EXAMPLE 2. Compute the quotient $\dfrac{1+5i}{3+2i}$.

SOLUTION To rationalize the denominator, we multiply by $(3-2i)/(3-2i)$:

$$\frac{1+5i}{3+2i}=\frac{1+5i}{3+2i}\cdot\frac{3-2i}{3-2i}=\frac{3-2i+15i+10}{9-6i+6i+4i^2}=\frac{13+13i}{13}=1+i.$$

This can be checked by multiplying $(1+i)(3+2i)$ and seeing if the result is $1+5i$. ◆

Now, All Polynomials Have Roots

The complex numbers provide the equation $x^2+1=0$ with two solutions, $x=i$ and $x=-i$. More generally, if $f(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_0$ is a polynomial of degree $n\geq 1$, with real or complex coefficients, then there are n complex numbers z such that $f(z)=0$. This fact is the **Fundamental Theorem of Algebra**. Its proof requires advanced mathematics.

EXAMPLE 3. Solve the quadratic equation $z^2-4z+5=0$.

SOLUTION By the quadratic formula, the solutions are

$$z=\frac{-(-4)\pm\sqrt{(-4)^2-4\cdot 1\cdot 5}}{2\cdot 1}=\frac{4\pm\sqrt{-4}}{2}=\frac{4\pm 2i}{2}=2\pm i.$$

They can be checked by substitution. Checking $z=2+i$,

$$z^2-4z+5=(2+i)^2-4(2+i)+5=(4+4i+i^2)-8-4i+5=4+4i-1-8-4i+5=0+0i=0.$$

It checks. The solution $z=2-i$ can be checked similarly. ◆

REALITY CHECK: The fact that all polynomial equations have solutions does not mean that these solutions are easy to find. This is just like the fact that every continuous function has an antiderivative, but it might not be expressible in terms of elementary functions.

The Geometry of Multiplication of Complex Numbers

The geometric relation between z_1, z_2 and their product z_1z_2 can be described using the magnitude and argument of a complex number. A complex number z other than the origin is at a positive distance r from the origin and has a polar angle θ relative to the positive x-axis. The distance r is called the **magnitude of** z, and θ is called an **argument of** z. (A common synonym for magnitude is **amplitude**.) A complex number has an infinity of arguments differing from each other by an integer multiple of 2π. The complex number 0, which lies at the origin, has magnitude 0 and any angle as argument.

We can think of magnitude and argument of z as polar coordinates r and θ, with the restriction that r is nonnegative. The magnitude of z is denoted $|z|$. The symbol $\arg(z)$ denotes any of the arguments of z, it being understood that if θ is an argument of z, then so is $\theta+2n\pi$ for any integer n.

EXAMPLE 4.
(a) Draw all complex numbers with magnitude 3.
(b) Draw the complex number z of magnitude 3 and argument $\pi/6$.

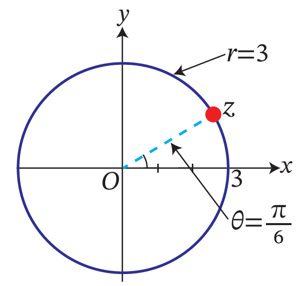


Figure 12.5.4

SOLUTION
(a) The complex numbers of magnitude 3 form a circle of radius 3 with center at 0; the (blue) circle in Figure 12.5.4.
(b) The complex number of magnitude 3 and argument $\pi/6$ is shown (in red) in Figure 12.5.4.

By the Pythagorean theorem, $|x+yi| = \sqrt{x^2+y^2}$. Each complex number $z = x+yi$ other than 0 can be written as the product of a positive real number and a complex number of magnitude 1. To show this, let $z = x+yi$ have magnitude r and argument θ. Using the relation between polar and rectangular coordinates, we have

$$z = r\cos(\theta) + r\sin(\theta)i = r(\cos(\theta) + i\sin(\theta)).$$

The number r is a positive real number. The magnitude of $\cos(\theta) + i\sin(\theta)$ is $\sqrt{\cos(\theta)^2 + \sin(\theta)^2} = 1$.

The next theorem describes how to multiply two complex numbers given in polar form, that is, in terms of their magnitudes and arguments.

Theorem 12.5.2: Geometry of Complex Multiplication

Assume that z_1 has magnitude r_1 and argument θ_1 and that z_2 has magnitude r_2 and argument θ_2. Then $z_1 z_2$ has magnitude $r_1 r_2$ and argument $\theta_1+\theta_2$.

To multiply two complex numbers, (i) multiply their magnitudes and (ii) add their arguments.

Proof of Theorem 12.5.2
By direct computation of the product of two complex numbers:

$$\begin{aligned} z_1 z_2 &= r_1(\cos(\theta_1) + i\sin(\theta_1))\, r_2(\cos(\theta_2) + i\sin(\theta_2)) \\ &= r_1 r_2 (\cos(\theta_1) + i\sin(\theta_1))(\cos(\theta_2) + i\sin(\theta_2)) \\ &= r_1 r_2 (\cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2) + i(\sin(\theta_1)\cos(\theta_2) + \cos(\theta_1)\sin(\theta_2))) \\ &= r_1 r_2(\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)) \end{aligned}$$

The last step uses the identities $\sin(\theta_1+\theta_2) = \sin(\theta_1)\cos(\theta_2) + \cos(\theta_1)\sin(\theta_2)$ and $\cos(\theta_1+\theta_2) = \cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2)$. Thus, the magnitude of $z_1 z_2$ is $r_1 r_2$ and its argument is $\theta_1+\theta_2$. This proves the theorem.

EXAMPLE 5. Find $z_1 z_2$ for z_1 and z_2 in Figure 12.5.5(a).

SOLUTION z_1 has magnitude 2 and argument $\pi/6$; z_2 has magnitude 3 and argument $\pi/4$. Thus $z_1 z_2$ has magnitude $2\cdot 3 = 6$ and argument $\pi/6 + \pi/4 = 5\pi/12$. (See Figure 12.5.5(a).)

EXAMPLE 6. Using the geometric description of multiplication, find the product of the real numbers -2 and -3.

SOLUTION The number -2 has magnitude 2 and argument π. The number -3 has magnitude 3 and argument π. Therefore $(-2)\cdot(-3)$ has magnitude $2\cdot 3 = 6$ and argument $\pi+\pi = 2\pi$. The complex number with magnitude 6 and argument 2π is 6. Thus $(-2)\cdot(-3) = 6$, in agreement with the statement that the product of two negative numbers is positive. (See Figure 12.5.5(b).)

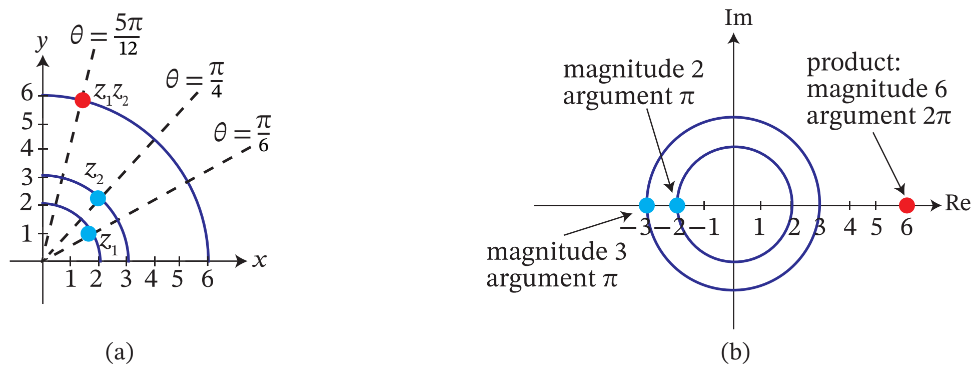


Figure 12.5.5

The Geometry of Division of Complex Numbers

Division of complex numbers in polar form is similar to multiplication, except that the magnitudes are divided and the arguments are subtracted:

$$\frac{r_1(\cos(\theta_1)+i\sin(\theta_1))}{r_2(\cos(\theta_2)+i\sin(\theta_2))} = \frac{r_1}{r_2}(\cos(\theta_1-\theta_2)+i\sin(\theta_1-\theta_2)).$$

The proof of this fact can be found in Exercise 31.

EXAMPLE 7. If $z_1 = 6\left(\cos\left(\frac{\pi}{2}\right)+i\sin\left(\frac{\pi}{2}\right)\right)$ and $z_2 = 3\left(\cos\left(\frac{\pi}{6}\right)+i\sin\left(\frac{\pi}{6}\right)\right)$, find (a) z_1z_2 and (b) $\frac{z_1}{z_2}$.

SOLUTION The graphical representation of these computations is presented in Figure 12.5.6(a).

(a) The product of z_1 and z_2 is

$$\begin{aligned} z_1z_2 &= 6\cdot 3\left(\cos\left(\frac{\pi}{2}+\frac{\pi}{6}\right)+i\sin\left(\frac{\pi}{2}+\frac{\pi}{6}\right)\right)\\ &= 18\left(\cos\left(\frac{2\pi}{3}\right)+i\sin\left(\frac{2\pi}{3}\right)\right)\\ &= 18\left(-\frac{1}{2}+\frac{\sqrt{3}}{2}i\right)\\ &= -9+9\sqrt{3}i. \end{aligned}$$

(b) The quotient of z_1 and z_2 is

$$\begin{aligned} \frac{z_1}{z_2} &= \frac{6}{3}\left(\cos\left(\frac{\pi}{2}-\frac{\pi}{6}\right)+i\sin\left(\frac{\pi}{2}-\frac{\pi}{6}\right)\right)\\ &= 2\left(\cos\left(\frac{\pi}{3}\right)+i\sin\left(\frac{\pi}{3}\right)\right)\\ &= 2\left(\frac{1}{2}+\frac{\sqrt{3}}{2}i\right)\\ &= 1+\sqrt{3}i. \end{aligned}$$

◆

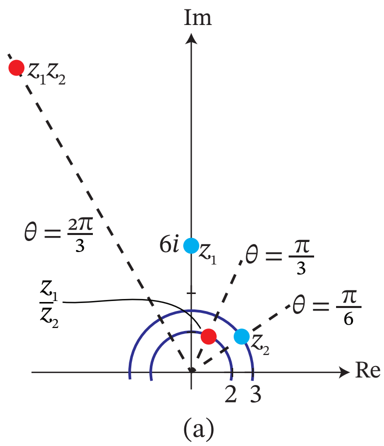
EXAMPLE 8. Compute $(1+i)(3+2i)$ arithmetically, and check the answer using magnitudes and arguments.

SOLUTION The graphical representation of these computations is presented in Figure 12.5.6(b).

$$(1+i)(3+2i) = 3+2i+3i+2i^2 = 3+2i+3i-2 = 1+5i.$$

To check, we verify that $|1+5i| = |1+i||3+2i|$. We have $|1+5i| = \sqrt{1^2+5^2} = \sqrt{26}$, $|1+i| = \sqrt{1^2+1^2} = \sqrt{2}$, and $|3+2i| = \sqrt{3^2+2^2} = \sqrt{13}$. Since $\sqrt{26} = \sqrt{2}\sqrt{13}$, the magnitude of $1+5i$ is the product of the magnitudes of $1+i$ and $3+2i$.

For the argument, $\arg(1+5i) = \arctan(5) \approx 1.3734$, $\arg(1+i) = \arctan(1) \approx 0.7854$, and $\arg(3+2i) = \arctan(2/3) \approx 0.5880$. Since $0.7854+0.5880 = 1.3734$, the argument of $1+5i$ is presumably the sum of the arguments of $1+i$ and $3+2i$. (See also Figure 12.5.6(b).) Exercise 61 outlines another way to check the arguments. ◆



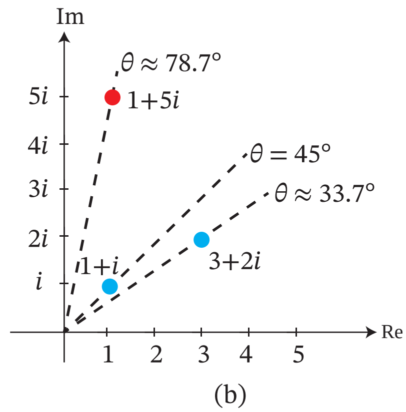


Figure 12.5.6

The Geometry of Powers of Complex Numbers

When the polar coordinates of z are known, it is easy to compute powers of z: z^2, z^3, z^4, ... Let z have magnitude r and argument θ. Then $z^2 = z \cdot z$ has magnitude $r \cdot r = r^2$ and argument $\theta + \theta = 2\theta$. So, to square a complex number, square its magnitude and double its argument (angle).

More generally, to compute z^n for any positive integer n, find $|z|^n$ and multiply the argument of z by n. We have

Theorem 12.5.3: DeMoivre's Law

If $z = r(\cos(\theta) + i\sin(\theta))$ and n is a positive integer, then

$$z^n = \left(r(\cos(\theta) + i\sin(\theta))\right)^n = r^n(\cos(n\theta) + i\sin(n\theta))$$

Example 9 illustrates DeMoivre's Law.

EXAMPLE 9. Let z have magnitude 1 and argument $\frac{2\pi}{5}$. Compute and sketch z, z^2, z^3, z^4, z^5, and z^6.

SOLUTION Since $|z| = 1$, it follows that $|z^2| = |z|^2 = 1^2 = 1$. Similarly, for all positive integers n, $|z^n| = 1$. That means that every positive integer power of z, z^n is a point on the unit circle with center at the origin.

The argument of z^2 is twice the argument of z: $2(2\pi/5) = 4\pi/5$. Similarly, $\arg(z^3) = 6\pi/5$, $\arg(z^4) = 8\pi/5$, $\arg(z^5) = 10\pi/5 = 2\pi$, and $\arg(z^6) = 12\pi/5$. Because z^5 has magnitude 1 and argument 2π, $z^5 = 1$. The argument of z^6 is $2\pi + 2\pi/5$. Because z^6 has the same magnitude as z does, and its argument differs from the argument of z by a multiple of 2π, z^6 and z lie on the same ray from the origin. Also, both lie on the unit circle with center at the origin. Therefore, $z^6 = z$. Or algebraically, $z^6 = z^{5+1} = z^5 \cdot z = 1 \cdot z = z$. Figure 12.5.7 shows that the powers of z are the vertices of a regular pentagon. ◆

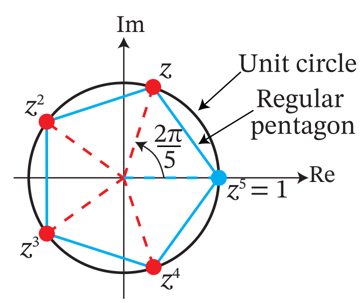


Figure 12.5.7

The equation $x^5 = 1$ has only one real root, $x = 1$. However, it has four complex roots. For instance, the number z shown in Figure 12.5.7 is a solution of $x^5 = 1$ since $z^5 = 1$. Another root is z^2, since $\left(z^2\right)^5 = z^{10} = \left(z^5\right)^2 = 1^2 = 1$. Similarly, z^3 and z^4 are roots of $x^5 = 1$. There are five roots: 1, z, z^2, z^3, and z^4.

The powers of i will be needed in the next section. They are $i^2 = -1$, $i^3 = i^2 \cdot i = (-1)i = -i$, $i^4 = i^3 \cdot i = (-i)i = -i^2 = 1$, $i^5 = i^4 \cdot i = i$, and so on. They repeat in blocks of four: for any n, $i^{n+4} = i^n$.

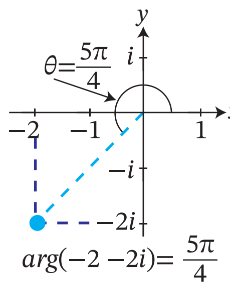


Figure 12.5.8

It is often useful to express a complex number $z = x + yi$ in polar form. Its magnitude is $|z| = \sqrt{x^2 + y^2}$. To find θ, it is best to sketch z to see in which quadrant it lies. Although $\tan(\theta) = y/x$ we cannot say that $\theta = \arctan(y/x)$, since $\arctan(u)$ lies between $-\pi/2$ and $\pi/2$. However, the angle of z may lie in the second or third quadrant

For instance, to put $z = -2 - 2i$ in polar form, first sketch z, as in Figure 12.5.8. We have $|z| = \sqrt{(-2)^2 + (-2)^2} = \sqrt{8}$ and $\arg(z) = 5\pi/4$. Thus

$$z = \sqrt{8}\left(\cos\left(\frac{5\pi}{4}\right) + i\sin\left(\frac{5\pi}{4}\right)\right).$$

The Geometry of Roots of Complex Numbers

A complex number z, other than 0, has exactly n n^{th} roots for a positive integer n. (There are an endless number of ways to write these roots, but there are only n distinct complex numbers.) They can be found by expressing z in polar coordinates. If $z = r(\cos(\theta) + i\sin(\theta))$, that is, has magnitude r and argument θ, then one n^{th} root of z is

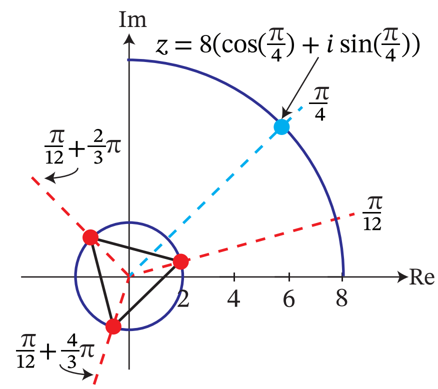


Figure 12.5.9

$$\sqrt[n]{z} = z^{1/n} = r^{1/n}\left(\cos\left(\frac{\theta}{n}\right) + i\sin\left(\frac{\theta}{n}\right)\right).$$

To check that this is an n^{th} root of z, raise it to the n^{th} power.

To find the other n^{th} roots of z, first change the argument z from θ to $\theta + 2k\pi$, where $k = 1, 2, \ldots, n-1$. Then

$$r^{1/n}\left(\cos\left(\frac{\theta + 2k\pi}{n}\right) + i\sin\left(\frac{\theta + 2k\pi}{n}\right)\right)$$

is also an n^{th} root of z. (Why?)

For instance, let $z = 8(\cos(\pi/4) + i\sin(\pi/4))$. Then the three cube roots of z have magnitude $8^{1/3} = 2$. Their arguments are

$$\frac{\pi/4}{3} = \frac{\pi}{12}, \qquad \frac{\pi/4 + 2\pi}{3} = \frac{\pi}{12} + \frac{2\pi}{3}, \qquad \frac{\pi/4 + 4\pi}{3} = \frac{\pi}{12} + \frac{4\pi}{3}.$$

The roots are shown in Figure 12.5.9, along with z; they form an equilateral triangle.

Observation 12.5.4: ***Geometry of Roots of a Complex Number***

The n roots of the equation $z^n = a$ are the vertices of a regular polygon with n sides.

Summary

The real numbers are a small part of the complex numbers, which fill the xy-plane. Complex numbers are added using a parallelogram law. To multiply them, multiply their magnitudes and add their arguments. The complex numbers provide a geometric explanation for the rule that the product of two negative numbers is positive. We also saw how to raise a complex number to a power and how to take its roots. Just as we view points on the x-axis as numbers that can be added and multiplied, we now can view points in the xy-plane as numbers that can be added and multiplied.

EXERCISES for Section 12.5

In Exercises 1 to 4 express the complex numbers in the form $x + yi$.

1. (a) $(5-2i)^2$ (b) $(2+3i)(2-3i)$ (c) $\dfrac{1}{2-i}$ (d) $\dfrac{3+2i}{4-i}$

2. (a) $(2+3i)^2$ (b) $(1+i)(3-i)$ (c) $\dfrac{4}{3-i}$ (d) $\dfrac{1+5i}{2-3i}$

3. (a) $(1+3i)^2$ (b) $(1+i)(1-i)$ (c) i^{-3} (d) $\dfrac{4+\sqrt{2}i}{2+i}$

4. (a) $(1+i)^3$ (b) $(3+i)(3-i)$ (c) $\dfrac{i}{1-i}$ (d) $\dfrac{5+2i}{5-2i}$

In Exercises 5 to 8 express the number in polar form $r(\cos(\theta) + i\sin(\theta))$ with θ in $[0, 2\pi)$.

5. $\sqrt{3}+i$ **6.** $\sqrt{3}-i$ **7.** $\sqrt{2}+\sqrt{2}i$ **8.** $-4+4i$

In Exercises 9 to 12 express the number in both polar and rectangular forms.

9. $(-1+i)^{10}$ **10.** $(\sqrt{3}+i)^4$ **11.** $(2+2i)^8$ **12.** $1-\sqrt{3}i)^7$

---

13. Rationalize the denominator in each fraction. That is, express the fraction as a fraction whose denominator does not have a square root or i.

(a) $\dfrac{1}{1+\sqrt{2}}$ (b) $\dfrac{1}{2-i}$ (c) $\dfrac{2-\sqrt{3}}{\sqrt{3}+2}$ (d) $\dfrac{3+2i}{i-3}$

14. For each equation, (i) find all solutions, (ii) plot all solutions in the complex plane, and (iii) check that each solution satisfies the equation.

(a) $x^2+x+1=0$ (b) $x^2-3x+5=0$ (c) $2x^2+x+1=0$ (d) $3x^2+4x+5=0$

15. This problem deals with the solutions of $f(x) = 2x^3-6x^2+7x-3 = 0$.

(a) Verify that $x = 1$ is a solution of $f(x) = 0$.
(b) Use the quadratic formula to find the other two solutions of $f(x) = 0$.
(c) Verify that each solution found in (b) satisfies $f(x) = 0$.
(d) Plot the solutions.

Each of Exercises 16 to 19 present two complex numbers, z_1 and z_2. For each pair of numbers, (a) plot z_1 and z_2, (b) find z_1z_2 using the polar form, (c) write z_1 and z_2 in the rectangular form $x + yi$, (d) using (c), compute z_1z_2, and (e) check that (b) and (d) give the same point.

16. Let z_1 have magnitude 2 and argument $\dfrac{\pi}{6}$, and let z_2 have magnitude 3 and argument $\dfrac{\pi}{3}$.

17. Let z_1 have magnitude 2 and argument $\dfrac{\pi}{4}$, and let z_2 have magnitude 3 and argument $\dfrac{3\pi}{4}$.

18. Let z_1 have magnitude 2 and argument $\dfrac{\pi}{6}$, and let z_2 have magnitude $\dfrac{1}{2}$ and argument $\dfrac{5\pi}{6}$.

19. Let z_1 have magnitude 2 and argument $\dfrac{\pi}{4}$, and let z_2 have magnitude 3 and argument $\dfrac{\pi}{4}$.

---

20. The complex number z has argument $\frac{\pi}{3}$ and magnitude 1. Find and plot (a) z^2, (b) z^3, (c) z^4, and (d) $1/\overline{z}$.

21. Find and plot (a) i^3, (b) i^4, (c) i^5, and (d) i^{73}.

22. Let z have magnitude 2 and argument $\frac{\pi}{6}$.

(a) What are the magnitude and argument of z^2, z^3, and z^4?
(b) Sketch z, z^2, z^3, and z^4.
(c) What are the magnitude and argument of z^n as $n \to \infty$?

23. Let z have magnitude 0.9 and argument $\frac{\pi}{4}$.
(a) Find and plot z^2, z^3, z^4, z^5, and z^6. (b) What happens to z^n as $n \to \infty$?

24. Find and plot all solutions of $z^5 = 32\left(\cos\left(\frac{\pi}{4}\right) + i\sin\left(\frac{\pi}{4}\right)\right)$.

25. Find and plot all solutions of $z^4 = 8 + 8\sqrt{3}i$.

26. Let z have magnitude r and argument θ. Let w have magnitude $1/r$ and argument $-\theta$. Show that $zw = 1$. (w is called the **reciprocal of** z, and is denoted z^{-1} or $1/z$.)

27. Find z^{-1} if $z = 4 + 4i$.

28. (a) By substitution, verify that $x = 2 + 3i$ is a solution of $x^2 - 4x + 13 = 0$.
(b) Use the quadratic formula to find all solutions of $x^2 - 4x + 13 = 0$.

29. Write each complex number in polar form: (a) $5 + 5i$, (b) $-\frac{1}{2} - \frac{\sqrt{3}}{2}i$, (c) $-\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i$, (d) $3 + 4i$, and (e) $\frac{1}{4i}$.

30. Write in rectangular form as simply as possible: In (d), express the answer to at least three decimal places.

(a) $3\left(\cos\left(\frac{3\pi}{4}\right) + i\sin\left(\frac{3\pi}{4}\right)\right)$, (b) $2\left(\cos\left(\frac{\pi}{6}\right) + i\sin\left(\frac{\pi}{6}\right)\right)$, (c) $10\left(\cos(\pi) + i\sin(\pi)\right)$, and (d) $\frac{1}{5}\left(\cos(22^\circ) + i\sin(22^\circ)\right)$.

31. Let z_1 have magnitude r_1 and argument θ_1, and let z_2 have magnitude r_2 and argument θ_2. Explain why (a) the magnitude of $\frac{z_1}{z_2}$ is $\frac{r_1}{r_2}$ and (b) the argument of $\frac{z_1}{z_2}$ is $\theta_1 - \theta_2$.

32. Compute the quotient $\dfrac{\cos\left(\frac{5\pi}{4}\right) + i\sin\left(\frac{5\pi}{4}\right)}{\cos\left(\frac{3\pi}{4}\right) + i\sin\left(\frac{3\pi}{4}\right)}$ in two ways: (a) by the result of Exercise 31 and (b) by rationalizing the denominator.

33. Compute (a) $(2+3i)(1+i)$, (b) $\frac{2+3i}{1+i}$, (c) $(7-3i)\overline{(7-3i)}$, and (d) $3(\cos(42^\circ)+i\sin(42^\circ))\cdot 5(\cos(168^\circ)+i\sin(168^\circ))$.

34. Compute (a) $\dfrac{\sqrt{8}(\cos(147^\circ) + i\sin(147^\circ))}{\sqrt{2}(\cos(57^\circ) + i\sin(57^\circ))}$, (b) $\frac{1}{3-i}$, (c) $\left((\cos(52^\circ) + i\sin(52^\circ)\right)^{-1}$, and (d) $\left(\cos\left(\frac{\pi}{6}\right) + i\sin\left(\frac{\pi}{6}\right)\right)^{12}$.

35. Compute (a) $(4+3i)(4-3i)$, (b) $\frac{3+5i}{-2+i}$, (c) $\frac{1}{2+i}$, and (d) $\left(\cos(\left(\frac{\pi}{12}\right) + i\sin(\left(\frac{\pi}{12}\right)\right)^{20}$.

36. Compute (a) $(r(\cos(\theta) + i\sin(\theta))^{-1}$, (b) $\mathrm{Re}\left((r(\cos(\theta) + i\sin(\theta)))^{10}\right)$, and (c) $\dfrac{3\left(\cos\left(\frac{\pi}{6}\right) + i\sin\left(\frac{\pi}{6}\right)\right)}{5 - 12i}$.

37. Find and plot all solutions of $z^3 = i$.

38. Plot all complex numbers z such that (a) $z^6 = 1$, (b) $z^6 = 64$, and (c) $z^6 = -1$.

39. (a) Why is the symbol $\sqrt{-4}$ ambiguous? (b) Plot all solutions of $z^2 = -4$.

40. If z_k has argument θ_k and magnitude r_k, $k = 1, 2$, write each of the following in polar form $r(\cos(\theta) + i\sin(\theta))$:
(a) z_1^2 (b) $\frac{1}{z_1}$ (c) $\overline{z_1}$ (d) $z_1 z_2$ (e) $\frac{z_1}{z_2}$ (f) $\frac{1}{\overline{z_1}}$

41. Plot the six sixth roots of (a) 1, (b) 64, (c) i, (d) -1, and (e) $\frac{-1}{2}+\frac{\sqrt{3}}{2}i$.

42. Use DeMoivre's law to find formulas for (a) $\cos(3\theta)$ and (b) $\sin(3\theta)$ in terms of $\cos(\theta)$ and $\sin(\theta)$.

43. Suppose $|z_1| = 1$ and $|z_2| = 1$. (a) How large can $|z_1 + z_2|$ be? (b) What can be said about $|z_1 z_2|$?

44. Show that (a) $\overline{z_1 z_2} = \overline{z_1} \cdot \overline{z_2}$ and (b) $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$.

45. Let $z = \frac{1+i}{\sqrt{2}}$.
(a) Compute z^2 algebraically, (b) compute z^2 by putting z into polar form, and (c) plot z, z^2, z^3, z^4, and z^5.

46. Let a, b, and c be complex numbers with $a\left(b^2-4ac\right) \neq 0$. Show that $ax^2+bx+c=0$ has two distinct roots.

47. Find and plot the roots of $x^2+ix+3-i=0$.

48. For each equation, compute the roots of the equation and plot them on the same axes:
(a) $x^2-3x+2=0$ (b) $x^2-3x+2.25=0$ (c) $x^2-3x+2.5=0$ (d) $x^2-3x+1.5=0$

49. The complex number $z = t+i$ (t a real number) lies on the line $y=1$.
(a) Plot z^2 for $t = 0, 1, -1$, and for at least two other values of t.
(b) Find the equation in rectangular coordinates of the curve on which z^2 lies.

50. The complex number $z = t+\frac{i}{t}$ $(t>0)$ lies on the curve $y=\frac{1}{x}$.
(a) Plot z^2 for $t = 1, 2, 3$, and for at least two other (positive) values of t.
(b) Determine an equation of the curve on which z^2 lies.

51. Let C be the same curve as in the preceding exercise.
(a) Sketch $1/\overline{z}$ for at least four choices of z on C.
(b) Find an equation in rectangular coordinates of the curve swept out by $\frac{1}{z}$ for points on C.

52. (a) Draw the curve on which $z = t+ti$ lies.
(b) Draw the curve on which z^2 lies.
(c) Give equations in rectangular coordinates for both curves.

53. For each complex number, z, plot (on the same set of axes) the z, its conjugate $(\overline{z})$, and its reciprocal $\left(\frac{1}{z}\right)$:
(a) $z = 1+\sqrt{3}i$, (b) $z=\frac{1+i}{\sqrt{2}}$, (c) $z=3$, and (d) $z=2i$. (e) Give a verbal explanation (no equations and no graphs) of the relationships among z, $\overline{z}$, and $\frac{1}{z}$.

54. For what complex numbers z is $\overline{z} = 1/z$?

55. Let z be a point on the line $x+y=1$.
(a) Plot z^2 for at least 5 points on the line.
(b) For z on the line $x+y=1$ write z^2 in the form $u+vi$.
(c) Find an equation in u and v for the curve on which z^2 lies.
(d) What type of curve is the curve in (c)?

56. Let $z=\frac{1+i}{2}$. (a) Sketch the numbers z^n for $n = 1, 2, 3, 4$, and 5. (b) What happens to z^n as $n\to\infty$?

57. Let $z=1+i$. (a) Sketch the numbers $z^n/n!$ for $n = 1, 2, 3, 4$, and 5. (b) What happens to $z^n/n!$ as $n\to\infty$?

58. (a) Graph $r=\cos(\theta)$ in polar coordinates.
(b) Pick five points on the curve. Viewing each as complex numbers z, plot z^2.
(c) As z runs through the curve in (a), what curve does z^2 sweep out?

The partial fraction representation of a rational function is simpler when we have complex numbers available because no second-degree polynomial $ax^2 + bx + c$ terms are needed. Exercise 59 indicates why this is so.

59. Let z_1 and z_2 be the roots of $ax^2 + bx + c = 0$, $a \neq 0$.

(a) Using the quadratic formula (or other means), show that $z_1 + z_2 = -b/a$ and $z_1 z_2 = c/a$.

(b) From (a) deduce that $ax^2 + bx + c = a(x - z_1)(x - z_2)$.

(c) Using (b) show that $\dfrac{1}{ax^2 + bx + c} = \dfrac{1}{a(z_1 - z_2)}\left(\dfrac{1}{x - z_1} - \dfrac{1}{x - z_2}\right)$.

Observation 12.5.5: *Partial Fractions with Complex-Valued Roots*

Exercise 59(c) shows that the theory of partial fractions, described in Section 8.4, becomes much simpler when complex numbers are allowed because then only partial fractions of the form $k/(ax+b)^n$ are needed.

---

60. Let $f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$, where the coefficients are real.

(a) Show that if c is a root of $f(x) = 0$, then so is $\overline{c}$.

(b) Show that if c is a root of f and is not real, then $(x - c)(x - \overline{c})$ divides $f(x)$.

(c) Using the fundamental theorem of algebra, show that any fourth-degree polynomial with real coefficients can be expressed as the product of polynomials of degree at most two with real coefficients.

61. Example 8 provides persuasive evidence that $\arctan(5) = \arctan(1) + \arctan\left(\dfrac{2}{3}\right)$. Establish this equation by evaluating the tangent of both sides.

62. The function $f(z)$ has complex-valued inputs and outputs. Its derivative $f'(z)$ is defined the same way as the derivative of a real-valued function, namely: $f'(z_0) = \lim_{\Delta z \to 0} \Delta f(z)/\Delta z$, where $\Delta f(z) = f(z + \Delta z) - f(z)$. Assume that $f'(z_0)$ is not 0. Show that the image of a small disk with center z_0 is approximately a disk with center $f(z_0)$. $\Delta f(z) \approx f'(z_0)\Delta z$, that is, $\Delta f(z)$ is obtained approximately by stretching Δz by the factor $|f'(z_0)|$, and by rotating it by the angle $\arg(f'(z_0))$.

Historical Note: Beyond the Complex Numbers

When we went from the real numbers to the complex numbers we gained roots for all polynomials, but we also lost something.

If x and y are real numbers then there is an ordering: either $x < y$, $y < x$, or $y = x$. Also, $x < y$ and $y < z$ imply $x < z$. That is not the case for complex numbers.

It is possible to introduce an arithmetic in four-dimensional space, called the **quaternions**. (See Exercises 73 and 74 in Section 14.4.)

However, its multiplication does not obey the commutative law, $xy = yx$. In any dimension one can introduce an addition by the parallelogram law, which amounts to adding two points by adding their corresponding coordinates. It is the multiplication that gives trouble. For instance, there is no way to define addition and multiplication in three-dimensional space that satisfies the laws of algebra, such as the distributive law, $x(y + z) = xy + xz$, even if we do not demand that multiplication be commutative.

12.6 The Relation Between Exponential and Trigonometric Functions

Leonard Euler discovered in 1743 that the trigonometric functions can be expressed in terms of the exponential function e^z, where z is complex. This section retraces his discovery. In particular, we show that

$$e^{i\theta} = \cos(\theta) + i\sin(\theta), \qquad \cos(\theta) = \frac{e^{i\theta}+e^{-i\theta}}{2}, \quad \text{and} \quad \sin(\theta) = \frac{e^{i\theta}-e^{-i\theta}}{2i}.$$

Series with Complex-Valued Terms

To relate the exponential function with the trigonometric functions, we will use infinite series $\sum_{k=0}^{\infty} z_k$, where the z_k's are complex numbers. Such a series is said to converge to S if its n^{th} partial sum S_n approaches S in the sense that $|S-S_n| \to 0$ as $n \to \infty$. (Recall that $|z|$ refers to the magnitude of a complex number z.) It is shown in Exercise 37 that if $\sum_{k=0}^{\infty} |z_k|$ (a series with real-valued terms) converges, so does $\sum_{k=0}^{\infty} z_k$, and the complex-valued series $\sum_{k=0}^{\infty} z_k$ is said to **converge absolutely**. As discussed previously in Section 11.5, if a series converges absolutely, we may rearrange the terms in any order without changing the sum. (This includes rearranging the real and complex parts of the terms.)

Let $z_k = x_k + i y_k$, where x_k and y_k are real. If $\sum_{k=0}^{\infty} z_k$ converges, so do $\sum_{k=0}^{\infty} x_k$ and $\sum_{k=0}^{\infty} y_k$. If $\sum_{k=0}^{\infty} z_k = S = a+bi$, then $\sum_{k=0}^{\infty} x_k = a$ and $\sum_{k=0}^{\infty} y_k = b$. The real-valued series $\sum_{k=0}^{\infty} x_k$ is called the real part of $\sum_{k=0}^{\infty} z_k$ and $\sum_{k=0}^{\infty} y_k$ is the imaginary part of $\sum_{k=0}^{\infty} z_k$.

EXAMPLE 1. Determine for which complex numbers z, $\displaystyle\sum_{k=0}^{\infty} \frac{z^k}{k!}$ converges.

SOLUTION First, note that $\left|z^k\right| = |z|^k$. The series $\sum_{k=0}^{\infty} |z|^k/k!$ has real terms and is the Maclaurin series for $e^{|z|}$, which converges for all real numbers. Since the real-valued series $\sum_{k=0}^{\infty} \left|z^k\right|/k!$ converges for all real numbers, $\sum_{k=0}^{\infty} z^k/k!$ converges absolutely for all complex numbers z. Therefore, $\sum_{k=0}^{\infty} z^k/k!$ converges absolutely for all z, it converges for all z. ◆

Defining e^z for Complex Numbers

The Maclaurin series for e^x when x is real suggests the following definition:

Definition: e^z *for complex-valued numbers* z

Let z be a complex number. Define e^z to be the sum of the (absolutely convergent) series $\displaystyle\sum_{k=0}^{\infty} \frac{z^k}{k!}$.

When z is real, $z = x$, e^z is the real-valued exponential function e^x. It can be shown by multiplying the series for e^{z_1} and e^{z_2} that $e^{z_1+z_2} = e^{z_1}e^{z_2}$ in accordance with the basic law of exponents.

Euler's Formula: The Link between $e^{i\theta}$, $\cos(\theta)$, and $\sin(\theta)$

The following theorem of Euler provides the link between the exponential function e^z and the trigonometric functions $\cos(\theta)$ and $\sin(\theta)$.

Formula 12.6.1: *Euler's Formula*

Let θ be a real number. Then $e^{i\theta} = \cos(\theta) + i\sin(\theta)$.

Proof of Euler's Formula (Formula 12.6.1)
First, recall the that the series for e^x, $\sin(x)$, and $\cos(x)$ all are absolutely convergent for all real numbers x:

$$e^x = 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots+\frac{x^k}{k!}+\cdots, \quad \sin(x) = x-\frac{x^3}{3!}+\cdots+\frac{(-1)^k x^{2k+1}}{(2k+1)!}+\cdots, \text{and} \quad \cos(x) = 1-\frac{x^2}{2!}+\cdots+\frac{(-1)^k x^{2k}}{(2k)!}+\cdots.$$

Starting with the definition of the exponential function e^z for a complex-valued number z,

Recall that $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, $i^5 = i$, ...

$$\begin{aligned} e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \cdots. \\ &= 1 + i\theta + \frac{i^2\theta^2}{2!} + \frac{i^3\theta^3}{3!} + \frac{i^4\theta^4}{4!} + \cdots. \\ &= 1 + i\theta - \frac{\theta^2}{2!} - \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} + \cdots. \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right) + i\left(\theta - \frac{\theta^3}{3!} + \cdots\right) \quad \text{(rearranging)} \\ &= \cos(\theta) + i\sin(\theta). \end{aligned}$$

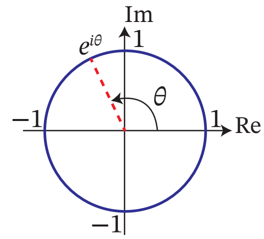


Figure 12.6.1

Figure 12.6.1 shows $e^{i\theta}$ is the complex number on the unit circle with center at the origin and argument θ. ●

Recall, from Section 4.1 (Exercises 71 to 74), that the hyperbolic functions $\cosh(x)$ and $\sinh(x)$ are defined in terms of the exponential function:

$$\cosh(x) = \frac{1}{2}\left(e^x + e^{-x}\right) \quad \text{and} \quad \sinh(x) = \frac{1}{2}\left(e^x - e^{-x}\right).$$

With the aid of Theorem 12.6.1, $\cos(\theta)$ and $\sin(\theta)$ may be expressed in terms of the exponential function.

Theorem 12.6.1: Relationship between Exponential and Trigonometric Functions

Let θ be a real number. Then $\cos(\theta) = \frac{1}{2}\left(e^{i\theta} + e^{-i\theta}\right)$ *and* $\sin(\theta) = \frac{1}{2i}\left(e^{i\theta} - e^{-i\theta}\right)$.

Proof of Theorem 12.6.1
From Euler's formula,

$$e^{i\theta} = \cos(\theta) + i\sin(\theta), \tag{12.6.1}$$

replacing θ by $-\theta$ gives

$$e^{-i\theta} = \cos(-\theta) + i\sin(-\theta).$$

Since cosine is an even function and sine is an odd function,

$$e^{-i\theta} = \cos(\theta) - i\sin(\theta). \tag{12.6.2}$$

The sum of (12.6.1) and (12.6.2) yields $e^{i\theta} + e^{-i\theta} = 2\cos(\theta)$. So

$$\cos(\theta) = \frac{1}{2}\left(e^{i\theta} + e^{-i\theta}\right).$$

Subtraction of (12.6.2) from (12.6.1) yields $e^{i\theta} - e^{-i\theta} = 2i\sin(\theta)$, from which we conclude

$$\sin(\theta) = \frac{1}{2i}\left(e^{i\theta} - e^{-i\theta}\right).$$

These two formulas complete the proof of Theorem 12.6.1. ●

Theorem 12.6.1 shows that the trigonometric functions can be similarly defined in terms of the exponential function, with the aid of complex numbers. That means that once we have the complex numbers we can bypass right triangles and unit circles when defining $\sin(\theta)$ and $\cos(\theta)$. We can even find the derivative of $\sin(\theta)$, for example, without using any geometry, as the following calculation shows.

$$\frac{d}{d\theta}\sin(\theta) = \left(\frac{e^{i\theta}-e^{-i\theta}}{2i}\right)' = \frac{ie^{i\theta}+ie^{-i\theta}}{2i} = \frac{e^{i\theta}+e^{-i\theta}}{2} = \cos(\theta).$$

NOTE: In advanced courses it is proved that the rules for differentiation can be extended to complex-valued functions of complex-valued variables.

Historical Note

It was around 1740 when Euler discovered the connection between the exponential and trigonometric functions is very similar to Maxwell's 1864 discovery of the connection between light and electricity. Maxwell's work is summarized in Chapter 18.

There is an old saying: "God created the complex numbers; anything less is the work of man." It appears this is an urban myth. What is well-documented is that The German mathematician Leopold Kronecker said, in 1886, in a lecture for the Berliner Naturforscher Versammlung, "God made the integers, all else is the work of man."

The license plate of New York University mathematician Martin Davis, whose e-mail signature is "eipye, add one, get zero," is shown in Figure 12.6.2.

Figure 12.6.2

Plotting e^z

When $z = x + yi$, Euler's formula tells us $e^z = e^{x+yi} = e^x e^{yi} = e^x(\cos(y) + i\sin(y))$.

Observation 12.6.2: *Magnitude and Argument of* e^z

The magnitude of e^{x+yi} is e^x and its argument is y.

Observation 12.6.3: *A Foundational Equation in Mathematics:* $e^{i\pi} = -1$

Euler's formula, Formula 12.6.1, asserts, for instance, that $e^{i\pi} = \cos(\pi) + i\sin(\pi) = -1 + i\cdot 0 = -1$.

This remarkable equation links e (the fundamental number in calculus), π (the fundamental number in trigonometry), i (the fundamental complex number), and the negative number -1. Its history recalls the struggles of hundreds of mathematicians to create the number system that we now take for granted. It is as important in mathematics as $F = ma$ or $E = mc^2$ in physics.

EXAMPLE 2. Compute and plot (a) $e^{2+(\pi/6)i}$, (b) $e^{2+\pi i}$, and (c) $e^{2+3\pi i}$.

SOLUTION

(a) $e^{2+(\pi/6)i}$ has magnitude e^2 and argument $\pi/6$.
(b) $e^{2+\pi i}$ has magnitude e^2 and argument π and therefore is $-e^2$.
(c) $e^{2+3\pi i}$ has magnitude e^2 and argument 3π, so it is the same number as the number in (b), $-e^2$.

These three complex numbers are plotted in Figure 12.6.3.

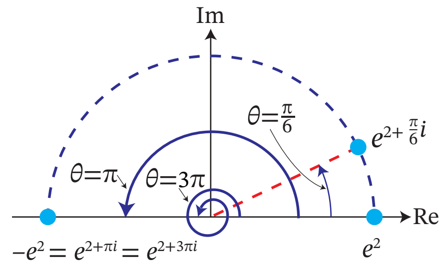


Figure 12.6.3

♦

The next example illustrates a computation in alternating currents. Electrical engineers frequently use j as the symbol for i so they can use i for current.

EXAMPLE 3. Find the real part of $100e^{j(\pi/6)}e^{j\omega t}$. Here t is time, ω is frequency, and j is the mathematician's i.

SOLUTION

$$100e^{j(\pi/6)}e^{j\omega t} = 100e^{j(\pi/6)+j\omega t} = 100e^{j(\pi/6+\omega t)} = 100\left(\cos\left(\frac{\pi}{6}+\omega t\right)+i\sin\left(\frac{\pi}{6}+\omega t\right)\right).$$

Thus, $\mathrm{Re}\left(100e^{j(\pi/6)}e^{j\omega t}\right) = 100\cos\left(\frac{\pi}{6}+\omega t\right)$. ♦

It is sometimes convenient to convert a real-valued problem into one involving complex numbers. For example, in the next example we convert $\cos(\theta)$ to $\mathrm{Re}\left(e^{i\theta}\right)$.

EXAMPLE 4. Evaluate $\displaystyle\sum_{k=0}^{\infty}\frac{\cos(k\theta)}{2^k}$.

SOLUTION From Euler's formula $e^{ik\theta}=\cos(k\theta)+i\sin(k\theta)$ we observe that $\cos(k\theta)=\mathrm{Re}\left(e^{ik\theta}\right)$. Then, we find

$$\sum_{k=0}^{\infty}\frac{\cos(k\theta)}{2^k} = \sum_{k=0}^{\infty}\mathrm{Re}\left(\frac{e^{ik\theta}}{2^k}\right) = \mathrm{Re}\left(\sum_{k=0}^{\infty}\frac{e^{ik\theta}}{2^k}\right). \tag{12.6.3}$$

To simplify the complex-valued expression inside the parentheses, notice that

$$\frac{e^{ik\theta}}{2^k} = \left(\frac{e^{i\theta}}{2}\right)^k.$$

Because $\left|e^{i\theta}/2\right| = 1/2 < 1$, this geometric series converges with sum

$$\begin{aligned}
\sum_{k=0}^{\infty}\frac{e^{ik\theta}}{2^k} &= \frac{1}{1-\left(\frac{e^{i\theta}}{2}\right)} && \text{(sum of geometric series with base } e^{i\theta}/2\text{)}\\
&= \frac{2}{2-\cos(\theta)-i\sin(\theta)} && \text{(Euler's formula and simplify)}\\
&= \frac{2}{2-\cos(\theta)-i\sin(\theta)}\,\frac{2-\cos(\theta)+i\sin(\theta)}{2-\cos(\theta)+i\sin(\theta)} && \text{(rationalize the denominator)}\\
&= \frac{2(2-\cos(\theta)+i\sin(\theta))}{(2-\cos(\theta))^2+(\sin(\theta))^2} && \text{(expand denominator)}\\
&= \frac{2(2-\cos(\theta)+i\sin(\theta))}{5-4\cos(\theta)} && \text{(simplify).}
\end{aligned} \tag{12.6.4}$$

Combining (12.6.4) with (12.6.3) gives

$$\sum_{k=0}^{\infty}\frac{\cos(k\theta)}{2^k}=\mathrm{Re}\left(\sum_{k=0}^{\infty}\frac{e^{ik\theta}}{2^k}\right)=\mathrm{Re}\left(\frac{2(2-\cos(\theta)+i\sin(\theta))}{5-4\cos(\theta)}\right)=\frac{2(2-\cos(\theta))}{5-4\cos(\theta)}.$$

♦

Summary

Using power series, we obtained the fundamental relation $e^{i\theta}=\cos(\theta)+i\sin(\theta)$ and showed that $\cos(\theta)$ and $\sin(\theta)$ can be expressed in terms of the exponential function. We may define even x^n, $x>0$, in terms of the exponential function as $e^{n\ln(x)}$. Similarly, a^x, $a>0$, can be defined as $e^{x\ln(a)}$. (While not discussed in this section, $\ln(x)$ is the inverse of e^x, so it too is obtained from the exponential function.) These observations are the primary evidence supporting the claim that the most fundamental function in calculus is e^x, where x is real or complex.

Historical Note: Why R is called the Radius of Convergence

The following question and answer illustrate why R is called the radius of convergence of a power series.

Q: What is the radius of convergence of the Taylor series in powers of $z-3$ associated with $1/(1+z^2)$?

A: It is the distance from 3 to the nearest complex number at which $1/(1+z^2)$ "blows up," that is, when $1+z^2=0$. This occurs when z is i or $-i$. The distance from 3 to i is $|3-i|=\sqrt{3^2+1^2}=\sqrt{10}$. Also, $|3-(-i)|=\sqrt{10}$. Therefore $R=\sqrt{10}$ and the Taylor series is guaranteed to converge for all complex numbers whose distance from 3 does not exceed $\sqrt{10}$. (See Figure 12.6.4.)

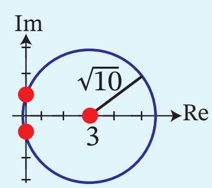


Figure 12.6.4

EXERCISES for Section 12.6

In Exercises 1 to 4 express each number in the form $re^{i\theta}$ for real numbers $r>0$ and $-\pi<\theta\le\pi$.

1. $\frac{e^2}{\sqrt{2}}-\frac{e^2}{\sqrt{2}}i$

2. $3\left(\cos\left(\frac{\pi}{4}\right)+i\sin\left(\frac{\pi}{4}\right)\right)$

3. $5\left(\cos\left(\frac{\pi}{6}\right)+i\sin\left(\frac{\pi}{6}\right)\right)\cdot 3\left(\cos\left(\frac{\pi}{2}\right)+i\sin\left(\frac{\pi}{2}\right)\right)$

4. $7\left(\cos\left(\frac{7\pi}{3}\right)+i\sin\left(\frac{7\pi}{3}\right)\right)$

In Exercises 5 to 10 plot the numbers and state their real and imaginary parts.

5. $e^{5\pi i/4}$

6. $5e^{\pi i/4}$

7. $2e^{\pi i/4}+3e^{\pi i/6}$

8. e^{2+3i}

9. $e^{\pi i/6}e^{3\pi i/4}$

10. $2e^{\pi i}\cdot 3e^{-\pi i/3}$

In Exercises 11 to 18 plot $\exp(z)$ for z equal to

11. 2

12. $\frac{\pi i}{2}$

13. $2-\frac{\pi i}{3}$

14. $-1+\frac{17\pi i}{6}$

15. $\frac{\pi}{4}+3\pi i$

16. $1+\frac{9\pi i}{4}$

17. $2-\frac{\pi i}{3}$

18. $-1+\frac{17\pi i}{6}$

19. Let $z=e^{a+bi}$, where a and b are real constants. (In (f), assume a and b are positive.) Find
(a) $|z|$ (b) $\bar{z}$ (c) z^{-1} (d) $\mathrm{Re}(z)$ (e) $\mathrm{Im}(z)$ (f) $\arg(z)$

20. How far is $\exp(x + yi)$ from (a) the x-axis, (b) the y-axis, and (c) the origin.

21. For what values of a and b is $\lim_{n\to\infty}\left(e^{a+ib}\right)^n = 0$?

22. Find all complex numbers z such that $e^z = 1$.

23. Find all complex numbers z such that $e^z = -1$.

24. Find all z such that $e^z = 3+4i$.

Observation 12.6.4: *Exponential Function Does Not Have an Inverse*

As a result of Exercises 22 to 24, we see that e^z, with z a complex-valued variable, is not one-to-one, so it does not have an inverse. In fact, for every nonzero complex number w there are an infinite number of complex numbers z with $e^z = w$.

Exercises 25 to 28 are related.

25. (a) Find $\left|e^{3+4i}\right|$. (b) Plot e^{3+4i}.

26. (a) Plot all complex numbers of the form e^{x+4i}, x real. (b) Plot all complex numbers of the form e^{3+yi}, y real.

27. If z lies on the line $y = 1$, where does $\exp(z)$ lie?

28. If z lies on the line $x = 1$, where does $\exp(z)$ lie?

---

29. In a physics book there is the remark: "Using the fact that $\left(e^{-i\omega_0 t}\right)^*\left(e^{-i\omega_0 t}\right) = 1$, we can easily evaluate the probability density for these standard waves." Justify this equation. (Assume ω_0 is real and z^* is the conjugate of z.)

30. Find a formula for the trigonometric sum $\sum_{m=0}^{n-1}\cos(m\theta)$.

31. Find a formula for the trigonometric sum $\sum_{m=1}^{n-1}\sin(m\theta)$.

32. Assuming that $e^{z_1+z_2} = e^{z_1}e^{z_2}$ for complex numbers z_1 and z_2, obtain the trigonometric identities for $\cos(A+B)$ and $\sin(A+B)$.

In Exercises 33 and 34 show that the series converges absolutely and evaluate it.

33. $\sum_{k=0}^{\infty}\frac{\cos(k\theta)}{k!}$.

34. $\sum_{k=0}^{\infty}\frac{\sin(k\theta)}{k!}$.

---

35. Find all solutions to (a) $e^z = e^{-z}$, (b) $e^{iz} = e^{-iz}$, and (c) $\sin(z) = 0$.

36. Let z be a complex number and θ a real number. What is the geometric relationship between z and $e^{i\theta}z$? Experiment, conjecture, and explain.

37. This problem shows that if $\sum_{k=0}^{\infty} |z_k|$ converges, so does $\sum_{k=0}^{\infty} z_k$. Let $z_k = x_k + i y_k$ and assume that $\sum_{k=0}^{\infty} |z_k|$ converges.

(a) Show that $\sum_{k=0}^{\infty} |x_k|$ and $\sum_{k=0}^{\infty} |x_{k+1}|$ converge.

(b) Show that $\sum_{k=0}^{\infty} x_k$ and $\sum_{k=0}^{\infty} y_k$ converge.

(c) Show that $\sum_{k=0}^{\infty} (x_k + i y_k)$ converges.

38. Let $f(z)$ be a polynomial with real coefficients.
(a) Show that if $f(a) = 0$, then $f(\overline{a}) = 0$. (b) Show that $\overline{e^z} = e^{\overline{z}}$. (c) Show that $\overline{\sin(z)} = \sin(\overline{z})$.

NOTE: (a) shows that roots of f occur in conjugate pairs: both $z = a$ and $z = \overline{a}$ are solutions of $f(z) = 0$.

39. When z is real, $|\sin(z)| \le 1$ and $|\cos(z)| \le 1$. Do these inequalities hold for complex z?

40. Does the equation $\cos^2(z) + \sin^2(z) = 1$ hold for complex z?

41. Let $z = \frac{1+i}{\sqrt{2}}$. (a) Plot z, $\frac{z^2}{2!}$, $\frac{z^3}{3!}$, and $\frac{z^4}{4!}$. (b) Plot $1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!}$. (c) Plot $e^{(1+i)/\sqrt{2}}$ on the xy-plane.

NOTE: The point plotted in (b) is an estimate for $\exp\left(\frac{1+i}{\sqrt{2}}\right)$.

42. An integral table lists $\int xe^{ax}\,dx = \frac{1}{a^2}e^{ax}(ax-1)$. At first glance, finding $\int xe^{ax}\cos(bx)\,dx$ may appear to be a much harder problem. However, by noticing that $\cos(bx) = \mathrm{Re}\left(e^{ibx}\right)$, it reduces to a simpler problem. Following this approach, find $\int xe^{ax}\cos(bx)\,dx$. (The formula for $\int xe^{ax}\,dx$ holds when a is complex.)

Exercises 43 and 44 investigate the connections between hyperbolic and trigonometric functions. In Section 4.1 we define $\cosh(x) = \frac{1}{2}\left(e^x + e^{-x}\right)$ and $\sinh(x) = \frac{1}{2}\left(e^x - e^{-x}\right)$. The same definitions are used when x is complex. In view of Theorem 12.6.1, define sine and cosine for complex z by $\sin(z) = \frac{1}{2i}\left(e^{iz} - e^{-iz}\right)$ and $\cos(z) = \frac{1}{2}\left(e^{iz} + e^{-iz}\right)$.

43. Show that, for all complex numbers z, (a) $\cosh(z) = \cos(iz)$ and (b) $\sinh(z) = -i\sin(iz)$.

44. Show that, for all complex numbers z, (a) $\sin(z) = i\sinh(iz)$, (b) $\cos(z) = \cosh(iz)$, and (c) $\cosh(z)^2 - \sinh(z)^2 = 1$.

---

45. Sam is at it again, but Jane is not around to hear this idea.

SAM: I don't need power series to define e^z. I just write z as $x + yi$ and define e^{x+yi} to be $e^x(\cos(y) + i\sin(y))$. That's all there is to it.

JANE: Explain yourself.

SAM: If I call this function $E(z)$, then it's easy to check that $E(z_1 + z_2) = E(z_1)E(z_2)$. Moreover, if z is real, then $y = 0$ and $E(z) = e^x$, agreeing with our familiar $\exp(x)$.

(a) Is Sam right?

(b) Does his $E(z)$ obey the basic law of exponents, as he claims?

(c) Jane asks him, "But where did you get the idea for that definition? It seems to float in out of thin air." What is Sam's answer?

46. Later, Jane returns and Sam shares another idea with her.

SAM: I can show that $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ without using Taylor series.

Jane: That would be nice.

Sam: I differentiate the quotient $\dfrac{e^{i\theta}}{\cos(\theta)+i\sin(\theta)}$ and get 0. So it's a constant. Then it's easy to show the constant is 1. That does it.

Jane: But you used that the derivative of e^z is e^z.

Sam: I did, but that follows from the definition of e^z as $\displaystyle\sum_{k=0}^{\infty}\frac{z^k}{k!}$.

Jane: You may be right, but once again, why did you think of $\cos(\theta)+i\sin(\theta)$?

Sam: That fellow, Euler.

Check Sam's calculations. Is his reasoning correc?

12.7 Fourier Series

Section 12.3 introduced power series, sums of terms of the form ax^n, where n is a nonnegative integer and a is a number. We found that in order for a power series can be used to represent a function it must have derivatives of all orders. Now sums of terms of the form $a\cos(kx)$ and $b\sin(kx)$, where a, b, and k are numbers, will be used. This method applies to a broader class of functions, even to the absolute value function, $f(x)=|x|$, which is not differentiable at 0, and to some functions that are not even continuous. The sums, called **Fourier series**, are used in heat conduction, electric circuits, and the theory of sound and mechanical vibrations.

The use of sine and cosine, which are periodic functions, may seem surprising. However, if one thinks in terms of sound, it is quite plausible. A tuning fork produces a pure pitch at a specific frequency. A collection of tuning forks with different pitches can, when struck simultaneously, approximate the sound made by an orchestra. A tuning fork vibrates like $\sin(kt)$ or $\cos(kt)$, where t is time. One set at concert A vibrates at the rate of 440 cycles per second, that is, 440 Hertz (Hz). In this case the sound wave is expressed as $\sin(440(2\pi t))$, for, as t increases by $1/440$ second, the argument $440(2\pi t)$ increases by 2π, enabling the function to complete one cycle.

Historical Note: The Beginnings of Fourier Series

The name Joseph Fourier (1768—1830) is attached to trigonometric series because he explored and applied them in his classic *Analytic Theory of Heat*, published in 1822. He came upon the formulas for the coefficients indirectly, starting with the Maclaurin series for $\sin(x)$ and $\cos(x)$. For the details, see Morris Kline's *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, New York, 1972 (especially pages 671– 675), but see further references in its index.

In the nineteenth and twentieth centuries mathematicians developed a variety of conditions that imply the series converges to the function. The most significant development is due to Lennart Carleson in 1966, which settled a famous and longstanding conjecture.

Periodic Functions

The functions $\cos(x)$ and $\sin(x)$ have period 2π, as $\cos(x+2\pi)=\cos(x)$. Changing the input by 2π does not change the output. It follows that $\cos(x-2\pi)=\cos(x)$, $\cos(x+4\pi)=\cos(x)$, and for any integer n, $\cos(x)$ has $n(2\pi)$ as a period. A function's **natural period**, also called the **period**, is its shortest period. When we say that $\cos(x)$ has period 2π we are saying that the natural period of $\cos(x)$ is 2π.

EXAMPLE 1. Find the period of (a) $\cos(3\pi x)$ and (b) $\cos\left(\dfrac{k\pi x}{L}\right)$, where k is a positive integer and $L>0$.

SOLUTION To determine the period of a function $f(x)$, we ask "By how much must x change for the argument (the input) of $f(x)$ to change by 2π?"

(a) For $3\pi x$ to change by 2π, we solve $3\pi x=2\pi$, obtaining $x=2/3$. Thus $\cos(3\pi x)$ has period $2/3$.

(b) For $\cos(k\pi x/L)$ the same reasoning as in (a) shows that the period is $2L/k$. The larger L is, the longer the period. The larger k is, the shorter the period. For each k, $2L$ is among its periods. ♦

Fourier Series for Functions with Period 2π

We first treat functions that have period 2π. Its values are determined by its values on an interval of length 2π. We choose $(-\pi,\pi]$. Later, we consider the general case where the period is $2L$.

Definition: *Coefficients in Fourier Series*

Let $f(x)$ be a function with period 2π. The **Fourier series** for $f(x)$ is

$$\frac{a_0}{2}+\sum_{k=1}^{\infty}\left(a_k\cos(kx)+b_k\sin(kx)\right) \tag{12.7.1}$$

where the **Fourier coefficients** for $f(x)$ are

$$a_k=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)\,dx \qquad k=0,1,2,\ldots \tag{12.7.2}$$

$$b_k=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(kx)\,dx \qquad k=1,2,\ldots. \tag{12.7.3}$$

provided the integrals in (12.7.2) and (12.7.3) exist.

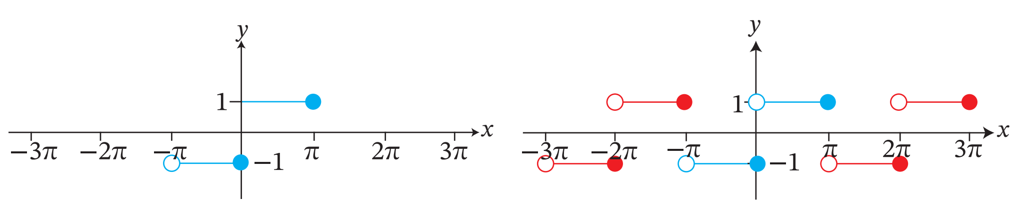
Historical Note: Terminology for Fourier Coefficients

The formulas (12.7.2) and (12.7.3) for the Fourier coefficients, a_k and b_k, are known by various names, including **Euler's formulas** and **Euler-Fourier formulas**. Euler published them in 1777, and Fourier used trigonometric series in his study of heat conduction in the early 1800's but was unaware of Euler's formulas.

After computing two Fourier series, we will show why the coefficients are given by (12.7.2) and (12.7.3).

EXAMPLE 2. Find the Fourier series for $f(x)=\begin{cases}-1 & -\pi<x\le 0\\ 1 & 0<x\le\pi.\end{cases}$

To make $f(x)$ have period 2π, repeat the graph on every interval of the form $(-\pi+2n\pi,\pi+2n\pi]$, where n is an integer. The graph of $f(x)$ is shown in Figure 12.7.1(a) and its extension is shown in Figure 12.7.1(b).



(a) (b)

Figure 12.7.1

SOLUTION First, compute a_0 separately. Because the definition of $f(x)$ comes in two pieces, one for $-\pi < x \le 0$ and one for $0 < x \le \pi$, so the definite integral over $-\pi \le x \le \pi$ into two separate pieces, as follows:

$$\begin{aligned}
a_0 &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx && (\text{ definition of } a_0\text{ })\\
&= \frac{1}{\pi}\int_{-\pi}^{0} f(x)\,dx + \frac{1}{\pi}\int_{0}^{\pi} f(x)\,dx && (\text{ splitting integral into two pieces })\\
&= \frac{1}{\pi}\int_{-\pi}^{0} -1\,dx + \frac{1}{\pi}\int_{0}^{\pi} 1\,dx && (\text{ definition of } f(x)\text{ })\\
&= \frac{1}{\pi}(-\pi) + \frac{1}{\pi}(\pi) && (\text{ recognize definite integrals as areas of rectangles })\\
&= 0.
\end{aligned}$$

Similarly, for $k \ge 1$:

$$\begin{aligned}
a_k &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)\,dx && (\text{ definition of } a_k\text{ })\\
&= \frac{1}{\pi}\int_{-\pi}^{0} f(x)\cos(kx)\,dx + \frac{1}{\pi}\int_{0}^{\pi} f(x)\cos(kx)\,dx && (\text{ splitting integral into two pieces })\\
&= \frac{1}{\pi}\int_{-\pi}^{0} (-\cos(kx))\,dx + \frac{1}{\pi}\int_{0}^{\pi} \cos(kx)\,dx && (\text{ definition of } f(x)\text{ })\\
&= \frac{1}{\pi}\left.\frac{-\sin(kx)}{k}\right|_{-\pi}^{0} + \frac{1}{\pi}\left.\frac{\sin(kx)}{k}\right|_{0}^{\pi} && (\text{ FTC I })\\
&= 0 + 0 = 0.
\end{aligned}$$

and

$$\begin{aligned}
b_k &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(kx)\,dx && (\text{ definition of } b_k\text{ })\\
&= \frac{1}{\pi}\int_{-\pi}^{0} f(x)\sin(kx)\,dx + \frac{1}{\pi}\int_{0}^{\pi} f(x)\sin(kx)\,dx && (\text{ splitting integral into two pieces })\\
&= \frac{1}{\pi}\int_{-\pi}^{0} (-\sin(kx))\,dx + \frac{1}{\pi}\int_{0}^{\pi} \sin(kx)\,dx && (\text{ definition of } f(x)\text{ })\\
&= \frac{1}{\pi}\left.\frac{\cos(kx)}{k}\right|_{-\pi}^{0} + \frac{1}{\pi}\left.\frac{-\cos(kx)}{k}\right|_{0}^{\pi} && (\text{ FTC I })\\
&= \frac{1}{\pi}\left(\frac{1-\cos(-k\pi)}{k}\right) + \frac{1}{\pi}\left(\frac{-\cos(k\pi)+1}{k}\right).
\end{aligned}$$

Because $\cos(-k\pi) = \cos(k\pi)$, we have

$$b_k = \frac{1}{k\pi}\left((1-\cos(k\pi)) + (1-\cos(k\pi))\right) = \frac{2(1-\cos(k\pi))}{k\pi}.$$

When k is even, $1-\cos(k\pi) = 1-1 = 0$. When k is odd, $1-\cos(k\pi) = 1-(-1) = 2$. Thus

$$b_k = \begin{cases} 0 & \text{when } k \text{ is even} \\ \dfrac{4}{k\pi} & \text{when } k \text{ is odd.} \end{cases}$$

The Fourier series (12.7.1) has only terms involving $\sin(kx)$ with k odd. The first three nonzero terms in the Fourier series representation of $f(x)$ are

$$\frac{4}{\pi}\sin(x) + \frac{4}{3\pi}\sin(3x) + \frac{4}{5\pi}\sin(5x) + \dots.$$

When $x = \pi/2$, $f(x) = 1$ and we have

$$\begin{aligned} 1 &= \frac{4}{\pi}\sin\left(\frac{\pi}{2}\right) + \frac{4}{3\pi}\sin\left(\frac{3\pi}{2}\right) + \frac{4}{5\pi}\sin\left(\frac{5\pi}{2}\right) + \dots \\ &= \frac{4}{\pi} - \frac{4}{3\pi} + \frac{4}{5\pi} - \dots \end{aligned}$$

Thus

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \dots,$$

which was obtained previously in Exercise 22 in Section 12.4 using the Maclaurin series for $\arctan(x)$. ♦

Observation 12.7.1: *Fourier Series for an Odd Function*

When $f(x)$ is an odd function, all $a_k = 0$; only sines appear in the Fourier series for an odd function.

That $f(x)$ in Example 2 is defined on a full period is convenient. In many applications the function is given only on one half of the period. For example, $f(x) = x$ for $0 \le x < \pi$ (see Figure 12.7.2(a)). Because $f(x)$ is not periodic, we replace $f(x)$ with a function $g(x)$ that has period 2π and coincides with $f(x)$ on its domain, that is, on $[0,\pi)$. Two extensions of $f(x)$ are shown in Figure 12.7.2(b) and (c). Both have period 2π; one extension is odd, the other is even. The Fourier series for the odd extension will involve only sines, which are odd functions. The Fourier series for the even extension will involve cosines, which are even functions.

RECALL: If $T(x) = T(-x)$, then $T(x)$ is an even function and so $\int_{-\pi}^{\pi} T(x)\,dx = 2\int_0^{\pi} T(x)\,dx$.

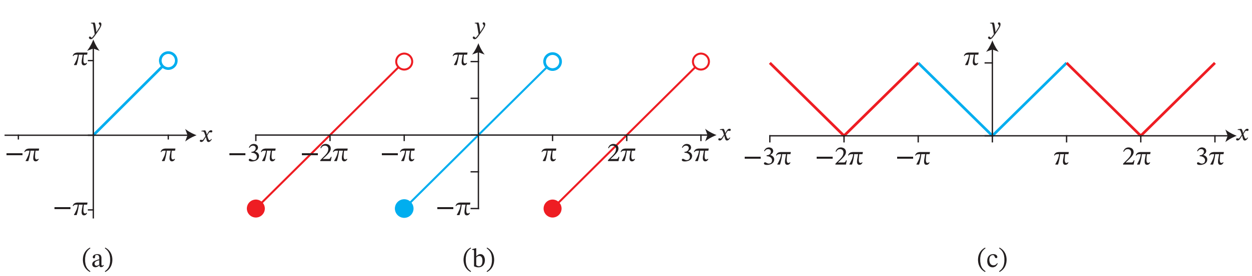


Figure 12.7.2

EXAMPLE 3. Find the Fourier series for the triangular wave with period 2π shown in Figure 12.7.2(c).

SOLUTION Let $T(x)$ denote the triangular wave whose graph appears in Figure 12.7.2(c). To compute its Fourier series we need to know its definition on an interval with length 2π. From Figure 12.7.2(c) we have $T(x) = |x|$ for x in $(-\pi, \pi]$. Or, as a piecewise-defined function,

$$T(x) = \begin{cases} x & \text{for } 0 \le x \le \pi \\ -x & \text{for } -\pi < x < 0 \end{cases}.$$

Because $T(x)$ is an even function, $T(x)\sin(kx)$ will be an odd function for all $k = 1, 2, 3, \ldots$. Therefore, all $b_k = 0$. Now, turning to the coefficient of the constant term ($k = 0$):

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} T(x)\,dx = \frac{2}{\pi}\int_0^{\pi} x\,dx = \frac{1}{\pi}\,x^2\Big|_0^{\pi} = \pi.$$

Observation 12.7.2: ***Geometric Interpretation of the Constant Term***

The constant term in the Fourier series for $f(x)$ is $\dfrac{a_0}{2} = \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} f(x)dx$, the average value of $f(x)$ over $[-\pi, \pi]$.

And, finally, the coefficients of the cosine terms are

$$\begin{aligned}
a_k &= \frac{1}{\pi}\int_{-\pi}^{\pi} T(x)\cos(kx)\,dx && (\text{ definition of } a_k\text{ }) \\
&= \frac{2}{\pi}\int_0^{\pi} x\cos(kx)\,dx && (\text{ because integrand is even}) \\
&= \frac{2}{\pi}\left(\frac{x}{k}\sin(kx)\Big|_0^{\pi} - \frac{1}{k}\int_0^{\pi}\sin(kx)\,dx\right) && (\text{ integrate by parts }) \\
&= \frac{2}{\pi}\left(0 + \frac{1}{k^2}\cos(kx)\Big|_0^{\pi}\right) && (\text{ FTC I }) \\
&= \frac{2}{k^2\pi}(\cos(k\pi) - 1) && (\cos(k\pi) = (-1)^k\text{ }) \\
&= \frac{2((-1)^k - 1)}{k^2\pi}.
\end{aligned}$$

When k is an even integer, $a_k = 2((-1)^k - 1)/(k^2\pi) = 0$. When k is an odd integer, $a_k = 2((-1)^k - 1)/(k\pi)^2 = -4/(k^2\pi)$.

The first three nonzero terms in the Fourier series for the triangular wave is

$$T(x) = \frac{\pi}{2} - \frac{4}{\pi}\left(\cos(x) + \frac{1}{9}\cos(3x) + \frac{1}{25}\cos(5x) + \ldots\right).$$

◆

Observation 12.7.3: ***Fourier Series for an Even Function***

When $f(x)$ is an even function, all $b_k = 0$; only cosines appear in the Fourier series for an even function.

Figure 12.7.3 shows the partial Fourier sums for the triangular wave with 1, 2, and 5 terms.

In advanced courses it is proved that the partial sums converge to the function. As is easy to check, replacing x by 0 in $T(x)$ shows that the sum of the reciprocals of the squares of all the positive odd integers is $\pi^2/8$.

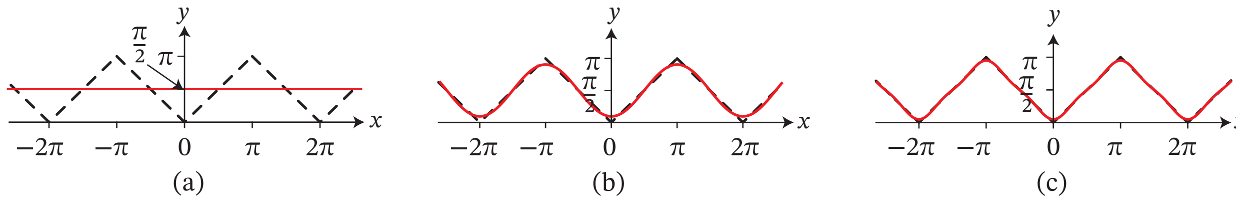


Figure 12.7.3

The Origins of the Formulas for a_k and b_k

We will derive the formulas for the Fourier coefficients when the period is 2π. Exercises 13 and 14 outline the argument for the general case when the period is $2L$.

The keys are the following definite integrals:

$$\int_{-\pi}^{\pi} \sin(kx)\sin(mx)\,dx = \begin{cases} \pi & \text{if } m = k,\ k = 1, 2, \dots \\ 0 & \text{if } m \neq k,\ k = 1, 2, \dots \end{cases}$$

$$\int_{-\pi}^{\pi} \cos(kx)\cos(mx)\,dx = \begin{cases} \pi & \text{if } m = k,\ k = 1, 2, \dots \\ 0 & \text{if } m \neq k,\ k = 1, 2, \dots \end{cases}$$

$$\int_{-\pi}^{\pi} \sin(kx)\cos(mx)\,dx = 0 \quad \text{for } m = 1, 2, \dots \text{ and } k = 1, 2, \dots.$$

The third one requires almost no work to evaluate. The integrand is an odd function because the product of an odd function (sin) and an even function (cos) will always be an odd function. The other two depend on trigonometric identities, and were developed in Exercises 17 to 19 in Section 8.5.

ASSUMPTION: It is permissible to integrate series term-by-term. That is, $\int_{-\pi}^{\pi} \left(\sum_{k=1}^{\infty} f_k(x) \right) dx = \sum_{k=1}^{\infty} \left(\int_{-\pi}^{\pi} f_k(x)\,dx \right)$.

The formula for a_m, $m = 1, 2, \dots$, is found by multiplying $f(x)$ by $\cos(mx)$ and integrating term by term over one period of length 2π:

$$\begin{aligned} \int_{-\pi}^{\pi} f(x)\cos(mx)\,dx &= \int_{-\pi}^{\pi} \left(\frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k\cos(kx) + b_k\sin(kx)\right) \right) \cos(mx)\,dx \\ &= \frac{a_0}{2}\int_{-\pi}^{\pi} \cos(mx)\,dx + \sum_{k=1}^{\infty} \left(a_k \int_{-\pi}^{\pi} \cos(kx)\cos(mx)\,dx + b_k \int_{-\pi}^{\pi} \sin(kx)\cos(mx)\,dx \right). \end{aligned}$$

Each integral in this last expression is zero except the coefficient of a_m. This gives

$$\int_{-\pi}^{\pi} f(x)\cos(mx)\,dx = a_m \int_{-\pi}^{\pi} (\cos(kx))^2\,dx = a_m\pi.$$

Solving for a_m, we find that

$$a_m = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(mx)\,dx.$$

The derivation of the formulas for a_0 and for b_k are similar. (See Exercises 13 and 14.)

Remarks on the Underlying Theory

As a Taylor series for a function may not represent the function, the Fourier series for a function may not represent it, even if the function is continuous. However, there are theorems that assure us that for many functions met in applications the series does converge to the function. We will state one such theorem.

Definition: *Jump Discontinuity*

A **jump discontinuity** occurs in the graph of $y = f(x)$ when the left-hand limit $\lim\limits_{x\to a^-} f(x)$ and right-hand limit $\lim\limits_{x\to a^-} f(x)$ both exist but are not equal.

Theorem 12.7.4: Convergence Theory for Fourier Series

Let $f(x)$ have period $2L$. Assume that in the interval $(-L, L]$ $f(x)$ is differentiable except at a finite number of points, where there are jump discontinuities, and at $-L$ the right-hand limit of $f(x)$ exists and at L the left-hand limit of $f(x)$ exists. Then

(i) if the function is continuous at a, its Fourier series converges to $f(a)$.

(ii) if $f(x)$ has a jump discontinuity at a, then the series converges to the average of the left- and right-hand limits at a.

(iii) at the endpoints, L and $-L$, the Fourier series converges to the average of $\lim\limits_{x\to -L^+} f(x)$ and $\lim\limits_{x\to L^-} f(x)$.

NOTE: Theorem 12.7.4 does not mention second-order or higher-order derivatives. In fact, the first derivative might not exist at a finite number of points in the interval.

Summary

While Taylor series are useful for dealing with a function that is smooth (having derivatives of all orders), Fourier series can represent a function that is not even continuous. The Taylor coefficients are expressed with derivatives and Fourier coefficients are expressed with integrals.

Even nonperiodic functions can be represented by Fourier series. For instance, to deal with x^2 on, say, $[0, 100)$ extend its domain to the whole x-axis by defining a function of period 100 that agrees with x^2 on $[0, 100)$.

EXERCISES for Section 12.7

The following list of integration formulas will be useful when evaluating some of the integrals in the exercises.

$$\int x\sin(ax)\,dx = \frac{1}{a^2}\sin(ax) - \frac{x}{a}\cos(ax)$$
$$\int x\cos(ax)\,dx = \frac{1}{a^2}\cos(ax) + \frac{x}{a}\sin(ax)$$
$$\int x^2\sin(ax)\,dx = \frac{2}{a^3}\cos(ax) + \frac{2x}{a^2}\sin(ax) - \frac{x^2}{a}\cos(ax)$$
$$\int x^2\cos(ax)\,dx = \frac{-2}{a^3}\sin(ax) + \frac{2x}{a^2}\cos(ax) + \frac{x^2}{a}\sin(ax)$$
$$\int \sin(x)\sin(ax)\,dx = \frac{1}{2(a-1)}\sin((a-1)x) - \frac{1}{2(a+1)}\sin((a+1)x)$$
$$\int \sin(x)\cos(ax)\,dx = \frac{1}{2(a-1)}\cos((a-1)x) - \frac{1}{2(a+1)}\cos((a+1)x)$$
$$\int \cos(x)\sin(ax)\,dx = \frac{-1}{2(a-1)}\cos((a-1)x) - \frac{1}{2(a+1)}\cos((a+1)x)$$
$$\int \cos(x)\cos(ax)\,dx = \frac{1}{2(a-1)}\sin((a-1)x) + \frac{1}{2(a+1)}\sin((a+1)x)$$

In Exercises 1 and 2 give the period of the function.

1. (a) $\tan(x)$, (b) $\sin\left(\frac{\pi x}{3}\right)$, (c) $\cos(3x)$, and (d) $\sin(2\pi x)$.

2. (a) $\sin\left(\frac{x}{5}\right)$, (b) $\cos\left(\frac{2\pi x}{5}\right)$, (c) $\frac{2}{\cos^2(x)}$, and (d) $\sin\left(\frac{x}{3}\right)$.

---

3. If $f(x)$ is periodic with period 2π and is odd, its Fourier expansion contains no cosines.

(a) Show that in this case $a_n = 0$ and $b_n = \frac{2}{\pi}\int\limits_0^\pi f(x)\sin(2nx)\,dx$.

(b) Show that for x in $[0,\pi]$ the Fourier series for $f(x) = \sin(x)$ is simply $\sin(x)$.

4. If $f(x)$ is periodic with period 2π and is even, its Fourier expansion contains no sines.

(a) Show that in this case $a_n = \frac{2}{\pi}\int\limits_0^\pi f(x)\cos(2nx)\,dx$ and $b_n = 0$.

(b) Let $f(x) = |\sin(x)|$ for x in $(-\pi,\pi]$. Show that for x in $[0,\pi]$,

$$\sin(x) = \frac{2}{\pi} - \frac{4}{\pi}\left(\frac{\cos(2x)}{2^2-1} + \frac{\cos(4x)}{4^2-1} + \frac{\cos(6x)}{6^2-1} + \cdots\right). \quad (12.7.4)$$

(c) Show that the series in (b) is absolutely convergent.
(d) Substituting 0 for x, what equation results?
(e) Without referring to Fourier series, check that the equation in (d) holds.

5. (a) Replacing x in (12.7.4) with $\frac{\pi}{2}$ yields the equation

$$1 = \frac{2}{\pi} - \frac{4}{\pi}\left(\frac{-1}{2^2-1} + \frac{1}{4^2-1} - \frac{1}{6^2-1} + \frac{1}{8^2-1} - \frac{1}{10^2-1} + \cdots\right).$$

(b) Show that the series in parentheses in (a) converges absolutely.

(c) Using the identity $\frac{2}{a^2-1} = \frac{1}{a-1} - \frac{1}{a+1}$, show that

$$1 = \frac{2}{\pi} - \frac{4}{\pi}\left(\frac{-1}{2} + \frac{1}{3} - \frac{1}{5} + \frac{1}{7} - \frac{1}{9} + \frac{1}{11} - \frac{1}{13} + \cdots\right).$$

(d) Does the series in parentheses in (c) converge? converge absolutely?

(e) Using the equation in (c), show that $1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = \frac{\pi}{4}$.

(f) Obtain the result in (e) using the Maclaurin series associated with $\arctan(x)$.

6. Let $f(x) = 0$ for x in $(-\pi, 0)$ and $f(x) = \pi$ for x in $[0, \pi]$, and have period 2π.

(a) Is f an even function, odd function, or neither?

(b) Graph $f(x)$ for x in $[-2\pi, 4\pi]$.

(c) Show that $f(x) = \frac{\pi}{2} + 2\left(\frac{\sin(x)}{1} + \frac{\sin(3x)}{3} + \frac{\sin(5x)}{5} + \cdots\right)$.

(d) At $x = 0$ the series has the value $\pi/2$, but $f(0) = \pi$. Explain why this does not contradict Theorem 12.7.4.

Exercises 7 and 8 describe two ways to obtain Fourier series for a function defined on $[0, \pi)$. One uses only cosines, the other, only sines.

7. Let $f(x) = |x|$ for x in $(-\pi, \pi]$ and have period 2π.

(a) Graph $f(x)$ for x in $[-2\pi, 4\pi]$.

(b) Show that $f(x) = \frac{\pi}{2} - \frac{4}{\pi}\left(\frac{\cos(x)}{1^2} + \frac{\cos(3x)}{3^2} + \frac{\cos(5x)}{5^2} + \cdots\right)$.

(c) Deduce that $\frac{\pi^2}{8} = \frac{1}{1^2} + \frac{1}{3^2} + \frac{1}{5^2} + \cdots$.

8. Let $f(x) = x$ for x in $(-\pi, \pi]$ and have period 2π. NOTE: This function is known as the **sawtooth function**.

(a) Graph $f(x)$ for x in $[-2\pi, 4\pi]$.

(b) Show that $f(x) = 2\left(\frac{1}{1}\sin(x) - \frac{1}{2}\sin(2x) + \frac{1}{3}\sin(3x) + \cdots\right)$.

(c) Deduce that $\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} + \cdots$.

---

9. Let $f(x) = x^2$ for x in $(-\pi, \pi]$ and have period 2π.

(a) Find $f(\pi)$, $f(2\pi)$, $f(3\pi)$, $f(-\pi)$, $f(-2\pi)$, and $f(-3\pi)$.

(b) Graph $f(x)$ for x in $[-4\pi, 4\pi]$.

(c) Why will the Fourier series for $f(x)$ have no sine terms?

(d) Show that $f(x) = \frac{\pi^3}{4} + 4\sum_{k=1}^{\infty} \frac{(-1)^k}{k^2}\cos(kx)$.

(e) Deduce that $\frac{\pi^2}{12} = \frac{1}{1^2} - \frac{1}{2^2} + \frac{1}{3^2} - \frac{1}{4^2} + \cdots$.

(f) Find the Fourier series for $f(x)$.

In Exercises 10 and 11 assume the function has period 2π.

(a) Sketch the graph of $f(x)$ over an interval at least as long as two periods of the function,
(b) Compute the Fourier series for the indicated function, and
(c) List all points where the function and the sum of its Fourier series do not agree.

10. $f(x) = \begin{cases} 0 & \text{for } -\pi \le x < 0 \\ \sin(x) & \text{for } 0 \le x < \pi \end{cases}$

11. $f(x) = \begin{cases} 1 & \text{for } -\pi \le x < 0 \\ \cos(x) & \text{for } 0 \le x < \pi \end{cases}$

---

12. In Section 11.5, Example 3, it is claimed that the series $\frac{\cos(x)}{1^2} + \frac{\cos(2x)}{2^2} + \frac{\cos(3x)}{3^2} + \cdots + \frac{\cos(kx)}{k^2} + \cdots$ converges to $\frac{1}{12}(3x^2 - 6\pi x + 2\pi^2)$ for $0 \le x \le 2\pi$. Use Fourier series to verify this.

Exercises 13 and 14 complete the derivation of the Fourier series for a function with period 2π. That is, of (12.7.1) with coefficients given by (12.7.2) and (12.7.3).

13. Derive (12.7.2).

14. Derive (12.7.3).

---

15. Let $f(x) = -x^2$ for x in $[-\pi, 0)$, x^2 for x in $[0, \pi)$, and have period 2π.

(a) Find $f(\pi)$, $f(2\pi)$, $f(-\pi)$, and $f(-2\pi)$.
(b) Graph $f(x)$ for x in $[-4\pi, 4\pi]$.
(c) Show that f is almost an odd function. For which x is $f(x) \ne -f(x)$?
(d) Show that the Fourier series for $f(x)$ is

$$2\frac{\pi^2 - 4}{\pi}\sin(x) - \pi\sin(2x) + 2\frac{9\pi^2 - 4}{27\pi}\sin(3x) - \frac{\pi}{2}\sin(4x)$$
$$+2\frac{25\pi^2 - 4}{125\pi}\sin(5x) - \frac{\pi}{3}\sin(6x) + 2\frac{49\pi^2 - 4}{343\pi}\sin(7x) - \frac{\pi}{4}\sin(8x) + \cdots.$$

(e) Why are there no cosine terms?

16. Find the Fourier series for the function with period 2π whose graph is shown in Figure 12.7.4.

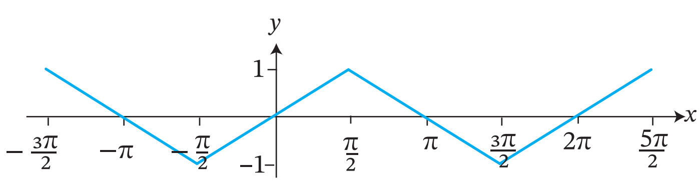


Figure 12.7.4

Exercises 17 to 19 are related. Exercise 17 shows how to find the Fourier series for a function of period $2L$. Exercises 18 and 19 apply this result to functions with periods 4 and 2, respectively.

17. (a) Let $g(x)$ be a function with period $2L$ Define $f(z) = g\left(\frac{Lz}{\pi}\right)$. Show that $f(z)$ has period 2π.

(b) Assume $f(z) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(kz) + \sum_{k=1}^{\infty} b_k \sin(kx)$ where a_k is given by (12.7.2) and b_k is given by (12.7.3).

Use the substitution $z = \dfrac{\pi x}{L}$ to show that the Fourier series for $g(x)$ is

$$g(x) = \frac{c_0}{2} + \sum_{k=1}^{\infty} c_k \cos\left(\frac{k\pi}{L} z\right) + \sum_{k=1}^{\infty} d_k \sin\left(\frac{k\pi}{L} x\right) \tag{12.7.5}$$

where

$$c_k = \frac{1}{L}\int_{-L}^{L} g(x)\, \cos\left(\frac{k\pi}{L} z\right)\, dx \qquad k = 0, 1, 2, \ldots \tag{12.7.6}$$

$$d_k = \frac{1}{L}\int_{-L}^{L} g(x)\, \sin\left(\frac{k\pi}{L} z\right)\, dx \qquad k = 1, 2, \ldots . \tag{12.7.7}$$

18. Let $f(x) = 0$ for x in $(-2, 0)$ and $f(x) = 1$ for x in $[0, 2]$, and has period 4. Use (12.7.5), (12.7.6), and (12.7.7) to show that

$$f(x) = \frac{1}{2} + \frac{2}{\pi}\left(\sin\left(\frac{\pi x}{2}\right) + \frac{1}{3}\sin\left(\frac{3\pi x}{2}\right) + \frac{1}{5}\sin\left(\frac{5\pi x}{2}\right) + \cdots\right).$$

19. Let $f(x) = |x|$ for x in $(-1, 1]$ and have period 2. Use (12.7.5), (12.7.6), and (12.7.7) to show that

$$|x| = \frac{1}{2} - \frac{4}{\pi^2}\sum_{k=1}^{\infty} \frac{\cos((2k-1)\pi x)}{(2k-1)^2}.$$

Complex numbers helped show a close tie between exponential and trigonometric functions. They also reveal a relation between power series and Fourier series. Exercise 20 helps make this connection.

20. A Taylor series $\sum_{k=0}^{\infty} a_k z^k$ does not look like a Fourier series. However, when a_k is written as $b_k + i c_k$ and z is expressed as $r(\cos(\theta) + i\sin(\theta))$, where r is constant, the connection becomes clear. To check that this is so, write the series in the form $A + Bi$ where A and B are real. What two Fourier series appear as the real and imaginary parts, A and B?

12.S Chapter Summary

When Taylor polynomials were first encountered in Section 5.5, they suggested the power series for a function that has derivatives of all orders at a, namely

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

This power series always converges when x is a and may converge for other values of x, but not necessarily to $f(x)$. The power series for e^x, $\sin(x)$, and $\cos(x)$ converge to the function for all values of x.

The error in using a partial sum up through the term involving $(x-a)^n$ to estimate $f(x)$ is given by Lagrange's formula,

$$f(x) - P_n(x; a) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \qquad \text{for some } c \text{ between } x \text{ and } a$$

where

$$P_n(x; a) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k$$

is the degree n Taylor polynomial for $f(x)$ based at $x = a$.

For some functions, such as $\tan(x)$, it is not easy to find a general formula for the k^{th} derivative. We can obtain a few terms of its Maclaurin series by dividing the series for $\sin(x)$ by the series for $\cos(x)$.

Table 12.S.1 contains some commonly encountered functions with nice closed-form Maclaurin series and their corresponding interval of convergence. The series for $\ln(1+x)$ and $\arctan(x)$ can be obtained by integrating the geometric series representations of their derivatives: $1/(1+x)$ and $1/(1+x^2)$, respectively. The series for $1/(1-x)^2$ is obtained by differentiating the geometric series representation of $1/(1-x)$.

Function	Maclaurin Series	Interval of Convergence	How Found?
e^x	$\sum_{k=0}^{\infty} \frac{x^k}{k!}$	$-\infty < x < \infty$	Taylor's theorem
$\sin(x)$	$\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}$	$-\infty < x < \infty$	Taylor's theorem
$\cos(x)$	$\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!}$	$-\infty < x < \infty$	Taylor's theorem
$\frac{1}{1-x}$	$\sum_{k=0}^{\infty} x^k$	$-1 < x < 1$	Geometric series
$\frac{1}{(1-x)^2}$	$\sum_{k=0}^{\infty} kx^{k-1}$	$-1 < x < 1$	Differentiate $(1-x)^{-1}$
$(1+x)^r$	$1 + rx + \frac{r(r-1)}{2!}x^2 + \frac{r(r-1)(r-2)}{3!}x^3 + \cdots$	$-1 < x < 1$	Taylor's theorem (binomial series)
$\ln(1+x)$	$\sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^k}{k}$	$-1 < x \le 1$	Integrate $1/(1+x)$
$\arctan(x)$	$\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1}$	$-1 < x \le 1$	Integrate $1/(1+x^2)$
$\arcsin(x)$	$x + \frac{1}{2}\frac{x^3}{3} + \frac{1\cdot 3}{2\cdot 4}\frac{x^5}{5} + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\frac{x^7}{7} + \cdots$	$-1 \le x \le 1$	Integrate $(1-x^2)^{-1/2}$

Table 12.S.1

Each power series $\sum_{k=0}^{\infty} a_k (x-a)^k$ has a radius of convergence, R. For $|x-a| < R$, the series converges absolutely and for $|x-a| > R$ the series does not converge. If it converges for all x, then $R = \infty$. For $|x-a| < R$, we may differentiate and integrate the series term by term. While the radius of convergence can be determined from the radius of convergence of the underlying function, convergence (or divergence) at endpoints must always be checked separately.

Estimating an integrand $f(x)$ by a partial sum of a power series lets us estimate $\int_a^b f(x)\,dx$. Power series are of use in finding indeterminate limits of the type zero-over-zero.

Maclaurin series, combined with complex numbers, reveal a fundamental relation between exponential and trigonometric functions:

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

Many other important truths, not covered in this chapter, are similarly revealed with the aid of complex numbers. For example, if we allow complex coefficients, every polynomial can be written as the product of first-degree polynomials, thus simplifying the partial fractions of Section 8.4.

The final section introduced Fourier series. A Fourier series replaces the Taylor series' infinite sum of powers of $x-a$ with coefficients given by derivatives with an infinite sum of trigonometric functions ($\sin(kx)$ and $\cos(kx)$) with coefficients given by integrals. Fourier series apply to a larger class of functions than do Taylor series. However, the functions must be periodic. For a nonperiodic function, we restrict the domain to an interval $(-L,L)$ and extend the function to have period $2L$

EXERCISES for Section 12.S

1. What are the polar coordinates of e^{x+yi}?

2. The following problem arises in statistics. Let a_1, a_2, a_3, ... approach a. Show that $\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n$ is e^a.

CONTRIBUTED BY: Frank Saminiego

3. Let $M(x)$ be the Maclaurin series for $f(x)$ for x in $(-a,a)$. Show that $M\left(x^2\right)$ is the Maclaurin series for $g(x) = f\left(x^2\right)$ for x in $(-\sqrt{a},\sqrt{a})$.

4. The integral $\displaystyle\int_0^{2\pi}\frac{1-\cos(x)}{x}\,dx$ occurs in the theory of antennas.

(a) Show that it is not an improper integral.
(b) Show that there is a continuous function whose domain is $[0,2\pi]$ that coincides with the integrand when x is not 0.
(c) The integrand does not have an elementary antiderivative. How many terms are needed in the Maclaurin series for the integrand to obtain an approximation to the integral that is accurate to 3 decimal places?

Exercises 5 to 7 use complex numbers to find the average value of the logarithm of a certain function. Exercise 5 is related to Exercise 85 in Section 8.S.

5. Let a point P be a distance $a\neq 1$ from the center of a unit circle. Show that the average value of the natural logarithm of the distance from P to points on the unit circle is $\displaystyle\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2}\ln\left(1+a^2-2a\cos(\theta)\right)\,d\theta$.

6. Let $z_0, z_1, \dots, z_{n-1}$ be the n n^{th} roots of 1. It can be shown that $(z-z_0)(z-z_1)(z-z_2)\cdots(z-z_{n-1}) = z^n-1$. Check that the equation holds when (a) $n=2$, (b) $n=3$, and (c) $n=4$.

7. Let $z_0, z_1, \dots, z_{n-1}$ be the n n^{th} roots of 1.

(a) Why is $\displaystyle\frac{1}{n}\sum_{i=0}^{n-1}\ln|a-z_i|$ an estimate of the average distance from P to the unit circle?
(b) Show that the average in (a) equals $\displaystyle\frac{1}{n}\ln|a^n-1|$.
(c) If $0<a<1$, show $\displaystyle\lim_{n\to\infty}\frac{1}{n}\ln|a^n-1| = 0$.
(d) If $a>1$, show $\displaystyle\lim_{n\to\infty}\frac{1}{n}\ln|a^n-1| = \ln(a)$.

NOTE: The case when $a=1$ is not covered by parts (c) and (d).

(e) If $a=1$, choose Q to be a point on the unit circle whose polar angle is not a rational multiple of π. (So no z_i coincides with Q.) Then argue as in parts (c) or (d).
(f) Use the results in (c), (d), and (e) to evaluate the integral $\displaystyle\int_0^{2\pi}\ln(1+a^2-2a\cos(\theta)\,d\theta$ for any $0<a<\infty$.

---

8. Find (a) $\lim_{x\to\infty} \frac{xe^x}{e^{x^2}}$ and (b) $\lim_{x\to 0} \frac{x(e^{\sqrt{x}}-1)}{e^{x^2}-1}$.

9. Does $\sum_{k=1}^{\infty} (1-\cos(1/k))$ converge or diverge? Explain.

10. Assume that $f(x)$ has a continuous fourth derivative. Let M_4 be the maximum of $\left|f^{(4)}(x)\right|$ for x in $[-1,1]$. Show

$$\left|\int_{-1}^{1} f(x)\,dx - f\left(\frac{1}{\sqrt{3}}\right) - f\left(\frac{-1}{\sqrt{3}}\right)\right| \le \frac{7M_4}{270}.$$

(Use the representation $f(x) = f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 + \frac{1}{6}f^{(3)}(0)x^3 + \frac{1}{24}f^{(4)}(c)x^4$, where c depends on x.

See also Exercise 40 in Section 6.5, where Gaussian quadrature was introduced.

11. Justify this statement, found in a biological monograph:

> Expanding the equation $a\cdot\ln(x+p) + b\cdot\ln(y+q) = M$, we obtain
>
> $$a\left(\ln(p) + \frac{x}{p} - \frac{x^2}{2p^2} + \frac{x^3}{3p^3} - \cdots\right) + b\left(\ln(q) + \frac{y}{q} - \frac{y^2}{2q^2} + \frac{y^3}{3q^3} - \cdots\right) = M.$$

12. Estimate $\int_1^3 e^{-x^2}\,dx$ using a Taylor series at $x=2$ for e^{-x^2}.

13. Explain why both $\cos(x)$ and $\sin(x)$ can be expressed in terms of the exponential function e^z.

14. State some of the advantages of complex numbers over real numbers.

15. Why is the radius of convergence called the radius of convergence rather than the interval of convergence?

16. Starting with $1 + x + x^2 + \cdots + x^n + \cdots = \frac{1}{1-x}$ obtain the Maclaurin series for

(a) $\frac{1}{(1-x)^2}$ (b) $\frac{1}{1+x}$ (c) $\frac{1}{1+x^2}$ (d) $\ln(1+x)$ (e) $\arctan(x)$

17. Find the radius of convergence for the series in Exercise 16

18. Show that the series in Exercise 16 converge to the appropriate function.

19. The function $f(z) = \frac{1}{z}$ maps part of the hyperbola $xy = 1$ in the first quadrant into a curve C.

(a) Find and sketch at least four points on C. Then sketch C.
(b) What is the polar equation for $xy = 1$?
(c) What is the polar equation for C?
(d) Check that the image of the point (x, y) is $\left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right)$.

20. The function $f(z) = \frac{1}{z}$ maps the parabola $y = x^2$ into a curve C.

(a) Find and sketch at least four points on C. Then sketch C.
(b) What is the polar equation for $y = x^2$?
(c) What is the polar equation for C?
(d) Find the rectangular equation for C.

21. (a) Graph the circle $r = \sqrt{2}\cos(\theta)$. (b) Show that $f(z) = z^2$ maps the circle into the cardioid $r = 1 + \cos(\theta)$.

22. Suppose f is a function with the property that $f^{(n)}(x)$ is small in the sense that $|f^{(n)}(x)| \le |(x+100)^n|$ for all x. Show that the Maclaurin series represents $f(x)$ for all x.

Exercises 23 and 24 treat the complex logarithms of a complex number. They show that $z = \ln(w)$ is not single-valued.

23. Let w be a nonzero complex number. Show that there are an infinite number of complex numbers z such that $e^z = w$.

24. When $e^z = w$, we write $z = \ln(w)$ although $\ln(w)$ is not uniquely defined. If b is a nonzero complex number and q is a complex number, define b^q to be $e^{q\ln(b)}$. Since $\ln(b)$ is not unique, b^q is usually not unique. List all possible values of (a) $(-1)^i$, (b) $10^{1/2}$, and (c) 10^3. See also Exercise 23.

---

25. It was shown in Section 3.4 that $\lim_{x\to 0}(1+x)^{1/x} = e$. Let $E(x) = e - (1+x)^{1/x}$.

(a) Show that $\dfrac{E(x)}{(e/2)x}$ approaches 1 as $x \to 0$. (b) Is it more efficient to estimate e with $\lim_{x\to 0}(1+x)^{1/x}$ or $\sum_{k=0}^{\infty}\dfrac{1}{k!}$?

Before working the Exercises 26 and 27, review CIE 2 ("How Banks Multiply Money") at the end of Chapter 1.

26. In CIE 2, geometric series were used to determine that the multiplier, M, equals the reciprocal of the reserve fraction, r.

(a) Some economics texts offer a different approach to this result Their explanation begins with Sam's deposit of A. This generates an endless sequence of new deposits A_1, A_2, A_4, ... (Note that we use A_n both to name the deposit and to denote the amount of the deposit.)

The top row (all blue) represents Sam's deposit, $A = A_1$. The second row represents the amount left after the reserve $R_1 = rA_1$ (in dark turquoise) is deducted: $A_1 - R_1 = A_2$ (blue). The reserve for A_2 is $R_2 = rA_2$ (light turquoise segment in row three). QUESTION: Is R_2 less than R_1? (Why?)

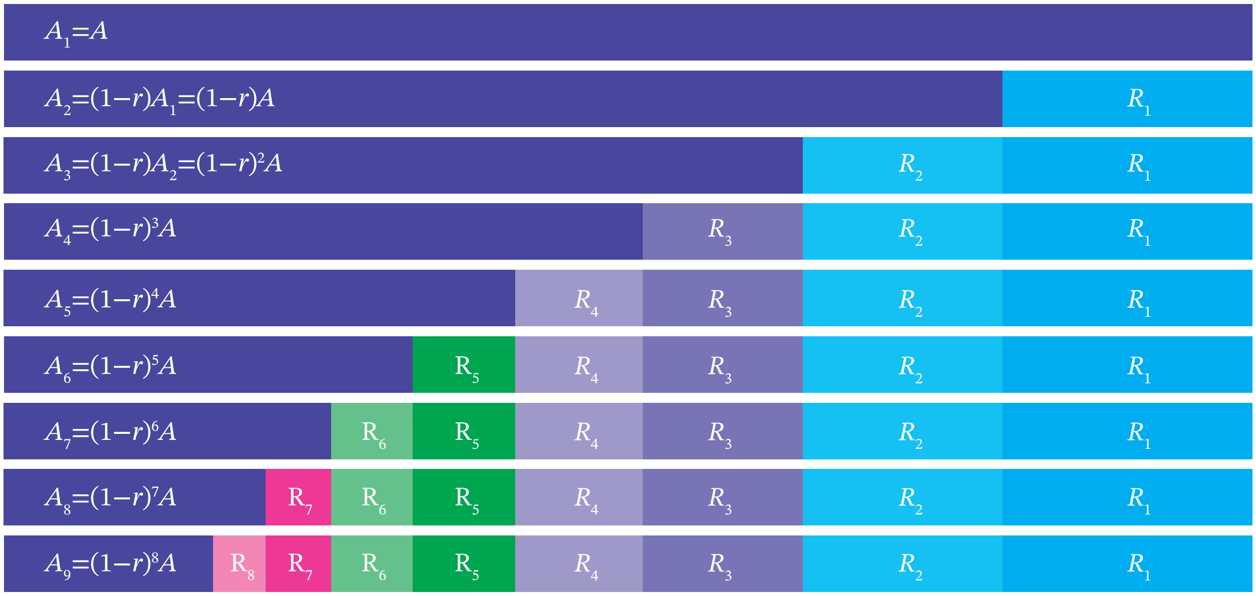


Figure 12.S.1

(b) Deposits $A_3, A_4, \ldots, A_8$ are created the same way, with reserves $R_3, R_4, \ldots, R_8$. The next step obtains A_9 from A_8, by deducting R_8. Note that $A_1, A_2, \ldots, A_9$ can be read directly from Figure 12.S.1. For instance, A_4 is obtained by deducting R_1, R_2, and R_3.

The reserves do not overlap. However, the A_n do overlap, though the deposits they represent usually do not.

Now, it is assumed that as the process goes on the deposits become arbitrarily small. That implies the final row is eventually filled up by all of the reserves $R_1, R_2, R_3, \ldots$. Thus, all of the reserves add up to Sam's deposit, A:

$$R_1 + R_2 + R_3 + \cdots = r(A_1 + A_2 + A_3 + \cdots) = A,$$

from which it follows that $M = \frac{1}{r}$.

QUESTION: This conclusion is correct, but one detail was overlooked. Fill in the missing detail needed to complete the derivation of the multiplier.

27. (a) Show that $A_{i+1} = (1-r)A_i$ for each $i = 1, 2, \ldots$.

(b) Show that $\lim\limits_{i\to\infty} A_i = 0$.

(c) Use the relation $A_i - A_{i+1} = R_i$ to show that $A_1 - A_{n+1} = \sum\limits_{i=1}^{n} R_i$.

(d) Show that $\sum\limits_{i=1}^{\infty} R_i = A_1$.

(e) Show that $\sum\limits_{i=1}^{\infty} rA_i = A_1$.

(f) Recall that the multiplier is $M = \frac{\sum i = 1^{\infty} A_i}{A_1}$. Show that the M equals $\frac{1}{r}$.

Calculus is Everywhere # 14

Sparse Traffic

Customers arriving at a checkout counter, cars traveling on a one-way road, raindrops falling on a street, and cosmic rays entering the atmosphere all illustrate the same mathematical idea that arises in sparse traffic involving independent events. We will develop the mathematics, which is the basis of the study of waiting time, whether of customers at a checkout counter or telephone calls at a switchboard.

Some Probability Theory

The probability that an event occurs is measured by a number p, which can be anywhere from 0 to 1. If $p = 1$ the event will certainly occur with negligible exceptions and if $p = 0$ then it will not occur with negligible exceptions. The probability that a penny turns up heads is $p = 1/2$ and that a die turns up 2 is $p = 1/6$. (The phrase "certainly occurs with negligible exceptions" means, roughly, that the times the event does not occur are so rare that we may disregard them, and similarly for "not occur with negligible exceptions" in the case $p = 0$.)

The probability that two events that have no effect on each other both occur is the product of their probabilities. For example, the probability of getting heads when tossing a penny and a 2 when tossing the die is $p = (1/2)(1/6) = 1/12$.

The probability that exactly one of several mutually exclusive events occurs is the sum of their probabilities. For instance, the probability of getting a 2 or a 3 with a die is $1/6 + 1/6 = 1/3$.

With that introduction, we will analyze sparse traffic on a one-way road. We will assume that the cars enter the road independently of each other, travel at the same speed, and for simplicity that each car is a point.

The Model

To construct a model we introduce the functions $P_0, P_1, P_2, \dots, P_n, \dots$ where $P_n(x)$ is the probability that an interval of length x contains exactly n cars (independently of the location of the interval). Thus $P_0(x)$ is the probability that an interval of length x is empty. We shall assume that

$$P_0(x) + P_1(x) + \cdots + P_n(x) + \cdots = 1 \qquad \text{for any } x.$$

We also shall assume that $P_0(0) = 1$, that is, the probability is 1 that a given point contains no cars.

We make two major assumptions:

(I) The probability that exactly one car is in a short section of the road is approximately proportional to the length of the section. That is, there is a number k such that

$$\lim_{\Delta x \to 0} \frac{P_1(\Delta x)}{\Delta x} = k.$$

(II) The probability that there is more than one car in a short section of the road is negligible, even when compared to the length of the section. That is,

$$\lim_{\Delta x \to 0} \frac{P_2(\Delta x) + P_3(\Delta x) + P_4(\Delta x) + \cdots}{\Delta x} = 0.$$

We shall now put assumptions ((I)) and ((II)) into more useful forms. If we let

$$\epsilon = \frac{P_1(\Delta x)}{\Delta x} - k,$$

where ϵ depends on Δx, ((I)) tells us that $\lim_{\Delta x \to 0} \epsilon = 0$. Thus, solving for $P_1(\Delta x)$, we see that assumption ((I)) can be phrased as

$$P_1(\Delta x) = k\Delta x + \epsilon \Delta x, \quad \text{where } \epsilon \to 0 \text{ as } \Delta x \to 0. \tag{C.14.1}$$

Since $P_0(\Delta x) + P_1(\Delta x) + \cdots + P_n(\Delta x) + \cdots = 1$, ((II)) may be expressed as

$$\lim_{\Delta x \to 0} \frac{1 - P_0(\Delta x) - P_1(\Delta x)}{\Delta x} = 0$$

or, using ((I)),

$$\lim_{\Delta x \to 0} \frac{1 - P_0(\Delta x)}{\Delta x} = k.$$

As we obtained (C.14.1), we may deduce that $1 - P_0(\Delta x) = k\Delta x + \delta \Delta x$, where $\delta \to 0$ as $\Delta x \to 0$. This leads us to conclude that

$$P_0(\Delta x) = 1 - k\Delta x - \delta \Delta x, \quad \text{where } \delta \to 0 \text{ as } \Delta x \to 0. \tag{C.14.2}$$

On the basis of ((I)) and ((II)), as expressed in (C.14.1) and (C.14.2), we shall obtain an explicit formula for P_n.

Let us find P_0 first. A section of length $x + \Delta x$ is vacant if its left-hand part of length x is vacant and its right-hand part of length Δx is also. Since cars move independently of each other, the probability that the whole interval of length $x + \Delta x$ is empty is the product of the probabilities that the intervals of lengths x and Δx are both empty. (See Figure C.14.1.) Thus,

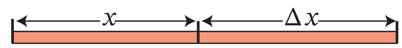


Figure C.14.1

$$P_0(x + \Delta x) = P_0(x) P_0(\Delta x). \tag{C.14.3}$$

Recalling (C.14.2), we write (C.14.3) as $P_0(x + \Delta x) = P_0(x)(1 - k\Delta x - \delta \Delta x)$ which gives

$$\frac{P_0(x + \Delta x) - P_0(x)}{\Delta x} = -(k + \delta) P_0(x). \tag{C.14.4}$$

Taking limits of both sides of (C.14.4) as $\Delta x \to 0$, we obtain

$$P_0'(x) = -k P_0(x).$$

(Recall that $\delta \to 0$ as $\Delta x \to 0$.) It follows that there is a constant A such that $P_0(x) = Ae^{-kx}$. Since $1 = P_0(0) = Ae^{-k0} = A$, we conclude that $A = 1$, hence

$$P_0(t) = e^{-kx}.$$

This is reasonable for e^{-kx} is a decreasing function of x, so that the larger an interval is, the less likely that it is empty.

To determine P_1, we examine $P_1(x + \Delta x)$ and relate it to $P_0(x)$, $P_0(\Delta x)$, $P_1(x)$, and $P_1(\Delta x)$, to find an equation involving the derivative of P_1. Again, imagine an interval of length $x + \Delta x$ cut into two intervals, the left-hand subinterval of length x and the right-hand subinterval of length Δx. Then there is precisely one car in the whole interval if either there is exactly one car in the left-hand interval and none in the right-hand subinterval or there is none in the left-hand subinterval and exactly one in the right-hand subinterval. (See Figure C.14.2.) Thus we have

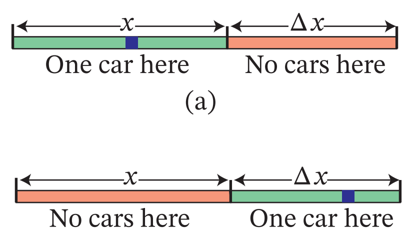


Figure C.14.2

$$P_1(x + \Delta x) = P_1(x) P_0(\Delta x) + P_0(x) P_1(\Delta x). \tag{C.14.5}$$

In view of (C.14.1) and (C.14.2), we may write (C.14.5) as

$$P_1(x + \Delta x) = P_1(x)(1 - k\Delta x - \delta \Delta x) + P_0(x)(k\Delta x + \epsilon \Delta x)$$

or

$$\frac{P_1(x + \Delta x) - P_1(x)}{\Delta x} = -(k + \delta) P_1(x) + (k + \epsilon) P_0(x).$$

Letting $\Delta x \to 0$ and remembering that $\delta \to 0$ and $\epsilon \to 0$ as $\Delta x \to 0$, we obtain $P_1'(x) = -kP_1(x) + kP_0(x)$. Recalling that $P_0(x) = e^{-kx}$, we deduce that

$$P_1'(x) = -kP_1(x) + ke^{-kx}. \tag{C.14.6}$$

This equation helps us find an explicit formula for $P_1(x)$. Since $P_0(x)$ involves e^{-kx} and so does (C.14.6), it is reasonable to guess that $P_1(x)$ does also. Therefore write $P_1(x)$ as $g(x)e^{-kx}$ and determine the form of $g(x)$. (Since we have the identity $P_1(x) = \left(P_1(x)e^{kx}\right)e^{-kx}$, we know that $g(x)$ exists.)

According to (C.14.6) we have $\left(g(x)e^{-kx}\right)' = -kg(x)e^{-kx} + ke^{-kx}$; hence

$$g'(x)e^{-kx} - g(x)(ke^{-kx}) = -kg(x)e^{-kx} + ke^{-kx}$$

from which it follows that $g'(x) = k$. Hence $g(x) = kx + c_1$, where c_1 is a constant. Thus $P_1(x) = (kx + c_1)e^{-kx}$. Since $P_1(0) = 0$, we have $P_1(0) = (k\cdot 0 + c_1)e^{-k\cdot 0} = c_1$ and hence $c_1 = 0$. Thus we have shown that

$$P_1(x) = kxe^{-kx},$$

and P_1 is completely determined.

To obtain P_2 we argue as we did in obtaining P_1. Instead of (C.14.5) we have

$$P_2(x+\Delta x) = P_2(x)P_0(\Delta x) + P_1(x)P_1(\Delta x) + P_0(x)P_2(\Delta x),$$

an equation that records the three ways in which two cars in a section of length $x + \Delta x$ can be situated in a section of length x and a section of length Δx. (See Figure C.14.3.)

Similar reasoning shows that

$$P_2(x) = \frac{k^2x^2}{2}e^{-kx}.$$

(See Exercise 8.) Applying the same reasoning inductively leads to

$$P_n(x) = \frac{(kx)^n}{n!}e^{-kx}. \tag{C.14.7}$$

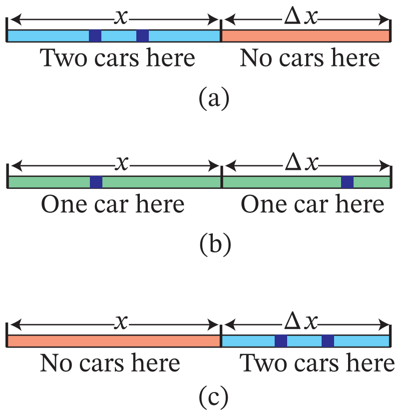


Figure C.14.3

We have obtained in (C.14.7) the formulas on which the rest of our analysis will be based. The formulas refer to a road section of any length, though the assumptions ((I)) and ((II)) refer only to short sections. What has enabled us to go from the microscopic to the macroscopic is the assumption that the traffic in a section is independent of the traffic in other sections. Equations (C.14.7) are known as the **Poisson formulas**.

The Meaning of k

The constant k was defined in terms of arbitrarily short intervals, at the microscopic level. How would we compute k in terms of observable data, at the macroscopic level? It turns out that k records the traffic density. We will show that the average number of cars in an interval of length x is kx.

The average number of cars in a section of length x is defined as $\sum_{n=0}^{\infty} nP_n(x)$. This weights each number of events (n) with its likelihood of occurring ($P_n(x)$). This average is

$$\sum_{n=0}^{\infty} nP_n(x) = \sum_{n=1}^{\infty} n\frac{(kx)^n e^{-kx}}{n!} = kxe^{-kx}\sum_{n=1}^{\infty}\frac{(kx)^{n-1}}{(n-1)!} = kxe^{-kx}e^{kx} = kx. \tag{C.14.8}$$

Thus the expected number of cars in a section is proportional to the length of the section. This shows that k measures the traffic density, the number of cars per unit length of road.

To estimate k, divide the number of cars in a long section of the road by its length.

EXAMPLE 1. (Traffic at a checkout counter.) Customers arrive at a checkout counter at the rate of 15 per hour. What is the probability that exactly five customers will arrive in a 20-minute period?

SOLUTION We assume that the probability of a customer arriving in a short interval of time is roughly proportional to the duration of that interval and that there is a negligible probability that more than one arrives. Conditions ((I)) and ((II)) hold, if we replace length of section by length of time. The probability of exactly n customers arriving in a period of x minutes is given by (C.14.7). The customer density is one per 4 minutes, so $k = 1/4$, and thus the probability that exactly five customers arrive during a 20-minute period, $P_5(20)$, is

$$\left(\frac{1}{4}\cdot 20\right)^5 \frac{e^{-(1/4)\cdot 20}}{5!} = \frac{5^5 e^{-5}}{120} \approx 0.17547.$$

◆

EXERCISES for CIE C.14

1. (a) Why would one expect that $P_0(a+b) = P_0(a)\cdot P_0(b)$ for any a and b?
(b) Verify that $P_0(x) = e^{-kx}$ satisfies the equation.

2. A cloud chamber registers an average of four cosmic rays per second.
(a) What is the probability that no cosmic rays are registered for 6 seconds?
(b) What is the probability that exactly two are registered in the next 4 seconds?

3. Telephone calls arrive at a rate of three calls per minute. What is the probability that no calls arrive in
(a) in 30 seconds? (b) in 1 minute? (c) in 3 minutes?

4. In a factory there are, on the average, two accidents per week. Let $P_n(x)$ denote the probability that there are exactly n accidents in an interval of time of length x weeks.
(a) Why is it reasonable to assume that there is a constant k such that $P_0(x)$, $P_1(x)$, ... satisfy assumptions ((I)) and ((II))?
(b) Assuming the conditions are satisfied, show that $P_n(x) = \frac{1}{n!}(kx)^n e^{-kx}$.
(c) Why is $k = 2$?
(d) Compute $P_0(1)$, $P_1(1)$, $P_2(1)$, $P_3(1)$, and $P_4(1)$.

5. An author makes an average of one mistake per ten pages. Let $P_n(x)$ be the probability that x pages (x need not be an integer) has exactly n errors.
(a) Why would one expect $P_n(x) = \left(\frac{x}{10}\right)^n \frac{1}{n!} e^{-x/10}$?
(b) Approximately how many pages would be error-free in a 1200-page book?

6. In a light rainfall on one square foot of pavement there is an average of 3 raindrops. Let $P_n(x)$ be the probability that there are n raindrops on x square feet.
(a) Check that assumptions ((I)) and ((II)) are likely to hold.
(b) Find the probability that an area of 5 square feet has exactly two raindrops.
(c) What is the most likely number of raindrops on an area of one square foot?

7. Write x^2 in the form $g(x)e^{-kx}$.

8. Show that $P_2(x) = \frac{k^2x^2}{2}e^{-kx}$.

9. (a) Why would one expect $P_3(a+b) = P_0(a)P_3(b) + P_1(a)P_2(b) + P_2(a)P_1(b) + P_3(a)P_0(b)$?
(b) Do the functions defined in (C.14.7) satisfy the equation in (a)?

10. (a) Why would one expect $\lim_{n\to\infty} P_n(x) = 0$?
(b) Show that the functions defined in (C.14.7) have that limit.

11. We obtained $P_0(x) = e^{-kx}$ and $P_1(x) = kxe^{-kx}$. Verify that $\lim_{\Delta x\to 0}\frac{P_1(\Delta x)}{\Delta x} = k$ and $\lim_{\Delta x\to 0}\frac{P_0(\Delta x)}{\Delta x} = 1-k$. Hence show that $\lim_{\Delta x\to 0}\frac{P_2(\Delta x)+P_3(\Delta x)+\cdots}{\Delta} x = 0$, and that assumptions ((I)) and ((II)) are satisfied.

12. (a) What length of road is most likely to contain exactly one car? That is, what x maximizes $P_1(x)$?
(b) What length of road is most likely to contain three cars?

13. For $x \geq 0$, $\sum_{n=0}^{\infty} P_n(x)$ should equal 1 because it is certain that there is some number of cars is in a section of length x (maybe 0 cars). Check that $\sum_{n=0}^{\infty} P_n(x) = 1$.

NOTE: This provides a probabilistic argument that $e^u = \sum_{n=0}^{\infty} \frac{u^n}{n!}$ for $u \geq 0$.

Calculus is Everywhere # 15

Hubbert's Peak

In the CIE for Chapter 6, Hubbert combined calculus with counting squares. Later he developed specific functions and used more calculus techniques in "Oil and Gas Supply Modeling", NBS Special Publication 631, U.S. Department of Commerce, National Bureau of Standards, May, 1982.

NOTE: NBS is now the National Institute of Standards and Technology (NIST).

In his approach, Q_∞ denotes the total amount of oil reserves at the time oil is first extracted and t denotes time. The derivative dQ/dt is the rate at which oil is extracted, where $Q(t)$ denotes the amount extracted up to time t. Hubbert assumes $Q(0) = 0$ and $Q'(0) = 0$. He wants to obtain a formula for $Q(t)$. He wrote:

> The curve of dQ/dt versus Q between 0 and Q_∞ can be represented by the Maclaurin series
>
> $$\frac{dQ}{dt} = c_0 + c_1 Q + c_2 Q^2 + c_3 Q^3 + \cdots.$$
>
> Since, when $Q = 0$, [then] $dQ/dt = 0$, [and] it follows that $c_0 = 0$.
> Since the curve [of dQ/dt] must return to 0 when $Q = Q_\infty$, the minimum number of terms that permit this, and the simplest form of the equation, becomes the second degree equation
>
> $$\frac{dQ}{dt} = c_1 Q + c_2 Q^2.$$
>
> By letting $a = c_1$ and $b = -c_2$, this can be rewritten as
>
> $$\frac{dQ}{dt} = aQ - bQ^2.$$
>
> Since when $Q = Q_\infty$, [then] $dQ/dt = 0$, [and] $aQ_\infty - bQ_\infty^2 = 0$ or $b = a/Q_\infty$ and [so]
>
> $$\frac{dQ}{dt} = a\left(Q - \frac{Q^2}{Q_\infty}\right).$$
>
> ... [The rate of change of production, dQ/dt,]
> is the equation of a parabola .... The maximum value occurs when the slope is 0, or when
>
> $$a - \frac{2a}{Q_\infty} Q = 0, \qquad \text{or} \qquad Q = \frac{Q_\infty}{2}.$$
>
> It is to be emphasized that the curve of dQ/dt versus Q does not have to be a parabola, but that a parabola is the simplest mathematical form that this curve can assume. We may regard the parabolic form as a sort of idealization for all such actual data curves.

He then points out that

$$\frac{dQ/dt}{Q} = a - \frac{aQ}{Q_\infty}.$$

Because the rate of production, dQ/dt, and the total amount produced up to time t, namely, $Q(t)$ are observable, the line can be drawn and its intercepts read off the graph. (The intercepts are $(0, a)$ and $(Q_\infty, 0)$.)

Hubbert then compares this with data, which it approximates fairly well.

The equation $dQ/dt = a\left(Q - Q^2/Q_\infty\right)$ can be written as

$$\frac{dQ}{dt} = \frac{a}{Q_\infty} Q\left(Q_\infty - Q\right),$$

which says that the rate of production is proportional both to the amount already produced and to the reserves $Q_\infty - Q$. This is related to the **logistic differential equation** describing bounded growth. (See Exercises 36 to 38 in Section 5.7.)

This approach, which is more formal than the one in CIE 8 at the end of Chapter 6, concludes that as Q approaches Q_∞, the rate of production will decline, approaching 0.

Chapter 13

Introduction to Differential Equations

This chapter is a brief introduction to one of the major applications of calculus, differential equations. These are equations that involve derivatives of an unknown function. The goal is usually to find the unknown function or, at least, to determine some of its properties.

The presentations given here are only an introduction to some of the main themes of a full introductory course in differential equations. Having this exposure before taking such a course should be a benefit to students.

As Section 13.1 reminds us, we have already met such equations, for instance the equation describing natural growth and decay, $\frac{dP}{dt} = kP$. Section 13.2 shows how to solve certain differential equations that involve only the unknown function and its first derivative. Sections 13.3 and 13.4 are concerned with solving an important class of differential equations that involve a function and its first and second derivatives. The last section, Section 13.5, presents a numerical method for finding an approximate solution to a differential equation.

13.1 Introduction and Review: Slope Fields and Separable Equations

A **differential equation** is an equation that involves the derivatives of an unknown function.

A differential equation for a function of one variable is called an **ordinary differential equation**. This is in contrast to differential equations for functions of more than one variable, that are called **partial differential equations**. This chapter deals only with ordinary differential equations.

We have already met ordinary differential equations on at least three occasions:

- Section 5.7 treated the differential equation that describes natural growth or decay: $P' = kP$.
- Section 3.6 had the equation for an antiderivative $F(x)$ of a known function $f(x)$: $dF/dx = f(x)$.
- The study of motion with constant acceleration in Section 3.7 was based on solving the equation $y'' = a$, where a is a constant.

There are a few examples of partial differential equations in the Calculus is Everywhere projects at the end of Chapters 16 and 17.

In this section we present a way to visualize solutions to differential equations and a method for solving differential equations of a special form. But, first, we introduce some of the basic vocabulary used to talk about differential equations.

Terminology

A **solution** of a differential equation is a function that satisfies the differential equation. To **solve** a differential equation means to find all functions that satisfy the differential equation (and any other conditions that might be specified). For instance, all solutions of $P' = kP$ are $P(t) = Ae^{kt}$ for any constant A. For a given continuous function, $f(x)$, the solutions of $F'(x) = f(x)$ are the antiderivatives of $f(x)$: $F(x) = \int f(x)\,dx + C$. And, the solution of $y'' = a$

which also satisfies $y(0) = s$ and $y'(0) = v$ is $y(t) = at^2/2 + vt + s$. The **general solution** of a differential equation is the collection of all solutions of the differential equation.

Each of these can be verified by checking that the derivatives of the solution satisfy the differential equation. For example, when $P(t) = Ae^{kt}$, then $P'(t) = kAe^{kt}$ and $kP(t) = kAe^{kt}$ and so the differential equation, $P' = kP$, is satisfied.

The **order** of a differential equation is the order of the highest-order derivative that appears in it. Examples of first-order differential equations include $P' = kP$, $y' + y^2 = 1$, $A' = 2A + 3x$, and $r^2 + a^2 + 2arr' = c^2$ (see CIE 20, The Path of the Rear Wheel of a Scooter, at the end of Chapter 15). Newton's second law of motion is a source for many second-order differential equations: $y'' = a$, $u'' + 3u' + 2u = t$, and $r'' = -GM/r^2$. Many differential equations met in applications have orders one or two. However, higher-order derivatives can appear, as in the equation used in modeling the bending of a beam, such as a diving board: $y'''' - y'' = -W$ (where $y = y(x)$ is the deflection of the beam at a distance x along the beam and W is the weight on the beam, assumed to be uniformly distributed). In Exercises 29 and 30 in Section 13.4 we will see that the general solution is

$$y(x) = c_1 e^x + c_2 e^{-x} + c_3 + c_4 x + \frac{W}{2} x^2,$$

where c_1, c_2, c_3, and c_4 are constants determined by properties of the beam.

The number of constants that appear in the solution usually equals the order of the differential equation. For a second-order differential equation for projectile motion, the **initial conditions** could be $y(t_0) = y_0$ (initial position) and $y'(t_0) = v_0$ (initial velocity). For example, in Example 4 of Section 3.7, the statements that the initial speed is 64 feet per second and the cliff is 96 feet above a beach translate into initial conditions $y(0) = 96$ and $y'(0) = 64$. A differential equation together with a set of initial conditions is called an **initial value problem**.

In some problems the constraints are given at two (or more) different values of the independent variable. For example, in a second-order differential equation the position at two times might be known, $y(t_1)$ and $y(t_2)$. These are called **boundary conditions**. (This is like determining an orbit of a comet from its positions at two different times.) A **boundary value problem** is a differential equation together with a set of boundary conditions. Most of the problems we will encounter will be initial value problems.

The general form for a **first-order linear differential equation** is

$$a_1(t)\frac{dy}{dt} + a_0(t)y = f(t).$$

A **second-order linear differential equation** allows one additional term

$$a_2(t)\frac{d^2y}{dt^2} + a_1(t)\frac{dy}{dt} + a_0(t)y = f(t).$$

More generally, an n^{th}**-order linear differential equation** has the form

$$a_n(t)\frac{d^ny}{dt^n} + a_{n-1}(t)\frac{d^{n-1}y}{dt^{n-1}} + \cdots + a_1(t)\frac{dy}{dt} + a_0(t)y = f(t).$$

A linear differential equation is said to be **homogeneous** if the right-hand side, $f(t)$, is identically zero, otherwise it is called **nonhomogeneous**.

The coefficient $a_n(t)$ of the highest-order term is assumed to not be the zero function. An n^{th}-order linear differential equation is in **standard form** when the coefficient of the highest-order term is $a_n(t) = 1$.

Differential equations such as $y' = y^2$ and $y'' + yy' = 1$ that are not linear are called **nonlinear**.

The model for natural growth or decay, $P' = kP$, is first-order, linear, and homogeneous. CIE 16, Origin of the Equation for Flow Through a Narrow Pipe, at the end of this chapter, presents a situation that leads to a homogeneous second-order linear differential equation. First- and second-order linear differential equations appear in many other applications.

TERMINOLOGY: The word "linear" reminds us of the equation of a line, $Ax + By = C$, where each variable appears to the first power. This does not mean that the solutions of a linear differential equation are lines.

One property of homogeneous linear differential equations is that any multiple of a solution is still a solution of the same differential equation. Our first example illustrates this fundamental difference between linear and nonlinear differential equations.

EXAMPLE 1. The differential equation $y' = 2y$ is linear and the differential equation $y' = 2\sqrt{y}$ is nonlinear.

(a) Verify that $y(t) = e^{2t}$ is a solution of $y' = 2y$.
(b) Find all values of A for which $y(t) = Ae^{2t}$ is a solution of $y' = 2y$.
(c) Verify that $y(t) = t^2$ is a solution of $y' = 2\sqrt{y}$ for $t > 0$, but not for $t < 0$.
(d) Find all values of B for which $y(t) = Bt^2$ is a solution of $y' = 2\sqrt{y}$ for $t > 0$.

SOLUTION

(a) When $y(t) = e^{2t}$, its derivative is $y' = 2e^{2t}$ which is exactly twice the original function. That is, $y' = 2y$, which means $y(t) = e^{2t}$ is a solution of this differential equation.
(b) When $y(t) = Ae^{2t}$, its derivative is $y' = 2Ae^{2t}$ which is exactly twice the original function. That is, $y' = 2y$ for any choice of the constant A.
(c) When $y(t) = t^2$, its derivative is $y' = 2t$ and $2\sqrt{y} = 2\sqrt{t^2} = 2|t|$. Since, when $t > 0$, $2\sqrt{y} = 2|t| = 2t$, $y(t) = t^2$ is a solution of $y' = 2\sqrt{y}$ for $t > 0$. But, when $t < 0$, then $2\sqrt{y} = 2|t| = -2t$ and $y(t) = t^2$ is not a solution of $y' = 2\sqrt{y}$ for $t < 0$.
(d) When $y(t) = Bt^2$, its derivative is $y' = 2Bt$ and, assuming $t > 0$, $2\sqrt{y} = 2\sqrt{Bt^2} = 2\sqrt{B}t$. So, $y' = 2\sqrt{y}$ only when $B = \sqrt{B}$, that is, only when $B = 0$ or $B = 1$. ◆

Section 13.2 will develop a method for solving any first-order linear differential equation. Solutions of second-order linear differential equations with **constant coefficients** (every coefficient of a derivative term is a constant) will be presented in Sections 13.3 and 13.4.

But, first, in the remainder of this section we will develop a way to visualize solutions of first-order differential equations that can be written as $y' = f(t, y)$ and a general solution method for another collection of first-order differential equations that can be solved by computing a couple antiderivatives.

Slope Fields and Solutions of First-Order Differential Equations

In Section 3.6 slope fields were introduced by the example $dy/dt = \sqrt{1+t^3}$ to picture the antiderivatives of $\sqrt{1+t^3}$. While that discussion was restricted to differential equations of the form $dy/dt = f(t)$, slope fields can be drawn when dy/dt is given in terms of t and y. At a point (t, y) we draw a segment whose slope is the derivative evaluated at (t, y).

EXAMPLE 2. Sketch the slope field for $\dfrac{dy}{dt} = y - t$.

SOLUTION Select a viewing window, say $-1 \le t \le 4$ and $-1 \le y \le 4$. At each point (t, y) on an appropriate collection of gridpoints in the viewing window, sketch a segment whose slope is $y - t$, see Figure 13.1.1(a). For instance, at $(2, 3)$ the slope is 1 (shown in blue).

When enough slopes are included in the slope field, the segments start to form curves that always adjust to match the slope of each segment that it touches, as in Figure 13.1.1(b). Each curve is the graph of a solution to the differential equation $y' = y - t$. ◆

Slope fields can be drawn only for a first-order differential equation of order one and only when it can be written as $dy/dt = f(t, y)$. The function $f(t, y)$ can involve both t and y or just one of them.

Creating a slope field by hand is very time-consuming and requires a lot of attention to detail if it is going to be done well. Fortunately, there are automatic slope field plotters that produce high-quality plots on almost any device with a graphical display: on calculators, on the web, even as an app for a smartphone.

A slope field can provide a first glimpse of solutions to a differential equation, showing where solutions are increasing ($y' > 0$) or decreasing ($y' < 0$), where the solution might have a local maximum or minimum ($y' = 0$). They can also show how solutions behave as functions of the independent variable. For example, in Figure 13.1.1(b),

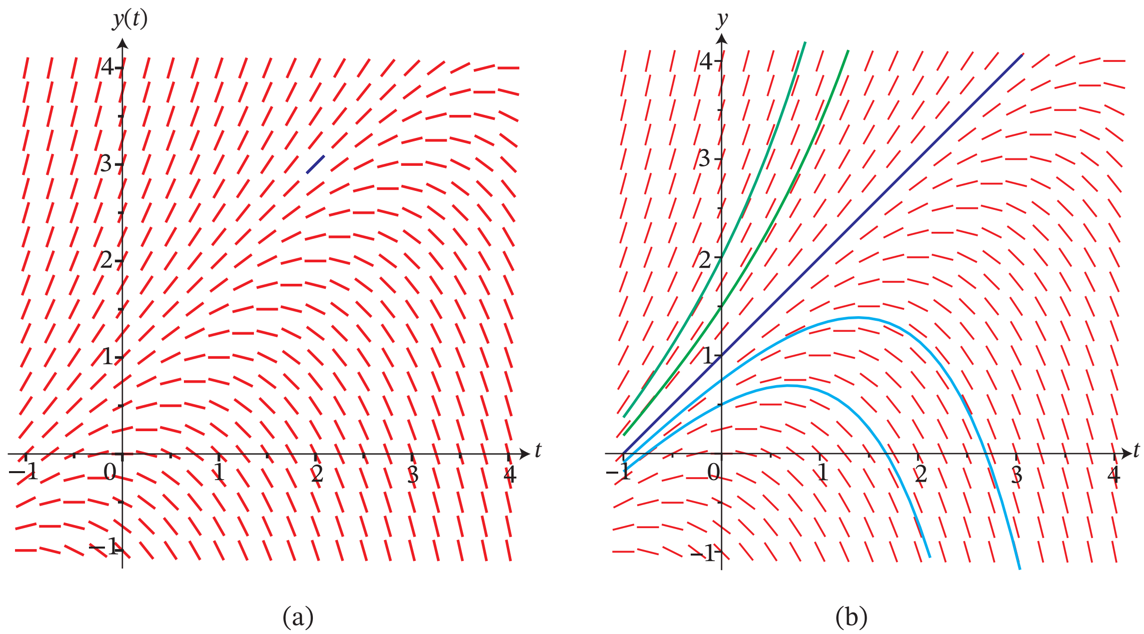


Figure 13.1.1

there appears to be one group of solutions (in green) that are always increasing, with increasing slope as t increases, and another group of solutions (in cyan) that change from increasing to decreasing. Moreover, these two groups of solutions are separated by a solution (in blue) that is a straight line. This observation can be difficult to make from an explicit formula for the solution.

Separable Differential Equations

In Section 5.7 we solved the differential equation

$$\frac{dP}{dt} = kP \qquad (P > 0). \tag{13.1.1}$$

The first step was to divide (13.1.1) by P:

$$\frac{\frac{dP}{dt}}{P} = k.$$

Both sides of (13.1.1) can be rewritten as derivatives:

$$\frac{d}{dt}\left(\ln(P(t))\right) = \frac{d}{dt}(kt),$$

and therefore there is a constant C such that $\ln(P(t)) = kt + C$. Thus $P(t) = e^{kt+C} = e^C e^{kt}$. Renaming the constant e^C by A, we conclude that

$$P(t) = Ae^{kt}.$$

In general, when a first-order differential equation can be written in a way that the slope is always equal to the product of a function of x and another function of y, $y' = f(x)g(y)$, dividing both sides of the differential equation by $g(y)$ separates the variables:

$$\frac{y'}{g(y)} = f(x).$$

Because all of the y's appear on one side of the equation and all of the x's on the other side, we say the variables have been separated.

The next step is to integrate to find an antiderivative (with respect to x) of each side:

$$\int \frac{y'(x)}{g(y(x))}\,dx = \int f(x)\,dx.$$

Notice that $y'(x)\,dx$ in the left-hand side is the differential dy:

$$\int \frac{dy}{g(y)} = \int f(x)\,dx. \tag{13.1.2}$$

To go further, compute both indefinite integrals in (13.1.2). Do not forget to include the constant of integration. For each value of this constant, the resulting formula is a solution of the differential equation.

Observation 13.1.1: *Dealing with Differentials*

Equation (13.1.2) can be obtained directly from the differential equation by treating $y' = dy/dx$ as a fraction when separating variables.

From here, what you do depends on the specific problem. If an initial condition is provided, it can be inserted into the general solution to determine a specific value of the constant of integration.

EXAMPLE 3. (a) Find the general solution of $\dfrac{dy}{dx} = -x^3y^3$. (b) Solve the initial value problem $y' = -x^3y^3$, $y(0) = \dfrac{-1}{4}$.

SOLUTION

(a) Making use of Observation 13.1.1, this differential equation can be separated by multiplying it by dx and, assuming $y \neq 0$, dividing by y^3:

$$\frac{dy}{y^3} = -x^3\,dx.$$

It is easier to find an antiderivative (with respect to y) of the left-hand side after rewriting the integrand with a negative exponent: $y^{-3}\,dy = -x^3\,dx$. Now it is easy to find an antiderivative of each side:

$$\frac{1}{-2}y^{-2} + C_1 = \frac{-1}{4}x^4 + C_2$$

where C_1 and C_2 are arbitrary constants. or

$$y^{-2} = \frac{x^4}{2} - 2(C_2 - C_1).$$

The expression $-2(C_2 - C_1)$ is really just another arbitrary constant, which we will now give the name C.

A formula for solutions is $y^{-2} = x^4/2 + C$ for any constant C. In this case it is possible to go further, to solve for y as a function of x. To do this, first take the reciprocal of both sides:

$$y^2(x) = \frac{1}{\frac{x^4}{2} + C} = \frac{2}{x^4 + 2C} \tag{13.1.3}$$

Taking the square root of both sides yields two formulas for the solution for each choice of the constant C:

$$y(x) = +\sqrt{\frac{2}{x^4+2C}} \qquad \text{and} \qquad y(x) = -\sqrt{\frac{2}{x^4+2C}}. \tag{13.1.4}$$

What this means is that for each value of C there are actually two solutions, one with a positive value of y and one with a negative value of y. See also Exercise 13.

There is one more solution that needs to be listed separately. When we separated variables, we had to assume $y \neq 0$. What happens when $y = 0$? The constant function $y(x) = 0$ satisfies the differential equation, so it is also a solution.

More generally, if $g(y_0) = 0$ then $y(x) = y_0$ is a solution of $y' = g(y)f(x)$. These constant functions that are not produced by the general process but still solve the differential equation are called **extraneous solutions**.

(b) We are asked to find the solution that satisfies the initial condition $y(0) = -1/4$.

Working with (13.1.4), we use the fact that the initial value is negative to select the solution formula with the negative sign: $y(x) = -\sqrt{2/(x^4+2C)}$. Evaluating this solution when $x = 0$ and $y(0) = -1/4$ leads to the requirement that

$$\frac{-1}{4} = -\sqrt{\frac{2}{0^4+2C}}.$$

This can be simplified by squaring both sides to $1/16 = 1/C$, from which we conclude $C = 16$. The solution of this initial value problem is

$$y(x) = -\sqrt{\frac{2}{x^4+32}}.$$

The problem of finding C to satisfy the initial condition can also be solved using (13.1.3). Evaluating this solution when $x = 0$ and $y(0) = -1/4$ yields the equation

$$\left(\frac{-1}{4}\right)^2 = \frac{2}{0^4+2C}$$

which simplifies to $1/16 = 1/C$. The only possible choice for C is $C = 16$. This gives $y^2(x) = 2/(x^4+32)$ as a solution to the initial value problem. Solving for $y(x)$ means taking the square root of both sides, and noting that only the solution with a negative sign gives a solution that satisfies the initial condition. (The other formula give $y(0) = +1/4$.) ♦

Summary

This introduction to differential equations focused on terminology, slope fields, and the solution of separable differential equations.

Slope fields provide a graphical visualization of solutions of any first-order differential equation that can be written as $y' = f(t,y)$. A slope field also reinforces the fact that, in general, there are an infinite number of solutions to a differential equation. When an initial condition is provided, say $y(t_0) = y_0$, this information can be used to identify a solution that passes through the point (t_0, y_0).

A separable differential equation can be written in the form $y' = f(x)g(y)$. The general process of solving a separable differential equation can be summarized as follows:

Algorithm: Solution Process for Solving a Separable Differential Equation

In general, to solve $\dfrac{dy}{dx} = f(x)g(y)$, follow these steps:

1. Multiply by the differential dx and divide by $g(y)$: $\dfrac{dy}{g(y)} = f(x)\,dx$.
2. Integrate both sides: $\displaystyle\int \frac{dy}{g(y)} = \int f(x)\,dx$ (remember to add the constant of integration, say C).
3. If an initial condition is provided, apply it to solve for C.
4. Attempt to manipulate this solution into a formula for y as a function of x.
5. Lastly, identify any extraneous solutions: $y(x) = y_0$ for any number y_0 that makes $g(y_0) = 0$.

EXERCISES for Section 13.1

1. Using as few mathematical symbols as possible, describe what is meant by a linear differential equation.

In Exercises 2 to 9

(a) state the order of the differential equation.

(b) is the differential equation linear or nonlinear?

(c) if it is linear, is it homogeneous or nonhomogeneous?

(d) write each linear differential equation in standard form.

2. $(y'')^3 + (y')^2 = y$

3. $3ty'' + t^2y' = y$

4. $t^5y''' + \cos(t^2)y + 3y = 0$

5. $t^5(y''')^4 + \cos(y^2)y + 3 = 0$

6. $\sqrt{1+t^3}\,y' + (\cos(y))^4y + 3t^2 = 0$

7. $\sqrt{1+t^3}\,y' + y + 3t^2 = 0$

8. $y'' + \sqrt{1+t^2}(y-3) = 0$

9. $(y'')^3\sqrt{1+(y')^2} = y^6$

In Exercises 10 to 13 sketch the slope field for the differential equation.

10. $y' = -y$

11. $y' = 1 - y$

12. $y' = 2t - y$

13. $y' = x^3y^3$

In Exercises 14 to 19 determine if the differential equation is separable or not separable.

Do not attempt to find the solution.

14. $3t^2y' + 6t = t^3 + \sin(t)$

15. $y' = \dfrac{\sin(t)}{y^3}$

16. $y' = \sin\left(\dfrac{t}{y^3}\right)$

17. $y' = \dfrac{e^{xy^2}}{y}$

18. $y' = t + y$

19. $t^3 + t^2y' = \ln(t)$

In Exercises 20 to 41 solve the separable differential equation.

20. $\dfrac{dy}{dt} = t$

21. $\dfrac{dy}{dt} = t^2$

22. $\dfrac{dy}{dt} = y$

23. $\dfrac{dy}{dt} = y^2$

24. $\dfrac{dy}{dx} = \dfrac{\sin(x)}{\cos(y)}$

25. $\dfrac{dy}{dt} = \sec^2(y)$

26. $\dfrac{dy}{dt} = \dfrac{t}{y}$

27. $\dfrac{dy}{dt} = \dfrac{y}{t}$

28. $\dfrac{dy}{dt} = \dfrac{ty}{t^2+1}$

29. $\dfrac{dy}{dx} = \dfrac{\cos(y)}{e^x}$

30. $\dfrac{dy}{dt} = \sqrt{\dfrac{1-y^2}{t}},\ t > 0$

31. $\dfrac{dy}{dx} = \dfrac{-e^{y^2}}{2xy}$

32. $\dfrac{dy}{dx} = x\ln(x)\left(4+y^2\right)$

33. $\dfrac{dy}{dx} = \dfrac{e^x\sin(3x)}{\sqrt{9-4y^2}}$

34. $\dfrac{dy}{dy} = \sin(3t)\sec(2y)$

35. $\dfrac{dy}{dt} = y^2\ln(t)$

36. $\dfrac{dy}{dt} = \dfrac{\tan(2t)}{e^{2y}}$

37. $\dfrac{dy}{dt} = y\sqrt{1+3t}$

38. $\sec(\theta)\dfrac{d\theta}{dt} - t^2 = 3t$

39. $y^2\cos(\theta)\dfrac{dy}{d\theta} - \sin(\theta) = \cos(\theta)$

40. $\dfrac{d\theta}{dt} = \dfrac{\sin^2(2t)}{\cos^2(3\theta)}$

41. $\dfrac{dy}{dt} = \dfrac{e^y\sec^2(2t)}{y}$

42. (a) What is the general solution of $\frac{d^2y}{dt^2} = -16$? (b) Find the solution of $\frac{d^2y}{dt^2} = -16$ with $y(0) = 10$ and $y'(0) = 5$.

43. Show that for constants a, b, c, and d, $y = ae^x + be^{-x} + c + dx + \frac{W}{2}x^2$ satisfies the fourth-order equation for a weight-bearing beam, $\frac{d^4y}{dx^4} - \frac{d^2y}{dx^2} = -W$.

44. Assume that $f(x)$ and $g(x)$ are solutions of $2y'' + 3y' + y = 0$. Which of the following are also solutions?
(a) $3f(x)$ (b) $g(3x)$ (c) $f(x) + g(x)$ (d) $3f(x) - 2g(x)$

45. (a) Check that $y = e^t - 1$ is a solution of $y' = 1 + y$. (b) Find all solutions of $y' = 1 + y$.

46. (a) Check that $y = \tan(x)$ is a solution of $y' = 1 + y^2$. (b) Find all solutions of $y' = 1 + y^2$.

47. For which of (a) to (d) is $y = \sin(3x)$ a solution?
(a) $\frac{d^2y}{dx^2} = 9y$ (b) $\frac{d^2y}{dx^2} = -9y$ (c) $\frac{dy}{dx} = \sqrt{1-y^2}$ (d) $\frac{dy}{dx} = 3\sqrt{1-y^2}$

The general solution of $y' = y^3\left(1-y^2\right)$ can be found by separation of variables using partial fractions, but, as Exercise 48 shows, does not yield an explicit formula for y as a function of t.

Exercises 49 to 51 illustrate other ways of analyzing the solutions of this differential equation that do not require a formula for y as a function of t.

Exercises 52 and 53 involve the qualitative analysis of a general first order **autonomous differential equation** $y' = f(y)$. These ideas are then applied to the limited growth model in Exercise 54. Finding the explicit solution of the limited growth model is point of Exercise 55.

48. Find the general solution of $y' = y^3\left(1-y^2\right)$. Why is not possible to write this solution as an explicit formula for y as a function of t?

49.
(a) What constant functions are solutions of $y' = y^3\left(1-y^2\right)$?
(b) Sketch the slope field for $0 \le t \le 2$ and $-2 \le y \le 2$.
(c) Starting with the slope field created in (b), sketch the solution of this differential equation with $y(0) = 2$.
(d) Add to the graph created in (c) the solution of this differential equation with $y(0) = \frac{1}{2}$.
(e) When $y(0) = 2$, does $\lim_{t\to\infty} y(t)$ seem to exist?
(f) When $y(0) = \frac{1}{2}$, does $\lim_{t\to\infty} y(t)$ seem to exist?

50. This approach to examining solutions of $y' = y^3\left(1-y^2\right)$, $y(0) = 2$, does not use a slope field.
(a) Show why $y(t)$ is never less than 1. (b) Is $y(t)$ ever increasing? (c) Why does $\lim_{t\to\infty} y(t)$ exist? (d) Find $\lim_{t\to\infty} y(t)$.

51. The approach used in Exercise 50 is easily adapted to analyze the solution of $y' = y^3\left(1-y^2\right)$, $y(0) = \frac{1}{2}$.
(a) Show why $y(t)$ is never more than 1. (b) Is $y(t)$ ever decreasing? (c) Why does $\lim_{t\to\infty} y(t)$ exist? (d) Find $\lim_{t\to\infty} y(t)$.

52. Consider the autonomous differential equation $y' = f(y)$ where f is a differentiable function. Assume the sign of $y'(t)$ is always opposite the sign of $y(t)$.

(a) Sketch what the graph of $y(t)$ might look like if $y(0) = 1$.
(b) Sketch what the graph of $y(t)$ might look like if $y(0) = -1$.
(c) Must the sign of $y''(t)$ always be opposite the sign of $y'(t)$?
(d) What is the relationship between $y''(t)$ and $y(t)$?

53. Consider the autonomous differential equation $y' = f(y)$ where f is a differentiable function. Assume that the sign of $y''(t)$ is always opposite the sign of $y(t)$.

(a) What might the graph of $y(t)$ look like? (b) Give a specific example of such a function.

54. The differential equation $P' = kP(M-P)$ where $P = P(t)$, k, and M are positive constants occurs in the theory of limited growth. It can be analyzed without solving. Assume that $t \geq 0$. (The solution is found in Exercise 55.)

(a) Show that $P(t) = 0$ and $P(t) = M$ are constant solutions to this equation.
(b) Show that P is increasing if and only if $0 < P(t) < M$.
(c) To determine the concavity of solutions it is necessary to know the sign of P''. Find an expression for P'' that involves P and P' (and k and M).
(d) Use the original differential equation to obtain an expression for P'' that involves P, but not P'.
(e) Explain why solutions have inflection points when $P = \dfrac{M}{2}$.
(f) Assume that $P(0) = \dfrac{M}{4}$. Show that $P(t)$ is never greater than M.

55. Find the solution to the problem analyzed in Exercise 54, that is, $P' = kP(M-P)$ with $P(0) = M/4$.

Your answer should be a formula for P as a function of t.

---

56. (a) Give an example of a solution to $\left(\dfrac{dy}{dx}\right)^2 = -y^2$. (b) Find all solutions of the equation in (a).

57. Find a first-order differential equation that has $y = e^{x^2}$ as a solution.

58. (a) Show directly that if $\dfrac{dy}{dx} = \sqrt{1-y^2}$ then $\dfrac{d^2y}{dx^2} = -y$.
(b) Show that for every constant k, $y = \sin(x+k)$ satisfies both differential equations in (a) for $|x+k| \leq \pi/2$.

13.2 First-Order Linear Differential Equations

Linear differential equations appear in many applications. For example, see CIE 16, Origin of the Equation for Flow Through a Narrow Pipe, at the end of this chapter.

The first example sets the stage for the presentation of a general method for solving a **first-order linear differential equation**.

EXAMPLE 1. Find the general solution of $t^2\dfrac{dy}{dt} + 2ty = e^{3t}$ for $t > 0$.

SOLUTION The two terms on the left-hand side of the differential equation are the derivative of the product of t^2 and $y(t)$:

$$\frac{d}{dt}\left(t^2 y(t)\right) = t^2 y'(t) + 2ty(t). \tag{13.2.1}$$

Because the differential equation says the right-hand side of (13.2.1) is equal to e^{3t}, this gives us a new, simpler, — and equivalent — differential equation:

$$\frac{d}{dt}\left(t^2 y(t)\right) = e^{3t}. \tag{13.2.2}$$

Integrating both sides of (13.2.2) with respect to t yields

$$t^2 y(t) = \frac{1}{3}e^{3t} + C.$$

Since the problem specifies that $t > 0$, it is permissible to divide by t^2:

$$y(t) = \frac{1}{3t^2}e^{3t} + \frac{C}{t^2}.$$

♦

The general method we will now describe assumes the differential equation is in **standard form**:

$$\frac{dy}{dt} + a(t)y = f(t). \tag{13.2.3}$$

The solution in Example 1 depended upon the observation that the two terms on the left-hand side of the differential equation could be combined into a single derivative of a product. This happens infrequently. But, if you multiply the general first-order linear differential equation in standard form, (13.2.3), by an **integrating factor**

$$v(t) = e^{\int a(t)\,dt}$$

this produces an equivalent differential equation

$$v(t)\frac{dy}{dt} + v(t)a(t)y = v(t)f(t) \tag{13.2.4}$$

with the property that the two terms on the left-hand side are the derivative of the product $v(t)y(t)$.

The key step in verifying this critical fact is the relationship between v and its first derivative, v':

$$\begin{aligned} v'(t) &= e^{\int a(t)\,dt}\frac{d}{dt}\left(\int a(t)\,dt\right) && (\text{ chain rule }) \\ &= e^{\int a(t)\,dt}a(t) && (\text{ FTC II }) \\ &= a(t)v(t) && (\text{ definition of } v(t)\text{ }). \end{aligned}$$

Thus, since the two terms on the left-hand side of (13.2.4) are the derivative of the product of the integrating factor and the solution of (13.2.3),

$$\begin{aligned} \frac{d}{dt}\big(v(t)y(t)\big) &= v(t)y'(t) + v'(t)y(t) && (\text{ product rule }) \\ &= v(t)y'(t) + v(t)a(t)y(t) && (\ v'(t) = a(t)v(t)\ ) \\ &= v(t)\big(y'(t) + a(t)y(t)\big) && (\text{ factoring }) \\ &= v(t)f(t) && (\ y' + a(t)y = f(t) \text{ by the original DE }). \end{aligned}$$

A solution of the differential equation $(v(t)y)' = v(t)f(t)$ is obtained by integrating both sides with respect to t. Then, dividing by $v(t)$ produces an explicit formula for the solution of the general first-order linear differential equation. All of this can be summarized in a single formula.

Formula 13.2.1: *General Solution of First-Order Linear Differential Equations*

The general solution of a first-order linear differential equation in standard form, $y' + a(t)y = f(t)$, is

$$y(t) = \frac{1}{v(t)}\int v(t)f(t)\,dt \tag{13.2.5}$$

where an integrating factor is $v(t) = e^{\int a(t)\,dt}$.

Observation 13.2.1: *Repeating Example 1 using Formula 13.2.1*

To use Formula 13.2.1 to solve Example 1, the differential equation needs to be written in standard form. To do this, divide the differential equation by t^2 (which is not zero because $t > 0$) to obtain

$$y' + \frac{2}{t}y = \frac{1}{t^2}e^{3t}. \tag{13.2.6}$$

With $a(t) = 2/t$, an integrating factor is $v(t) = e^{2\ln(t)} = e^{\ln(t^2)} = t^2$.

NOTE: Because all we need is an integrating factor, not all integrating factors, it was not necessary to include a constant of integration when finding an antiderivative of $a(t) = 2/t$.

Now, by (13.2.5), with $a(t) = 2/t$, $f(t) = e^{3t}/t^2$, and $v(t) = t^2$, the general solution is

$$y(t) = \frac{1}{t^2}\int t^2\left(\frac{1}{t^2}e^{3t}\right)\,dt = \frac{1}{t^2}\int e^{3t}\,dt = \frac{1}{t^2}\left(\frac{1}{3}e^{3t} + C\right)\,dt = \frac{1}{3t^2}e^{3t} + \frac{C}{t^2},$$

which is exactly what was found in Example 1.

EXAMPLE 2. Find all solutions of

$$\frac{dy}{dt} + 4t^3y = \cos(t)e^{-t^4}. \tag{13.2.7}$$

SOLUTION First, note that this differential equation is not separable, but it is linear (and first-order). Because the coefficient of the y' term is 1, it is in standard form.

While we could simply apply Formula 13.2.1, it is a more enlightening to work through the basic steps, as follows.

Since $a(t) = 4t^3$, an integrating factor is $v(t) = e^{t^4}$. Then,

$$\begin{aligned}
e^{t^4}y' + 4t^3e^{t^4}y &= \cos(t) && (\text{ multiply (13.2.7) by } v(t)\text{ })\\
\frac{d}{dt}\left(e^{t^4}y(t)\right) &= \cos(t) && (\text{ combine two terms into one derivative of a product })\\
e^{t^4}y(t) &= \sin(t) + C && (\text{ integrate both sides with respect to } t\text{ })\\
y(t) &= \sin(t)e^{-t^4} + Ce^{-t^4} && (\text{ solve for } y(t)\text{ }).
\end{aligned}$$

♦

Observation 13.2.2:

This process can be used to solve any first-order linear differential equation provided that we can evaluate two integrals: one involved in finding the integrating factor, $\int p(t)\,dt$, and also $\int v(t)f(t)\,dt$.

When an initial condition is provided, the first step is still to find the general solution. Then, the initial condition is used to determine the specific solution that also satisfies the initial condition.

EXAMPLE 3. Find the solution of the initial value problem $(t^2+1)\frac{dy}{dt} - ty = t$, $y(0) = 2$.

SOLUTION To put this linear differential equation in standard form, divide by $t^2 + 1$, obtaining

$$\frac{dy}{dt} - \frac{t}{t^2+1}y = \frac{t}{t^2+1}. \tag{13.2.8}$$

With $a(t) = -t/(t^2+1)$, $\int a(t)\,dt = -\ln(t^2+1)/2$ and so an integrating factor is

$$v(t) = e^{\int a(t)\,dt} = e^{\frac{-1}{2}\ln(t^2+1)} = e^{\ln\left((t^2+1)^{-1/2}\right)} = \left(t^2+1\right)^{-1/2}.$$

Then, proceeding as in Example 2,

$$\begin{aligned}
\left(t^2+1\right)^{-1/2} y' - t\left(t^2+1\right)^{-3/2} y' &= t\left(t^2+1\right)^{-3/2} && \text{(multiply (13.2.8) by } v(t) \text{)}\\
\frac{d}{dt}\left(\left(t^2+1\right)^{-1/2} y(t)\right) &= t\left(t^2+1\right)^{-3/2} && \text{(combine two terms into one derivative of a product)}\\
\left(t^2+1\right)^{-1/2} y(t) &= \int t\left(t^2+1\right)^{-3/2}\,dt && \text{(integrate both sides with respect to } t \text{)}\\
&= -\left(t^2+1\right)^{-1/2} + C && \text{(substitution: } u = t^2+1 \text{)}\\
y(t) &= -1 + C\left(t^2+1\right)^{1/2} && \text{(solve for } y(t) \text{).}
\end{aligned}$$

To satisfy the initial condition, $y(0) = 2$, evaluate the general solution when $t = 0$ and $y(0) = 2$ to find: $2 = -1 + C$, so $C = 3$. The solution of the initial value problem is $y(t) = -1 + 3\sqrt{t^2+1}$. ♦

Summary

The general solution of any first-order linear differential equation in standard form, $y' + a(t)y = f(t)$, can be found using Formula 13.2.1.

In practice, however, it is often easier to find an integrating factor, $v(t) = e^{\int a(t)\,dt}$, multiply the differential equation by $v(t)$, verify that the two terms on the left-hand side can be combined into the derivative of $v(t)y(t)$, integrate both sides of the resulting equation (remembering to include a constant of integration), and then solving for $y(t)$ by dividing by $v(t)$.

EXERCISES for Section 13.2

1. (a) What is the standard form for a homogeneous first-order linear differential equation?
(b) Show that every homogeneous first-order linear differential equation is separable.
(c) What does this mean about finding solutions of homogeneous first-order linear differential equations?

2. Verify that $y(t) = \dfrac{e^{3t}}{3t^2} + \dfrac{C}{t^2}$ is the solution of the differential equation in Example 1.

3. Check that $y(t) = \sin(t)e^{-t^4} + Ce^{-t^4}$ is the solution of the differential equation in Example 2.

4. Confirm that $y(t) = -1 + 3\left(t^2+1\right)^{1/2}$ is the solution of the initial value problem in Example 3.

In Exercises 5 to 19 find the general solution of the differential equation. If an initial condition is provided, find the corresponding solution that also satisfies the given initial conditions.

5. $y' + 2y = 1$, $y(0) = 0$

6. $y' = y - t$, $y(0) = 1/2$

7. $y' - y = 3e^{2t}$

8. $ty' + 7y = 5t^2$

9. $2ty' + y = 10\sqrt{t}$

10. $ty' + 2y = 5t$, $y(2) = 4$

11. $ty' - y = t^3$, $y(1) = 7$

12. $2ty' - 5y = 12t^2$

13. $(1+t)y' + y = \sin(t)$

14. $ty' - 2y = t^3\sec(t)\tan(t)$

15. $y' = 1 + t + y + ty$

16. $y' = 2ty + 9t^2e^{t^2}$, $y(0) = 0$

17. $ty' + (2t-3)y = 6t^6$

18. $ty' + 2y = \ln(t)$

19. $ty' + 4y = t^{-2}e^t$

13.3 Second-Order Linear Differential Equations: Homogeneous Case

Second-order linear differential equations are considerably harder to solve than first-order linear differential equations. They become much easier to solve when the coefficients are constants.

In this section we learn how to find the general solution of homogeneous second-order linear differential equations with constant coefficients:

$$a\frac{d^2y}{dt^2}+b\frac{dy}{dt}+cy=0 \tag{13.3.1}$$

where a, b, and c are constants. The corresponding nonhomogeneous second-order linear differential equations will be addressed in Section 13.4.

The assumption that all coefficients are constants may seem restrictive, but (13.3.1) is broad enough to be of use in many applications.

Before finding solutions of (13.3.1), we spend a little time to develop some important properties of solutions of linear differential equations.

New Solutions from Old

If $y(t)$ is a solution to (13.3.1), then so is $ky(t)$ for any constant k. To see this, check that ky satisfies (13.3.1) whenever y does:

$$\begin{aligned} a(ky)''+b(ky)'+c(ky) &= aky''+bky'+cky \\ &= k\cdot(ay''+by'+cy) \\ &= k\cdot 0 = 0. \end{aligned}$$

This result is not true if the differential equation is nonlinear. For example, $Y(t)=t^2$ is a solution of $y'=2\sqrt{y}$ for $t>0$ because $Y'=2t$ and $2\sqrt{Y}=2t$. But, if $y(t)=4Y(t)=4t^2$, then $y'=8t$ and $2\sqrt{y}=2(2t)=4t$, so y' and $2\sqrt{y}$ are not equal.

If y_1 and y_2 are solutions of (13.3.1), then so is y_1+y_2. We show this by direct substitution of $y(t)=y_1(t)+y_2(t)$ into the differential equation:

$$\begin{aligned} ay''+by'+cy &= a(y_1+y_2)''+b(y_1+y_2)'+c(y_1+y_2) \\ &= ay_1''+ay_2''+by_1'+by_2'+cy_1+cy_2 \\ &= (ay_1''+by_1'+cy_1)+(ay_2''+by_2'+cy_2) \\ &= 0+0=0. \end{aligned}$$

So $y(t)=y_1(t)+y_2(t)$ is also a solution of the same homogeneous second-order linear differential equation.

Because $y_1-y_2=y_1+(-1)y_2$, the difference y_1-y_2 of solutions is also a solution of the homogeneous equation. These observations about the structure of solutions to a homogeneous second-order linear differential equation are commonly referred to as the **superposition principle**.

Theorem 13.3.1: Superposition Principle for Solutions of $ay''+by'+cy=0$

If y_1 and y_2 are solutions of the homogeneous second-order linear differential equation (13.3.1), then so is $k_1y_1+k_2y_2$ for any values of the constants k_1 and k_2.

Proof of Theorem 13.3.1

The observations made immediately preceding the statement of Theorem 13.3.1 show that if y_1 and y_2 are solutions of $ay''+by'+cy=0$ and k_1 and k_2 are any constants, then k_1y_1 and k_2y_2 are solutions of the same differential equation. These same observations also show that their sum $k_1y_1+k_2y_2$ is a solution of the same linear and homogeneous differential equation.

Observation 13.3.2: ***Being Homogeneous and Linear are Essential for Superposition***

That the right-hand side of (13.3.1) is 0 is critical to the superposition principle. If the right-hand side was anything other than 0, the reasoning would not go through for all constants k_1 and k_2.

The superposition principle remains true if the coefficients are allowed to be functions of t, but the differential equation must be linear. The theorem is also true for homogeneous linear differential equations of any order, not just when the order is 2.

The Main Idea: Try $y = e^{rt}$

When r is a constant the derivatives of e^{rt} are multiples of e^{rt}. This suggests looking for solutions of (13.3.1) in that form.

If $y = e^{rt}$, we have

$$y = e^{rt}, \qquad y' = re^{rt}, \quad \text{and} \quad y'' = r^2 e^{rt}.$$

Substituting $y = e^{rt}$ into (13.3.1) we see that the differential equation is satisfied only when r satisfies a specific equation:

$$0 = a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + cy = ar^2e^{rt} + bre^{rt} + ce^{rt} = (ar^2 + br + c)e^{rt}.$$

Since e^{rt} is never zero we conclude that $y = e^{rt}$ is a solution to (13.3.1) when r is a root of the quadratic equation

$$ar^2 + br + c = 0. \tag{13.3.2}$$

The quadratic polynomial $ar^2 + br + c$, is called the **characteristic polynomial** of the differential equation (13.3.3).

The roots of (13.3.2) are, by the quadratic formula,

$$r_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \qquad \text{and} \qquad r_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

The sign of $b^2 - 4ac$ determines whether there are two distinct real roots ($b^2 - 4ac > 0$), a single repeated real root ($b^2 - 4ac = 0$), or a pair of complex conjugate roots ($b^2 - 4ac < 0$).

Each case will be treated separately.

Case 1: **Two distinct real roots** ($b^2 - 4ac > 0$)

The functions $y_1(t) = e^{r_1 t}$ and $y_2 = e^{r_2 t}$ are both solutions to (13.3.1). By Theorem 13.3.1,

$$y = k_1 y_1 + k_2 y_2 = k_1 e^{r_1 t} + k_2 e^{r_2 t}$$

is a solution of (13.3.1) for any constants k_1 and k_2.

It is proved in a more advanced course that $y = k_1 e^{r_1 t} + k_2 e^{r_2 t}$ is the general solution of the differential equation, and that there are no other solutions.

EXAMPLE 1. Find the general solution of $\dfrac{d^2y}{dt^2} - \dfrac{dy}{dt} - 2y = 0$.

SOLUTION The characteristic equation is $r^2 - r - 2 = 0$. Since $b^2 - 4ac = (-1)^2 - 4(1)(-2) = 9 > 0$ there are two real roots. Instead of using the quadratic formula, this characteristic equation can be factored: $r^2 - r - 2 = (r-2)(r+1)$. Thus the roots are $r_1 = 2$ and $r_2 = -1$ and the general solution of $y'' - y' = 2y = 0$ is

$$y = c_1 e^{2t} + c_2 e^{-t}.$$

◆

Case 2: **One real repeated root** ($b^2 - 4ac = 0$)

When $b^2 - 4ac = 0$, then $r_1 = -b/(2a)$ and $r_2 = -b/(2a)$. There is only one exponential solution, namely $y_1 = e^{-bt/(2a)}$. A second solution, $y_2 = te^{-b/(2a)}$ is found by a procedure called **variation of parameters** (see Exercise 21).

The general solution of $ay'' + by' + cy = 0$ when the characteristic equation has only one real root ($b^2 - 4ac = 0$) is

$$y = c_1 e^{-bt/(2a)} + c_2 te^{-bt/(2a)} = (c_1 + c_2 t)e^{-bt/(2a)}.$$

Exercise 16 invites you to check that y_2 is a solution by substituting it into (13.3.1) and using the fact that $b^2 - 4ac = 0$ (and $a \neq 0$). Exercise 21 contains the complete derivation of this second solution.

EXAMPLE 2. Find the solution of $\frac{d^2y}{dt^2} + 6\frac{dy}{dt} + 9y = 0$ with $y(0) = 2$ and $y'(0) = -2$.

SOLUTION The first step to solving this initial value problem is to find the general solution of the differential equation $y''+6y'+9y = 0$. The characteristic equation is $r^2+6r+9 = 0$. Since $b^2-4ac = (6)^2-4(1)(9) = 0$ there is only a single repeated real root. Because $r^2 + 6r + 9 = (r+3)^2$ we see that $r_1 = r_2 = -3$. In this case, the general solution is

$$y = c_1 e^{-3t} + c_2 te^{-3t}.$$

Now, looking at the first initial condition: $y(0) = 2$ is satisfied when $2 = c_1 e^{-3\cdot 0} + c_2 0 e^{-3\cdot 0} = c_1$.
The second initial condition, $y'(0) = -2$, requires a formula for the derivative of the general solution:

$$y'(t) = -3c_1 e^{-3t} + c_2(e^{-3t} - 3te^{-3t}).$$

Then, inserting $t = 0$ and $y'(0) = -2$ yields: $-2 = -3c_1 + c_2$. We already know $c_1 = 2$, so $-2 = -3(2) + c_2$. Thus, $c_2 = -2 + 6 = 4$. The solution to the initial value problem is $y(t) = 2e^{-3t} + 4te^{-3t} = (2+4t)e^{-3t}$. ♦

Case 3: **Complex conjugate roots** ($b^2 - 4ac < 0$)

If $\sqrt{b^2 - 4ac}$ is negative, then the roots of the characteristic polynomial are complex-valued. In fact, they are complex conjugates. Write the two complex-valued roots as $r_1 = p + iq$ and $r_2 = p - iq$ where $p = -b/(2a)$ and $q = \sqrt{4ac - b^2}/(2a)$. Then, the general solution of $ay'' + by' + cy = 0$ is

$$y(t) = c_1 e^{pt}\cos(qt) + c_2 e^{pt}\sin(qt). \tag{13.3.3}$$

The sudden appearance of trigonometric functions in (13.3.3) is explained by recalling Euler's formula: $e^{i\theta} = \cos(\theta) + i\sin(\theta)$. The two complex-valued roots create two complex-valued solutions:

$$u_1(t) = e^{r_1 t} = e^{(p+iq)t} = e^{pt}e^{iqt} \;\; = e^{pt}\left(\cos(qt) + i\sin(qt)\right) \quad = e^{pt}\cos(qt) + ie^{pt}\sin(qt)$$

and

$$u_2(t) = e^{r_2 t} = e^{(p-iq)t} = e^{pt}e^{-iqt} = e^{pt}\left(\cos(-qt) + i\sin(-qt)\right) = e^{pt}\cos(qt) - ie^{pt}\sin(qt).$$

The principle of superposition (Theorem 13.3.1) holds even when the solutions and constants are complex-valued. Two real-valued solutions are

$$y_1(t) = \frac{1}{2}u_1(t) + \frac{1}{2}u_2(t) = e^{pt}\cos(qt) \qquad \text{and} \qquad y_1(t) = \frac{1}{2i}u_1(t) - \frac{1}{2i}u_2(t) = e^{pt}\sin(qt).$$

The real-valued general solution is formed by adding real-valued constant multiples of y_1 and y_2.

EXAMPLE 3. Find the general solution of $\dfrac{d^2y}{dt^2} + 2\dfrac{dy}{dt} + 10y = 0$.

SOLUTION The characteristic equation is $r^2 + 2r + 10 = 0$. Since $b^2 - 4ac = (2)^2 - 4(1)(10) = -36 < 0$ there are complex conjugate roots $r_1 = (-2 + \sqrt{-36})/2 = -1 + 3i$ and $r_2 = (-2 - \sqrt{36})/2 = -1 - 3i$. The general solution is

$$y(t) = c_1 e^{-t}\cos(3t) + c_2 e^{-t}\sin(3t).$$

◆

The three cases cover all possibilities for homogeneous second-order linear differential equations with constant coefficients. Solutions to nonhomogeneous linear DEs with constant coefficients will be addressed in the next section.

Summary

A homogeneous second-order linear DE with constant coefficients has the form $ay'' + by' + cy = 0$ where a, b, and c are constants (and a is not zero). Its general solution is given in the following table. The three cases depend on whether $b^2 - 4ac$ is positive, zero, or negative.

Condition	Classification	Roots	General Solution
$b^2 - 4ac > 0$	distinct real roots	r_1, r_2	$c_1 e^{r_1 t} + c_2 e^{r_2 t}$
$b^2 - 4ac = 0$	repeated real root	$r_1 = -\dfrac{b}{2a}$	$c_1 e^{r_1 t} + c_2 t e^{r_1 t}$
$b^2 - 4ac < 0$	complex conjugate roots	$r_1 = p \pm qi$	$c_1 e^{pt}\cos(qt) + C_2 e^{pt}\sin(qt)$

Table 13.3.1

EXERCISES for Section 13.3

1. Which second-order linear DEs $ay'' + by' + cy = f(t)$ have the zero function $y(t) = 0$ as a solution?

2. Verify by substitution that $y_1 = e^{-t}\cos(3t)$ and $y_2 = e^{-t}\sin(3t)$ are solutions to the differential equation in Example 3.

3. The differential equation $y'' = 0$ was first encountered in Section 3.7 (Exercise 25) where it is found that the general solution is $y(t) = a + bt$ for constants a and b. Explain how this is consistent with Case 2 (see page 709).

In Exercises 4 to 9 find the general solution of the given differential equation.

4. $y'' + 5y' + 6y = 0$

5. $y'' - y' - 6y = 0$

6. $y'' + 9y = 0$

7. $y'' - 4y' + 4y = 0$

8. $y'' - 2y' + 5y = 0$

9. $y'' + 10y' + 25y = 0$

In Exercises 10 to 15 solve the initial value problem. That is, find the solution of the differential equation that satisfies the given initial conditions. The differential equations are the same as in Exercises 4 to 9.

10. $y'' + 5y' + 6y = 0$, $y(0) = 0$, $y'(0) = 2$

11. $y'' - y' - 6y = 0$, $y(0) = 1$, $y'(0) = 2$

12. $y'' + 9y = 0$, $y(0) = 1$, $y'(0) = 3$

13. $y'' - 4y' + 4y = 0$, $y(0) = 0$, $y'(0) = -1$

14. $y'' - 2y' + 5y = 0$, $y(0) = 0$, $y'(0) = 0$

15. $y'' + 10y' + 25y = 0$, $y(0) = 4$, $y'(0) = 0$

---

16. Suppose $ay'' + by' + cy = 0$ with $b^2 - 4ac = 0$.
(a) Verify that $y = e^{-bt/(2a)}$ is a solution of $ay'' + by' + cy = 0$.
(b) Verify that $y = te^{-bt/(2a)}$ is a solution of $ay'' + by' + cy = 0$.

17. Suppose $ay'' + by' + cy = 0$ with $b^2 - 4ac < 0$. Let $p = \dfrac{-b}{2a}$ and $q = \dfrac{\sqrt{4ac - b^2}}{2a}$.
(a) Verify that $y = e^{pt}\cos(qt)$ is a solution of $ay'' + by' + cy = 0$.
(b) Verify that $y = e^{pt}\sin(qt)$ is a solution of $ay'' + by' + cy = 0$.

18. Let k be a positive constant.
(a) Find the general solution of $y'' + k^2 y = 0$. (b) Check that your answer in (a) satisfies $y'' + k^2 y = 0$.

19. Let k be a positive constant.
(a) Find the general solution of $y'' - k^2 y = 0$. (b) Check that the solution found in (a) satisfy $y'' - k^2 y = 0$.

20. Let k be a positive constant.
(a) Find the general solution of $y'' + ky' = 0$. (b) Check that the solution found in (a) satisfies $y'' + ky' = 0$.

21. When $b^2 - 4ac = 0$ (and $a \neq 0$) we found that one solution of $ay'' + by' + cy = 0$ is $y_1(t) = Ae^{-bt/(2a)}$ and claimed that a second solution is $y_2(t) = te^{-bt/(2a)}$. We show how this second solution was obtained.
(a) Verify that $y_1(t) = Ae^{-bt/(2a)}$ is a solution of $ay'' + by' + cy = 0$ for any constant A.
(b) The **reduction of order** method suggests looking for a second solution of $ay'' + by' + cy = 0$ in the form $y(t) = A(t)e^{-bt/(2a)}$. Find the second-order differential equation for $A(t)$ that makes $y_2(t) = A(t)e^{-bt/(2a)}$ a solution of $ay'' + by' + cy = 0$.
(c) Solve the differential equation found in (b).
(d) What is the resulting solution y_2?

22. **SAM:** In Example 2 the authors say that the general solution is $y = c_1 e^{-3t} + c_2 te^{-3t}$.
JANE: What's your point?
SAM: They missed the obvious solution $y = 0$. I am going to send them an e-mail.
Write the authors' response.

13.4 Second-Order Linear Differential Equations: Nonhomogeneous Case

We consider the nonhomogeneous second-order linear differential equation with constant coefficients

$$ay'' + by' + cy = f(t), \tag{13.4.1}$$

with the goal of developing a method to find all solutions for many of the most common forcing functions $f(t)$.

In Section 13.3 we learned how to find the general solution of a homogeneous second-order linear differential equations

$$ay'' + by' + cy = 0. \tag{13.4.2}$$

Now, we see a need to extend this to nonhomogeneous second-order linear differential equations and learn a method known as the **method of undetermined coefficients**. This method is also called the **method of intelligent guessing**. But, first, we need to learn a little more about the principle of superposition.

The Principle of Superposition, Revisited

Our current knowledge about superposition was restricted to homogeneous linear differential equations, with a comment that these results do not apply when the differential equation is nonhomogeneous. (See Theorem 13.3.1 in Section 13.3.) The following theorem summarizes the principal of superposition for nonhomogeneous linear differential equations.

Theorem 13.4.1: Superposition for Nonhomogeneous Second-Order Linear Differential Equations

Let u be a solution of $ay''+by'+cy=f(t)$ and let v be a solution of $ay''+by'+cy=g(t)$. Then

(a) $y=u+v$ *is a solution of* $ay''+by'+cy=f(t)+g(t)$,

(b) $y=ku$ *is a solution of* $ay''+by'+cy=kf(t)$ *where k is a constant, and*

(c) $y=u-v$ *is a solution of* $ay''+by'+cy=f(t)-g(t)$.

Proof of Theorem 13.4.1

We prove (a). The proofs of (b) and (c) are similar.

The assumptions on u and v tell us that

$$au''+bu'+cu=f(t) \qquad \text{and} \qquad av''+bv'+cv=g(t).$$

Then

$$\begin{aligned} a(u+v)''+b(u+v)'+c(u+v) &= a(u''+v'')+b(u'+v')+c(u+v) \\ &= (au''+bu'+cu)+(av''+bv'+cv) \\ &= f(t)+g(t). \end{aligned}$$

We will use Theorem 13.4.1(c) in the proof of the following theorem.

Theorem 13.4.2: Structure of Solutions of $ay''+by'+cy=f(t)$

The general solution of $ay''+by'+cy=f(t)$ can be written as $y(t)=y_h(t)+y_p(t)$ where $y_h(t)$ is the general solution of the associated homogeneous differential equation, $ay''+by'+cy=0$, and $y_p(t)$ is any solution of the nonhomogeneous differential equation $ay''+by'+cy=f(t)$.

TERMINOLOGY: *The solution y_h is called the* ***homogeneous solution*** *of the nonhomogeneous differential equation $ay''+by'+cy=f(t)$. (Recall that y_h involves two constants of integration.) The solution y_p is called a* ***particular solution*** *of the nonhomogeneous differential equation $ay''+by'+cy=f(t)$.*

Proof of Theorem 13.4.2

Let both y_p and y_q be solutions of (13.4.1). That is,

$$ay_p''+by_p'+cy_p=f(t) \qquad \text{and} \qquad ay_q''+by_q'+cy_q=f(t).$$

By part (c) of Theorem 13.4.1, $y=y_q-y_p$ is a solution of

$$ay''+by'+cy=f(t)-f(t)=0.$$

Therefore, y_q-y_p must be a solution of the homogeneous differential equation.

Since every solution of the homogeneous equation (13.4.2) can be written with some choice of the two arbitrary constants in the homogeneous solution, this means $y_q(t)-y_p(t)=y_h(t)$, or $y_q(t)=y_p(t)+y_h(t)$. That is, any particular solution can be written as the sum of one particular solution and a homogeneous solution.

Observation 13.4.3: *Usage of Theorem 13.4.2*

Theorem 13.4.2 tells us that if we know one solution of (13.4.1) and all solutions of the associated homogeneous equation (13.4.2), we can construct all solutions of (13.4.1).

Section 13.3 described ways to find all solutions of the homogeneous equation (13.4.2). In this section we describe the Method of Undetermined Coefficients as a way to find a particular solution of (13.4.1).

Intelligent Guessing: The Method of Undetermined Coefficients

A few examples illustrate the method of undetermined coefficients.

EXAMPLE 1. Find a particular solution of $y'' - 3y' + 2y = 3e^{-2t}$.

SOLUTION Because the derivatives of e^{-2t} have the form "constant times e^{-2t}" we try $y = Ae^{-2t}$, hoping that some value of the constant A will provide a solution. Substituting $y' = -2Ae^{-2t}$ and $y'' = 4Ae^{-2t}$ into the differential equation gives

$$y'' - 3y' + 2y = 4Ae^{-2t} - 3\left(-2Ae^{-2t}\right) + 2\left(Ae^{-2t}\right) = (4 - 3(-2) + 2)Ae^{-2t} = 12Ae^{3t}.$$

The guess $y = Ae^{-2t}$ will be a solution of the differential equation if we can find a value of A such that $12Ae^{-2t} = 3e^{-2t}$. Because e^{-2t} is never 0 this reduces to

$$12A = 3,$$

so

$$A = \frac{1}{4}.$$

This process produces $y_p = (1/4)e^{3t}$ as a particular solution of $y' - 3y' + 2y = 5e^{3t}$. ◆

The next example replaces the exponential on the right-hand side of the differential equation with a quadratic power and a constant term.

EXAMPLE 2. Find a particular solution of $y'' - 3y' + 2y = 6t^2 - 4$.

SOLUTION Based on our success in Example 1 where we just changed the coefficient on the right-hand side, maybe we can find a particular solution in the form $y = At^2 + B$.

If $y = At^2 + B$ is going to be a particular solution of $y'' - 3y' + 2y = 6t^2 - 4$, then, because $y' = 2At$ and $y'' = 2A$:

$$y'' - 3y' + 2y = (2A) - 3(2At) + 2(At^2 + B) = 2At^2 - 6At + (2A + 2B).$$

must be equal to the right-hand side of the differential equation, $6t^2 - 4$. For these two quadratic polynomials to be equal, their constant, linear, and quadratic coefficients must match: $2A = 6$, $-6A = 0$, and $2A + 2B = -4$. But, this set of three equations in the two unknowns A and B does not have a solution.

The original guess was not appropriate. While it has all of the terms in the right-hand side, it did not have all of the terms that appeared in the derivatives of the guess. Maybe it would be better to choose a general quadratic polynomial as the form of particular solution. To test this idea, choose $y = At^2 + Bt + C$. When, as before, the first and second derivatives of the guess, $y' = 2At + B$ and $y'' = 2A$, are inserted into the differential equation, we find

$$y'' - 3y' + 2y = (2A) - 3(2At + B) + 2(At^2 + Bt + C) = 2At^2 + (-6A + 2B)t + (2A - 3B + 2C).$$

Now, matching the coefficients of like powers of t, we find: $2A = 6$, $-6A + 2B = 0$, and $2A - 3B + 2C = -4$. The search for a solution of this system of three equations for the three coefficients, A, B, and C, starts with finding $A = 3$, then $2B = 6A = 18$ so $B = 9$, and $2C = -4 - 2A + 3B = 17$ so $C = 17/2$.

A particular solution of $y'' - 3y' + 2y = 6t^2 - 4$ is $y_p(t) = 3t^2 + 9t + 17/2$. ◆

Observation 13.4.4: ***Lesson Learned from Example 2***

When the right-hand side of a nonhomogeneous linear differential equation involves powers of t, the guess needs to be a general polynomial with degree equal to the highest power on the right-hand side of the differential equation.

EXAMPLE 3. Find a particular solution of $y'' - 3y' + 2y = 8e^{-2t} + 3t^2 - 2$.

SOLUTION The polynomial part of this right-hand side, $3t^2-2$, is exactly half of the right-hand side in Example 2, $6t^2-4$. This means, by Theorem 13.4.1(b), a particular solution for $3t^2-2$ is one-half of the particular solution found in Example 2: $y_{p_1}(t) = (3t^2+9t+17/2)/2$.

Moreover, the first term is a multiple of the right-hand side in Example 1: $8e^{-2t} = (8/3)3e^{-2t}$. As a result, a particular solution for $8e^{-2t}$ is the particular solution found in Example 2 multiplied by $8/3$: $y_{p_2}(t) = (8/3)(1/4)e^{-2t}$.

Now, by Theorem 13.4.1(a), a particular solution of

$$y'' - 3y' + 2y = 8e^{-2t} + 3t^2 - 2 = \frac{8}{3}\left(3e^{-2t}\right) + \frac{1}{2}\left(6t^2-4\right)$$

is

$$\begin{aligned} y_p(t) &= y_{p_1}(t) + y_{p_2}(t) \\ &= \frac{8}{3}\left(\frac{1}{4}e^{-2t}\right) + \frac{1}{2}\left(3t^2+9t+\frac{17}{2}\right) \\ &= \frac{2}{3}e^{-2t} + \frac{3}{2}t^2 + \frac{9}{2}t + \frac{17}{4}. \end{aligned}$$

♦

Now that we have some experience with exponentials and polynomials, the next example considers a right-hand side involving a trigonometric term.

EXAMPLE 4. Find a particular solution of $y'' - 3y' + 2y = 40\cos(2t)$.

SOLUTION An obvious first guess for a particular solution might be $y = A\cos(2t)$. When this guess, along with $y' = -2A\sin(2t)$ and $y'' = -4A\cos(2t)$, are substituted into the left-hand side of the differential equation we find:

$$y'' - 3y' + 2y = -4A\cos(2t) + 6A\sin(2t) + 2A\cos(2t) = -2A\cos(2t) + 6A\sin(2t).$$

For this expression to match the right-hand side of the differential equation, $40\cos(2t)$, would require both $-2A = 40$ and $6A = 0$. This system of two equations for the single unknown, A, does not have a solution. There is no solution of this differential equation in the form $y = A\cos(2t)$.

Observation 13.4.5: *Lesson Learned from a Bad Guess*

As we saw in Example 2, there is a problem when the expression found when the guess is inserted into the differential equation has more terms than the number of unknown coefficients in the guess.

In this case, the fact that guessing $y = A\cos(2t)$ produces an expression involving both $\cos(2t)$ and $\sin(2t)$ suggests modifying the form of a particular solution to be $y = A\cos(2t) + B\sin(2t)$. With this updated guess, the first and second derivatives are $y' = -2A\sin(2t) + 2B\cos(2t)$ and $y'' = -4A\cos(2t) - 4B\sin(2t)$, so

$$\begin{aligned} y'' - 3y' + 2y &= (-4A\cos(2t) - 4B\sin(2t)) - 3(-2A\sin(2t) + 2B\cos(2t)) + 2(A\cos(2t) + B\sin(2t)) \\ &= (-2A - 6B)\cos(2t) + (-2B + 6A))\sin(2t). \end{aligned}$$

This simplifies to the right-hand side of the differential equation when both $-2A - 6B = 40$ and $-2B + 6A = 0$.

It is reassuring to find a system of two equations for two unknowns. In fact, a little algebra allows us to conclude $A = -2$ and $B = -6$. Thus, a particular solution of the nonhomogeneous differential equation is

$$y_p(t) = -2\cos(2t) - 6\sin(2t).$$

♦

Observation 13.4.6: *Lesson Learned from Example 3*

When the right-hand side of a linear differential equation involves a trigonometric term, $\sin(kt)$ or $\cos(kt)$, the form for a particular solution needs to include both $\sin(kt)$ and $\cos(kt)$.

You should be starting to become comfortable with the general process of applying the method of intelligent guessing, and with coming up with an appropriate form for a particular solution of a nonhomogeneous second-order linear differential equation with constant coefficients: $ay''+by'+cy=f(t)$. Do not let the next example shake your confidence. It looks harmless enough, being the same as Example 1 except for the sign of the coefficient of the exponent in the exponential function.

EXAMPLE 5. Find a particular solution of $y''-3y'+2y=3e^{2t}$.

SOLUTION A reasonable guess is $y=Ae^{2t}$. Substitution into the left-hand side of the differential equation leads to

$$\begin{aligned} y''-3y'+2y &= 4Ae^{2t}-3\left(2Ae^{2t}\right)+2\left(Ae^{2t}\right) \\ &= (4A-6A+2A)e^{2t} \\ &= 0. \end{aligned}$$

Obviously, there is no choice of A that makes this expression equal to $3e^{2t}$. In fact, this guess is a solution of the corresponding homogeneous differential equation: $y''-3y'+2y=0$.

Observation 13.4.7: *Lessons Learned from Another Bad Guess*

1. The form for the particular solution of a nonhomogeneous linear differential equation should never include any terms that are solutions of the corresponding homogeneous differential equation.
2. The only way to ensure that the form of a particular solution does not include any solutions of the corresponding homogeneous differential equation is to find the homogeneous solution prior to looking for a particular solution. We were fortunate that this did not arise in Examples 1 to 4.

The characteristic equation for this differential equation is $r^2-3r+2=0$. Because $r^2-3r+2=(r-2)(r-1)$, the characteristic equation has two distinct real roots, and so the homogeneous solution is $y_h(t)=c_1e^t+c_2e^{2t}$ for any constants c_1 and c_2.

Recall how a second solution is found when the characteristic equation has a repeated real root: multiply the first solution by t. Applying this same thinking here, look for a particular solution in the form $y=Ate^{2t}$. The first two derivatives are $y'=Ae^{2t}+2Ate^{2t}$ and $y''=4Ae^{2t}+4Ate^{2t}$. Then the left-hand side of the differential equation

$$\begin{aligned} y''-3y'+2y &= \left(4Ae^{2t}+4Ate^{2t}\right)-3\left(Ae^{2t}+2Ate^{2t}\right)+2\left(Ate^{2t}\right) \\ &= Ae^{2t} \end{aligned}$$

is equal to the right-hand side of the differential equation when $A=3$. This worked! A particular solution of $y''-3y'+2y=3e^{2t}$ is $y_p(t)=3te^{2t}$ and the general solution is $y(t)=y_h(t)+y_p(t)=c_1e^t+c_2e^{2t}+3te^{2t}$. ♦

Observation 13.4.8: *Lesson Learned from Example 5*

When computing the first and second derivatives of the updated guess, it is concerning that this created new terms, as this was what caused problems with the initial guesses in Examples 2 and 4. The difference here is that the extra terms ultimately cancel because of the homogeneous solution.

The next example involves complex conjugate roots of the characteristic equation and multiple trigonometric functions in the right-hand side of the differential equation. Then, the last example has a characteristic equation with repeated real roots and a right-hand side that is also a solution of the corresponding homogeneous differential equation.

EXAMPLE 6. Find the solution of the initial value problem $y'' + 4y = \sin(t) + \cos(2t)$, $y(0) = 0$, $y'(0) = -1$.

SOLUTION Our first step is to find the homogeneous solution. The characteristic equation of the corresponding homogeneous differential equation is $r^2 + 4 = 0$, which has complex-conjugate roots: $r = \pm 2i$. The homogeneous solution is $y_h(t) = c_1 \cos(2t) + c_2 \sin(2t)$.

The right-hand side has one term with sin and one with cos, but they are with different arguments, so are fundamentally different terms. We treat them separately, and then add the two results to find a particular solution of $y'' + 4y = \sin(t) + \cos(2t)$.

For $y'' + 4y = \sin(t)$, an appropriate initial guess is $y = A\cos(t) + B\sin(t)$, which does not contain any terms in the homogeneous solution. Then $y' = -A\sin(t) + B\cos(t)$ and $y'' = -A\cos(t) - B\sin(t)$. Substituting these into the left-hand side of the differential equation yields:

$$\begin{aligned} y'' + 4y &= (-A\cos(t) - B\sin(t)) + 4\,(A\cos(t) + B\sin(t)) \\ &= 3A\cos(t) + 3B\sin(t). \end{aligned}$$

For this to reduce to $\sin(t)$ requires both $3A = 0$ and $3B = 1$, so $A = 0$ and $B = 1/3$. A particular solution with the right-hand side of $\sin(t)$ is $y_{p_1}(t) = \sin(t)/3$.

For the other term in the right-hand side, $\cos(2t)$, the initial guess is $y = C\cos(2t) + D\sin(2t)$. But, both of these terms are in the homogeneous solution, so this guess needs to be multiplied by t. Our updated guess is $y = Ct\cos(2t) + Dt\sin(2t)$. As neither of these terms are in the homogeneous solution, we are confident this will work. Then

$$y' = (C + 2Dt)\cos(2t) + (-2Ct + D)\sin(2t)$$

and

$$\begin{aligned} y'' &= (2D + 2(-2Ct + D))\cos(2t) + (-2(C + 2Dt) - 2C)\sin(2t) \\ &= (-4Ct + 4D)\cos(2t) + (-4C - 4Dt)\sin(2t). \end{aligned}$$

The left-hand side of the differential equation

$$\begin{aligned} y'' + 4y &= ((-4Ct + 4D)\cos(2t) + (-4C - 4Dt)\sin(2t)) + 4\,(Ct\cos(2t) + Dt\sin(2t)) \\ &= 4D\cos(2t) - 4C\sin(2t) \end{aligned}$$

matches $\cos(2t)$ when $4D = 1$ and $-4C = 0$. The only solution is $C = 0$ and $D = 1/4$ and a particular solution of $y'' + 4y = \cos(2t)$ is $y_{p_2}(t) = t\sin(2t)/4$.

The general solution of $y'' + 4y = \sin(t) + \cos(2t)$ is

$$y(t) = y_h(t) + y_{p_1}(t) + y_{p_2}(t) = c_1 \cos(2t) + c_2 \sin(2t) + \frac{1}{3}\sin(t) + \frac{t}{4}\sin(2t).$$

To satisfy the two initial conditions, first evaluate the general solution for $t = 0$: $y(0) = c_1 = 0$, so $c_1 = 0$. Next, compute the derivative of the general solution: $y'(t) = -2c_1 \sin(2t) + 2c_2 \cos(2t) + \cos(t)/3 + \sin(2t)/4 + t\cos(2t)/2$. When evaluated at $t = 0$, we find $y'(0) = 2c_2 + 1/3$. So $y'(0) = -1$ is satisfied when $2c_2 + 1/3 = -1$, or $c_2 = -2/3$. The solution of the initial value problem is

$$y(t) = -\frac{2}{3}\sin(2t) + \frac{1}{3}\sin(t) + \frac{t}{4}\sin(2t).$$

◆

EXAMPLE 7. Find the general solution of $y'' + 6y' + 9y = 9t + 18 + te^{-3t} - e^{-3t}$.

SOLUTION Remembering the lessons learned in Example 5, the first task is to find the homogeneous solution. The characteristic polynomial is $r^2 + 6r + 9 = (r+3)^2$ so the characteristic equation has $r = -3$ as a repeated real root. The homogeneous solution is $y_h(t) = c_1 e^{-3t} + c_2 te^{-3t}$.

The right-hand side has four terms, but the first two terms are powers of t, which can be considered together, and the other two are e^{-3t} multiplied by a power of t, which are also similar enough to be considered together. The plan is first to find a particular solution y_{p_1} with the first two terms: $y'' + 6' + 9y = 9t + 18$ and then to find a particular solution y_{p_2} with the two exponential terms: $y'' + 6y' + 9y = te^{-3t} - e^{-3t} = (t-1)e^{-3t}$.

To find y_{p_1}, when the right-hand side is $9t + 18$, an initial guess is also a linear function: $y = At + B$. As neither of these terms are a homogeneous solution, there is no reason to update it. Then $y' = A$ and $y'' = 0$ so

$$\begin{aligned} y'' + 6y' + 9y &= 0 + 6(A) + 9(At + B) \\ &= 9At + (6A + 9B) \end{aligned}$$

will match the left-hand side of $9t + 18$ when $9A = 9$ and $6A + 9B = 18$. The solution of these equations is $A = 1$ and $B = 4/3$ so $y_{p_1}(t) = t + 4/3$.

To find y_{p_2}, with a right-hand side of $(t-1)e^{-3t}$, an initial guess is $y = (Ct + D)e^{-3t}$. Both terms in this form are homogeneous solutions, so multiply by t: $y = (Ct^2 + Dt)e^{-3t}$. One of the terms in this guess is still a homogeneous solution, so multiply by t: $y = (Ct^3 + Dt^2)e^{-3t}$. Now, neither term is a homogeneous solution. Hopefully, we will find a system of two equations for C and D.

The first two derivatives are

$$\begin{aligned} y' &= (3Ct^2 + 2Dt)e^{-3t} - 3(Ct^3 + Dt^2)e^{-3t} \\ &= (-3Ct^3 + (3C - 3D)t^2 + 2Dt)e^{-3t} \end{aligned}$$

and

$$\begin{aligned} y'' &= (-9Ct^2 + 2(3C - 3D)t + 2D)e^{-3t} - 3(-3Ct^3 + (3C - 3D)t^2 + 2Dt)e^{-3t} \\ &= (9Ct^3 + (-18C + 9D)t^2 + (6C - 12D)t + 2D)e^{-3t} \end{aligned}$$

Then the left-hand side of the differential equation

$$\begin{aligned} y'' + 6y' + 9y = &\left((9Ct^3 + (-18C + 9D)t^2 + (6C - 12D)t + 2D)e^{-3t}\right) \\ &+ 6\left((-3Ct^3 + (3C - 3D)t^2 + 2Dt)e^{-3t}\right) \\ &+ 9\left((Ct^3 + Dt^2)e^{-3t}\right) \\ =&(6Ct + 2D)e^{-3t} \end{aligned}$$

is equal to the right-hand side, $(t-1)e^{-3t}$, when $6C = 1$ and $2D = -1$. So, $C = 1/6$ and $D = -1/2$ and this particular solution is $y_{p_2}(t) = (t^3/6 - t^2/2)e^{-3t}$.

A particular solution of $y'' + 6y' + 9y = 9t + 18 + te^{-3t} + e^{-3t}$ is

$$\begin{aligned} y_p(t) &= y_{p_1}(t) + y_{p_2}(t) \\ &= t + \frac{4}{3} + \left(\frac{t^3}{6} - \frac{t^2}{2}\right)e^{-3t} \end{aligned}$$

and the general solution is

$$\begin{aligned} y(t) &= y_h(t) + y_p(t) \\ &= c_1 e^{-3t} + c_2 te^{-3t} + t + \frac{4}{3} + \left(\frac{1}{6}t^3 - \frac{1}{2}t^2\right)e^{-3t} \\ &= \left(c_1 + c_2 t - \frac{1}{2}t^2 + \frac{1}{6}t^3\right)e^{-3t} + t + \frac{4}{3} \end{aligned}$$

◆

Summary

The lessons learned from the examples in this section are summarized in the following general procedure for finding the general solution of a nonhomogeneous second-order linear differential equation with constant coefficients.

Algorithm: Solution Process for Nonhomogeneous Second-Order Linear Differential Equations with Constant Coefficients

To solve a nonhomogeneous second-order linear differential equation with constant coefficients,

$$ay'' + by' + cy = f(t), \tag{13.4.3}$$

first solve the associated homogeneous equation by the methods of Section 13.3. Then find a particular solution y_p of (13.4.3) by intelligent guessing, guided by the form of $f(t)$.

The general solution of (13.4.3) has the form $y = y_h + y_p$, where y_h is the associated homogeneous equation, with arbitrary constants c_1 and c_2 are arbitrary constants, and y_p solves the nonhomogeneous equation.

The initial guesses for the particular solution are given in Table 13.4.1. Every solution of the associated homogeneous equation contributes nothing towards matching the nonhomogeneous term of the original differential equation. As a result, when the initial guess includes terms that are in the homogeneous solution, the initial guesses must be multiplied by the power of t that produces a guess that does not have any terms that solve the associated homogeneous equation.

description	right-hand side: $f(t)$	initial guess for y_p
constant	A	B
exponential	Ae^{kt}	Be^{kt}
trigonometric	$A\cos(kt)$ or $A\sin(kt)$	$B_1\cos(kt) + B_2\sin(kt)$
polynomial (degree n)	$P(t) = A_0 + A_1t + A_2t^2 + \cdots + A_nt^n$	$B_0 + B_1t + B_2t^2 + \cdots + B_nt^n$
exponential × trigonometric	$e^{kt}\cos(kt)$ or $e^{kt}\sin(kt)$	$e^{kt}(B_1\cos(kt) + B_2\sin(kt))$
polynomial × exponential	$P(t)e^{kt}$	$(B_0 + B_1t + B_2t^2 + \cdots + B_nt^n)e^{kt}$
polynomial × trigonometric	$P(t)\cos(kt)$ or $P(t)\sin(kt)$	$(B_0 + B_1t + B_2t^2 + \cdots + B_nt^n)\cos(kt)$ $+ (C_0 + C_1t + C_2t^2 + \cdots + C_nt^n)\sin(kt))$
polynomial × exponential × trigonometric	$P(t)e^{kt}\cos(kt)$ or $P(t)e^{kt}\sin(kt)$	$(B_0 + B_1t + B_2t^2 + \cdots + B_nt^n)e^{kt}\cos(kt)$ $+ (C_0 + C_1t + C_2t^2 + \cdots + C_nt^n)e^{kt}\sin(kt))$

Table 13.4.1

EXERCISES for Section 13.4

In Exercises 1 to 9, what is the form of the initial guess for y_p if $ay'' + by' + cy = f(t)$? Assume that no term in $f(t)$ is a solution of the associated homogeneous equation and that its characteristic polynomial has no repeated roots.

1. $f(t) = 3\sin(2t)$
2. $f(t) = 4\cos(5t)$
3. $f(t) = 2e^{-2t}$
4. $f(t) = 3e^{5t}\sin(4t)$
5. $f(t) = 3t\sin(2t) + 5\cos(3t)$
6. $f(t) = t^2 + 5$
7. $f(t) = 7t + \cos(t)$
8. $f(t) = e^{7t} + \sin(3t)$
9. $f(t) = t\sin(3t)e^{7t}$

---

10. Determine an appropriate guess for a particular solution of $y'' - 8y' + 15y = f(t)$. Do not find the actual particular solution.

(a) $f(t) = e^{2t}$
(b) $f(t) = 5\sin(3t)$
(c) $f(t) = e^{3t}$
(d) $f(t) = e^{-5t}$
(e) $f(t) = 2\sin(3t) + 4\cos(3t)$
(f) $f(t) = e^{2t} + 6\sin(3t) - 2e^{5t}$

11. Determine an appropriate guess for a particular solution of $y'' + 10y' + 25y = f(t)$. Do not find the actual particular solution.

(a) $f(t) = e^{2t}$
(b) $f(t) = 5\sin(3t)$
(c) $f(t) = e^{3t}$
(d) $f(t) = e^{-5t}$
(e) $f(t) = 2\sin(3t) + 4\cos(3t)$
(f) $f(t) = e^{2t} + 6\sin(3t) - 2e^{5t}$

12. Determine an appropriate guess for a particular solution of $y'' + 25y = f(t)$. Do not find the actual particular solution.

(a) $f(t) = \sin(4t)$
(b) $f(t) = t\cos(4t)$
(c) $f(t) = \sin(5t)$
(d) $f(t) = t\cos(5t)$
(e) $f(t) = e^{-5t} - 4te^{5t}$
(f) $f(t) = t\cos(4t) + 2t\cos(5t)$

13. Determine an appropriate guess for a particular solution of $y'' + y' - 2y = f(t)$. Do not find the actual particular solution.

(a) $f(t) = 3t + 1$
(b) $f(t) = e^{3t}$
(c) $f(t) = e^{-2t}$
(d) $f(t) = e^{t} + e^{2t}$
(e) $f(t) = e^{t}\sin(t)$
(f) $f(t) = -t^2e^{t}\sin(t)$

In Exercises 14 to 21 find a particular solution of the given nonhomogeneous differential equation.

14. $y'' + y' - 6y = e^{t}$

15. $y'' + y' - 6y = e^{2t}$

16. $y'' + 25y = 3\cos(4t)$

17. $y'' + 25y = 20\cos(5t)$

18. $y'' + y' - 12y = e^{3t}$

19. $y'' - 2y' + y = e^{t}$

20. $y'' - 2y' + 3y = 2t^2 + 12$

21. $3y'' + 2y' + y = e^{2t}(56\cos(t) - 28\sin(t))$

---

22. Find all solutions of $y'' - 2y' - 3 = f(t)$ for the given forcing function $f(t)$.
(a) $f(t) = e^{2t}$, (b) $f(t) = e^{t}$, and (c) $f(t) = e^{-3t}$.

23. Find all solutions of $y'' + 6y' + 9y = f(t)$ for the given forcing function $f(t)$.
(a) $f(t) = e^{t}$, (b) $f(t) = e^{3t}$, and (c) $f(t) = e^{-3t}$.

24. Find all solutions of $y'' + 9y = f(t)$ for the given forcing function $f(t)$.
(a) $f(t) = \cos(4t)$, (b) $f(t) = \sin(4t)$, and (c) $f(t) = e^{t}\sin(4t)$.

25. Find all solutions of $y'' - 4y' + 4y = f(t)$ for the given forcing function $f(t)$.
(a) $f(t) = 8$, (b) $f(t) = 4t + 4$, and (c) $f(t) = -4t^2 + 16t - 6$.

---

26. Prove part (b) of Theorem 13.4.1

27. Prove part (c) of Theorem 13.4.1

---

28. Which equations of the form $ay'' + by' + cy = f(t)$ have distinct solutions u and v such that $u - v$ is also a solution of the same differential equation?

Exercise 43 in Section 13.1 asked you to verify a solution to the fourth-order beam equation $y^{(4)} - y'' = -W$ (W a constant). Exercises 29 and 30 present two ways to find this solution.

29. The first approach takes advantage of the fact that the beam equation involves only $y^{(4)}$ and y''.

(a) Show that the substitution $u = y''$ reduces the fourth-order differential equation to the second-order differential equation $u'' - u = -W$.

(b) Find its general solution.

(c) Solve the second-order differential equation $y'' = u$, where u is a solution found in (b).

30. The second approach to solving the beam equation extends the ideas of Section 13.3 to higher-order homogeneous differential equations and then uses the ideas of this section to find a particular solution of the nonhomogeneous differential equation.

(a) Find all r that make $y = e^{rt}$ a solution to $y^{(4)} - y'' = 0$.

(b) Use the information in (a) and the ideas in Section 13.3 to find the general solution of $y^{(4)} - y'' = 0$.

(c) Use the method of intelligent guessing to find a particular solution of $y^{(4)} - y'' = -W$.

---

31. Suppose $y'' - 4y' + qy = 4e^{3t}$ where q is a constant.

(a) For which value(s) of q is Ae^{3t} not an intelligent guess for a nonhomogeneous solution?

(b) For the value(s) of q found in (a), what is a valid guess for a nonhomogeneous solution?

32. Suppose $y'' + py' + qy = 7e^{3t}$ where p and q are constants.

(a) For which value(s) of p and q do neither $y_p = Ae^{3t}$ nor $y_p = Ate^{3t}$ provide a solution of this nonhomogeneous differential equation?

(b) For which values of p and q does $y_p = At^2e^{3t}$ provide a solution of this nonhomogeneous differential equation?

(c) For the values of p and q found in (b), why does it not make sense to guess $y_p = (At^2 + Bt + C)e^{3t}$ as the form of a particular solution of this nonhomogeneous differential equation?

13.5 Euler's Method

We have looked only at solving differential equations of special forms: separable (Section 13.1), first-order linear (Section 13.2), and second-order linear with constant coefficients (Sections 13.3 and 13.4). A full course on differential equations will contain ways to solve additional types of differential equations, as well as ways to analyze differential equations without finding a formula for the solutions. One lesson learned in such a course is that most differential equations do not have solutions that can be expressed in terms of elementary functions, even seemingly simple equations such as

$$\frac{dy}{dt} = e^{-t^2} \qquad \text{or} \qquad \frac{dy}{dt} = \sqrt{1+y^3}.$$

Other differential equations that appear no less complicated can be solved exactly, but only after much work, for instance,

$$\frac{dy}{dt} = \sin(y) \qquad \text{and} \qquad \frac{dy}{dt} = 1 + e^y.$$

FLASHBACK: This should remind you of what we found when learning about elementary functions in Section 6.4. Seemingly simple functions like $\sin\left(x^2\right)$, e^{-x^2}, and $x^2e^{-x^2}$ have antiderivatives that are not elementary functions, while the antiderivatives of $\sin^2(x)$, xe^{-x^2}, and $x^3e^{-x^2}$ are elementary.

Euler's method is a technique for obtaining approximate solutions and involves calculations that are easily implemented on a calculator or computer. The method applies to the general first-order differential equation

$$\frac{dy}{dt} = h(t, y) \tag{13.5.1}$$

with an initial condition

$$y(t_0) = y_0. \tag{13.5.2}$$

The solution is a function $y = f(t)$, whose graph is a curve that we will denote by C. (See Figure 13.5.1(a).)

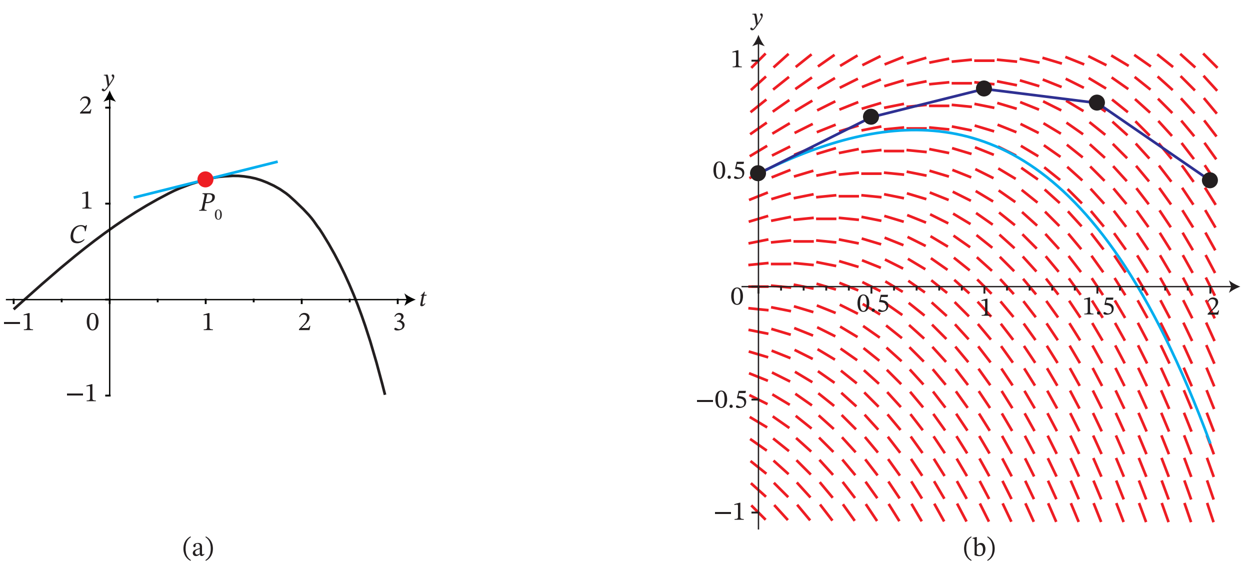


Figure 13.5.1

The point $P_0 = (t_0, y_0)$ is on C, and at P_0 the slope of C is $m_0 = y'(t_0) = h(t_0, y_0)$. We use the tangent line to C at P_0 to construct an estimate of the solution at a nearby value of the independent variable, say at $t = t_1$. That is, define y_1 by the condition that (t_1, y_1) lies on that tangent line: $y = y_0 + (t - t_0)m_0$.

Thus, the point P_1 has coordinates (t_1, y_1) where $y_1 = y_0 + (t_1 - t_0)m_0$. This is the linear approximation described in Section 5.5. While P_1 will not generally be on C it will be close to C when t_1 is near t_0.

Then we repeat the process, using P_1 instead of P_0. With $m_1 = h(t_1, y_1)$ as an estimate of the slope of C at P_1, we choose t_2 near t_1 and compute $y_2 = y_1 + (t_2 - t_1)m_1$. This process can be repeated to construct points $P_1, P_2, \ldots$ that approximate the curve C. (See Figure 13.5.1(b).)

With each step the estimate generally moves farther from the curve C. This tendency can be controlled by choosing t_i closer to t_{i-1}. However, even if the process is automated, using too many steps may result in a large error due to the accumulation of rounding errors.

EXAMPLE 1. Approximate the solution to $y' = y - t$, $y(0) = \dfrac{1}{2}$ on the interval $0 \le t \le 1$ using four equal steps.

SOLUTION From the differential equation and initial condition we define $h(t, y) = y - t$, $t_0 = 0$, and $y_0 = 1/2$. The right endpoint of the t-interval is $T = 1$ and the instructions are to use $n = 4$ equal steps. This means the step size will be $\Delta t = (T - t_0)/n = 1/4$.

Table 13.5.1 organizes the work involved in Euler's method. The first row is filled in using the initial condition.

i	t_i	$y_i = y_{i-1} + m_{i-1}\Delta t$	$m_i = y_i - t_i$	$P_i = (t_i, y_i)$
0	$t_0 = 0$	$y_0 = \frac{1}{2}$	$m_0 = \frac{1}{2} - 0 = \frac{1}{2}$	$P_0 = (0, 0.5)$
1	$t_1 = \frac{1}{4}$			
2	$t_2 = \frac{1}{2}$			
3	$t_3 = \frac{3}{4}$			
4	$t_4 = 1$			

Table 13.5.1

Each step of Euler's method is completed by filling in the next row in the table. Table 13.5.2 shows the calculations for the four steps of Euler's method.

i	t_i	$y_i = y_{i-1} + m_{i-1}\Delta t$	$m_i = y_i - t_i$	$P_i = (t_i, y_i)$
0	$t_0 = 0$	$y_0 = \frac{1}{2}$	$m_0 = \frac{1}{2} - 0 = \frac{1}{2}$	$P_0 = (0.00, 0.500)$
1	$t_1 = \frac{1}{4}$	$y_1 = \frac{1}{2} + \frac{1}{2}\frac{1}{4} = \frac{5}{8}$	$m_1 = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}$	$P_1 = (0.25, 0.625)$
2	$t_2 = \frac{1}{2}$	$y_2 = \frac{5}{8} + \frac{3}{8}\frac{1}{4} = \frac{23}{32}$	$m_2 = \frac{23}{32} - \frac{1}{2} = \frac{7}{32}$	$P_2 \approx (0.50, 0.719)$
3	$t_3 = \frac{3}{4}$	$y_3 = \frac{23}{32} + \frac{7}{32}\frac{1}{4} = \frac{99}{128}$	$m_3 = \frac{99}{128} - \frac{3}{4} = \frac{3}{128}$	$P_3 \approx (0.75, 0.773)$
4	$t_4 = 1$	$y_4 = \frac{99}{128} + \frac{3}{128}\frac{1}{4} = \frac{399}{512}$	$m_4 = \frac{399}{512} - \frac{4}{4} = \frac{-113}{512}$	$P_4 \approx (1.00, 0.779)$

Table 13.5.2

Euler's method, with four steps of size $\Delta t = 1/4$, provides the following approximate values of the solution: $y(0.25) \approx 0.625$, $y(0.50) \approx 0.719$, $y(0.75) \approx 0.773$, $y(1.00) \approx 0.779$. Note that while these computations are accurate to three decimal places, this does not mean they possess this accuracy as approximations to the exact solution of the initial value problem. ♦

Warning: Reporting of Numerical Results

While numerical results reported in this section are reported to three decimal places, the actual computations are done with more decimal places.

Increasing the number of steps reduces Δt, which should improve the accuracy of Euler's method. Doubling the number of steps in Example 1 reduces Δt to 1/8 and yields Table 13.5.3.

i	t_i	$y_i = y_{i-1} + m_{i-1}\Delta t$	$m_i = h(t_i, y_i)$	$P_i = (t_i, y_i)$
0	$t_0 = 0$	$y_0 = \frac{1}{2}$	$m_0 = \frac{1}{2}$	$P_0 = (0.00, 0.500)$
1	$t_1 = \frac{1}{8}$	$y_1 = 0.5625$	$m_1 = 0.4375$	$P_1 \approx (0.125, 0.563)$
2	$t_2 = \frac{1}{4}$	$y_2 \approx 0.617$	$m_2 \approx 0.367$	$P_2 \approx (0.250, 0.617)$
3	$t_3 = \frac{3}{8}$	$y_3 \approx 0.663$	$m_3 \approx 0.288$	$P_3 \approx (0.375, 0.663)$
4	$t_4 = \frac{1}{2}$	$y_4 \approx 0.699$	$m_4 \approx 0.199$	$P_4 \approx (0.500, 0.699)$
5	$t_5 = \frac{5}{8}$	$y_5 \approx 0.724$	$m_5 \approx 0.990$	$P_5 \approx (0.625, 0.724)$
6	$t_6 = \frac{3}{4}$	$y_6 \approx 0.736$	$m_6 \approx -0.014$	$P_6 \approx (0.750, 0.736)$
7	$t_7 = \frac{7}{8}$	$y_7 = 0.735$	$m_7 \approx -0.140$	$P_7 \approx (0.875, 0.735)$
8	$t_8 = 1$	$y_8 \approx 0.717$	$m_8 \approx -0.283$	$P_8 \approx (1.000, 0.717)$

Table 13.5.3

With $n = 8$ steps, the approximate value for $y(1)$ is 0.717. Doubling the number of steps again, to $n = 16$, further refines the estimate to $y(1) \approx 0.681$. And, with $n = 32$ steps, the estimate improves to $y(1) \approx 0.642$.

The solution to the first-order linear initial value problem in Example 1 is $y = 1 + t - e^t/2$. The curve C is the graph of $y = 1 + t - e^t/2$. Thus $y(1)$ is $2 - e/2 \approx 0.641$. Table 13.5.4 shows how the estimates of $y(1)$ approach 0.641 and Figure 13.5.2 shows the convergence visually.

n	Δt	Estimate of $y(1)$	Error
4	$\frac{1}{4} = 0.25$	0.779	$0.779 - 0.641 = 0.138$
8	$\frac{1}{8} = 0.125$	0.717	$0.717 - 0.641 = 0.076$
16	$\frac{1}{16} = 0.0625$	0.681	$0.681 - 0.641 = 0.040$
32	$\frac{1}{32} = 0.03125$	0.662	$0.662 - 0.641 = 0.021$

Table 13.5.4

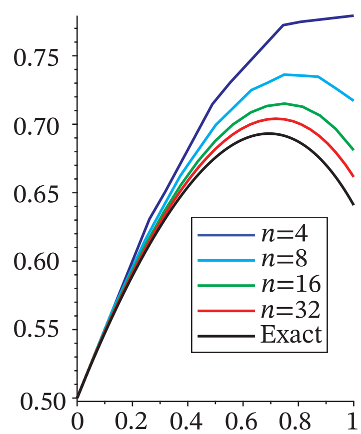


Figure 13.5.2

Table 13.5.4 shows that the error roughly halves each time the number of steps doubles.

Though we have focused on the approximate value of the solution at the right-hand endpoint of the interval under consideration, Euler's method provides estimates of the solution at equally spaced points with spacing Δt.

Thus, Euler's method can also be used to graph an approximate solution to an initial value problem. This is how the piecewise linear curves in Figure 13.5.2 were created. Notice that the curves become smoother as the step size Δt is reduced. Notice also that even when Δt is small, each step of Euler's method introduces more error into the approximate solution.

In practice, it is not unusual to use numerical methods with better convergence rates and better overall control of errors. Courses on differential equations or numerical analysis often introduce other methods for estimating the solution of a first-order initial value problem $y' = h(t, y)$, $y(t_0) = y_0$. Many are extensions and improvements of Euler's method designed to provide more accurate approximations without requiring too many more computations.

Summary

Euler's method for approximating the solution to the first-order initial value problem

$$y' = h(t, y), \qquad y(a) = y_0$$

on the interval $a \le t \le b$ can be summarized in the following algorithm:

Algorithm: Euler's Method for $y' = h(t, y)$, $y(t_0) = y_0$

The following algorithm uses Euler's method to estimate the value of the solution of the initial value problem $y' = h(t, y)$, $y(t_0) = y_0$ at $t = T$ using equal-width n steps. Upon completion, $y(T) \approx y_n$.

Given	$h(t, y)$, t_0, y_0	*(IVP parameters)*
Given	T, n	*(Euler's method parameters)*
Compute	$\Delta t = \dfrac{T - t_0}{n}$	*(step size)*
Compute	*for i = 1, 2, 3, ..., n do*	
	$m_i = h(t_{i-1}, y_{i-1})$	
	$t_i = t_{i-1} + \Delta t$	
	$y_i = y_{i-1} + m_i \Delta t$	
	end do	

EXERCISES for Section 13.5

In Exercises 1 to 3 estimate $y(1)$ for the problem considered in Example 1 using Euler's method with the indicated number of steps. Also, estimate the error. No calculator or computer is needed.

1. $n = 1$ ($\Delta t = 1.0$)

2. $n = 2$ ($\Delta t = 0.5$)

3. $n = 3$ ($\Delta t = 0.333$)

In Exercises 4 to 9 estimate $y(1)$ for $y' = \frac{1}{4}y(8 - y)$ with $y(0) = 1$ using Euler's method with the indicated number of steps. Use the fact that $y(1) = 4.10815$ to five decimal digits to estimate the error in each case. Exercises 4 to 7 can be done without a calculator.

4. $n = 1$ ($\Delta t = 1.0$)

5. $n = 2$ ($\Delta t = 0.5$)

6. $n = 3$ ($\Delta t = 0.333$)

7. $n = 4$ ($\Delta t = 0.25$)

8. $n = 8$ ($\Delta t = 0.125$)

9. $n = 16$ ($\Delta t = 0.0625$)

In Exercises 10 to 17, use Euler's method to estimate y at the right-hand endpoint of the given interval. Present your estimates both as a table and as a graph.

10. $y' = 2t - 3y$, $y(0) = 1$, $\Delta t = 0.2$, $0 \le t \le 1$

11. $y' = t + 4y$, $y(0) = \frac{1}{2}$, $\Delta t = 0.2$, $0 \le t \le 1$

12. $y' = 3ty$, $y(1) = 1$, $\Delta t = 0.1$, $1 \le t \le 2$

13. $y' = 2t^2 y$, $y(0) = 2$, $\Delta t = 0.1$, $0 \le t \le 1$

14. $y' = \cos(t) - \sin(t)$, $y(0) = 0$, $\Delta t = 0.1$, $0 \le t \le 1$

15. $y' = \tan(t)\sec(t)$, $y(0) = 0$, $\Delta t = 0.2$, $0 \le t \le 1$

16. $y' = y\ln(t)$, $y(2) = 1$, $\Delta t = 0.2$, $2 \le t \le 4$

17. $y' = e^t - y$, $y(0) = 1$, $\Delta t = 0.1$, $0 \le t \le 1$

---

18. In Example 1, Euler's method with $n = 4$ and $\Delta t = 0.25$ was used to estimate $y(1)$ for the initial value problem $y' = y - t$, $y(0) = \frac{1}{2}$. The estimates with $n = 8$ ($\Delta t = \frac{1}{8}$) and $n = 16$ ($\Delta t = 0.25$) were also given.

(a) Verify the estimate of $y(1)$ for $n = 8$.
(b) Verify the estimate of $y(1)$ for $n = 16$.
(c) Obtain estimates of $y(1)$ for $n = 32$, $n = 64$, and $n = 128$.
(d) Create a table showing each estimate's error, that is, the difference between it and the exact solution.
(e) What pattern emerges in this table?

19. **SAM:** I have a neat trick to save labor when using Euler's method.
JANE: Yes?
SAM: Say that with $n = 4$ I get an estimate E_4 and with $n = 8$ I get an estimate E_8. Then I predict that $2E_8 - E_4$ will be a much better estimate.
JANE: Please give me an example.
SAM: In this section, $E_4 = 0.779$ and $E_8 = 0.717$. So, my estimate $2E_8 - E_4 = 2(0.717) - 0.779$ is 0.655. That's pretty close to the correct value. It's even better than $E_{16} = 0.681$.
JANE: How did you ever get such a smart idea?
SAM: There's a clue in the book.

Sam has hit on a well-known method for improving Euler's estimates. In this method the error is smaller, being proportional to $(\Delta t)^2$ (which is smaller than Δt when $\Delta t < 1$).

Explain Sam's reasoning.

20. Assume $y'' = h(t, y, y')$ with $y(0) = y_0$, and $y'(0) = v_0$. Describe how to modify Euler's method to estimate $y(1)$.

13.S Chapter Summary

A slope field, or direction field, for the first-order differential equation $y' = f(t, y)$ provides a graphical way to visualize solutions of the differential equation. They can be useful to understand some of the qualitative properties of solutions, particularly when the solution is not available or is too complicated to analyze easily.

The chapter presented methods to find the general solution of the following four types of differential equations:

- § 13.1 – Separable differential equations: $y' = f(x)g(y)$
- § 13.2 – First-order linear differential equations: $y' + a(t)y = f(t)$
- § 13.3 – Homogeneous second-order linear equations with constant coefficients: $a_2 y'' + a_1 y' + a_0 y = 0$
- § 13.4 – Nonhomogeneous second-order linear equations with constant coefficients: $a_2 y'' + a_1 y' + a_0 y = f(t)$

Solution methods for several additional classes of differential equations are spread throughout the exercises in this chapter.

These solutions involve the same number of constants as the order of the differential equation. When appropriate initial conditions are specified, the specific values of these constants that satisfy the initial conditions can be found, producing the solution of that initial value problem.

For any differential equation of the form $y' = f(t, y)$, Section 13.5 introduced Euler's method as a way to compute an approximate numerical solution. Typically, reducing the step size used in Euler's method provides better approximations to the solution. As a rule-of-thumb, the error is proportional to the step size, Δt.

EXERCISES for Section 13.S

In Exercises 1 to 8 solve the given differential equation.

1. $y^3 \dfrac{dy}{dt} = (y^4 + 1)\sin(t)$

2. $\dfrac{dy}{dt} = 2t\sec(y)$

3. $(x^2+1)(\tan(y))y' = x$

4. $x^2y' = 1 - x^2 + y^2 - x^2y^2$

5. $y' + 2y = 3te^{-2t}$

6. $y' - 2ty = e^{t^2}$

7. $y'' + 2y' + 5y = e^t\sin(t)$

8. $y'' - 9y = 2t^2e^{3t} + 5$

In Exercises 9 to 14 find the solution of the differential equation that satisfies the initial value condition.

9. $2y\dfrac{dy}{dt} = \dfrac{t}{\sqrt{t^2-16}}$, $y(5) = 2$

10. $\tan(t)\dfrac{dy}{dt} = y$, $y\left(\dfrac{\pi}{2}\right) = \dfrac{\pi}{2}$

11. $y' + 2ty = t$, $y(0) = 2$

12. $(1+t)y' + y = \cos(t)$, $y(0) = 1$

13. $y'' - 2y' + 2y = t + 1$, $y(0) = 3$, $y'(0) = 0$

14. $y'' + 3y' + 2y = e^t$, $y(0) = 0$, $y'(0) = 3$

---

15. In this exercise we explore some of the differences between initial conditions and boundary conditions.

(a) Find the general solution to $y'' + y = 0$.
(b) Show that there is exactly one solution that satisfies the initial conditions $y(t_0) = a$, $y'(t_0) = b$ for any values of t_0, a and b.
(c) Show that there is exactly one solution that satisfies the boundary conditions $y(0) = a$, $y(T) = b$ (for any values of a and b) only when T is not an integer multiple of π.
(d) Show that there is no solution that satisfies the boundary conditions $y(0) = a$, $y(T) = b$ when T is an integer multiple of π and $b \neq a$.
(e) Show that there is an infinite number of solutions that satisfy the boundary conditions $y(0) = a$, $y(T) = b$ when T is an integer multiple of π and $b = a$.

16. Let M be a positive constant. Assume that a population $P(t)$ changes at a rate proportional to $M - P(t)$. That is, there is a positive constant k such that $\dfrac{dP}{dt} = k(M - P(t))$.

(a) Show that $Q(t) = M - P(t)$ grows or shrinks exponentially.
(b) Find $\lim_{t\to\infty} P(t)$.
(c) Interpret the constant M in terms of a population.

17. Assume that the outdoor temperature increases linearly as a function of t, $h(t) = t + 1$. The temperature of a house is T_0, at time $t = 0$. Then it warms up by Newton's law. That is, if the temperature in the house at time t is $T(t)$, then $T'(t) = k(t + 1 - T(t))$, where k is a positive constant.

(a) Find $T(t)$.
(b) Does the long-term behavior of the solution depend on the initial temperature?
(c) Is the graph of $T(t)$ asymptotic to the graph of the outdoor temperature?

18. Some species have a maximum sustainable population, which we call M. Assume the population changes at a rate proportional to itself, $P(t)$, and to the amount left to grow, $M-P(t)$. Then P is governed by a **logistic growth** model: $\frac{dP}{dt} = kP(t)(M-P(t))$.

This model has been seen before, in Exercises 36 to 38 in Section 5.7 and in the Hubbert's Peak CIE at the end of Chapter 12.

Let M and k be positive constants.

(a) Show that $\frac{1}{P(M-P)}$ can be written as $\frac{1}{M}\left(\frac{1}{P}+\frac{1}{M-P}\right)$.

(b) Use separation of variables to find all solutions to $\frac{dP}{dt} = kP(M-P)$.

(c) Use the result in (b) to find the solution of the initial value problem: $P' = kP(M-P)$, $P(0) = P_0$.

19. The differential equation $y' = 1 - y^{10}$ for $t > 0$ with $y(0) = 0$ is not linear.

(a) Show that $|y(t)| \le 1$ for all values of t.

(b) Show that $\lim_{t\to\infty} y(t)$ exists and that it is 1.

(c) When is the graph of y concave up? concave down?

(d) What might the graph of y look like?

The method of undetermined coefficients is effective when the right-hand side is one of the special forms we have discussed. This method is useless if the right-hand side contains functions such as $\tan(t)$ or $\ln(t+1)$ or $\frac{1}{t^2+1}$. Then the **method of variation of parameters** is more effective. Exercise 20 outlines the general process for variation of parameters and Exercise 21 works through it for a specific example. Additional problems for which variation of parameters is appropriate are in Exercises 22 to 29.

20. The method of variation of parameters for $y'' + by' + cy = f(t)$ is similar to procedures we have already used. When y_1 and y_2 are solutions to the associated homogeneous equation, the general solution of the homogeneous equation is $y_h(t) = c_1y_1(t) + c_2y_2(t)$ for any constants c_1 and c_2. We seek a solution of the form $y_p(t) = u_1(t)y_1(t) + u_2(t)y_2(t)$ where the unknown coefficients, $u_1(t)$ and $u_2(t)$, are found as follows:

(a) Compute y_p', which has four terms. To prepare for the computation of y_p'', assume $u_1'y_1 + u_2'y_2 = 0$. This leaves $y_p' = u_1y_1' + u_2y_2'$.

(b) Compute y_p'' by differentiating the y_p' found in (a).

(c) Show that when y_p, y_p', and y_p'' are substituted into the nonhomogeneous equation the resulting equation is $u_1'(t)y_1'(t) + u_2'(t)y_2'(t) = f(t)$.

We now have two linear equations for $u_1'(t)$ and $u_2'(t)$. After solving them, all that remains is to integrate $u_1'(t)$and $u_2'(t)$ to find $u_1(t)$ and $u_2(t)$.

21. Apply the method of variation of parameters to $y'' + y = \tan(t)$, as follows.

(a) Show that the general solution to the associated homogeneous equation is $y_h = c_1\sin(t) + c_2\cos(t)$.

(b) Show that $y_p = u_1(t)\sin(t) + u_2(t)\cos(t)$ is a solution to the nonhomogeneous equation when

$$u_1'(t)\sin(t) + u_2'(t)\cos(t) = 0 \qquad \text{and} \qquad u_1'(t)\cos(t) - u_2'(t)\sin(t) = \tan(t).$$

(c) Solve the equations in (b) to find

$$u_1'(t) = \sin(t) \qquad \text{and} \qquad u_2'(t) = -\tan(t)\sin(t).$$

(d) Integrate $u_1'(t)$ and $u_2'(t)$ to find

$$u_1(t) = -\cos(t) + K_1 \qquad \text{and} \qquad u_2(t) = \sin(t) - \ln|\sec(t) + \tan(t)|.$$

(e) Conclude that $y_p = -\cos(t)\ln|\sec(t)+\tan(t)|$.

In Exercises 22 to 29, use the method of variation of parameters to find a solution of the given nonhomogeneous differential equation.

22. $y'' + 3y' + 2y = 4e^t$

23. $y'' - 2y' + y = t^{-1}e^t$

24. $y'' + 4y' + 4y = e^{-2t}\ln(t)$ $(t > 0)$

25. $y'' + 5y' + 6y = 108t^2$

26. $y'' + 2y' + y = e^{-t}$

27. $y'' + 4y = \sec^2(2t)$

28. $y'' + 9y = \tan^2(3t)$

29. $y'' - 5y' - 24y = 1331t^2e^{8t}$

Calculus is Everywhere # 16
Origin of the Equation for Flow Through a Narrow Pipe

The next CIE, *Flow Through a Narrow Pipe: Poiseuille's Law*, will obtain a fourth-order law for the rate at which fluid flows through a narrow pipe. It does this by finding the solutions to the differential equation

$$\mu\frac{d^2v}{dr^2}+\frac{\mu}{r}\frac{dv}{dr}+A=0. \tag{C.16.1}$$

In this section we show how this equation is obtained (C.16.1) from basic physical assumptions.

The first assumptions are that the fluid is incompressible and has been flowing long enough to have reached a steady-state, that is, the flow is not changing with respect to time.

The Physics

In Sir Isaac Newton's monumental *Mathematical Principles of Natural Philosophy and the System of the World*, usually referred to as "The Principia," he stated three laws of motion. The second and third laws of motion will be used in the present discussion.

Newton's second law implies that if a moving object undergoes no acceleration, then the total force operating on that object is zero.

The third law reads, "To every action there is always opposed an equal reaction; or, the mutual actions of two bodies upon each other are always equal, and directed to contrary parts."

He follows this law by a commentary that begins, "Whatever draws or presses another is as much drawn or pressed by that other. If you press a stone with your finger, the finger is also pressed by the stone. If a horse draws a stone tied to a rope, the horse (if I may say so) will be equally drawn back towards the stone."

We first apply the third law.

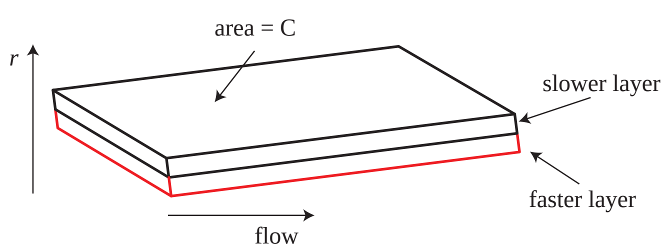


Figure C.16.1

Imagine two thin planar layers of fluid moving from left to right, as in Figure C.16.1. Fluid in the lower one is moving faster than the fluid in the upper one. The area of contact we call C. The velocity v of the fluid depends on r, shown in Figure C.16.1. We are assuming that v is a decreasing function of r.

The faster fluid exerts a force F_1 on the slower fluid, tending to speed it up. The slower fluid exerts a force F_2 on the faster fluid, tending to slow it down. By Newton's third law these two forces are equal but in opposite directions: $F_2=-F_1$.

The magnitude of F_1 and F_2 is proportional to the area C. It also is proportional to the difference in their velocities, which we will measure by the derivative dv/dr, viewing v as a function of r, as shown in Figure C.16.1. Thus

$$F_1 \text{ and } F_2 \text{ are proportional to } C \text{ times } \frac{dv}{dr}.$$

The constant of proportionality depends on the particular fluid. Call this positive constant μ. Because v is a decreasing function of r, dv/dr is negative, and we have

$$F_1=-\mu C\frac{dv}{dr}. \tag{C.16.2}$$

By Newton's third law,

$$F_2=\mu C\frac{dv}{dr}. \tag{C.16.3}$$

Equations (C.16.2) and (C.16.3) hold also when the layers are curved, for instance, when they are two concentric thin pipes, the case we will be using. The constant μ is called the **viscosity** of the fluid. The larger it is the more drag or pull one layer exerts on the other.

Observation C.16.1: *Viscosity in Everyday Life*

Viscosity is a measure of the internal friction of the fluid. The higher the viscosity, the harder it is to make the fluid flow. In the SI system of units, water has a low viscosity, 0.894 μPa-s (microPascal-seconds), while olive oil's viscosity is much larger (about 81 μPa-s), and blood is in between (about 3.5 μPa-s). Temperature affects viscosity. For instance, honey has a very high viscosity at room temperature but flows easily at high temperature.

The Mathematics

Now that we have the necessary physical principles, we are ready to consider a fluid moving from left to right through a narrow cylindrical pipe of inner radius R and length L, as in Figure C.16.2(a). Fluid at a distance r from the axis has the velocity $v = v(r)$ for $0 \leq r \leq R$, a decreasing function of r.

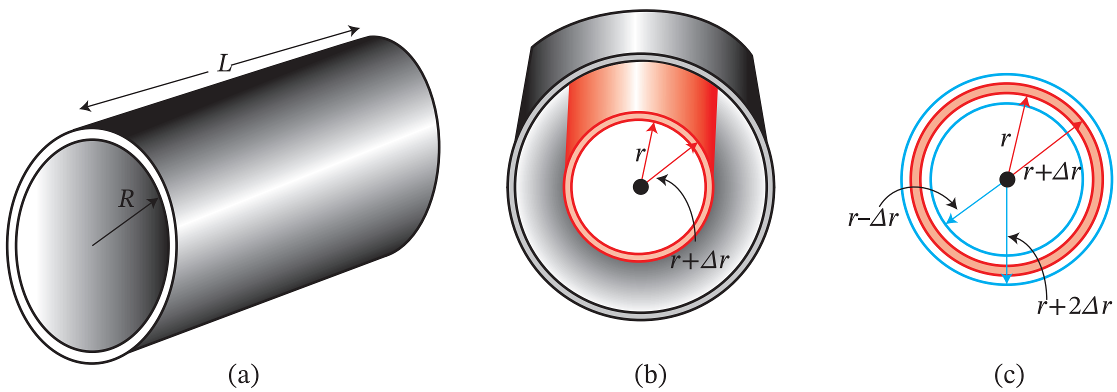


Figure C.16.2

Imagine breaking up the cylinder of radius R and length L into thin concentric pipes or straws, as we did with the shell technique in Section 7.5. Figure C.16.2(b) shows one such pipe. Its inner radius is r and its outer radius is $r + \Delta r$. The flow in this pipe is affected by the flow in the two pipes adjacent to it, shown in Figure C.16.2(c).

Four forces act on the middle pipe: the force at the left end, the force at the right end, the drag to the left due to slower fluid in the outer pipe, and the pull to the right of the faster fluid in the inner pipe. We calculate each of these forces.

The pressure at the left end, P_1, is higher than the pressure at the right end, P_2. Both pressures are assumed to be independent of position (and time), that is, they are constants.

The force against the left end is the product of the pressure there, P_1, and the area of the base of the middle pipe, which is $\pi(r+\Delta r)^2 - \pi r^2 = 2\pi\,(r + \Delta r/2)\ \Delta r$, as Exercise 1 shows. Thus this force is

$$P_1\left(2\pi\left(r + \frac{\Delta r}{2}\right)\ \Delta r\right).$$

Similarly, the force against the right end of the middle pipe is

$$-P_2\left(2\pi\left(r + \frac{\Delta r}{2}\right)\ \Delta r\right).$$

The force of the inner pipe on the middle pipe is

$$-\mu\,(\text{Area Shaded})\,\frac{dv}{dr} = -\mu\,(2\pi r L)\,\frac{dv}{dr}(r),$$

where dv/dr is evaluated at r. The minus sign is inserted because dv/dr is negative.

The force of the outer pipe on the middle pipe is

$$\mu(2\pi(r+\Delta r)L)\frac{dv}{dr}(r+\Delta r)$$

where dv/dr is evaluated at $r+\Delta r$.

Because the fluid is moving at a steady rate, not being accelerated, the sum of these forces is, by Newton's second law, zero:

$$0 = P_1\left(2\pi\left(r+\frac{\Delta r}{2}\right)\Delta r\right) - P_2\left(2\pi\left(r+\frac{\Delta r}{2}\right)\Delta r\right) - \mu(2\pi rL)\frac{dv}{dr}(r) + \mu(2\pi(r+\Delta r)L)\frac{dv}{dr}(r+\Delta r).$$

Define P to be the net pressure on the two ends, $P_1 - P_2$. Dividing by $2\pi L$ yields this simpler equation:

$$0 = \frac{P}{L}\left(r+\frac{\Delta r}{2}\right)\Delta r - \mu r\frac{dv}{dr}(r) + \mu(r+\Delta r)\frac{dv}{dr}(r+\Delta r) \tag{C.16.4}$$

The last factor of the last term in (C.16.4) can be approximated using a Taylor polynomial of degree one, that is, its linear approximation:

$$\frac{dv}{dr}(r+\Delta r) \approx \frac{dv}{dr} + \frac{d^2v}{dr^2}\Delta r, \tag{C.16.5}$$

where both dv/dr and d^2v/dr^2 are evaluated at r. With this approximation, (C.16.4) becomes

$$0 = \frac{P}{L}\left(r+\frac{\Delta r}{2}\right)\Delta r - \mu r\frac{dv}{dr} + \mu(r+\Delta r)\left(\frac{dv}{dr} + \frac{d^2v}{dr^2}\Delta r\right). \tag{C.16.6}$$

Expanding the product on the right-hand side of (C.16.6) gives

$$0 = \frac{P}{L}\left(r+\frac{\Delta r}{2}\right)\Delta r - \mu r\frac{dv}{dr} + \mu r\frac{dv}{dr} + \mu r\frac{d^2v}{dr^2}\Delta r + \mu\frac{dv}{dr}\Delta r + \mu\frac{d^2v}{dr^2}(\Delta r)^2. \tag{C.16.7}$$

When Δr is small, $(\Delta r)^2$ is even smaller than Δr. Therefore we may omit $(\Delta r)^2$, and, after a cancellation, reach:

$$0 = \frac{P}{L}r\,\Delta r + \mu r\frac{d^2v}{dr^2}\Delta r + \mu\frac{dv}{dr}\Delta r. \tag{C.16.8}$$

Dividing (C.16.8) by $r\Delta r$ gives us

$$0 = \frac{P}{L} + \mu\frac{d^2v}{dr^2} + \frac{\mu}{r}\frac{dv}{dr}.$$

This is equivalent to (C.16.1), where $A = P/L$. This completes our goal.

EXERCISES for CIE C.16

1. Show that the area of a ring of inner radius r and outer radius $r+\Delta r$ is exactly $2\pi\left(r+\frac{\Delta r}{2}\right)\Delta r$.

2. Often $2\pi r\Delta r$ is used as an approximation of the area of the ring in Exercise 1. Would the argument that derives (C.16.1) still go through with this approximation? Explain.

3. Why is there no minus sign in (C.16.2)?

4. Deduce (C.16.8) from (C.16.7).

5. Show that (C.16.1) can be written as $\frac{\mu}{r}\frac{d}{dx}\left(r\frac{dv}{dr}\right) + A = 0$.

6. In a Taylor expansion the coefficient of the second derivative is $\frac{1}{2}$. However, in (C.16.5) the coefficient of the second derivative is not $\frac{1}{2}$. Why not?

Calculus is Everywhere # 17

Flow Through a Narrow Pipe: Poiseuille's Law

Poiseuille is pronounced *pwa-záy*.

In the 1830's the physician Jean Poiseuille (1797–1869) investigated the flow of liquids in pipes whose diameters were as small as 0.015 mm to as large as 0.6 mm. He was motivated by the desire to understand the flow of blood in arteries. His experiments were not easy to carry out. For instance, calibrating a single pipe could take as long as twelve hours. In 1839 he deposited a sealed packet with the French Academy of Sciences describing his results, the standard way to establish priority at that time.

Poiseuille concluded that the flow is proportional to the fourth power of the inner radius (R^4), to the difference in pressures at the ends of the pipe (P), and inversely proportional to the length of the pipe (L), and thus proportional to R^4P/L. At the time it was thought that the flow would be proportional to R^3, not R^4. In 1860 Eduard Hagenbach confirmed Poiseuille's conjecture, deriving his formula mathematically from physical principles. The previous Calculus is Everywhere obtained the differential equation from Newton's laws, starting with a form of the Navier-Stokes equation, a fundamental equation in fluid dynamics.

Use R to denote the inner radius of the pipe. One might expect the flow to be proportional to the cross-sectional area πR^2, hence to the square of the radius. That assumes all the fluid flows at the same speed. But that is not so because fluid does not move at all along the surface and moves fastest along the axis.

On the interval $[-R, R]$ let $v(r)$ be the velocity of the fluid at a distance $|r|$ from the axis. Thus $v(-r) = v(r)$ and $v(R) = 0$. Because the maximum velocity occurs along the axis, $v'(0) = 0$.

Let μ denote the viscosity of the fluid (large for oil, small for water, in between for blood), and A be the pressure gradient, defined as P/L. We will find how $v(r)$ depends on r. Once we know that, we can measure the flow.

Our starting point is the differential equation derived in CIE 16 on the *Origins of the Equation for Flow in a Narrow Pipe* that appears immediately before this CIE. The differential equation is

$$\mu\frac{d^2v}{dr^2} + \frac{\mu}{r}\frac{dv}{dr} = -A. \tag{C.17.1}$$

In this CIE we find solutions of this differential equation.

Only the first and second derivatives of v appear, not v itself. Letting $w = dv/dr$ and dividing by μ, (C.17.1) becomes:

$$\frac{dw}{dr} + \frac{w}{r} = -\frac{A}{\mu}, \tag{C.17.2}$$

a first-order linear differential equation that can be solved using an integrating factor (see Section 13.2). A general solution method for problems of this form can be found in Exercise 5. However, here we will use a different approach to obtain the solution.

The terms on the left-hand side of (C.17.1) appear often in problems that involve viscosity. If we use the observation that the two terms on the left-hand side of (C.17.1) can be combined into a single term as follows:

$$\frac{dw}{dr} + \frac{w}{r} = \frac{1}{r}\frac{d}{dr}(rw(r)),$$

then (C.17.2) becomes

$$\frac{1}{r}\frac{d}{dr}(rw(r)) = \frac{-A}{\mu}.$$

Multiplying this equation by r results in

$$\frac{d}{dr}(rw(r)) = \frac{-A}{\mu}r,$$

which is easily integrated. There is a constant K such that

$$rw(r) = \frac{-A}{2\mu}r^2 + K \qquad \text{for all } r \text{ between 0 and } R.$$

Note that $w(0) = 0$ because the maximum velocity occurs at the axis. To find K replace r by 0, obtaining $0w(0) = K$, which shows that $K = 0$. Thus, replacing $w(r)$ with dv/dr brings us to

$$\frac{dv}{dr} = \frac{-A}{2\mu} r.$$

Another integration produces

$$v(r) = \frac{-Ar^2}{4\mu} + Q \qquad \text{for some constant } Q. \tag{C.17.3}$$

When $r = R$, v is 0, so (C.17.3) implies that

$$0 = \frac{-AR^2}{4\mu} + Q$$

and we have $Q = AR^2/(4\mu)$. Thus (C.17.3) becomes

$$v(r) = \frac{-A}{4\mu} r^2 + \frac{A}{4\mu} R^2.$$

Therefore the velocity of the fluid in the pipe is

$$v(r) = \frac{A}{4\mu}\left(R^2 - r^2\right). \tag{C.17.4}$$

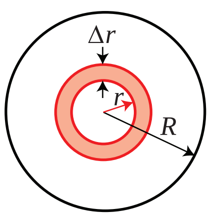


Figure C.17.1

Now that we know how the velocity varies with distance from the axis, we can calculate the total flow of the liquid.

A cross section of the liquid perpendicular to the axis of the pipe is a disk composed of narrow rings of width dr as shown in Figure C.17.1. The area of the ring with inner radius r is approximately $2\pi r\, dr$. Combining this with (C.17.4) shows that the rate at which fluid crosses the ring is approximately

$$\text{(Velocity)(Area of Ring)} = \frac{A}{4\mu}\left(R^2 - r^2\right)(2\pi r\, dr) = \frac{\pi A}{2\mu}\left(R^2 r - r^3\right)\, dr.$$

This fluid moves past the entire disk at the rate

$$\begin{aligned}
\int_0^R \frac{\pi A}{2\mu}\left(R^2 r - r^3\right)\, dr &= \frac{\pi A}{2\mu}\int_0^R \left(R^2 r - r^3\right) dr && \left(\text{since } \int c f(r)\, dr = c\int f(r)\, dr\right)\\
&= \frac{\pi A}{2\mu}\left(\frac{R^2 r^2}{2} - \frac{r^4}{4}\right)\Bigg|_0^R && (\text{by FTC I})\\
&= \frac{\pi A}{2\mu}\left(\frac{R^4}{2} - \frac{R^4}{4}\right) - 0\\
&= \frac{\pi A R^4}{8\mu} && (\text{simplification})\\
&= \frac{\pi R^4 P}{8\mu L} && (\text{since } A = P/L).
\end{aligned}$$

This says that the flow is proportional to $R^4 P/L$ and Poiseuille's conjecture is a consequence of physical principles.

Observation C.17.2: *Poiseuille's Law in Real Life*

The flow depends strongly on the radius R. Reducing the radius from R to $R/2$ reduces the flow by a factor of $2^4 = 16$. Even reducing R by 20% cuts the flow by almost 60%. That is why narrowing of the arteries is a serious medical condition.

Poiseuille's Law also explains how the muscles that circle the arteries can control blood flow by a slight tightening or relaxing.

Historical Note: Poiseuille's Experiments

For a detailed and fascinating description of Poiseuille's experiments, see "The History of Poiseuille's Law," by Salvatore P. Sutera and Rickard Skalak in *The Annual Review of Fluid Mechanics* **25** (1993), 1–19.

EXERCISES for CIE C.17

To solve a differential equation of the form $\frac{dy}{dt} + \frac{k}{t}y = q(t)$, with k a constant, multiply the differential equation by t^k. The left-hand side is then the derivative of $t^k y(t)$. Thus the equation can be written as

$$\frac{d}{dt}\left(t^k y(t)\right) = t^k q(t).$$

The solution is found by integrating both sides and then dividing by t^k to obtain an equation for $y(t)$.

This is a special case of the integrating factor method for solving first-order linear differential equations that was introduced in Section 13.2.

Use this method to find the solution of

1. $y' + \frac{3}{t}y = 3t$ **2.** $y' - \frac{1}{2t}y = 1$ **3.** $y' + \frac{1}{t}y = e^t$ **4.** $y' - \frac{4}{t}y = t^4 + 2t^3 + 1$

5. In this exercise we present a different method for solving (C.17.1) with $v(R) = 0$.

(a) Verify that $w = \frac{C}{r} - \frac{Ar}{2\mu}$ is a solution of (C.17.2) for any value of the constant C.
(b) Find the solution of (C.17.1) by integrating the solution of (C.17.2) given in (a).
(c) Find all solutions of (C.17.1) that also satisfy $v(R) = 0$.
(d) The solution in (c) should still involve one constant. Find a value that makes the solution well-defined (that is, is finite) for $r = 0$.
(e) Does the solution found in (d) agree with (C.17.4)? Be sure to explain any differences.

6. (a) Why is the area of the ring in Figure C.17.1 approximately $2\pi r\,dr$?
(b) What is the exact area of that ring?

7. We showed that the velocity is proportional to $R^2 - r^2$. If, instead, the velocity is proportional to $R - r$ what power of R would appear in the formula for the flow?

8. Define $v(r,R)$ to be the velocity of the fluid at a distance r from the axis of a pipe of inner radius R. Assume that there is a differentiable function $f(r)$ such that $v(r,R) = f(R) - f(r)$. Assume that there are constants k and $m \neq 2$ such that for all R, $\int_0^R \left(f(R) - f(r)\right) r\,dr = kR^m$. Show that (a) $v(R,R) = 0$ and (b) $v(r,R) = \frac{2km}{m-2}\left(R^{m-2} - r^{m-2}\right)$.

Chapter 14

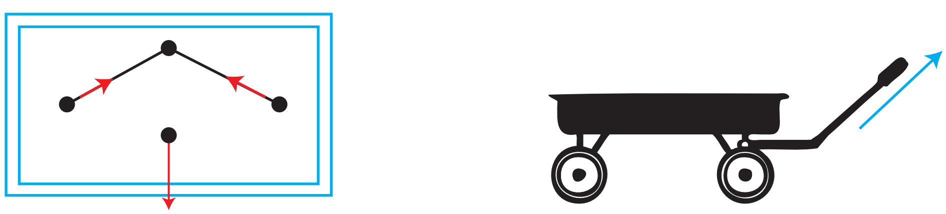
Vectors

This chapter is part of algebra, not calculus, because it involves no limits, derivatives, or integrals.

Section 14.1 introduces vectors, which are usually pictured as arrows. Section 14.2 examines the dot product, a number associated with a pair of vectors. Section 14.3 defines the cross product, a vector perpendicular to two vectors. Applications of vectors and the dot product in Section 14.4 include finding the distance from a point to a line or plane, and giving a parametric description of a line.

14.1 The Algebra of Vectors

When you hang a picture on wire you deal with three vectors: one describing the downward force of gravity and two describing the force of the wire pulling up, as in Figure 14.1.1(a)



(a) (b)

Figure 14.1.1

When you pull a wagon the force you use is represented by a vector, as in Figure 14.1.1(b). The harder you pull, the larger the vector.

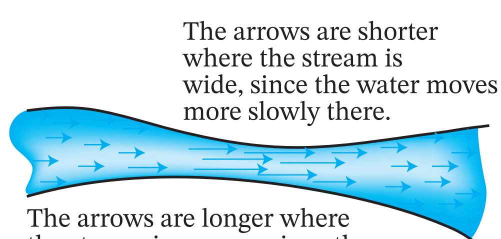


Figure 14.1.2

A vector has a direction and a magnitude. By comparison, a number (or scalar) has only magnitude. Numbers can be positive, negative, or zero, and are represented on a number line. Vectors are used to describe quantities in a two-dimensional plane or in in three (or higher) dimensional space. A vector can be thought of as an arrow, whose length and direction carry information. Vectors are of use in describing the flow of a fluid, as in Figure 14.1.2, or the wind, or the strength and direction of a magnetic field.

Vectors in the Plane

A **vector** in the xy-plane is an ordered pair of numbers x and y, denoted $\langle x, y\rangle$. Its **magnitude**, or length, is $\sqrt{x^2+y^2}$. Though the notation resembles that for a point, (x, y), we treat vectors differently. We can add them, subtract them, and multiply them by a number.

The vector $\langle 0,0\rangle$ is denoted **0** and is called the **zero vector**.

We can represent a nonzero vector $\langle x, y\rangle$ by an arrow whose tail is at $(0,0)$ and whose head (or tip) is at the point (x, y), as in Figure 14.1.3(a).

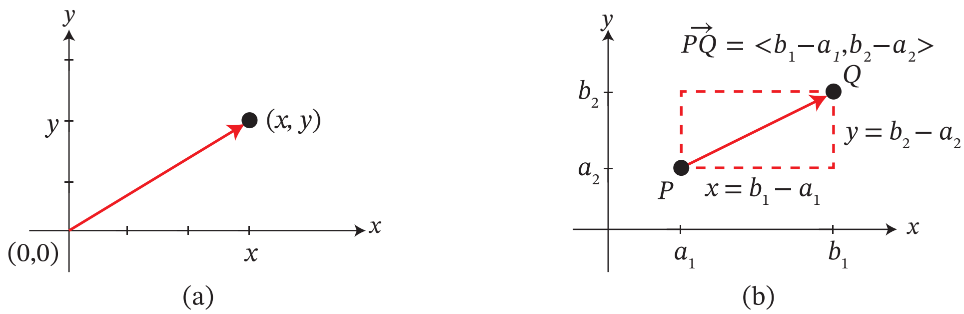


Figure 14.1.3

More generally, we can represent $\langle x, y\rangle$ using points $P = (a_1, a_2)$ and $Q = (b_1, b_2)$ if $b_1 - a_1 = x$ and $b_2 - a_2 = y$, as in Figure 14.1.3(b). We speak then of the vector from P to Q and denote it $\overrightarrow{PQ}$.

A vector $\langle x, y\rangle$ will be written with bold-face letters, such as **A**, **B**, **r**, **v**, and **a**. In writing they are denoted by putting an arrow on top of the letters, for instance $\overrightarrow{AB}$. A vector of length 1 is called a **unit vector** and is sometimes topped with a little hat, as in $\hat{\mathbf{r}}$, which is read as "r hat".

Table 14.1.1 summarizes the basic operations on vectors. In this table $\mathbf{A} = \langle a_1, a_2\rangle$ and $\mathbf{B} = \langle b_1, b_2\rangle$ are vectors and c is a number.

Operation	Definition	Geometry	Comment
$\mathbf{A}+\mathbf{B}$	$\langle a_1+b_1, a_2+b_2\rangle$	Figure 14.1.4	The tail of **B** is placed at the head of **A**
$-\mathbf{A}$	$\langle -a_1, -a_2\rangle$	Figure 14.1.5(a)	$-\mathbf{A}$ points in opposite direction of **A**
$\mathbf{A}-\mathbf{B}$	$\langle a_1-b_1, a_2-b_2\rangle$	Figure 14.1.5(b)	What is added to **B** to get **A**
$c\mathbf{A}$	$\langle ca_1, ca_2\rangle$	Figure 14.1.5(c)	Parallel to **A**, opposite direction if $c<0$, and $\lvert c\rvert$ times as long as **A**,
$\frac{1}{c}\mathbf{A}$	$\left\langle \frac{a_1}{c}, \frac{a_2}{c}\right\rangle$	Figure 14.1.5(d)	Parallel to **A**, opposite direction if $c<0$, and $1/\lvert c\rvert$ times as long as **A** ($c \neq 0$)

Table 14.1.1

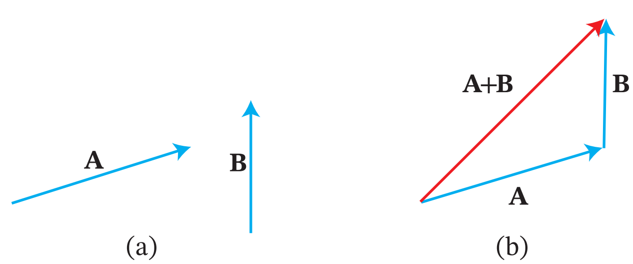


Figure 14.1.4

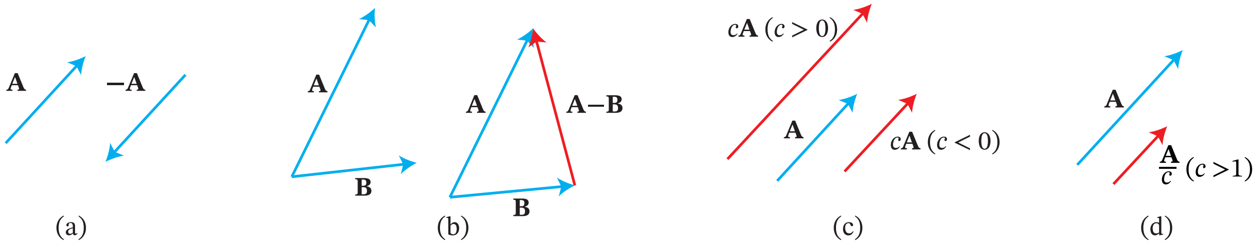


Figure 14.1.5

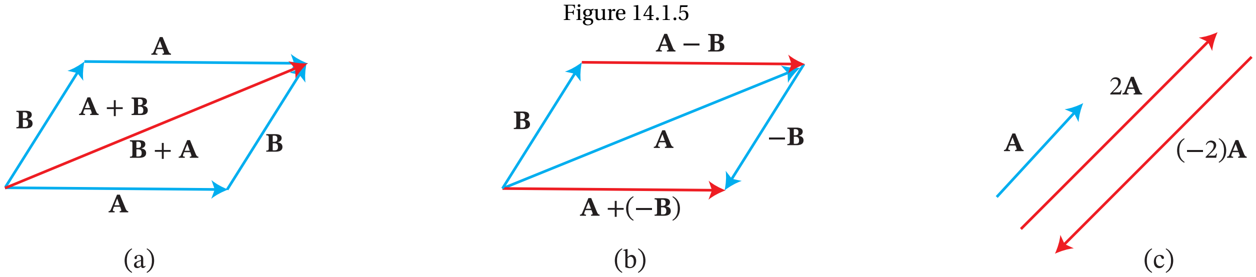


Figure 14.1.6

Figure 14.1.6(a) shows both $\mathbf{A}+\mathbf{B}$ and $\mathbf{B}+\mathbf{A}$; they are equal. In terms of arrows it makes sense.

The operation of addition obeys the usual rules of addition of numbers, $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$ and $\mathbf{A}+(\mathbf{B}+\mathbf{C})=(\mathbf{A}+\mathbf{B})+\mathbf{C}$.

Also, $\mathbf{A}-\mathbf{B}=\mathbf{A}+(-\mathbf{B})$ follows from the definitions. This property can also be seen in Figure 14.1.6(b) since $\mathbf{A}-\mathbf{B}$ and $\mathbf{A}+(-\mathbf{B})$ appear as opposite sides of a parallelogram.

When $\mathbf{A}=\langle x,y\rangle$ is a vector and c is a positive constant, $c\mathbf{A}$ is a multiple of $\mathbf{A}$ that points in the same direction as $\mathbf{A}$. If $c>1$, then $c\mathbf{A}$ is longer than $\mathbf{A}$ and if $0<c<1$, then $c\mathbf{A}$ is shorter than $\mathbf{A}$. If $c<0$, then $c\mathbf{A}=-|c|\mathbf{A}$ is the opposite of $|c|\mathbf{A}$. See Figure 14.1.6(c).

When referring to numbers, such as c, x, and y, in the context of vectors, we call them **scalars**. Thus in $c\mathbf{A}$ the scalar c is multiplying the vector $\mathbf{A}$.

Two vectors $\mathbf{A}$ and $\mathbf{B}$ are **parallel** if either $\mathbf{A}=c\mathbf{B}$ or $\mathbf{B}=c\mathbf{A}$ for some scalar c.

IMPORTANT FACT: This definition of parallel implies that the zero vector is parallel to every vector.

Usually when we are dealing with parallel vectors, we will know that one of them is nonzero. In that case we do not need to consider both equations in the definition. For example, if $\mathbf{B}$ is nonzero then $\mathbf{A}$ is parallel to $\mathbf{B}$ if and only if $\mathbf{A}=c\mathbf{B}$ for some scalar c, because the other equation, $\mathbf{B}=c\mathbf{A}$, would imply that $c\neq 0$ and $\mathbf{A}=(1/c)\mathbf{B}$.

EXAMPLE 1. Let $\mathbf{A}=\langle 1,2\rangle$, $\mathbf{B}=\langle 3,-1\rangle$ and $c=-2$. Compute $\mathbf{A}+\mathbf{B}$, $\mathbf{A}-\mathbf{B}$, and $c\mathbf{A}$. Draw the corresponding arrows.

SOLUTION By direct computation:

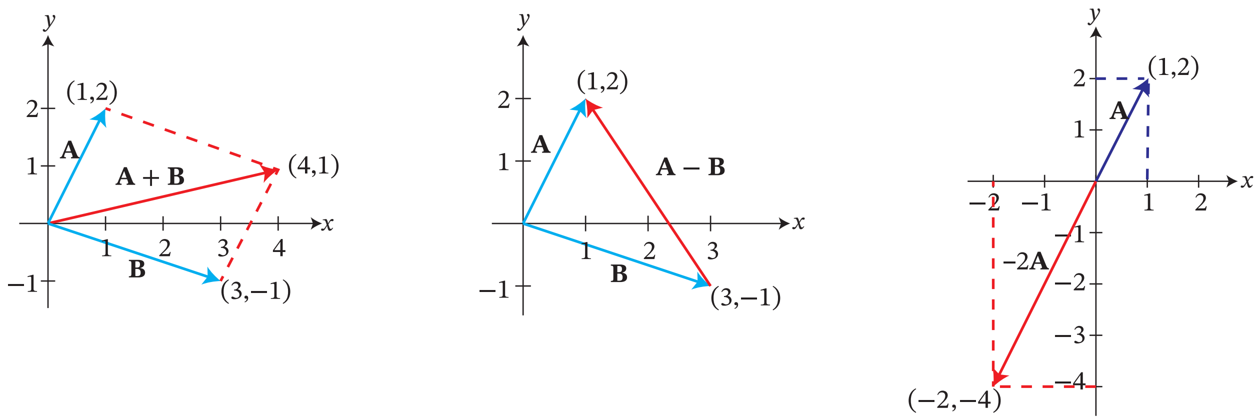
$$\begin{aligned}
\mathbf{A}+\mathbf{B} &= \langle 1,2\rangle+\langle 3,-1\rangle = \langle 1+3, 2+(-1)\rangle = \langle 4,1\rangle\\
\mathbf{A}-\mathbf{B} &= \langle 1,2\rangle-\langle 3,-1\rangle = \langle 1-3, 2-(-1)\rangle = \langle -2,3\rangle\\
c\mathbf{A} &= -2\langle 1,2\rangle \qquad = \langle -2(1), -2(2)\rangle \;\; = \langle -2,-4\rangle
\end{aligned}$$

The vectors $\mathbf{A}-\mathbf{B}$ and $\mathbf{A}+\mathbf{B}$ lie on the diagonals of a parallelogram. See Figure 14.1.7. ◆

Coordinates in Space

Before we can define vectors in space, we must introduce an appropriate coordinate system.

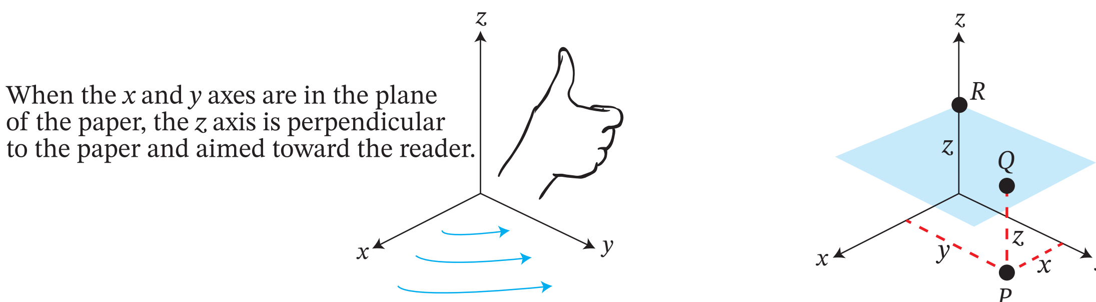
Pick a pair of perpendicular intersecting lines to serve as the x- and y-axes. The positive parts of these axes are indicated by arrows. These two lines determine the xy-**plane**. The line perpendicular to the xy-plane and meeting



(a) (b) (c)

Figure 14.1.7

the x- and y-axes at $(0,0)$ will be called the z-axis. The point where the three axes meet is called the **origin**. The 0 of the z-axis is at the origin. But which half of the z-axis will have the positive numbers and which half will have the negative numbers? It is customary to determine this by the **right-hand rule**. Moving in the xy-plane from the positive x-axis to the positive y-axis determines a sense of rotation around the z-axis. If the fingers of the right hand curl in that sense, the thumb points in the direction of the positive z-axis, as shown in Figure 14.1.8(a).



(a) (b)

Figure 14.1.8

A point Q in space is now described by three numbers. Two specify the x- and y-coordinates of the point P in the xy-plane directly below (or above) Q. Then the height of Q above (or below) the xy-plane is recorded by the z-coordinate of the point R where the plane through Q and parallel to the xy-plane meets the z-axis. The point Q is then denoted (x, y, z). See Figure 14.1.8(b).

Points (x, y, z) for which $z = 0$ lie in the xy-plane. The points (x, y, z) for which $x = 0$ lie in the plane determined by the y- and z-axes, which is called the yz-**plane**. Similarly, the equation $y = 0$ describes the xz-**plane**. The xy-, xz-, and yz-planes are called the **coordinate planes**.

EXAMPLE 2. Plot the point $(1,2,3)$.

SOLUTION One way is to first plot $(1,2)$ in the xy-plane. Then, on a line perpendicular to the xy-plane at that point, show the point $(1,2,3)$ as in Figure 14.1.9(a).

Another way is to draw a box that is 1 unit thick, 2 units wide, and 3 units tall as in Figure 14.1.9b; then the point $(1,2,3)$ is the corner diagonally opposite the corner at the origin, $(0,0,0)$. ◆

The axes in the xy-plane divide the plane into four quadrants, and the coordinate planes divide space into eight octants. The octant in which all three coordinates are positive is called the first octant. (The other seven octants do not have numeric names.)

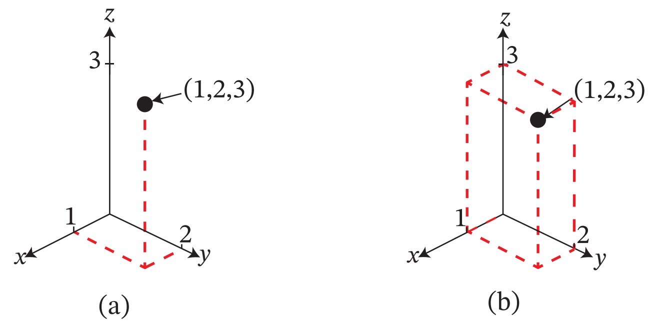


Figure 14.1.9

Vectors in Space

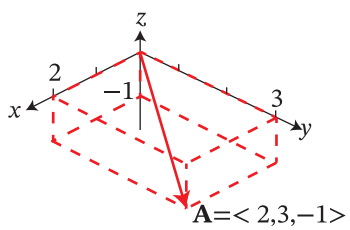


Figure 14.1.10

The only difference between a vector in space and a vector in the xy-plane is that it has three components, x, y, and z, and is written $\langle x, y, z\rangle$. Its **length**, or **magnitude**, is defined to be $\sqrt{x^2+y^2+z^2}$, the distance from the origin $(0,0,0)$ to (x,y,z).

The definitions of the sum and difference of vectors in space are so similar to the definitions for plane vectors that we will not list them. For instance, $\langle a_1, a_2, a_3\rangle + \langle b_1, b_2, b_3\rangle$ is $\langle a_1+b_1, a_2+b_2, a_3+b_3\rangle$. They are harder to draw, even though they can be suggested by an arrow. It may help to visualize a three-dimensional vector by drawing a box in which it is a main diagonal. For instance, to draw the vector $\langle 2,3,-1\rangle$ draw the box shown in Figure 14.1.10.

The Standard Unit Vectors

Three unit vectors indicate the directions of the positive x-, y-, and z-axes. They will be denoted $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, respectively. For instance, $\mathbf{i} = \langle 1,0,0\rangle$. The vector $\langle x,y,z\rangle$ can thus be written as $x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$.

EXAMPLE 3. Draw $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, and $\mathbf{i}+2\mathbf{j}+3\mathbf{k}$.

SOLUTION Figure 14.1.11(a) shows $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ and Figure 14.1.11(b) shows $\mathbf{i}+2\mathbf{j}+3\mathbf{k}$. ♦

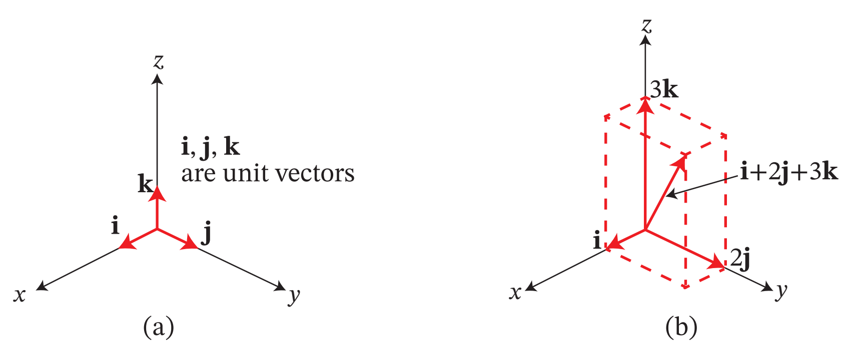


Figure 14.1.11

The magnitude of $\mathbf{A}$ is indicated by $|\mathbf{A}|$, a scalar.

Let $\mathbf{A}$ be a nonzero vector. Then $\mathbf{A}/|\mathbf{A}|$ is a unit vector in the direction of $\mathbf{A}$ because $\mathbf{A}/|\mathbf{A}|$ is the same as the vector $(1/|\mathbf{A}|)\mathbf{A}$, which has length $(1/|\mathbf{A}|)|\mathbf{A}| = 1$.

Example 4 shows how vectors can be used to establish geometric properties.

EXAMPLE 4. Prove that the line that joins the midpoints of two sides of a triangle is parallel to the third side and half as long.

SOLUTION Let the triangle have vertices P, Q, and R. Let the midpoint of side PQ be M and the midpoint of side PR be N as in Figure 14.1.12(a).

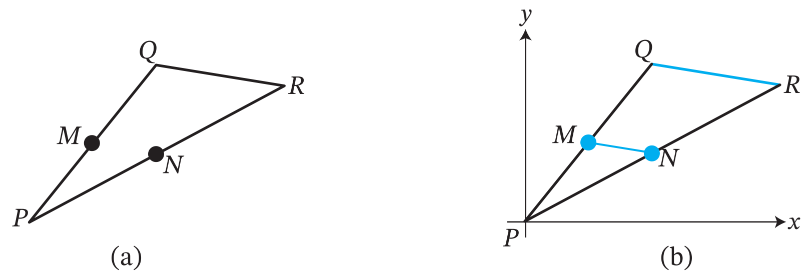


Figure 14.1.12

Introduce an xy-coordinate system in the plane of the triangle. Its origin could be anywhere in the plane, but we put it at P to simplify the calculations. (See Figure 14.1.12(b).)

We wish to show that the vector $\overrightarrow{MN}$ is $\overrightarrow{QR}/2$. To begin, compute $\overrightarrow{MN}$ and $\overrightarrow{QR}$ in terms of vectors involving P, Q, and R.

Because $\overrightarrow{PM} = \overrightarrow{PQ}/2$ and $\overrightarrow{PN} = \overrightarrow{PR}/2$ we have $\overrightarrow{MN} = \frac{1}{2}\overrightarrow{PR} - \frac{1}{2}\overrightarrow{PQ} = \frac{1}{2}(\overrightarrow{PR} - \overrightarrow{PQ}) = \frac{1}{2}\overrightarrow{QR}$. ♦

The next example shows the importance of thinking in terms of vectors. Not thinking that way, one of us had a picture frame fall and break a vase.

EXAMPLE 5. A picture weighing 10 pounds has a wire on the back, which rests on a picture hook, as shown in Figure 14.1.13(a). Find the force (tension) in the wire.

SOLUTION There are three vectors involved. One is straight down, with magnitude 10 pounds and two, $\mathbf{v}_1$ and $\mathbf{v}_2$, are along the wire, with unknown magnitude F: $|\mathbf{v}_1| = F = |\mathbf{v}_2|$.

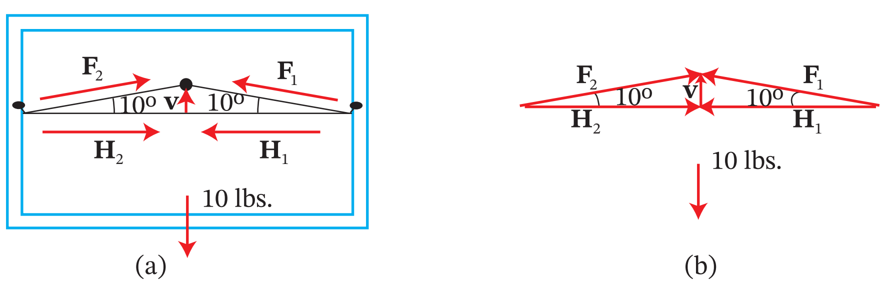


Figure 14.1.13

To balance the downward force of gravity, each end of the wire must have a vertical component of 5 pounds. Since the angle with the horizontal is 10° we must have $F \sin(10°) = 5$ pounds or $F = 5/\sin(10°) \approx 29$ pounds. That is greater than the weight of the painting and can pull the hook out of the frame. ♦

Summary

We introduced the notion of vectors $\langle x, y\rangle$ in the xy-plane or $\langle x, y, z\rangle$ in space and we defined their addition, subtraction, and multiplication of a vector by a scalar c.

We visualized vectors with the aid of arrows, which could be drawn anywhere in the xy-plane or in space. Each vector in the xy-plane can be written as $x\mathbf{i}+y\mathbf{j}$. Vectors in space can be written as $x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$.

EXERCISES for Section 14.1

1. Draw the vector $2\mathbf{i}+3\mathbf{j}$, placing its tail at (a) $(0,0)$, (b) $(-1,2)$, (c) $(1,1)$.

2. Draw the vector $-\mathbf{i}+2\mathbf{j}$, placing its tail at (a) $(0,0)$, (b) $(3,0)$, (c) $(-2,2)$.

In Exercises 3 to 6 draw the vector $\mathbf{A}$ and enough extra lines to show how it is situated in space.

3. $\mathbf{A}=2\mathbf{i}+\mathbf{j}+3\mathbf{k}$, (a) tail at $(0,0,0)$ and (b) tail at $(1,1,1)$.

4. $\mathbf{A}=\mathbf{i}+\mathbf{j}+\mathbf{k}$ (a) tail at $(0,0,0)$ and (b) tail at $(2,3,4)$.

5. $\mathbf{A}=-\mathbf{i}-2\mathbf{j}+2\mathbf{k}$ (a) tail at $(0,0,0)$ and (b) tail at $(1,1,-1)$.

6. $\mathbf{A}=\mathbf{j}+\mathbf{k}$ (a) tail at $(0,0,0)$ and (b) tail at $(-1,-1,-1)$.

In Exercises 7 to 10 plot the points P and Q, draw the vector $\overrightarrow{PQ}$, express it in the form $x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$, and find its length.

7. $P=(0,0,0), Q=(1,3,4)$

8. $P=(1,2,3), Q=(2,5,4)$

9. $P=(2,5,4), Q=(1,2,2)$

10. $P=(1,1,1), Q=(-1,3,-2)$

In Exercises 11 and 12 express the vector $\mathbf{A}$ in the form $x\mathbf{i}+y\mathbf{j}$. North is along the positive y-axis and east is along the positive x-axis.

11. (a) $|\mathbf{A}|=10$ and $\mathbf{A}$ points northwest
(b) $|\mathbf{A}|=6$ and $\mathbf{A}$ points south
(c) $|\mathbf{A}|=9$ and $\mathbf{A}$ points southeast
(d) $|\mathbf{A}|=5$ and $\mathbf{A}$ points east

12. (a) $|\mathbf{A}|=1$ and $\mathbf{A}$ points southwest
(b) $|\mathbf{A}|=2$ and $\mathbf{A}$ points west
(c) $|\mathbf{A}|=\sqrt{8}$ and $\mathbf{A}$ points northeast
(d) $|\mathbf{A}|=1/2$ and $\mathbf{A}$ points south

Exercises 13 and 14 are related.

13. The wind is 30 miles per hour to the northeast. An airplane is traveling 100 miles per hour relative to the air, and the vector from the tail of the plane to its front tip points to the south.

(a) What is the speed of the plane relative to the ground?
(b) What is the direction of the flight relative to the ground?

14. The jet stream is moving 200 miles per hour to the southeast. A plane with a speed of 550 miles per hour relative to the air is aimed to the northwest.

(a) Draw the vectors representing the wind and the plane relative to the air. (Choose a scale and make an accurate drawing.)
(b) Using your drawing, estimate the speed of the plane relative to the ground.
(c) Compute the speed exactly.

---

15. Compute $\mathbf{A}+\mathbf{B}$ and $\mathbf{A}-\mathbf{B}$ if (a) $\mathbf{A}=\langle -1,2,3\rangle$ and $\mathbf{B}=\langle 7,0,2\rangle$ and (b) $\mathbf{A}=3\mathbf{j}+4\mathbf{k}$ and $\mathbf{B}=6\mathbf{i}+7\mathbf{j}$.

16. Compute $\mathbf{A}+\mathbf{B}$ and $\mathbf{A}-\mathbf{B}$ if (a) $\mathbf{A}=\left\langle \frac{1}{2},\frac{1}{3},\frac{1}{6}\right\rangle$ and $\mathbf{B}=\left\langle 2,3,\frac{-1}{3}\right\rangle$ and (b) $\mathbf{A}=2\mathbf{i}+3\mathbf{j}+4\mathbf{k}$ and $\mathbf{B}=-\mathbf{i}+5\mathbf{j}+6\mathbf{k}$.

17. Compute and sketch $c\mathbf{A}$ if $\mathbf{A}=2\mathbf{i}+3\mathbf{j}+\mathbf{k}$ and c is (a) 2, (b) -2, (c) $\frac{1}{2}$, and (d) $-\frac{1}{2}$.

18. Express the vectors in the form $c(2\mathbf{i}+3\mathbf{j}+4\mathbf{k})$. (a) $\langle 4,6,8\rangle$, (b) $-2\mathbf{i}-3\mathbf{j}-4\mathbf{k}$, (c) $\mathbf{0}$, and (d) $\frac{2}{11}\mathbf{i}+\frac{3}{11}\mathbf{j}+\frac{4}{11}\mathbf{k}$.

19. If $|\mathbf{A}|=6$, find the length of (a) $-2\mathbf{A}$, (b) $\mathbf{A}/3$, (c) $\mathbf{A}/|\mathbf{A}|$, (d) $-\mathbf{A}$, and (e) $\mathbf{A}+2\mathbf{A}$.

20. If $|\mathbf{A}|=3$, find the length of (a) $-4\mathbf{A}$, (b) $13\mathbf{A}-7\mathbf{A}$, (c) $|\mathbf{A}/|\mathbf{A}||$, (d) $\mathbf{A}/0.05$, and (e) $\mathbf{A}-\mathbf{A}$.

21. Sketch a unit vector pointing in the same direction as $3\mathbf{i}+4\mathbf{j}$.

22. (a) Find a unit vector $\mathbf{u}$ with the same direction as $\mathbf{A}=\mathbf{i}+2\mathbf{j}+3\mathbf{k}$.
(b) Draw $\mathbf{A}$ and $\mathbf{u}$, with their tails at the origin.

23. (a) Find a unit vector $\mathbf{u}$ with the same direction as $\mathbf{A}=2\mathbf{i}-2\mathbf{j}+3\mathbf{k}$.
(b) Draw $\mathbf{A}$ and $\mathbf{u}$, with their tails at the origin.

24. Using the definition of addition of vectors $\mathbf{A}=\langle a_1,a_2,a_3\rangle$ and $\mathbf{B}=\langle b_1,b_2,b_3\rangle$, show the $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$ and $\mathbf{A}-\mathbf{B}=\mathbf{A}+(-\mathbf{B})$.

25. Using the definition of addition of vectors show that $\mathbf{A}+(\mathbf{B}+\mathbf{C})=(\mathbf{A}+\mathbf{B})+\mathbf{C}$.

26. *(Midpoint formula)* Let A and B be two points in space. Let M be their midpoint and O the origin. Let $\mathbf{A}=\overrightarrow{OA}$, $\mathbf{B}=\overrightarrow{OB}$, and $\mathbf{M}=\overrightarrow{OM}$. (a) Show that $\mathbf{M}=\mathbf{A}+\frac{1}{2}(\mathbf{B}-\mathbf{A})$. (b) Deduce that $\mathbf{M}=\frac{1}{2}(\mathbf{A}+\mathbf{B})$.

27. Let A and B be two distinct points in space. Let C be the point on the line segment AB that is twice as far from A as it is from B. Let $\mathbf{A}=\overrightarrow{OA}$, $\mathbf{B}=\overrightarrow{OB}$, and $\mathbf{C}=\overrightarrow{OC}$. Show that $\mathbf{C}=\frac{1}{3}\mathbf{A}+\frac{2}{3}\mathbf{B}$.

28. Show that $2\mathbf{i}+3\mathbf{j}+4\mathbf{k}$ and $6\mathbf{i}+9\mathbf{j}+12\mathbf{k}$ are parallel.

29. Show that $\mathbf{i}-3\mathbf{j}+6\mathbf{k}$ and $-2\mathbf{i}+6\mathbf{j}-12\mathbf{k}$ are parallel.

30. Write $\mathbf{A}$ and $\mathbf{B}$ in components, and obtain the rule by expressing both $c(\mathbf{A}+\mathbf{B})$ and $c\mathbf{A}+c\mathbf{B}$ in components.

This exercise outlines a proof of the distributive rule, $c(\mathbf{A}+\mathbf{B})=c\mathbf{A}+c\mathbf{B}$.

31. (a) Show that the vectors $\mathbf{u}_1=\frac{1}{2}\mathbf{i}+\frac{\sqrt{3}}{2}\mathbf{j}$ and $\mathbf{u}_2=\frac{\sqrt{3}}{2}\mathbf{i}-\frac{1}{2}\mathbf{j}$ are perpendicular unit vectors.
(b) Find scalars x and y such that $\mathbf{i}=x\mathbf{u}_1+y\mathbf{u}_2$.

32. (a) Show that the vectors $\mathbf{u}_1=\frac{\sqrt{2}}{2}\mathbf{i}+\frac{\sqrt{2}}{2}\mathbf{j}$ and $\mathbf{u}_2=-\frac{\sqrt{2}}{2}\mathbf{i}+\frac{\sqrt{2}}{2}\mathbf{j}$ are perpendicular unit vectors.
(b) Express $\mathbf{i}$ in the form of $x\mathbf{u}_1+y\mathbf{u}_2$.
(c) Express $\mathbf{j}$ in the form $x\mathbf{u}_1+y\mathbf{u}_2$.
(d) Express $-2\mathbf{i}+3\mathbf{j}$ in the form $x\mathbf{u}_1+y\mathbf{u}_2$.

33. (a) Draw a unit vector $\mathbf{u}$ tangent to the curve $y=\sin x$ at $\left(\frac{\pi}{4},\frac{\sqrt{2}}{2}\right)$.
(b) Express $\mathbf{u}$ in the form $x\mathbf{i}+y\mathbf{j}$.

34. (a) Draw a unit vector $\mathbf{u}$ tangent to the curve $y=x^3$ at $(1,1)$.
(b) Express $\mathbf{u}$ in the form $x\mathbf{i}+y\mathbf{j}$.

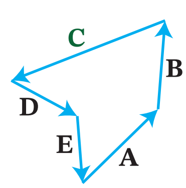


Figure 14.1.14

35. (a) What is the sum of the five vectors shown in Figure 14.1.14?
(b) Sketch the polygon whose sides, in order, are $\mathbf{A}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{E}$, $\mathbf{B}$.
(c) What is $\mathbf{A}+\mathbf{C}+\mathbf{D}+\mathbf{E}+\mathbf{B}$?

36. A rectangular box has sides of length x, y, and z. Show that its longest diagonal has length $\sqrt{x^2+y^2+z^2}$.

37. See Example 5 about hanging a picture. What would be the tension in the wire if it were at an angle of (a) 60° instead of 10° to the horizontal? (b) 5° instead of 10° to the horizontal?

38. (a) Draw the vectors $\mathbf{A} = 2\mathbf{i} + \mathbf{j}$, $\mathbf{B} = 4\mathbf{i} - \mathbf{j}$, and $\mathbf{C} = 5\mathbf{i} + 2\mathbf{j}$.
(b) Using the drawing in (a), show that there are scalars x and y such that $\mathbf{C} = x\mathbf{A} + y\mathbf{B}$.
(c) Estimate x and y from the drawing.
(d) Find x and y exactly.

39. (See Exercise 13.) Let $\mathbf{A}$ and $\mathbf{B}$ be two nonzero and nonparallel vectors in the xy-plane. Let $\mathbf{C}$ be a vector in the xy-plane. Show with a sketch that there are scalars x and y such that $\mathbf{C} = x\mathbf{A} + y\mathbf{B}$.

40. Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be three vectors that do not all lie in one plane. Let $\mathbf{D}$ be a vector in space. Show with a sketch that there are scalars x, y, and z such that $\mathbf{D} = x\mathbf{A} + y\mathbf{B} + z\mathbf{C}$.

41. Let A, B, and C be the vertices of a triangle. Let $\mathbf{A} = \overrightarrow{OA}$, $\mathbf{B} = \overrightarrow{OB}$, and $\mathbf{C} = \overrightarrow{OC}$.
(a) Let P be the point that is on the line segment joining A to the midpoint of the edge BC that is twice as far from A as from the midpoint. Show that $\overrightarrow{OP} = (\mathbf{A} + \mathbf{B} + \mathbf{C})/3$.
(b) Use (a) to show that the medians of a triangle are concurrent.

42. The midpoints of a quadrilateral in space are joined to form another quadrilateral. Prove that the second quadrilateral is a parallelogram.

Exercises 43 and 44 discuss a special case of the Cauchy-Schwarz inequality. This was introduced in the CIE at the end of Chapter 7 (see page 492) and will be proved in Section 16.7 (Exercise 29). It also appears in Exercises 4 and 5 of Section 9.S.

43. (a) Using a diagram, explain why $|\mathbf{A} + \mathbf{B}| \le |\mathbf{A}| + |\mathbf{B}|$. (This is called the **triangle inequality**.)
(b) For what pairs of vectors $\mathbf{A}$ and $\mathbf{B}$ is $|\mathbf{A} + \mathbf{B}| = |\mathbf{A}| + |\mathbf{B}|$?

44. (a) From Exercise 43 deduce that for real numbers x_1, y_1, x_2, and y_2, $x_1x_2 + y_1y_2 \le \sqrt{x_1^2+y_1^2}\sqrt{x_2^2+y_2^2}$.
(b) When does equality hold?

14.2 The Dot Product of Two Vectors

This section introduces the dot product or scalar product, a number defined for pairs of vectors by giving its definition, describing a major application, and developing its properties. First, we define the angle between two vectors.

Definition: *Angle between two nonzero vectors.*

Let $\mathbf{A}$ and $\mathbf{B}$ be two nonparallel and nonzero vectors. They determine a triangle and an angle θ, shown in Figure 14.2.1. The **angle between A and B** is θ, $0 < \theta < \pi$.

If $\mathbf{A}$ and $\mathbf{B}$ are parallel and nonzero, the angle between them is 0 if they have the same direction, or π if they have opposite directions.

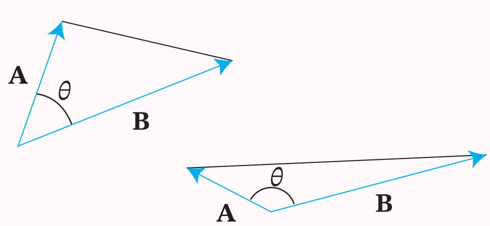


Figure 14.2.1

NOTE: The angle between the zero vector, $\mathbf{0}$, and any other vector is not defined.

NOTATION: The cosine of the angle between nonzero vectors $\mathbf{A}$ and $\mathbf{B}$ is denoted $\cos(\mathbf{A}, \mathbf{B})$.

The Dot Product

Definition: *Dot product*

Let **A** and **B** be two nonzero vectors. Their **dot product** is the number

$$\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}||\mathbf{B}|\cos(\mathbf{A},\mathbf{B}).$$

If **A** or **B** is **0**, their dot product is 0.

The dot product of two vectors is a scalar. For this reason it is also called the **scalar product** of **A** and **B**.

One of the major applications of the dot product is in the calculation of the work accomplished by a force.

Suppose a rock is pulled along level ground by a constant force **F** at an angle θ to the ground, shown in Figure 14.2.2. How much work does it accomplish in moving the rock from the tail to the head of **R**?

Figure 14.2.2

The force can be written as $\mathbf{F}_1+\mathbf{F}_2$, where $\mathbf{F}_1$ is in the direction of motion. The work accomplished by **F** is the amount accomplished by $\mathbf{F}_1$. The other component of the force, $\mathbf{F}_2$, accomplishes no work. The angle between **F** and $\mathbf{F}_1$ is denoted by θ.

The work accomplished by $\mathbf{F}_1$ is its magnitude $|\mathbf{F}_1|$ times the distance it moves the rock, $|\mathbf{R}|$. Because $|\mathbf{F}_1| = |\mathbf{F}|\cos(\theta)$, the work accomplished by **F** is $|\mathbf{F}_1||\mathbf{R}|\cos(\theta)$, the dot product of **F** and **R**,

$$\text{Work} = \underbrace{|\mathbf{F}|\cos(\theta)}_{\text{Force in Direction of }\mathbf{R}} \cdot \underbrace{|\mathbf{R}|}_{\text{Distance traveled}} = \mathbf{F}\cdot\mathbf{R}.$$

The dot product satisfies many useful properties that follow from the definition:

Theorem 14.2.1: Properties of the Dot Product

For any vectors **A** *and* **B** *and any scalar c,*

$$\begin{aligned}
\mathbf{A}\cdot\mathbf{B} &= \mathbf{B}\cdot\mathbf{A} && (\textit{ dot product is commutative })\\
|\mathbf{A}| &= \sqrt{\mathbf{A}\cdot\mathbf{A}} && (\textit{ length using dot product })\\
(c\mathbf{A})\cdot\mathbf{B} = c(\mathbf{A}\cdot\mathbf{B}) &= \mathbf{A}\cdot(c\mathbf{B}) && (\textit{ dot product is associative with scalar multiplication })\\
\mathbf{0}\cdot\mathbf{A} &= 0 && (\textit{ zero property of dot product }).
\end{aligned}$$

Partial Proof of Theorem 14.2.1

The commutativity of the dot product follows immediately the observation that the angle between **A** and **B** is the same as the angle between **B** and **A**, so $\cos(\mathbf{A},\mathbf{B}) = \cos(\mathbf{B},\mathbf{A})$. Then $\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}||\mathbf{B}|\cos(\mathbf{A},\mathbf{B}) = |\mathbf{A}||\mathbf{B}|\cos(\mathbf{A},\mathbf{B}) = \mathbf{B}\cdot\mathbf{A}$.

To establish that $|\mathbf{A}| = \sqrt{\mathbf{A}\cdot\mathbf{A}}$ it suffices to show that $|\mathbf{A}|^2 = \mathbf{A}\cdot\mathbf{A}$. Because every vector is parallel with itself, $\cos(\mathbf{A},\mathbf{A}) = 1$, hence $\mathbf{A}\cdot\mathbf{A} = |\mathbf{A}||\mathbf{A}|\cos(\mathbf{A},\mathbf{A}) = |\mathbf{A}|^2$. The other two properties can be proved by similar arguments.

Observation 14.2.2: *An Important Identity*

The fact that $\mathbf{A}\cdot\mathbf{A} = |\mathbf{A}|^2$ will be used many times.

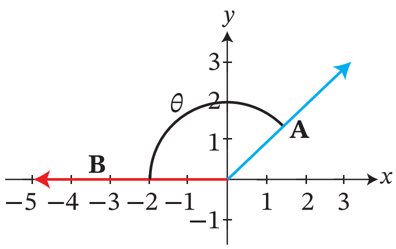


Figure 14.2.3

EXAMPLE 1. Find the dot product $\mathbf{A}\cdot\mathbf{B}$ if $\mathbf{A} = 3\mathbf{i} + 3\mathbf{j}$ and $\mathbf{B} = -5\mathbf{i}$.

SOLUTION Figure 14.2.3 shows that θ, the angle between $\mathbf{A}$ and $\mathbf{B}$, is $3\pi/4$. Also,

$$|\mathbf{A}| = \sqrt{3^2 + 3^2} = \sqrt{18}$$

and

$$|\mathbf{B}| = \sqrt{(-5)^2 + 0^2} = 5.$$

And so $\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}||\mathbf{B}|\cos\theta = \sqrt{18}\,(5)\left(-\sqrt{2}/2\right) = -15$. ♦

EXAMPLE 2. Find (a) $\mathbf{i}\cdot\mathbf{j}$, (b) $\mathbf{i}\cdot\mathbf{i}$, and (c) $2\mathbf{k}\cdot(-3\mathbf{k})$.

SOLUTION

(a) The angle between $\mathbf{i}$ and $\mathbf{j}$ is $\pi/2$. Thus $\mathbf{i}\cdot\mathbf{j} = |\mathbf{i}||\mathbf{j}|\cos\left(\frac{\pi}{2}\right) = 1\cdot 1\cdot 0 = 0$.

(b) The angle between $\mathbf{i}$ and $\mathbf{i}$ is 0. Thus $\mathbf{i}\cdot\mathbf{i} = |\mathbf{i}||\mathbf{i}|\cos(0) = 1\cdot 1\cdot 1 = 1$.

(c) The angle between $2\mathbf{k}$ and $-3\mathbf{k}$ is π. Thus $2\mathbf{k}\cdot(-3\mathbf{k}) = |2\mathbf{k}||-3\mathbf{k}|\cos(\pi) = 2\cdot 3\cdot(-1) = -6$.

♦

Computations like those in Example 2 show that $a\mathbf{i}\cdot b\mathbf{i} = ab$, $a\mathbf{j}\cdot b\mathbf{j} = ab$, and $a\mathbf{k}\cdot b\mathbf{k} = ab$, while $a\mathbf{i}\cdot b\mathbf{j} = 0$, $a\mathbf{i}\cdot b\mathbf{k} = 0$, and $a\mathbf{j}\cdot b\mathbf{k} = 0$. In particular, $\mathbf{i}\cdot\mathbf{i} = \mathbf{j}\cdot\mathbf{j} = \mathbf{k}\cdot\mathbf{k} = 1$, while $\mathbf{i}\cdot\mathbf{j} = \mathbf{i}\cdot\mathbf{k} = \mathbf{j}\cdot\mathbf{k} = 0$.

The Geometry of the Dot Product

Let $\mathbf{A}$ and $\mathbf{B}$ be nonzero vectors and θ the angle between them. Their dot product is $\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}||\mathbf{B}|\cos(\theta)$.

The quantities $|\mathbf{A}|$ and $|\mathbf{B}|$, being the lengths of nonzero vectors, are positive. However, $\cos(\theta)$ can be positive, zero, or negative. Nonzero vectors $\mathbf{A}$ and $\mathbf{B}$ are said to be **perpendicular** when $\cos(\mathbf{A},\mathbf{B}) = 0$, that is, when the angle between $\mathbf{A}$ and $\mathbf{B}$ is $\theta = \pi/2$. And, the zero vector is perpendicular to every vector. The dot product thus provides a way of telling whether $\mathbf{A}$ and $\mathbf{B}$ are perpendicular:

Theorem 14.2.3: A Test for Perpendicular Vectors

Let $\mathbf{A}$ *and* $\mathbf{B}$ *be vectors. If* $\mathbf{A}\cdot\mathbf{B} = 0$, *then* $\mathbf{A}$ *and* $\mathbf{B}$ *are perpendicular. Conversely, if* $\mathbf{A}$ *and* $\mathbf{B}$ *are perpendicular, then* $\mathbf{A}\cdot\mathbf{B} = 0$.

REMINDER: The zero vector is both parallel and perpendicular to every vector.

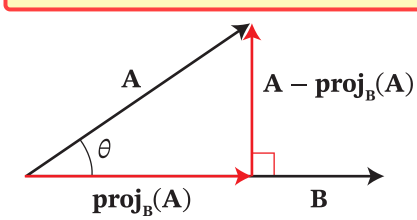


Figure 14.2.4

As Figure 14.2.4 shows, $\mathbf{A}$ can be expressed as the sum of a vector parallel to $\mathbf{B}$ and a vector perpendicular to $\mathbf{B}$. The one parallel to $\mathbf{B}$ is $|\mathbf{A}|\cos(\theta)$ times the unit vector $\mathbf{B}/|\mathbf{B}|$, that is

$$|\mathbf{A}|\cos(\theta)\frac{\mathbf{B}}{|\mathbf{B}|}$$

which can be rewritten as

$$\left(\mathbf{A}\cdot\frac{\mathbf{B}}{|\mathbf{B}|}\right)\left(\frac{\mathbf{B}}{|\mathbf{B}|}\right).$$

Definition: *Projection of* A *onto* B

Given a vector **A** and a nonzero vector **B**, as shown in Figure 14.2.4, the component of **A** parallel to **B** is called the **projection of A on B**:

$$\mathbf{proj_B}(\mathbf{A}) = \frac{\mathbf{A}\cdot\mathbf{B}}{|\mathbf{B}|}\frac{\mathbf{B}}{|\mathbf{B}|} = \frac{(\mathbf{A}\cdot\mathbf{B})\mathbf{B}}{|\mathbf{B}|^2}.$$

The component of **A** perpendicular to **B** is then $\mathbf{A} - \mathbf{proj_B}(\mathbf{A})$.

This definition is much simpler if the unit vector $\mathbf{B}/|\mathbf{B}|$ is denoted **u**. Then $\mathbf{proj_B}(\mathbf{A})$ is simply $(\mathbf{A}\cdot\mathbf{u})\mathbf{u}$.

The length of $\mathbf{proj_B}(\mathbf{A})$ is $|\mathbf{A}||\cos(\theta)|$, which equals $\dfrac{|\mathbf{A}\cdot\mathbf{B}|}{|\mathbf{B}|}$.

If θ is less than $\pi/2$, $\mathbf{proj_B}(\mathbf{A})$ points in the same direction as **B**. If $\pi/2 < \theta \le \pi$, then $\mathbf{proj_B}(\mathbf{A})$ points in the direction opposite to that of **B**. In either case, since $\mathbf{B}/|\mathbf{B}|$ is the unit vector in the direction of **B**; we have

Observation 14.2.4: *The Dot Product and the Angle Between Vectors*

Let **A** and **B** be nonzero vectors.

1. If $\mathbf{A}\cdot\mathbf{B}$ is positive, then the angle between the vectors is less than $\pi/2$ and $\mathbf{proj_B}(\mathbf{A})$ points in the same direction as **B**.
2. If $\mathbf{A}\cdot\mathbf{B}$ is negative, then the angle between the vectors is greater than $\pi/2$ and $\mathbf{proj_B}(\mathbf{A})$ points in the direction opposite that of **B**. (See Figure 14.2.5.)

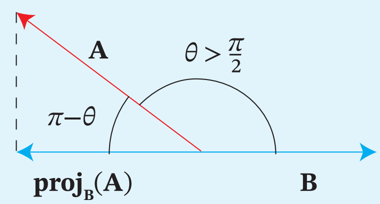


Figure 14.2.5

Computing A · B in Terms of Components

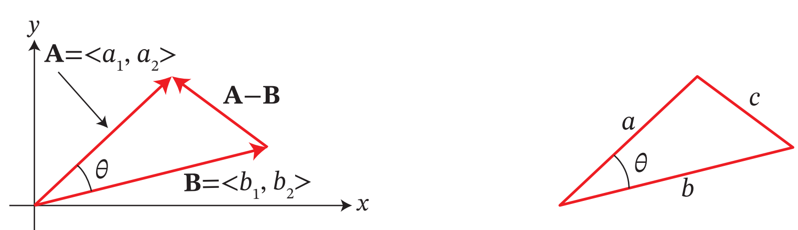
We defined $\mathbf{A}\cdot\mathbf{B}$, using the geometric interpretation of **A** and **B**. But what if **A** and **B** are given in terms of their components along **i**, **j**, and **k**, $\mathbf{A} = \langle a_1, a_2, a_3\rangle$ and $\mathbf{B} = \langle b_1, b_2, b_3\rangle$? Or, what if $\mathbf{A} = \langle a_1, a_2\rangle$ and $\mathbf{B} = \langle b_1, b_2\rangle$? How would we find $\mathbf{A}\cdot\mathbf{B}$ in those cases?

The answer turns out to be simple: in both cases the dot product is the sum of products of corresponding components.

Formula 14.2.1: *Component Form of the Dot Product*

1. If $\mathbf{A} = \langle a_1, a_2\rangle$ and $\mathbf{B} = \langle b_1, b_2\rangle$, then $\mathbf{A}\cdot\mathbf{B} = a_1b_1 + a_2b_2$.
2. If $\mathbf{A} = \langle a_1, a_2, a_3\rangle$ and $\mathbf{B} = \langle b_1, b_2, b_3\rangle$, then $\mathbf{A}\cdot\mathbf{B} = a_1b_1 + a_2b_2 + a_3b_3$.

The dot product is the sum of two or three numbers, the products of corresponding components.



(a) (b)

Figure 14.2.6

These formulas depend on the law of cosines, which asserts that in a triangle whose sides have lengths a, b, and c, and angle θ opposite the side with length c, as in Figure 14.2.6(b), $c^2 = a^2 + b^2 - 2ab\cos(\theta)$.

For simplicity we obtain the formula for vectors in the xy-plane. In terms of the triangle in Figure 14.2.6(a),

$$|\mathbf{A}-\mathbf{B}|^2 = |\mathbf{A}|^2 + |\mathbf{B}|^2 - 2|\mathbf{A}||\mathbf{B}|\cos(\theta),$$

which tells us that

$$|\mathbf{A}-\mathbf{B}|^2 = |\mathbf{A}|^2 + |\mathbf{B}|^2 - 2\mathbf{A}\cdot\mathbf{B}. \tag{14.2.1}$$

Translating (14.2.1) into components, we have

$$(a_1-b_1)^2 + (a_2-b_2)^2 = \left(a_1^2+a_2^2\right) + \left(b_1^2+b_2^2\right) - 2\mathbf{A}\cdot\mathbf{B}$$

or

$$a_1^2 - 2a_1b_1 + b_1^2 + a_2^2 - 2a_2b_2 + b_2^2 = a_1^2 + a_2^2 + b_1^2 + b_2^2 - 2\mathbf{A}\cdot\mathbf{B}.$$

Hence

$$-2(a_1b_1 + a_2b_2) = -2\mathbf{A}\cdot\mathbf{B}.$$

And, finally,

$$\mathbf{A}\cdot\mathbf{B} = a_1b_1 + a_2b_2.$$

The argument for space vectors is essentially the same, as answering Exercise 30 will show.

The formula for the dot product in terms of components can be used in a straightforward manner to prove the following distributive property of the dot product over addition. (See Exercise 31.)

Theorem 14.2.5: Distributive Property of the Dot Product

For any vectors $\mathbf{A}$, $\mathbf{B}$, *and* $\mathbf{C}$,

$$\mathbf{A}\cdot(\mathbf{B}+\mathbf{C}) = \mathbf{A}\cdot\mathbf{B} + \mathbf{A}\cdot\mathbf{C}. \tag{14.2.2}$$

EXAMPLE 3. Find $\cos(\mathbf{A},\mathbf{B})$ when $\mathbf{A} = \langle 6,3\rangle$ and $\mathbf{B} = \langle -1,1\rangle$.

SOLUTION We know that $\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}||\mathbf{B}|\cos(\mathbf{A},\mathbf{B})$, where

$$\mathbf{A}\cdot\mathbf{B} = 6\cdot(-1) + 3\cdot(1) = -3, \qquad |\mathbf{A}| = \sqrt{6^2+3^2} = \sqrt{45} = 3\sqrt{5}, \quad \text{and} \quad |\mathbf{B}| = \sqrt{(-1)^2+1^2} = \sqrt{2}.$$

Thus

$$-3 = 3\sqrt{5}\sqrt{2}\cos(\mathbf{A},\mathbf{B})$$

from which we conclude that $\cos(\mathbf{A},\mathbf{B}) = -1/\sqrt{10}$.

The angle between $\mathbf{A}$ and $\mathbf{B}$ is obtuse. A calculator would give a numerical estimate of its value. ♦

As Example 3 illustrates, the dot product can be used to find the cosine of the angle between two vectors and therefore the angle itself:

Formula 14.2.2: ***Cosine of the Angle between Two Nonzero Vectors***

Let $\mathbf{A}$ and $\mathbf{B}$ be two nonzero vectors. The cosine of the angle between $\mathbf{A}$ and $\mathbf{B}$ can be computed using the formula

$$\cos(\theta) = \cos(\mathbf{A},\mathbf{B}) = \frac{\mathbf{A}\cdot\mathbf{B}}{|\mathbf{A}||\mathbf{B}|}.$$

EXAMPLE 4.

(a) Find the projection of $\mathbf{A} = 3\mathbf{i} + 2\mathbf{j}$ on $\mathbf{B} = -3\mathbf{i} + 2\mathbf{j}$.
(b) Express **A** as the sum of a vector parallel to **B** and a vector perpendicular to **B**.

SOLUTION

(a) From the definition of the projection of **A** onto **B**:

$$\begin{aligned}\mathbf{proj_B}(\mathbf{A}) &= \frac{\mathbf{A}\cdot\mathbf{B}}{|\mathbf{B}|}\frac{\mathbf{B}}{|\mathbf{B}|}\\ &= \frac{(3\mathbf{i}+2\mathbf{j})\cdot(-3\mathbf{i}+2\mathbf{j})}{|-3\mathbf{i}+2\mathbf{j}|}\frac{-3\mathbf{i}+2\mathbf{j}}{|-3\mathbf{i}+2\mathbf{j}|}\\ &= \frac{(-9+4)}{\sqrt{13}}\frac{(-3\mathbf{i}+2\mathbf{j})}{\sqrt{13}}\\ &= \frac{-5}{13}(-3\mathbf{i}+2\mathbf{j}) = \frac{15}{13}\mathbf{i} - \frac{10}{13}\mathbf{j}.\end{aligned}$$

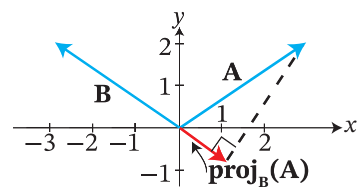


Figure 14.2.7

Figure 14.2.7 shows **A**, **B**, and $\mathbf{proj_B}(\mathbf{A})$.

In this case $\mathbf{A}\cdot\mathbf{B}$ is negative, the angle between **A** and **B** is obtuse, and $\mathbf{proj_B}(\mathbf{A})$ points in the direction opposite the direction of **B**, in fact, $\mathbf{proj_B}(\mathbf{A}) = -5\mathbf{B}/13$.

(b) The vector $\mathbf{proj_B}(\mathbf{A})$ found in (a) is parallel to **B**. Subtracting this from **A** produces a vector perpendicular to **B**:

$$\mathbf{A} - \mathbf{proj_B}(\mathbf{A}) = 3\mathbf{i} + 2\mathbf{j} - \left(\frac{15}{13}\mathbf{i} - \frac{10}{13}\mathbf{j}\right) = \frac{24}{13}\mathbf{i} + \frac{36}{13}\mathbf{j}.$$

(A quick computation verifies that $(\mathbf{A} - \mathbf{proj_B}(\mathbf{A}))\cdot\mathbf{B} = 0$.) Therefore, the vector $\mathbf{A} - \mathbf{proj_B}(\mathbf{A})$ is perpendicular to **B** and we have

$$\mathbf{A} = \mathbf{proj_B}(\mathbf{A}) + (\mathbf{A} - \mathbf{proj_B}(\mathbf{A})) = \underbrace{\left(\frac{15}{13}\mathbf{i} - \frac{10}{13}\mathbf{j}\right)}_{\text{parallel to } \mathbf{B}} + \underbrace{\left(\frac{24}{13}\mathbf{i} + \frac{36}{13}\mathbf{j}\right)}_{\text{perpendicular to } \mathbf{B}}.$$

◆

The scalar $\mathbf{A}\cdot(\mathbf{B}/|\mathbf{B}|)$ is the **scalar component of A in the direction of B**, denoted $\text{comp}_{\mathbf{B}}(\mathbf{A})$. It can be positive, negative, or zero. Its absolute value is the length of $\mathbf{proj_B}(\mathbf{A})$.

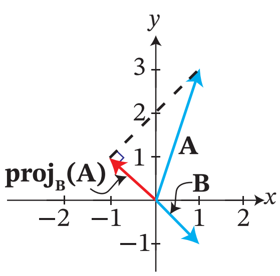


Figure 14.2.8

EXAMPLE 5. Find $\mathbf{proj_B}(\mathbf{A})$ and $\text{comp}_{\mathbf{B}}(\mathbf{A})$ when $\mathbf{A} = \mathbf{i} + 3\mathbf{j}$ and $\mathbf{B} = \mathbf{i} - \mathbf{j}$.

SOLUTION Since $|\mathbf{B}| = \sqrt{1^2 + (-1)^2} = \sqrt{2}$ and $\mathbf{A}\cdot\mathbf{B} = 1 - 3 = -2$,

$$\mathbf{proj_B}(\mathbf{A}) = \frac{\mathbf{A}\cdot\mathbf{B}}{|\mathbf{B}|}\frac{\mathbf{B}}{|\mathbf{B}|} = \frac{-2}{\sqrt{2}}\frac{(\mathbf{i}-\mathbf{j})}{\sqrt{2}} = -\mathbf{i} + \mathbf{j}$$

and

$$\text{comp}_{\mathbf{B}}(\mathbf{A}) = (\mathbf{A}\cdot\mathbf{B})/|\mathbf{B}| = -2/\sqrt{2} = -\sqrt{2}.$$

This agrees with Figure 14.2.8. ◆

EXAMPLE 6. Show that the vectors $\langle 1, 2, 1\rangle$ and $\langle 2, -3, 4\rangle$ are perpendicular.

SOLUTION All we need to do is compute the dot product of **A** and **B** and see if it is 0. We have

$$\mathbf{A}\cdot\mathbf{B} = (1)(2) + (2)(-3) + (1)(4) = 2 - 6 + 4 = 0,$$

Consequently, the vectors **A** and **B** are perpendicular. ◆

Summary

The dot (scalar) product of two vectors **A** and **B** is defined geometrically as $|\mathbf{A}||\mathbf{B}|\cos(\theta)$, where θ is the angle between **A** and **B**. For two vectors **A** and **B** in the plane, the formula for $\mathbf{A}\cdot\mathbf{B}$ in terms of their components is:

$$\langle a_1, a_2\rangle \cdot \langle b_1, b_2\rangle = a_1 b_1 + a_2 b_2$$

and there is a similar formula for the dot product of two space vectors.

The dot product enables us to express a vector **A** as the sum of a vector parallel to **B**, $\mathbf{proj_B}(\mathbf{A})$, and one perpendicular to **B**, $\mathbf{A} - \mathbf{proj_B}(\mathbf{A})$.

When the dot product of two vectors is 0 they are perpendicular.

The zero vector, **0**, is considered to be perpendicular to every vector.

The dot product can be used to find the angle θ between two vectors: $\cos(\theta) = \cos(\mathbf{A},\mathbf{B}) = \dfrac{\mathbf{A}\cdot\mathbf{B}}{|\mathbf{A}||\mathbf{B}|}$.

Application 14.2.6: *The Dot Product in Business and Statistics*

Imagine that a fast food restaurant sells 30 hamburgers, 20 salads, 15 soft drinks, and 13 orders of French fries. This is recorded by the four-dimensional "vector" $\langle 30, 20, 15, 13\rangle$. A hamburger sells for \$1.99, a salad for \$1.50, a soft drink for \$1.00, and an order of French fries for \$1.10. The "price vector" is $\langle 1.99, 1.50, 1.00, 1.10\rangle$. The dot product of these two vectors,

$$\langle 30, 20, 15, 13\rangle \cdot \langle 1.99, 1.50, 1.00, 1.10\rangle = 30(1.99) + 20(1.50) + 15(1.00) + 13(1.10),$$

would be the total amount paid for all items.

Descriptions of the economy use "production vectors," "cost vectors," "price vectors," and "profit vectors" with many more than the four components of our restaurant example.

In statistics the **coefficient of correlation** is defined in terms of a dot product. For instance, say you have the heights and weights of n persons. Let the height of the ith person be h_i and the weight be w_i. Let h be the average of the n heights and w be the average of the n weights. Let $\mathbf{H} = \langle h_1 - h, h_2 - h, \cdots, h_n - h\rangle$ and $\mathbf{W} = \langle w_1 - w, w_2 - w, \cdots, w_n - w\rangle$. The coefficient of correlation between the heights and weights is defined to be

$$\frac{\mathbf{H}\cdot\mathbf{W}}{|\mathbf{H}||\mathbf{W}|}.$$

In analogy with vectors in the plane or space,

$$\mathbf{H}\cdot\mathbf{W} = \sum_{i=1}^{n}(h_i - h)(w_i - w), \qquad |\mathbf{H}| = \sqrt{\sum_{i=1}^{n}(h_i - h)^2}, \text{and} \qquad |\mathbf{W}| = \sqrt{\sum_{i=1}^{n}(w_i - w)^2}.$$

It turns out that the coefficient of correlation is simply the cosine of the angle between the vectors **H** and **W** in n-dimensional space. The closer it is to 1, the closer the angle between the vectors is to 0, and the better the weights and heights "correlate."

EXERCISES for Section 14.2

In Exercises 1 to 4 compute $\mathbf{A} \cdot \mathbf{B}$.

1. $\mathbf{A}$ has length 3, $\mathbf{B}$ has length 4, and the angle between them is $\frac{\pi}{4}$.

2. $\mathbf{A}$ has length 2, $\mathbf{B}$ has length 3, and the angle between them is $\frac{3\pi}{4}$.

3. $\mathbf{A}$ has length 5, $\mathbf{B}$ has length $\frac{1}{2}$, and the angle between them is $\frac{\pi}{2}$.

4. $\mathbf{A}$ is the zero vector $\mathbf{0}$, and $\mathbf{B}$ has length 5.

In Exercises 5 to 8 compute $\mathbf{A} \cdot \mathbf{B}$.

5. $\mathbf{A} = -2\mathbf{i} + 3\mathbf{j}$, $\mathbf{B} = 4\mathbf{i} + 4\mathbf{j}$

6. $\mathbf{A} = 0.3\mathbf{i} + 0.5\mathbf{j}$, $\mathbf{B} = 2\mathbf{i} - 1.5\mathbf{j}$

7. $\mathbf{A} = 2\mathbf{i} - 3\mathbf{j} - \mathbf{k}$, $\mathbf{B} = 3\mathbf{i} + 4\mathbf{j} - \mathbf{k}$

8. $\mathbf{A} = \mathbf{i} + \mathbf{j} + \mathbf{k}$, $\mathbf{B} = 2\mathbf{i} + 3\mathbf{j} - 5\mathbf{k}$

---

9. (a) Draw the vectors $7\mathbf{i} + 12\mathbf{j}$ and $9\mathbf{i} - 5\mathbf{j}$.
(b) Do they seem to be perpendicular?
(c) Determine whether they are perpendicular by examining their dot product.

10. (a) Draw the vectors $\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$ and $\mathbf{i} + \mathbf{j} - \mathbf{k}$.
(b) Do they seem to be perpendicular?
(c) Determine whether they are perpendicular by examining their dot product.

11. (a) Estimate the angle between $\mathbf{A} = 3\mathbf{i} + 4\mathbf{j}$ and $\mathbf{B} = 5\mathbf{i} + 12\mathbf{j}$ by drawing them.
(b) Find the angle between them.

12. Let $P = (6,1)$, $Q = (3,2)$, $R = (1,3)$, and $S = (4,5)$.
(a) Draw the vectors $\overrightarrow{PQ}$ and $\overrightarrow{RS}$.
(b) Using the diagram estimate the angle between $\overrightarrow{PQ}$ and $\overrightarrow{RS}$.
(c) Using the dot product, find $\cos(\overrightarrow{PQ}, \overrightarrow{RS})$.
(d) Using a calculator, find the angle.

13. Find the angle between $2\mathbf{i} - 4\mathbf{j} + 6\mathbf{k}$ and $\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$.

14. Find the angle between $\mathbf{i} + \mathbf{j} + 3\mathbf{k}$ and $3\mathbf{i} + 6\mathbf{j} - 3\mathbf{k}$.

15. Find the angle between $\overrightarrow{AB}$ and $\overrightarrow{CD}$ if $A = (1,3)$, $B = (7,4)$, $C = (2,8)$, and $D = (1,-5)$.

16. Find the angle between $\overrightarrow{AB}$ and $\overrightarrow{CD}$ if $A = (1,2,-5)$, $B = (1,0,1)$, $C = (0,-1,3)$, and $D = (2,1,4)$.

17. Find the length of the projection of $-4\mathbf{i} + 5\mathbf{j}$ on the line through $(2,-1)$ and $(6,1)$.
(a) By making a drawing and estimating the length by eye.
(b) By using the dot product.

18. (a) Find a vector $\mathbf{C}$ parallel to $\mathbf{i} + 2\mathbf{j}$ and a vector $\mathbf{D}$ perpendicular to $\mathbf{i} + 2\mathbf{j}$ such that $-3\mathbf{i} + 4\mathbf{j} = \mathbf{C} + \mathbf{D}$.
(b) Draw the vectors to check that your answer is reasonable.

19. (a) Find a vector $\mathbf{C}$ parallel to $2\mathbf{i} - \mathbf{j}$ and a vector $\mathbf{D}$ perpendicular to $2\mathbf{i} - \mathbf{j}$ such that $3\mathbf{i} + 4\mathbf{j} = \mathbf{C} + \mathbf{D}$.
(b) Draw the vectors to check that your answer is reasonable.

20. What is $\mathbf{proj_B}(\mathbf{A})$ if $\mathbf{A} = 2\mathbf{i} + \mathbf{j} - 3\mathbf{k}$ and $\mathbf{B} = \mathbf{i} + \mathbf{j} + \mathbf{k}$?

21. Express the vector $\mathbf{i} + \mathbf{j} + \mathbf{k}$ as the sum of a vector parallel to $\mathbf{i} - \mathbf{j} + 2\mathbf{k}$ and a vector perpendicular to $\mathbf{i} - \mathbf{j} + 2\mathbf{k}$.

22. Give an example of a nonzero vector in the xy-plane that is perpendicular to $3\mathbf{i} - 2\mathbf{j}$.

23. Give an example of a nonzero vector that is perpendicular to $5\mathbf{i} - 3\mathbf{j} + 4\mathbf{k}$.

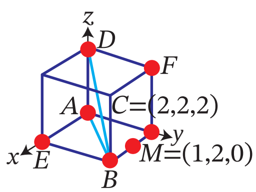


Figure 14.2.9

Exercises 24 to 29 refer to the cube in Figure 14.2.9.

24. Find $\cos(\overrightarrow{AB}, \overrightarrow{BD})$.

25. Find $\cos(\overrightarrow{AF}, \overrightarrow{BD})$.

26. Find $\cos(\overrightarrow{AB}, \overrightarrow{AM})$.

27. Find $\cos(\overrightarrow{MD}, \overrightarrow{MF})$.

28. Find $\cos(\overrightarrow{AE}, \overrightarrow{CF})$.

29. Find $\cos(\overrightarrow{EF}, \overrightarrow{BD})$.

30. Prove that $\langle a_1, a_2, a_3\rangle \cdot \langle b_1, b_2, b_3\rangle = a_1b_1 + a_2b_2 + a_3b_3$.

Exercises 31 and 32 deal with the **distributive property for the dot product** (Theorem 14.2.5)

$$\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C}.$$

31. (a) Prove Theorem 14.2.5 when the vectors **A**, **B**, and **C** are vectors in the xy-plane.
(b) Prove Theorem 14.2.5 when the vectors **A**, **B**, and **C** are vectors in space.

32. Show that (14.2.2) implies $\mathbf{A} \cdot (\mathbf{B} + \mathbf{C} + \mathbf{D}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C} + \mathbf{A} \cdot \mathbf{D}$.

---

33. If $\mathbf{A} \cdot \mathbf{B} = \mathbf{A} \cdot \mathbf{C}$ and $\mathbf{A}$ is not $\mathbf{0}$, must $\mathbf{B} = \mathbf{C}$? Explain your answer.

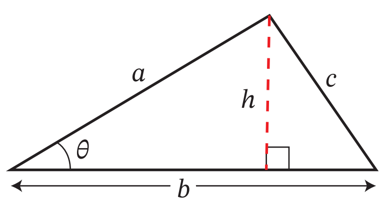


Figure 14.2.10

34. We found the dot product in terms of components by using the Law of Cosines. We now see why the Law of Cosines is true. The proof consists of two applications of the Pythagorean Theorem. Figure 14.2.10 shows a triangle with sides a, b, c, with angle θ opposite side c. (We illustrate the case where θ is less than $\frac{\pi}{2}$.)
(a) Show that $h^2 = a^2 - a^2\cos^2(\theta)$.
(b) Show that $h^2 = c^2 - (b - a\cos(\theta))^2$.
(c) By equating the two expressions for h^2, obtain the Law of Cosines.
(d) Carry out the proof when θ is greater than $\pi/2$.

35. (a) $\mathbf{A}$ is a vector in the xy-plane and $\mathbf{u}_1$ and $\mathbf{u}_2$ are perpendicular unit vectors in the xy-plane. If $\mathbf{A} \cdot \mathbf{u}_1 = 0$ and $\mathbf{A} \cdot \mathbf{u}_2 = 0$, must $\mathbf{A} = \mathbf{0}$?
(b) $\mathbf{v}_1$ and $\mathbf{v}_2$ are nonparallel unit vectors in the xy-plane and $\mathbf{A}$ lies in the xy-plane. If $\mathbf{A} \cdot \mathbf{v}_1 = 0$ and $\mathbf{A} \cdot \mathbf{v}_2 = 0$, must $\mathbf{A} = \mathbf{0}$?

36. JANE: I don't like the way the author found how to express **A** as the sum of a vector parallel to **B** and a vector perpendicular to **B**.

SAM: It was O.K. for me. But I had to memorize a formula.

JANE: My goal is to memorize nothing. I write $\mathbf{A} = x\mathbf{B} + \mathbf{C}$, when **C** is perpendicular to **A**. Then I dot with **B**, getting $\mathbf{A} \cdot \mathbf{B} = x\mathbf{B} \cdot \mathbf{B} + \mathbf{C} \cdot \mathbf{B}$. Since **C** is perpendicular to **B**, $\mathbf{C} \cdot \mathbf{B} = 0$, and lo and behold, I have $x = \dfrac{\mathbf{A} \cdot \mathbf{B}}{\mathbf{B} \cdot \mathbf{B}}$. So the vector parallel to **B** is $\dfrac{\mathbf{A} \cdot \mathbf{B}}{\mathbf{B} \cdot \mathbf{B}}\mathbf{B}$.

SAM: Cool. So why did the author go through all that stuff?

Jane: Maybe they wanted to reinforce the definition of the dot product and the role of the angle.
Sam: O.K. But how do I get the vector **C** perpendicular to **B**?
Jane: Simple...

Complete Jane's reply.

37. By taking the dot product of the unit vectors $\mathbf{u}_1 = \cos(\theta_1)\mathbf{i} + \sin(\theta_1)\mathbf{j}$ and $\mathbf{u}_2 = \cos(\theta_2)\mathbf{i} + \sin(\theta_2)\mathbf{j}$, prove that

$$\cos(\theta_1 - \theta_2) = \cos(\theta_1)\cos(\theta_2) + \sin(\theta_1)\sin(\theta_2).$$

38. A tetrahedron, not necessarily regular, has six edges. Show that the line segment joining the midpoints of two opposite edges is perpendicular to the line segment joining another pair of opposite edges if and only if the remaining two edges have the same length.

39. The output of a firm that manufactures x_1 washing machines, x_2 refrigerators, x_3 dishwashers, x_4 stoves, and x_5 clothes dryers is recorded by the production vector $\mathbf{P} = \langle x_1, x_2, x_3, x_4, x_5\rangle$. Similarly, the cost vector $\mathbf{C} = \langle y_1, y_2, y_3, y_4, y_5\rangle$ records the cost of producing each item.

(a) What is the economic significance of $\mathbf{P}\cdot\mathbf{C} = \langle 20, 0, 7, 9, 15\rangle \cdot \langle 50, 70, 30, 20, 10\rangle$?
(b) If the firm doubles the production of all items, what is its new production vector?

40. Assume that the profit from selling a washing machine is P_1 and that P_2, P_3, P_4, and P_5 are defined analogously for the firm of Exercise 39. What does it mean to the firm to have $\langle P_1, P_2, P_3, P_4, P_5\rangle$ perpendicular to $\langle x_1, x_2, x_3, x_4, x_5\rangle$?

41. Prove that $\mathbf{A}\cdot\mathbf{B} = \mathbf{B}\cdot\mathbf{A}$

(a) using the geometric definition of the dot product,
(b) using the formula for the dot product in terms of components.

42. Assume that $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ are unit vectors such that any two are perpendicular.

(a) Draw a picture that shows for any vector **A** that there are scalars x, y, and z such that $\mathbf{A} = x\mathbf{u}_1 + y\mathbf{u}_2 + z\mathbf{u}_3$.
(b) Find **B** such that $\mathbf{A}\cdot\mathbf{B} = x$.
(c) Find **C** such that $\mathbf{A}\cdot\mathbf{C} = x - z$.

43. For vectors **A** and **B**, with $\mathbf{B} \neq \mathbf{0}$, a vector parallel to **B** is $\mathbf{proj_B}(\mathbf{A})$. We then used a picture to suggest that $\mathbf{A} - \mathbf{proj_B}(\mathbf{A})$ is perpendicular to **B**. Using the dot product, show that $\mathbf{A} - \mathbf{proj_B}(\mathbf{A})$ and **B** are perpendicular.

44. A force **F** of 10 newtons is parallel to $2\mathbf{i} + 3\mathbf{j} + \mathbf{k}$ and pushes an object on a ramp in a straight line from the point $(3, 1, 5)$ to the point $(4, 3, 7)$, where coordinates are measured in meters. How much work does the force accomplish?

45. Show that if the two diagonals of a parallelogram are perpendicular, then its four sides have the same length (forming a rhombus).

46. How far is the point $(2, 3, 5)$ from the line through the origin and $(1, -1, 2)$? (Use the dot product, not calculus.)

47. How far is the point $(1, 2, 3)$ from the line through $(1, 4, 2)$ and $(2, 1, -4)$? (Use the dot product, not calculus.)

48. Some molecules, such as methane, consist of four atoms arranged as the vertices of a tetrahedron, the points labeled O, A, B, and C in Figure 14.2.11.

(a) Show that O, A, B, and C are the vertices of a regular tetrahedron.
(b) Chemists are interested in the angle $\theta = \angle ADB$, where D is the center of the tetrahedron. Show that $\cos(\theta) = \dfrac{-1}{3}$.
(c) Find θ (approximately).

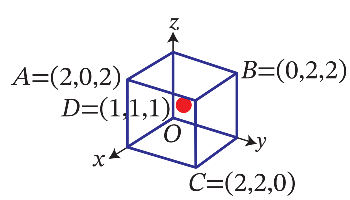


Figure 14.2.11

14.3 The Cross Product of Two Vectors

The dot product of two vectors is a scalar. In this section we define a product of two vectors that is a vector perpendicular to both.

Definition of the Cross Product

When $\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $\mathbf{B} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$ are two nonzero vectors that are not parallel we will construct a vector $\mathbf{C}$ that is perpendicular to both $\mathbf{A}$ and $\mathbf{B}$. The vector $\mathbf{C}$ is not unique since any vector parallel to $\mathbf{C}$ is also perpendicular to $\mathbf{A}$ and $\mathbf{B}$.

For $\mathbf{C} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ we want $\mathbf{C}\cdot\mathbf{A}$ and $\mathbf{C}\cdot\mathbf{B}$ to be 0. This gives us the equations

$$a_1x + a_2y + a_3z = 0 \tag{14.3.1}$$
$$b_1x + b_2y + b_3z = 0 \tag{14.3.2}$$

We eliminate x by subtracting b_1 times (14.3.1) from a_1 times (14.3.2), as follows:

$$a_1b_1x + a_1b_2y + a_1b_3z = 0 \qquad (\ a_1 \text{ times (14.3.2)}\ ) \tag{14.3.3}$$
$$b_1a_1x + b_1a_2y + b_1a_3z = 0 \qquad (\ b_1 \text{ times (14.3.1)}\ ) \tag{14.3.4}$$

Subtracting (14.3.4) from (14.3.3) gives

$$(a_1b_2 - a_2b_1)y + (a_1b_3 - a_3b_1)z = 0 \tag{14.3.5}$$

There are an infinite number of solutions of (14.3.5). One solution is

$$y = -(a_1b_3 - a_3b_1), \quad z = a_1b_2 - a_2b_1.$$

This is like solving $2y + 3z = 0$ by reversing the equation's coefficients and changing one sign: $y = -3$, $z = 2$.

Then, to find x, substitute these values for y and z into (14.3.1). As is shown in Exercise 30, some algebraic manipulations lead to

$$x = a_2b_3 - a_3b_2.$$

So the vector

$$\mathbf{A}\times\mathbf{B} = (a_2b_3 - a_3b_2)\mathbf{i} - (a_1b_3 - a_3b_1)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k} \tag{14.3.6}$$

is perpendicular to both $\mathbf{A}$ and $\mathbf{B}$. It is called the **vector product of A and B** or the **cross product of A and B**. This vector is defined even if $\mathbf{A}$ and $\mathbf{B}$ are parallel or if one (or both) of them is the zero vector, $\mathbf{0}$.

The vector $\mathbf{A}\times\mathbf{B}$ is called a cross product because the symbol $\times$ is commonly called a "cross" in mathematics. The name "vector product" reminds us that $\mathbf{A}\times\mathbf{B}$ is a vector. This is consistent with the use of both dot product and scalar product for $\mathbf{A}\cdot\mathbf{B}$, which is a scalar (number).

Determinants and the Cross Product

The expression (14.3.6) for $\mathbf{A}\times\mathbf{B}$ is not easy to memorize. Determinants provide a convenient memory aid.

Four numbers arranged in a square form a **matrix of order** 2, for instance

$$\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}.$$

Its determinant is the number

$$\det\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} = a_1b_2 - a_2b_1.$$

Each term in (14.3.6) involves the determinant of a matrix of order 2, namely

$$a_2b_3 - a_3b_2 = \det\begin{pmatrix} a_2 & a_3 \\ b_2 & b_3 \end{pmatrix},$$
$$-(a_1b_3 - a_3b_1) = -\det\begin{pmatrix} a_1 & a_3 \\ b_1 & b_3 \end{pmatrix},$$

and

$$a_1b_2 - a_2b_1 = \det\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}.$$

Nine numbers arranged in a square form a **matrix of order** 3, for instance

$$M = \begin{pmatrix} c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix}.$$

Its determinant is defined with the aid of determinants of order 2:

$$\det M = c_1\det\begin{pmatrix} a_2 & a_3 \\ b_2 & b_3 \end{pmatrix} - c_2\det\begin{pmatrix} a_1 & a_3 \\ b_1 & b_3 \end{pmatrix} + c_3\det\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2. \end{pmatrix}$$

The coefficient of each c_i is plus or minus the determinant of the matrix of order 2 that remains when the row and column in which c_i appears are deleted. For example,

$$\begin{pmatrix} c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix}$$

shows how, since c_1 is in the first row and first column, deleting the first row and first column of M produces the matrix of order 2 whose determinant is multiplied by c_1.

Therefore we can write (14.3.6) as a determinant of a matrix, and we have

$$\mathbf{A}\times\mathbf{B} = \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} \qquad (14.3.7)$$

Definition: *Cross Product (Vector Product) in Terms of Determinants.*

The **cross product**, or **vector product**, of the vectors $\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $\mathbf{B} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$ is expressed as the determinant of a 3×3 matrix:

$$\begin{aligned}\mathbf{A}\times\mathbf{B} &= \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} \\ &= \mathbf{i}\det\begin{pmatrix} a_2 & a_3 \\ b_2 & b_3 \end{pmatrix} - \mathbf{j}\det\begin{pmatrix} a_1 & a_3 \\ b_1 & b_3 \end{pmatrix} + \mathbf{k}\det\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \\ &= (a_2b_3 - a_3b_2)\mathbf{i} - (a_1b_3 - a_3b_1)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k}.\end{aligned}$$

Equation (14.3.8) shows how the determinant for $\mathbf{A}\times\mathbf{B}$ can be evaluated by expanding the 3×3 matrix along its first row:

$$\mathbf{A}\times\mathbf{B} = \mathbf{i}\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} - \mathbf{j}\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} + \mathbf{k}\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} \qquad (14.3.8)$$

EXAMPLE 1. Compute $\mathbf{A}\times\mathbf{B}$ if $\mathbf{A} = 2\mathbf{i}-\mathbf{j}+3\mathbf{k}$ and $\mathbf{B} = 3\mathbf{i}+4\mathbf{j}+\mathbf{k}$.

SOLUTION By definition,

$$\mathbf{A}\times\mathbf{B} = \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 2 & -1 & 3 \\ 3 & 4 & 1 \end{pmatrix} = \mathbf{i}\det\begin{pmatrix} -1 & 3 \\ 4 & 1 \end{pmatrix} - \mathbf{j}\det\begin{pmatrix} 2 & 3 \\ 3 & 1 \end{pmatrix} + \mathbf{k}\det\begin{pmatrix} 2 & -1 \\ 3 & 4 \end{pmatrix} = -13\mathbf{i}+7\mathbf{j}+11\mathbf{k}$$

◆

Theorem 14.3.1: Properties of the Cross Product

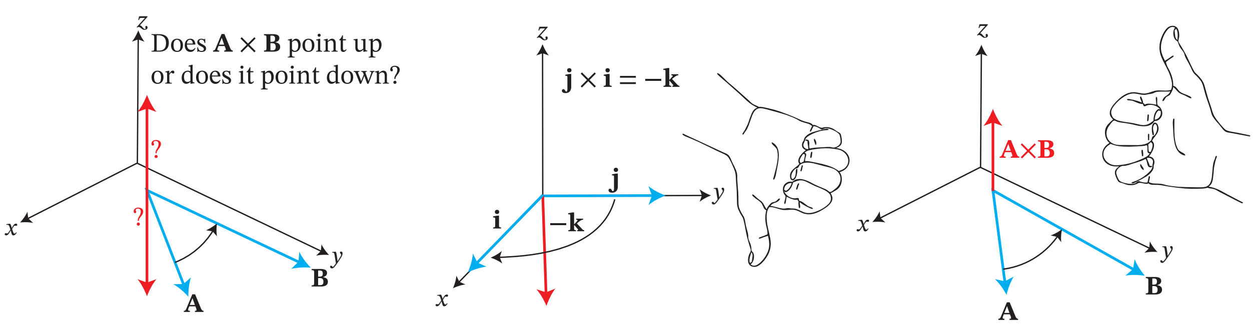
For any vectors $\mathbf{A}$, $\mathbf{B}$, *and* vC, *the cross product has these properties:*

1. $\mathbf{A}\times\mathbf{B}$ *is perpendicular to both* $\mathbf{A}$ *and* $\mathbf{B}$, *that is,* $(\mathbf{A}\times\mathbf{B})\cdot\mathbf{A} = 0$ *and* $(\mathbf{A}\times\mathbf{B})\cdot\mathbf{B} = 0$.
2. $\mathbf{A}\times\mathbf{B} = -(\mathbf{B}\times\mathbf{A})$, *so the cross product is not commutative.*
3. $\mathbf{A}\times\mathbf{B} = \mathbf{0}$ *if* $\mathbf{A}$ *and* $\mathbf{B}$ *are parallel, or if at least one of them is* $\mathbf{0}$.
4. $\mathbf{A}\times(\mathbf{B}+\mathbf{C}) = \mathbf{A}\times\mathbf{B}+\mathbf{A}\times\mathbf{C}$, *the cross product distributes over addition.*

The first property holds because that is how we constructed the cross product. The second and third are established in Exercises 16 and 17. The fourth can be verified by straightforward computations, using (14.3.7).

What is the Direction of $\mathbf{A}\times\mathbf{B}$?

We constructed $\mathbf{A}\times\mathbf{B}$ to be perpendicular to $\mathbf{A}$ and $\mathbf{B}$, but which of the two possible directions does it have? (See Figure 14.3.1(a).)



(a) (b) (c)

Figure 14.3.1

To see which, take a specific case, say $\mathbf{j}\times\mathbf{i}$:

$$\mathbf{j}\times\mathbf{i} = \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = 0\mathbf{i}-0\mathbf{j}-\mathbf{k} = -\mathbf{k}.$$

Figure 14.3.1(b) shows that if you curl the fingers of your right hand from $\mathbf{j}$ towards $\mathbf{i}$, your thumb points in the direction $-\mathbf{k}$. This suggests that the direction of $\mathbf{A}\times\mathbf{B}$ is given by the **right hand rule**:

Observation 14.3.2: *Right-Hand Rule for Finding the Direction of* $\mathbf{A}\times\mathbf{B}$

Orient your right hand so the fingers curl from $\mathbf{A}$ to $\mathbf{B}$. Then, the thumb points in the direction of $\mathbf{A}\times\mathbf{B}$. (See Figure 14.3.1(c).)

The next example gives a geometric application of the cross product.

EXAMPLE 2. Find a vector perpendicular to the plane determined by the three points $P = (3,1,2)$, $Q = (0,1,6$, and $R = (-1,4,3)$.

SOLUTION The vectors $\overrightarrow{PQ}$ and $\overrightarrow{PR}$ lie in a plane (see Figure 14.3.2). The vector $\mathbf{N} = \overrightarrow{PQ} \times \overrightarrow{PR}$, being perpendicular to both $\overrightarrow{PQ}$ and $\overrightarrow{PR}$, is perpendicular to the plane. Because $\overrightarrow{PQ} = -3\mathbf{i} + 0\mathbf{j} + 4\mathbf{k}$ and $\overrightarrow{PR} = -4\mathbf{i} + 3\mathbf{j} + 1\mathbf{k}$, it follows that

$$\mathbf{N} = \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -3 & 0 & 4 \\ -4 & 3 & 1 \end{pmatrix} = -12\mathbf{i} - 13\mathbf{j} - 9\mathbf{k}.$$

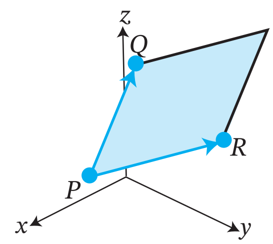


Figure 14.3.2

♦

Length of $\mathbf{A} \times \mathbf{B}$

To find the length of $\mathbf{A} \times \mathbf{B}$, $|\mathbf{A} \times \mathbf{B}|$, we will compute the square of its length:

$$\begin{aligned}
|\mathbf{A} \times \mathbf{B}|^2 &= \left|(a_2b_3 - a_3b_2)\mathbf{i} + (a_3b_1 - a_1b_3)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k}\right|^2 \\
&= (a_2b_3 - a_3b_2)^2 + (a_3b_1 - a_1b_3)^2 + (a_1b_2 - a_2b_1)^2 \\
&= a_2^2b_3^2 + a_3^2b_2^2 + a_3^2b_1^2 + a_1^2b_3^2 + a_1^2b_2^2 + a_2^2b_1^2 \\
&\quad - 2(a_2a_3b_2b_3 + a_1a_3b_1b_3 + a_1a_2b_1b_2) \\
&= (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_3)^2.
\end{aligned}$$

The first term is $|\mathbf{A}|^2|\mathbf{B}|^2$ and the second term is the square of $\mathbf{A} \cdot \mathbf{B}$. Denoting the angle between $\mathbf{A}$ and $\mathbf{B}$ by θ, we have:

$$\begin{aligned}
|\mathbf{A} \times \mathbf{B}|^2 &= (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_3)^2 \\
&= |\mathbf{A}|^2|\mathbf{B}|^2 - (\mathbf{A} \cdot \mathbf{B})^2 \\
&= |\mathbf{A}|^2|\mathbf{B}|^2 - (|\mathbf{A}||\mathbf{B}|\cos(\theta))^2 \\
&= |\mathbf{A}|^2|\mathbf{B}|^2(1 - \cos^2(\theta)) \\
&= |\mathbf{A}|^2|\mathbf{B}|^2\sin^2(\theta).
\end{aligned}$$

Then, because $\sin(\theta)$ is not negative for θ in $[0, \pi]$,

$$|\mathbf{A} \times \mathbf{B}| = |\mathbf{A}||\mathbf{B}|\sin(\theta). \tag{14.3.9}$$

Figure 14.3.3

We can now give a geometric meaning for the length of $\mathbf{A} \times \mathbf{B}$. The area of a parallelogram (see Figure 14.3.3(a)) is the product of the length of its base and its height (see Figure 14.3.3(b)). Thus,

$$\text{Area of the parallelogram spanned by } \mathbf{A} \text{ and } \mathbf{B} = \underbrace{|\mathbf{A}|}_{\text{base}}\ \underbrace{|\mathbf{B}|\sin(\theta)}_{\text{height}}.$$

Formula 14.3.1: ***The Length of* $\mathbf{A}\times\mathbf{B}$**

Let **A** and **B** be nonzero vectors and θ the angle between them. Then

$$|\mathbf{A}\times\mathbf{B}| = |\mathbf{A}||\mathbf{B}|\sin(\theta).$$

The length of $\mathbf{A}\times\mathbf{B}$ is the area of the parallelogram spanned by **A** and **B**.

EXAMPLE 3. Find the area of the parallelogram spanned by $\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j}$ and $\mathbf{B} = b_1\mathbf{i} + b_2\mathbf{j}$.

SOLUTION Write **A** as $a_1\mathbf{i} + a_2\mathbf{j} + 0\mathbf{k}$ and **B** as $b_\mathbf{i} + b_2\mathbf{j} + 0\mathbf{k}$. The area of this parallelogram is the length of $\mathbf{A}\times\mathbf{B}$. We first compute $\mathbf{A}\times\mathbf{B}$.

$$\mathbf{A}\times\mathbf{B} = \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & 0 \\ b_1 & b_2 & 0 \end{pmatrix} = (a_1b_2 - a_2b_1)\mathbf{k}.$$

The area is therefore $|a_1b_2 - a_2b_1|$. It is the absolute value of the determinant

$$\det\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}.$$

◆

We now have a completely geometric description of the cross product:

Observation 14.3.3: ***Geometric Description of* $\mathbf{A}\times\mathbf{B}$**

For any nonzero vectors **A** and **B**, $\mathbf{A}\times\mathbf{B}$ is the vector perpendicular to **A** and **B**, whose direction is given by the right-hand rule, and whose length is the area of the parallelogram that **A** and **B** span.

The Scalar Triple Product

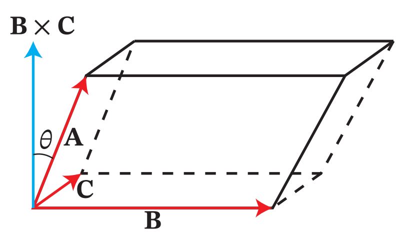


Figure 14.3.4

The scalar $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$ is called the **scalar triple product**. It has an important geometric meaning. (The vector $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$ is called the **vector triple product**. See Exercises 22, 23, and 24.)

Three vectors **A**, **B**, and **C** that do not lie in a plane span a parallelepiped, as shown in Figure 14.3.4. The angle between $\mathbf{B}\times\mathbf{C}$ and **A** is θ (which could be greater than $\pi/2$). The area of the base of the parallelogram is $|\mathbf{B}\times\mathbf{C}|$. The height of the parallelepiped is $|\mathbf{A}||\cos(\theta)|$. Thus its volume is

$$\underbrace{|\mathbf{A}||\cos(\theta)|}_{\text{height}}\ \underbrace{|\mathbf{B}\times\mathbf{C}|}_{\text{area of base}}.$$

Notice that, except for the absolute values around $\cos(\theta)$, this is the dot product of **A** and $(\mathbf{B}\times\mathbf{C})$. This connection provides a geometric interpretation of the scalar triple product.

Observation 14.3.4: ***Geometric Interpretation of* $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$**

For vectors **A**, **B**, and **C**, the scalar triple product $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$ is plus or minus the volume of the parallelepiped spanned by **A**, **B**, and **C**. If **A**, **B**, and **C** lie in a plane this volume is 0.

The scalar triple product can be expressed as a determinant. The dot product of **A** and **B** × **C** is

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = a_1 \det\begin{pmatrix} b_2 & b_3 \\ c_2 & c_3 \end{pmatrix} + a_2\left(-\det\begin{pmatrix} b_1 & b_3 \\ c_1 & c_3 \end{pmatrix}\right) + a_3 \det\begin{pmatrix} b_1 & b_2 \\ c_1 & c_2 \end{pmatrix} \tag{14.3.10}$$

which shows, by the definition of the determinant of a matrix of order 3, that

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = \det\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$$

This determinant is plus or minus the volume of the parallelepiped spanned by **A**, **B**, and **C**.

This is an analogue of the two-dimensional case where $\det\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}$ is plus or minus the area of the parallelogram spanned by the vectors $\langle a_1, a_2\rangle$ and $\langle b_1, b_2\rangle$.

Summary

We constructed a vector **C** perpendicular to vectors $\mathbf{A} = \langle a_1, a_2, a_3\rangle$ and $\mathbf{B} = \langle b_1, b_2, b_3\rangle$ by requiring that $\mathbf{C}\cdot\mathbf{A} = 0$ and $\mathbf{C}\cdot\mathbf{B} = 0$. The vector **C** with these properties is the cross product of **A** and **B**:

$$\mathbf{C} = \mathbf{A}\times\mathbf{B} = \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3. \end{pmatrix}$$

The cross product of **A** and **B** also may be described as the vector whose length is the area of the parallelogram spanned by **A** and **B** and whose direction is given by the right-hand rule (the fingers curling from **A** towards **B**).

The cross product, also known as the vector product, has the properties:

1. $\mathbf{A}\times\mathbf{B} = -(\mathbf{B}\times\mathbf{A})$ (anticommutative)
2. $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$ is not usually equal to $(\mathbf{A}\times\mathbf{B})\times\mathbf{C}$ (not associative)
3. $\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = (\mathbf{A}\cdot\mathbf{C})\mathbf{B} - (\mathbf{A}\cdot\mathbf{B})\mathbf{C}$ (**Jacobi's Identity**: see Exercise 22.)
4. $(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{A}\times\mathbf{B}) = (\mathbf{A}\cdot\mathbf{A})(\mathbf{B}\cdot\mathbf{B}) - (\mathbf{A}\cdot\mathbf{B})(\mathbf{A}\cdot\mathbf{B})$ (appeared in finding the length of $\mathbf{A}\times\mathbf{B}$)
5. $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = \pm$ volume of parallelepiped spanned by **A**, **B**, and **C**

EXERCISES for Section 14.3

In Exercises 1 to 4 compute and sketch $\mathbf{A}\times\mathbf{B}$.

1. $\mathbf{A} = \mathbf{k}$, $\mathbf{B} = \mathbf{j}$

2. $\mathbf{A} = \mathbf{i}+\mathbf{j}$, $\mathbf{B} = \mathbf{i}-\mathbf{j}$

3. $\mathbf{A} = \mathbf{i}+\mathbf{j}+\mathbf{k}$, $\mathbf{B} = \mathbf{i}+\mathbf{j}$

4. $\mathbf{A} = \mathbf{k}$, $\mathbf{B} = \mathbf{i}+\mathbf{j}$

In Exercises 5 and 6, find $\mathbf{A}\times\mathbf{B}$ and check that it is perpendicular to **A** and **B**.

5. $\mathbf{A} = 2\mathbf{i}-3\mathbf{j}+\mathbf{k}$, $\mathbf{B} = \mathbf{i}+\mathbf{j}+2\mathbf{k}$

6. $\mathbf{A} = \mathbf{i}-\mathbf{j}$, $\mathbf{B} = \mathbf{j}+4\mathbf{k}$

In Exercises 7 to 10 use the cross product to find the area of

7. the parallelogram three of whose vertices are $(0,0,0)$, $(1,5,4)$, and $(2,-1,3)$.

8. the parallelogram three of whose vertices are $(1,2,-1)$, $(2,1,4)$, and $(3,5,2)$.

9. the triangle two of whose sides are $\mathbf{i}+\mathbf{j}$ and $3\mathbf{i}-\mathbf{j}$.

10. the triangle two of whose sides are $\mathbf{i}+2\mathbf{j}+3\mathbf{k}$ and $2\mathbf{i}-\mathbf{j}+2\mathbf{k}$.

In Exercises 11 to 14 find the volume of the parallelepiped spanned by

11. $\langle 2,1,3\rangle$, $\langle 3,-1,2\rangle$, $\langle 4,0,3\rangle$

12. $3\mathbf{i}+4\mathbf{j}+3\mathbf{k}$, $2\mathbf{i}+3\mathbf{j}+4\mathbf{k}$, $\mathbf{i}-\mathbf{j}-\mathbf{k}$.

13. $\overrightarrow{PQ}$, $\overrightarrow{PR}$, $\overrightarrow{PS}$, where $P=(1,1,1)$, $Q=(2,1,-2)$, $R=(3,5,2)$, and $S=(1,-1,2)$.

14. $\overrightarrow{PQ}$, $\overrightarrow{PR}$, $\overrightarrow{PS}$, where $P=(0,0,0)$, $Q=(3,3,2)$, $R=(1,4,-1)$, and $S=(1,2,3)$.

---

15. Evaluate $\mathbf{A}\cdot(\mathbf{A}\times\mathbf{B})$.

16. Prove that $\mathbf{B}\times\mathbf{A}=-(\mathbf{A}\times\mathbf{B})$ in two ways (a) using the algebraic definition of the cross product and (b) using the geometric description of the cross product.

17. Show that if $\mathbf{B}=c\mathbf{A}$, then $\mathbf{A}\times\mathbf{B}=\mathbf{0}$ (a) using the algebraic definition of the cross product and (b) using the geometric description of the cross product.

18. Show that $(0,0,0)$, (x_1,y_1,z_1), (x_2,y_2,z_2) and (x_3,y_3,z_3) lie on a plane if and only if $\det\begin{pmatrix} x_1 & y_1 & z_1\\ x_2 & y_2 & z_2\\ x_3 & y_3 & z_3\end{pmatrix}=0$.

19. (a) If $\mathbf{B}$ is parallel to $\mathbf{C}$, is $\mathbf{A}\times\mathbf{B}$ parallel to $\mathbf{A}\times\mathbf{C}$?
(b) If $\mathbf{B}$ is perpendicular to $\mathbf{C}$, is $\mathbf{A}\times\mathbf{B}$ perpendicular to $\mathbf{A}\times\mathbf{C}$?

20. Let $\mathbf{A}$ be a nonzero vector. If $\mathbf{A}\times\mathbf{B}=\mathbf{0}$ and $\mathbf{A}\cdot\mathbf{B}=0$, must $\mathbf{B}=\mathbf{0}$?

21. (a) Give an example of a vector perpendicular to $3\mathbf{i}-\mathbf{j}+\mathbf{k}$.
(b) Give an example of a unit vector perpendicular to $3\mathbf{i}-\mathbf{j}+\mathbf{k}$.

Exercises 22 to 25 involve the vector triple product.

22. The purpose of this exercise is to motivate the equation known as **Jacobi's Identity**, $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=(\mathbf{A}\cdot\mathbf{C})\mathbf{B}-(\mathbf{A}\cdot\mathbf{B})\mathbf{C}$. Let $\mathbf{B}$ and $\mathbf{C}$ be nonzero, nonparallel vectors and $\mathbf{A}$ a vector that is perpendicular neither to $\mathbf{B}$ nor $\mathbf{C}$.
(a) Why are there scalars x and y such that $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=x\mathbf{B}+y\mathbf{C}$?
(b) Why is $0=x(\mathbf{A}\cdot\mathbf{B})+y(\mathbf{A}\cdot\mathbf{C})$?
(c) Using (b), show that there is a scalar z such that $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=z\left((\mathbf{A}\cdot\mathbf{C})\mathbf{B}-(\mathbf{A}\cdot\mathbf{B})\mathbf{C}\right)$.
(d) It would be nice if there were a simple geometric way to show that z is a constant and equals 1. We could show that $z=1$ by writing $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ in components and making a tedious calculation. Try to find a simple way to determine that $z=1$. (The authors do not know of such a way.)

23. Show that $\mathbf{A}\times(\mathbf{A}\times\mathbf{B})=(\mathbf{A}\cdot\mathbf{B})\mathbf{A}-(\mathbf{A}\cdot\mathbf{A})\mathbf{B}$.

24. Show that $(\mathbf{A}\times\mathbf{B})\times(\mathbf{C}\times\mathbf{D})=((\mathbf{A}\times\mathbf{B})\cdot\mathbf{D})\mathbf{C}-((\mathbf{A}\times\mathbf{B})\cdot\mathbf{C})\mathbf{D}$.

25. Let $\mathbf{u}$ be a unit vector and $\mathbf{B}$ a vector. What happens as you keep crossing by $\mathbf{u}$, and form the sequence $\mathbf{B}$, $\mathbf{u}\times\mathbf{B}$, $\mathbf{u}\times(\mathbf{u}\times\mathbf{B})$ and so on? (See Exercise 23.)

---

26. (a) From the distributive law $\mathbf{A}\times(\mathbf{B}+\mathbf{C})=\mathbf{A}\times\mathbf{B}+\mathbf{A}\times\mathbf{C}$, and the property $\mathbf{D}\times\mathbf{E}=-\mathbf{E}\times\mathbf{D}$, deduce the distributive law $(\mathbf{B}+\mathbf{C})\times\mathbf{A}=\mathbf{B}\times\mathbf{A}+\mathbf{C}\times\mathbf{A}$.
(b) From the distributive law $\mathbf{A}\times(\mathbf{B}+\mathbf{C})=\mathbf{A}\times\mathbf{B}+\mathbf{A}\times\mathbf{C}$, show that $\mathbf{A}\times(\mathbf{B}+\mathbf{C}+\mathbf{D})=\mathbf{A}\times\mathbf{B}+\mathbf{A}\times\mathbf{C}+\mathbf{A}\times\mathbf{D}$.

27. Check that $-13\mathbf{i}+7\mathbf{j}+11\mathbf{k}$ in Example 1 is perpendicular to $\mathbf{A}$ and to $\mathbf{B}$.

28. Show, using (14.3.7), that $\mathbf{0}\times\mathbf{B}=\mathbf{0}$.

29. Show, using (14.3.7), that $\mathbf{B}\times\mathbf{A}=-\mathbf{A}\times\mathbf{B}$.

30. Using the values for y and z found when solving equations (14.3.3) and (14.3.4), find x.

31. Using (14.3.7), show that if $\mathbf{B}$ is parallel to $\mathbf{A}$, then $\mathbf{A}\times\mathbf{B}=\mathbf{0}$.

32. In finding $|\mathbf{A}\times\mathbf{B}|^2$ we stated that

$$a_2^2b_3^2+a_3^2b_2^2+a_3^2b_1^2+a_1^2b_3^2+a_1^2b_2^2+a_2^2b_1^2-2(a_2a_3b_2b_3+a_1a_3b_1b_3+a_1a_2b_1b_2)$$

equals

$$(a_1^2+a_2^2+a_3^2)(b_1^2+b_2^2+b_3^2)-(a_1b_1+a_2b_2+a_3b_3)^2.$$

Verify that this is correct.

33. (a) How could you use cross products to find a vector perpendicular to $2\mathbf{i}+3\mathbf{j}+4\mathbf{k}$? Give an example.
(b) How could you use cross products to find two vectors perpendicular to $2\mathbf{i}+3\mathbf{j}+4\mathbf{k}$ and to each other? Give an example.

34. To understand why you cannot omit the parentheses in $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$, let $\mathbf{A}$ and $\mathbf{B}$ be nonzero, nonparallel vectors. Show that $\mathbf{A}\times(\mathbf{A}\times\mathbf{B})$ is never equal to $(\mathbf{A}\times\mathbf{A})\times\mathbf{B}$. This shows that the cross product is not associative.

35. *(Crystallography)* A crystal is described by three vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. They span a fundamental parallelepiped, whose copies fill out a crystal lattice. (See Figure 14.3.5.) Atoms are at the corners. To study the diffraction of x-rays and light through a crystal, crystallographers work with the reciprocal lattice, whose fundamental parallelepiped is spanned by three vectors, $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$. The vector $\mathbf{k}_1$ is perpendicular to the parallelogram spanned by $\mathbf{v}_2$ and $\mathbf{v}_3$ and has a length equal to the reciprocal of the distance between the parallelogram and the opposite parallelogram of the fundamental parallelepiped. The vectors $\mathbf{k}_2$ and $\mathbf{k}_3$ are defined similarly in terms of the other four faces of the fundamental parallelepiped.

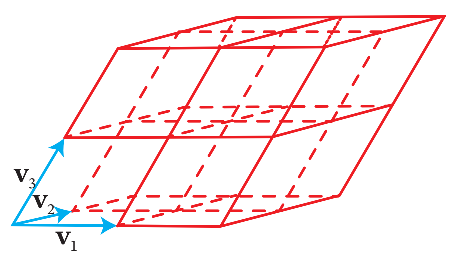


Figure 14.3.5

(a) Show that $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$ may be chosen to be

$$\mathbf{k}_1=\frac{\mathbf{v}_2\times\mathbf{v}_3}{\mathbf{v}_1\cdot(\mathbf{v}_2\times\mathbf{v}_3)},\qquad \mathbf{k}_2=\frac{\mathbf{v}_3\times\mathbf{v}_1}{\mathbf{v}_1\cdot(\mathbf{v}_2\times\mathbf{v}_3)},\qquad \mathbf{k}_3=\frac{\mathbf{v}_1\times\mathbf{v}_2}{\mathbf{v}_1\cdot(\mathbf{v}_2\times\mathbf{v}_3)}.$$

(b) Show that the volume of the parallelepiped determined by $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$ is the reciprocal of the volume of the one determined by $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$.
(c) Is the reciprocal of the reciprocal lattice the original lattice? For instance, is $\mathbf{v}_1=\dfrac{\mathbf{k}_2\times\mathbf{k}_3}{\mathbf{k}_1\cdot(\mathbf{k}_2\times\mathbf{k}_3)}$?

Exercises 37 and 38 provide a little deeper understanding of cross products, but require two additional properties of determinants. These properties are developed in Exercise 36.

36. Prove the following two properties of determinants:
(a) If two rows (or two columns) of a 3×3 matrix are switched with each other, then the determinant of the resulting matrix is the determinant of the original matrix with the sign changed.
(b) If two rows (or two columns) of a 3×3 matrix are identical, then the determinant of the matrix is 0.

37. (a) Explain using parallelepipeds why $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$ is plus or minus $\mathbf{B}\cdot(\mathbf{C}\times\mathbf{A})$.
(b) Using properties of determinants, decide if it is plus or minus.

38. In some expositions of the cross product, $\mathbf{A}\times\mathbf{B}$ is defined as the determinant of a matrix of order 3. If we start with this definition, use a property of determinants to show that $\mathbf{A}\times\mathbf{B}$ is perpendicular to both $\mathbf{A}$ and $\mathbf{B}$.

The purpose of Exercises 39 and 40 is to derive the distributive law for the cross product. This means that the distributive property cannot be used to answer these questions.,

39. Let $\mathbf{A}$ be a nonzero vector and $\mathbf{B}$ be a vector. Let $\mathbf{B}_1$ be the projection of $\mathbf{B}$ on a plane perpendicular to $\mathbf{A}$. Let $\mathbf{B}_2$ be obtained by rotating $\mathbf{B}_1$ through an angle of 90° in the direction given by the right-hand rule with thumb pointing in the same direction as $\mathbf{A}$. (a) Show that $\mathbf{A}\times\mathbf{B}=\mathbf{A}\times\mathbf{B}_1$. (b) Show that $\mathbf{A}\times\mathbf{B}=|\mathbf{A}|\mathbf{B}_2$.

40. Using Exercise 39(b), show that for $\mathbf{A}$ not $\mathbf{0}$, $\mathbf{A}\times(\mathbf{B}+\mathbf{C})=\mathbf{A}\times\mathbf{B}+\mathbf{A}\times\mathbf{C}$.

14.4 Applications of the Dot and Cross Products

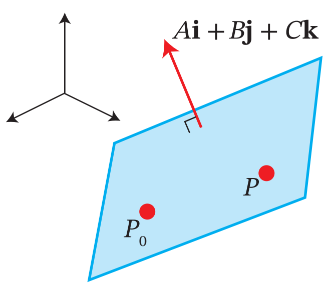


Figure 14.4.1

This section uses the dot product to deal with lines, planes, and the cross product to produce vectors perpendicular to planes.

Equation of a Plane

First we find an equation of the plane through the point $P_0=(x_0,y_0,z_0)$ and perpendicular to the vector $A\mathbf{i}+B\mathbf{j}+C\mathbf{k}$, shown in Figure 14.4.1.

Let $P=(x,y,z)$ be a point on the plane. The vector $\overrightarrow{P_0P}$ is perpendicular to $A\mathbf{i}+B\mathbf{j}+C\mathbf{k}$. (Slide it so that P_0 coincides with the tail of $A\mathbf{i}+B\mathbf{j}+C\mathbf{k}$.) Thus

$$(A\mathbf{i}+B\mathbf{j}+C\mathbf{k})\cdot((x-x_0)\mathbf{i}+(y-y_0)\mathbf{j}+(z-z_0)\mathbf{k})=0.$$

Formula 14.4.1: *Equation of a Plane*

An equation of the plane containing the point (x_0,y_0,z_0) and perpendicular to the vector $A\mathbf{i}+B\mathbf{j}+C\mathbf{k}$ is

$$A(x-x_0)+B(y-y_0)+C(z-z_0)=0. \qquad (14.4.1)$$

The vector $A\mathbf{i}+B\mathbf{j}+C\mathbf{k}$ is called a **normal** to the plane.

EXAMPLE 1. Find an equation of the plane through $(2,-3,4)$ and normal to $\mathbf{i}+2\mathbf{j}+3\mathbf{k}$.

SOLUTION By Formula 14.4.1, an equation of the plane through the point $(x_0,y_0,z_0)=(2,-3,4)$ and normal to the vector $\langle 1,2,3\rangle$ is

$$1(x-2)+2(y-(-3))+3(z-4)=0$$

which simplifies to

$$x+2y+3z-8=0.$$

◆

The graph of $Ax+By+Cz+D=0$, where not all of A, B, and C are 0, is a plane perpendicular to $A\mathbf{i}+B\mathbf{j}+C\mathbf{k}$. To show this, first pick a point (x_0,y_0,z_0) or which $Ax_0+By_0+Cz_0+D=0$. Subtracting this from the original equation gives

$$A(x-x_0)+B(y-y_0)+C(z-z_0)=0,$$

which is an equation of the plane through (x_0, y_0, z_0) perpendicular to $A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$.

The equation of a line in the xy-plane through a point and perpendicular to a vector in the plane is obtained by restricting Formula 14.4.1 with $z_0 = 0$ and $C = 0$:

Formula 14.4.2: *Equation of a Line in the Plane*

An equation of the line through (x_0, y_0) and perpendicular to the vector $A\mathbf{i} + B\mathbf{j}$ is

$$A(x - x_0) + B(y - y_0) = 0.$$

Distance from a Point (p, q) to the Line $Ax + By + C = 0$

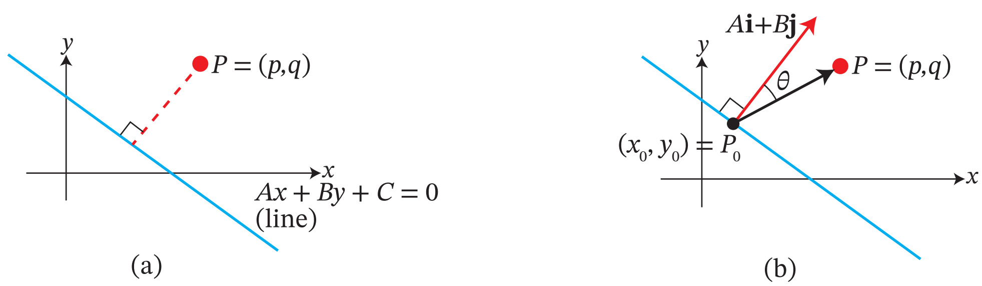


Figure 14.4.2

Let us find the distance from $P = (p, q)$ to the line in the xy-plane whose equation is $Ax + By + C = 0$, shown in Figure 14.4.2(a). Pick a point $P_0 = (x_0, y_0)$ on the line and place $A\mathbf{i} + B\mathbf{j}$ with its tail at P_0, as in Figure 14.4.2(b). If θ is the angle between $\overrightarrow{P_0P}$ and $A\mathbf{i} + B\mathbf{j}$, then the distance from P to the line is

$$\begin{aligned}
|\overrightarrow{P_0P}||\cos(\theta)| &= |\overrightarrow{P_0P}|\frac{\left|(A\mathbf{i} + B\mathbf{j}) \cdot ((p - x_0)\mathbf{i} + (q - y_0)\mathbf{j})\right|}{|\overrightarrow{P_0P}||A\mathbf{i} + B\mathbf{j}|} \\
&= \frac{\left|A(p - x_0) + B(q - y_0)\right|}{\sqrt{A^2 + B^2}} \\
&= \frac{\left|Ap + Bq - (Ax_0 + By_0)\right|}{\sqrt{A^2 + B^2}}.
\end{aligned}$$

Since $Ax_0 + By_0 + C = 0$, we have

Formula 14.4.3: *Distance from a Point to a Line*

The distance from (p, q) to the line $Ax + By + C = 0$ is

$$\frac{|Ap + Bq + C|}{\sqrt{A^2 + B^2}}.$$

Observation 14.4.1: *A Quick Way to Find the Distance from a Point to a Line*

Formula 14.4.3 is not as complicated as it might appear at first glance. The key is to write the equation of the line in the form $Ax + By + C = 0$.

To find the distance from a point (p, q) to the line $Ax + By + C = 0$, substitute the coordinates of the point (p, q) into $Ax + By + C$, take its absolute value, and divide by $\sqrt{A^2 + B^2}$.

EXAMPLE 2. How far is the point $(1,3)$ from the line $2x-4y=5$?

SOLUTION Write the equation in the form $2x-4y-5=0$. Then the distance is

$$\frac{|2(1)-4(3)-5|}{\sqrt{2^2+4^2}}=\frac{|-15|}{\sqrt{20}}=\frac{3\sqrt{5}}{2}.$$

♦

Distance from a Point (p,q,r) to the Plane $Ax+By+Cz+D=0$

The corresponding formula for the distance from a point $P=(p,q,r)$ to a plane is obtained in Exercise 55.

Formula 14.4.4: *Distance from a Point to a Plane*

The distance from (p,q,r) to the plane $Ax+By+Cz+D=0$ is

$$\frac{|Ap+Bq+Cr+D|}{\sqrt{A^2+B^2+C^2}}.$$

Using Vectors to Parameterize a Line

Let L be the line through the point $P_0=(x_0,y_0,z_0)$ parallel to the nonzero vector $\mathbf{B}$, shown in Figure 14.4.3(a). If P is any point on L, then the vector $\overrightarrow{P_0P}$, which is parallel to $\mathbf{B}$, is $t\mathbf{B}$ for some scalar t. See Figure 14.4.3(b). So we have $\overrightarrow{OP}=\overrightarrow{OP_0}+\overrightarrow{P_0P}=\overrightarrow{OP_0}+t\mathbf{B}$.

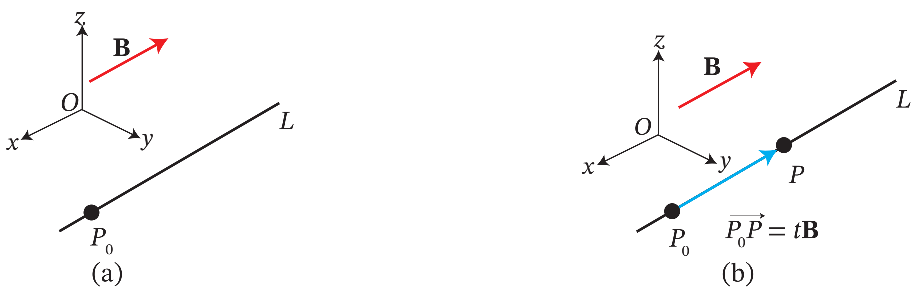


Figure 14.4.3

Formula 14.4.5: *Parametric Equation of a Line*

The line through P_0 parallel to the vector $\mathbf{B}$ is parameterized by $\overrightarrow{OP}=\overrightarrow{OP_0}+t\mathbf{B}$. As t varies, the vector from O to P varies, sweeping out the line L.

EXAMPLE 3. Parameterize the line L that passes through the point $(1,1,2)$ and is parallel to the vector $3\mathbf{i}+4\mathbf{j}+5\mathbf{k}$.

SOLUTION The vector from the origin to a point on the line is $\overrightarrow{OP_0}=\mathbf{i}+\mathbf{j}+2\mathbf{k}$ and the direction of the line is $\mathbf{B}=3\mathbf{i}+4\mathbf{j}+5\mathbf{k}$, so a vector from the origin to an arbitrary point P on the line is

$$\overrightarrow{OP}=\mathbf{i}+\mathbf{j}+2\mathbf{k}+t(3\mathbf{i}+4\mathbf{j}+5\mathbf{k})=(3t+1)\mathbf{i}+(4t+1)\mathbf{j}+(5t+2)\mathbf{k}.$$

If P is (x,y,z), then another representation of $\overrightarrow{OP}$ is the vector $x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$, where

$$x=3t+1,\qquad y=4t+1,\qquad \text{and}\qquad z=5t+2.$$

One vector equation does the work of three scalar equation. ♦

Describing the Direction of a Vector

The direction of a vector in the plane is described by the angle it makes with the positive x-axis. To specify the direction of a vector in space requires three angles, two of which almost determine the third.

Definition: ***Direction angles of a vector.***

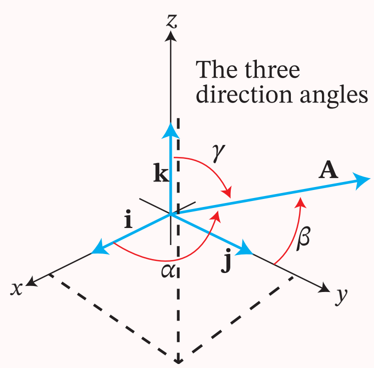


Figure 14.4.4

When $\mathbf{A}$ is a nonzero vector in space, then the angle between
$\mathbf{A}$ and $\mathbf{i}$ is denoted α (alpha),
$\mathbf{A}$ and $\mathbf{j}$ is denoted β (beta),
$\mathbf{A}$ and $\mathbf{k}$ is denoted γ (gamma).
The angles α, β and γ are called the **direction angles of A**. (See Figure 14.4.4.)

Definition: ***Direction cosines of a vector***

The **direction cosines** of a vector are the cosines of its direction angles, $\cos(\alpha)$, $\cos(\beta)$, and $\cos(\gamma)$.

EXAMPLE 4. The angle between $\mathbf{A}$ and $\mathbf{k}$ is $\pi/6$. Find γ and $\cos(\gamma)$ for (a) $\mathbf{A}$, and (b) $-\mathbf{A}$.

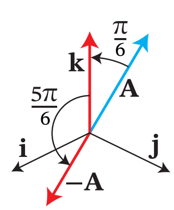


Figure 14.4.5

SOLUTION
(a) By definition, the direction angle γ for $\mathbf{A}$ is $\pi/6$. Then $\cos(\gamma)=\cos(\pi/6)=\sqrt{3}/2$.
(b) To find γ and $\cos(\gamma)$ for $-\mathbf{A}$, we draw Figure 14.4.5. For the opposite vector, $-\mathbf{A}$, the direction angle $\gamma=5\pi/6$ and so $\cos(\gamma)=\cos(5\pi/6)=-\sqrt{3}/2$. ♦

As Example 4 illustrates, if the direction angles of $\mathbf{A}$ are α, β, γ, then the direction angles of $-\mathbf{A}$ are $\pi-\alpha$, $\pi-\beta$, and $\pi-\gamma$. The direction cosines of $-\mathbf{A}$ are the negatives of the direction cosines of $\mathbf{A}$.

The three direction angles are not independent, as is shown by the next theorem. Two of the direction *cosines* determine the third direction *cosine* up to its sign.

Theorem 14.4.2: An Important Identity for Direction Cosines

If α, β, γ are the direction angles of **A**, *then* $\cos^2(\alpha)+\cos^2(\beta)+\cos^2(\gamma)=1$.

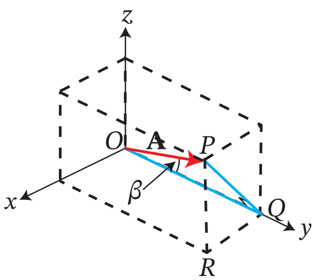


Figure 14.4.6

Proof of Theorem 14.4.2

It is no loss of generality to assume that **A** is a unit vector. Its component on the y-axis, for instance, is $\cos(\beta)$, as the right triangle $\triangle OPQ$ in Figure 14.4.6 shows. The vector **A** lies along the hypotenuse and $|OQ|=|\cos(\beta)|$. Also PQ is the hypotenuse of right triangle $\triangle PRQ$, whose sides are the components $\cos(\alpha)$ and $\cos(\gamma)$, respectively.

Since **A** is a unit vector, $|\mathbf{A}|^2=1$, and so $\cos^2(\alpha)+\cos^2(\beta)+\cos^2(\gamma)=1$. ●

Notice that Theorem 14.4.2 also tells us that if the direction cosines of a vector **A** are a, b, and c, then the vector $a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$ is the unit vector in the direction of **A**.

EXAMPLE 5. If **A** makes an angle of 60° with both the x-axis and the y-axis, what angle does it make with the z-axis?

SOLUTION Here $\alpha=60°$ and $\beta=60°$; hence $\cos(\alpha)=1/2$ and $\cos(\beta)=1/2$. Since $\cos^2(\alpha)+\cos^2(\beta)+\cos^2(\gamma)=1$, it follows that

$$\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^2+\cos^2(\gamma)=1$$

so that $\cos^2(\gamma)=1/2$. Thus, either

$$\cos(\gamma)=\frac{\sqrt{2}}{2}\quad\text{or}\quad\cos(\gamma)=-\frac{\sqrt{2}}{2}.$$

Hence

$$\gamma=\frac{\pi}{4}=45^\circ\quad\text{or}\quad\gamma=\frac{3\pi}{4}=135^\circ.$$

Figures 14.4.7(a) and (b) show the possibilities for **A**.

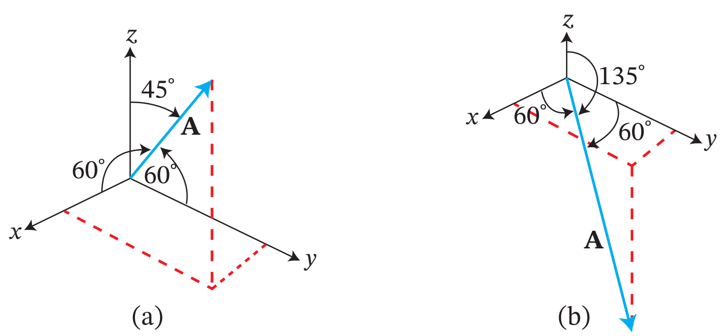


Figure 14.4.7

◆

EXAMPLE 6. The three points P, Q, and R, which do not lie on a line, determine a plane. So do the non-coplanar points P', Q', and R'. How would one determine whether the two planes are (a) perpendicular? (b) parallel?

SOLUTION The vector $\mathbf{A}=\overrightarrow{PQ}\times\overrightarrow{PR}$ is perpendicular to the first plane and $\mathbf{A}'=\overrightarrow{P'Q'}\times\overrightarrow{P'R'}$ is perpendicular to the second plane.

(a) The two planes are perpendicular if **A** is perpendicular to $\mathbf{A}'$, that is, $\mathbf{A}\cdot\mathbf{A}'=0$.

(b) The two planes are parallel if **A** is parallel to $\mathbf{A}'$, that is, $\mathbf{A}\times\mathbf{A}'=\mathbf{0}$. ◆

Dot Products and Flow

Assume that vector **v**, whose magnitude is v, describes the velocity of water flowing down a river, as in Figure 14.4.8(a). Hold a stick of length L on the surface of the water. The amount of water crossing the stick depends on its position. If the stick is parallel to **v**, no water crosses it. If the stick is not parallel to **v**, water crosses it. How does the angle at which we place the stick affect the amount of water that crosses it?

To answer this, we introduce a unit vector **n** perpendicular to the stick and record its position, as in Figure 14.4.8(b). Let the angle between **n** and **v** be θ.

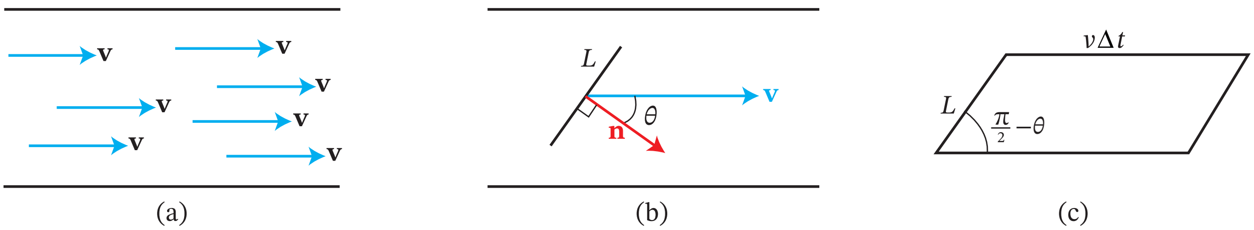


Figure 14.4.8

The amount of water that crosses the stick during time Δt is equal to the area of the parallelogram in Figure 14.4.8(c). The base of the parallelogram has length $v\Delta t$ (speed times time). The height is $L\cos(\theta)$. The area of the parallelogram is therefore

$$vL\cos(\theta)\Delta t.$$

Then $vL\cos(\theta)$ measures the amount of water that crosses the stick in one unit of time.

But $v\cos(\theta)$ is equal to $\mathbf{v}\cdot\mathbf{n}$. So $\mathbf{v}\cdot\mathbf{n}$ measures the tendency of water to cross the stick.

As a check, when the stick is parallel to **v**, $\theta = \pi/2$ and $\cos(\pi/2) = 0$. Then $\mathbf{v}\cdot\mathbf{n} = 0$ and no water crosses the stick. When the stick is perpendicular to **v**, $\theta = 0$, and $\mathbf{v}\cdot\mathbf{n} = v$.

Summary

The dot product was used to obtain an equation of a plane in space (or of a line in the xy-plane) and to find the distance from a point to a line or plane. The cross product enabled us to find a vector perpendicular to a plane. A vector parallel to a line is useful for parameterizing a line.

Direction angles and cosines of a vector were defined. We showed how the dot product describes the rate of flow across a line segment. This concept will be needed in Chapters 17 and 18, where we deal with flows across curves and surfaces.

EXERCISES for Section 14.4

In Exercises 1 to 4 find an equation of the line in the xy-plane through the point and perpendicular to the vector.

1. $(2,3)$, $4\mathbf{i}+5\mathbf{j}$ **2.** $(1,0)$, $2\mathbf{i}-\mathbf{j}$ **3.** $(4,5)$, $\mathbf{i}+3\mathbf{j}$ **4.** $(2,-1)$, $\mathbf{i}+3\mathbf{j}$

In Exercises 5 to 8 find a vector in the xy-plane that is perpendicular to the line.

5. $2x-3y+8=0$ **6.** $\pi x-\sqrt{2}y=7$ **7.** $y=3x+7$ **8.** $2(x-1)+5(y+2)=0$

In Exercises 9 to 12 find a parametric equation of the line in xyz-space with the indicated properties.

9. Contains $(2,3,4)$ and perpendicular to $4\mathbf{i}+5\mathbf{j}$ and $3\mathbf{i}+5\mathbf{j}+6\mathbf{j}$.

10. Belongs to both $x+2y-z=4$ and $-2x+2y+5z=-2$.

11. Contains $(4,0,5)$ and is parallel to $\mathbf{i}+3\mathbf{j}-2\mathbf{j}$.

12. Contains the two points $(2,-1,-1)$ and $(0,1,3)$.

---

13. Find a vector perpendicular to the plane through $(2,1,3)$, $(4,5,1)$ and $(-2,2,3)$.

14. How far is the point $(1,2,2)$ from the plane through $(0,0,0)$, $(3,5,-2)$, and $(2,-1,3)$?

15. How far is the point $(1,2,3)$ from the line through $(-2,-1,3)$, and $(4,1,2)$?

16. (a) Describe in words how to find an equation for the plane through $P_1=(x_1,y_1,z_1)$, $P_2=(x_2,y_2,z_2)$, and $P_3=(x_3,y_3,z_3)$.
(b) Find an equation for the plane through $(2,2,1)$, $(0,1,5)$ and $(2,-1,0)$.

17. (a) Describe in words how to decide whether the line through $P_1=(x_1,y_1,z_1)$ and $P_2=(x_2,y_2,z_2)$ is parallel to the line through $P_3=(x_3,y_3,z_3)$ and $P_4=(x_4,y_4,z_4)$.
(b) Is the line through $(1,2,-3)$ and $(5,9,4)$ parallel to the line through $(-1,-1,2)$ and $(1,3,5)$?

18. (a) Describe in words how to decide whether the line through $P_1=(x_1,y_1,z_1)$ and $P_2=(x_2,y_2,z_2)$ is parallel to the plane $Ax+By+Cz+D=0$.
(b) Is the line through $(1,-2,3)$ and $(5,3,0)$ parallel to the plane $2x-y+z+3=0$?

19. (a) Describe in words how to decide whether the line through P_1 and P_2 is parallel to the plane through Q_1, Q_2, and Q_3.
(b) Is the line through $(0,0,0)$ and $(1,1,-1)$ parallel to the plane through $(1,0,1)$, $(2,1,0)$, and $(1,3,4)$?

20. (a) Describe in words how to decide whether the plane through P_1, P_2, and P_3 is parallel to the plane through Q_1, Q_2, and Q_3.
(b) Is the plane through $(1,2,3)$, $(4,1,-1)$, and $(2,0,1)$ parallel to the plane through $(0,-1,5)$, $(3,-2,1)$, and $(-4,-8,7)$?

21. Find the parametric equations of the line through $(1,1,2)$ and perpendicular to the plane $3x-y+z=6$.

22. Find an equation of the plane through $(1,2,3)$ that contains the line given parametrically as $\overrightarrow{OP}=2\mathbf{i}-\mathbf{j}+3\mathbf{k}+t(3\mathbf{i}+2\mathbf{j}+\mathbf{k})$.

23. Is the point $(21,-3,28)$ on the line given parametrically as $\overrightarrow{OP}=\mathbf{i}+2\mathbf{j}+3\mathbf{k}+t(4\mathbf{i}-\mathbf{j}+5\mathbf{k})$?

24. Find the angle between the line through $(3,2,2)$ and $(4,3,1)$ and the line through $(3,2,2)$ and $(5,2,7)$.

25. The angle between two planes is the angle between their normal vectors. Find the angle between the planes $2x+3y+4z=11$ and $3x-y+2z=13$.

NOTE: Depending on the choice of normal vectors there are two angles between the normal vectors to two planes. We adopt the convention that the angle between two planes is always between 0 and π.

26. (a) How many unit vectors are perpendicular to the plane $Ax+By+Cz+D=0$?
(b) Find a unit vector perpendicular to the plane $3x-2y+4z+6=0$.

27. (a) Explain how to find a point on the plane $Ax+By+Cz+D=0$.
(b) Give the coordinates of a point that lies on the plane $3x-y+z+10=0$.

28. (a) Explain how to find a point that lies in both planes $A_1x+B_1y+C_1z+D_1=0$ and $A_2x+B_2y+C_2z+D_2=0$.
(b) Find a point that lies in both planes $3x+z+2=0$ and $x-y-z+5=0$.

29. (a) Explain how to find the area of the parallelogram spanned by two vectors $\mathbf{A}$ and $\mathbf{B}$ in space.
(b) Find the area of the parallelogram spanned by $\langle 2,3,1\rangle$ and $\langle 4,-1,5\rangle$.

In Exercises 30 to 33 find the distance from the point to the plane.

30. The point $(0,0,0)$ to the plane $2x-4y+3z+2=0$

31. The point $(1,2,3)$ to the plane $x+2y-3z+5=0$.

32. The point $(2,2,-1)$ to the plane that passes through $(1,4,3)$ and has a normal $2\mathbf{i}-7\mathbf{j}+2\mathbf{k}$.

33. The point $(0,0,0)$ to the plane that passes through $(4,1,0)$ and is perpendicular to the vector $\mathbf{i}+\mathbf{j}+\mathbf{k}$.

In Exercises 34 to 37 it will be useful to use the fact that if the direction cosines of a vector $\mathbf{v}$ are a, b, and c, then $\langle a,b,c\rangle$ is the unit vector in the direction of $\mathbf{v}$.

34. Find the direction cosines of $2\mathbf{i}+3\mathbf{j}+4\mathbf{k}$.

35. Find the direction cosines of the vector from $(1,3,2)$ to $(4,-1,5)$.

36. Let $P_0=(2,1,5)$ and $P_1=(3,0,4)$. Find the direction cosines and direction angles of (a) $\overrightarrow{P_0P_1}$ and (b) $\overrightarrow{P_1P_0}$.

37. Give parametric equations for the line through $\left(\frac{1}{2},\frac{1}{3},\frac{1}{2}\right)$ parallel to a vector with direction cosines $\alpha=\frac{2}{\sqrt{93}}$, $\beta=\frac{-5}{\sqrt{93}}$, and $\beta=\frac{8}{\sqrt{93}}$ (a) in scalar form and (b) in vector form.

---

38. Give parametric equations for the line through $(1,2,3)$ and $(4,5,7)$ (a) in scalar form and (b) in vector form.

39. Give parametric equations for the line through $(7,-1,5)$ and $(4,3,2)$.

40. A vector $\mathbf{A}$ has direction angles $\alpha=70°$ and $\beta=80°$. Find the third direction angle γ and show the possible angles for γ on a diagram.

41. Find where the line through $(1,2)$ and $(3,5)$ meets the line through $(1,-1)$ and $(2,3)$.

42. Where does the line through $(1,2,4)$ and $(2,1,-1)$ meet the plane $x+2y+5z=0$?

43. Give parametric equations for the line through $(1,3,-5)$ that is perpendicular to the plane $2x-3y+4z=11$.

44. How far is the point $(1,5)$ from the line through $(4,2)$ and $(3,7)$?

45. How far is the point $(1,2,-3)$ from the line through $(2,1,4)$ and $(1,5,-2)$?

46. Give parametric equations for the line through $(1,3,4)$ that is parallel to the line through $(2,4,6)$ and $(5,3,-2)$.

47. (a) If you know the coordinates of point P and parametric equations of line L, explain (in a few sentences, and no formulas and equations) how you would find parametric equations of the plane that contains P and L. (Assume P is not on L.)
(b) Implement the method described in (a) to find parametric equations for the plane through $(1,1,1)$ that contains the line parameterized by $x=2+t$, $y=3-t$, $z=4+2t$.

48. (a) Sketch points P, Q, R, and S, not all in one plane, such that $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ are not parallel. Explain why there is a unique pair of parallel planes one of which contains P and Q and one of which contains R and S.
(b) Express a normal vector to the planes in terms of P, Q, R, and S.

49. Find an equation for the plane through P_1 that is parallel to the nonparallel segments P_2P_3 and P_4P_5.

50. Find where the line through $P_0 = (2,1,3)$ and $P_1 = (4,-2,5)$ meets the plane whose equation is $2x+y-4z+5=0$.

51. Find where the line through $(1,2,1)$ and $(2,1,3)$ meets the plane that is perpendicular to $2\mathbf{i}+5\mathbf{j}+7\mathbf{k}$ and passes through the point $(1,-2,-3)$.

52. Are the points $(1,2,-3)$, $(1,6,2)$, and $(7,14,11)$ on a single line?

53. If α, β, and γ are direction angles of a vector, what is $\sin^2(\alpha)+\sin^2(\beta)+\sin^2(\gamma)$?

54. Find the angle between the line through $(1,3,2)$ and $(4,1,5)$ and the plane $x-y-2z+15=0$.

55. We showed that the distance from (p,q) to the line $Ax+By+C=0$ is $\frac{|Ap+Bq+C|}{\sqrt{A^2+B^2}}$. Show, using a similar argument, that the distance from (p,q,r) to the plane $Ax+By+Cx+D=0$ is $\frac{|Ap+Bq+Cr+D|}{\sqrt{A^2+B^2+C^2}}$.

56. How far apart are the planes $Ax+By+Cz+D=0$ and $Ax+By+Cz+E=0$? Explain.

57.
(a) Sketch a parabola and a line in the xy-plane that does not meet the parabola.
(b) Identify, graphically, the point on the parabola closest to the line.
(c) Using calculus, find the point on the parabola $y=x^2$ closest to the line $y=x-3$.
(d) Is the tangent to the parabola at the point found in (c) parallel to the line $y=x-3$? Explain.

58. The planes $A_1x+B_1y+C_1z+D_1=0$ and $A_2x+B_2y+C_2z+D_2=0$ intersect in a line L. Find the direction cosines of a vector parallel to L.

59. How far apart are the lines given parametrically by $2\mathbf{i}+\mathbf{j}-3\mathbf{k}+t(3\mathbf{i}-5\mathbf{j}+2\mathbf{k})$ and $3\mathbf{i}+\mathbf{j}+5\mathbf{k}+s(2\mathbf{i}+6\mathbf{j}+7\mathbf{k})$? We use different letters, s and t, for the parameters because they are independent of each other.

60.
(a) Use properties of determinants to show that the equation of a line through two distinct points (a_1,a_2) and (b_1,b_2) can be written as $\det\begin{pmatrix} x & y & 1 \\ a_1 & a_2 & 1 \\ b_1 & b_2 & 1 \end{pmatrix} = 0$.
(b) Assuming that all properties of determinants apply to any square matrix, the determinant of what 4×4 matrix would give an equation for the plane through three distinct points in space, say (a_1,a_2,a_3), (b_1,b_2,b_3), and (c_1,c_2,c_3)?

61. Does the line through $(5,7,10)$ and $(3,4,5)$ meet the line through $(1,4,0)$ and $(3,6,4)$? If so, where?

62. Develop a formula for determining the distance from $P_1=(x_1,y_1,z_1)$ to the line through $P_0=(x_0,y_0,z_0)$ that is parallel to $\mathbf{A}=a_1\mathbf{i}+z_2\mathbf{j}+a_3\mathbf{k}$. The formula should be in terms of $\overrightarrow{P_0P_1}$ and $\mathbf{A}$.

63. How far is $(1,2,-1)$ from the line through $(1,3,5)$ and $(2,1,-3)$?
(a) Solve by calculus, minimizing a function., and (b) Solve by vectors..

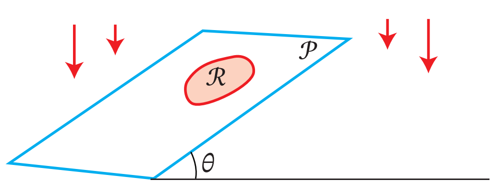


Figure 14.4.9

64. How small can the largest of three direction angles of a vector be?

65. The plane $\mathscr{P}$ in Figure 14.4.9 is tilted at an angle θ to a horizontal plane. A convex region $\mathscr{R}$ in $\mathscr{P}$ has area A. Show that if rays of light are perpendicular to the horizontal plane, the area of its projection on the horizontal plane is $A\cos(\theta)$. Show that the area of the shadow of $\mathscr{R}$ on that plane is $A\cos(\theta)$.

66.
(a) Let L_1 be the line through P_1 and Q_1 and let L_2 be the line through P_2 and Q_2. Assume that L_1 and L_2 are **skew lines** (that is, not parallel and not intersecting). How would you find the point R_1 on L_1 and point R_2 on L_2 such that $\overrightarrow{R_1R_2}$ is perpendicular to both L_1 and L_2?
(b) Find R_1 and R_2 when $P_1=(3,2,1)$, $Q_1=(1,1,1)$, $P_2=(0,2,0)$, $Q_2=(2,1,-1)$.

67. A square with side a lies in the plane $2x+3y+2z=8$. What is the area of its projection
(a) on the xy-plane? (b) on the yz-plane? (c) on the xz-plane?

68. A disk of radius a lies in the plane $2x+3y+4z=5$. What is the area of its projection on $2x+y-z=6$?

Exercises 69 to 71 show how vector ideas appear in different aspects of computer graphics.

69. How could one decide whether the origin and $P=(x_0,y_0,z_0)$ are on the same side or on opposite sides of the plane $Ax+By+Cz+D=0$?

70. How can one decide whether the points P and Q are on the same side or opposite sides of $Ax+By+Cz+D=0$?

71. Without drawing the triangle, devise a procedure for determining whether $P=(x,y)$ is inside the triangle whose three vertices are $P_1=(x_1,y_1)$, $P_2=(x_2,y_2)$, and $P_3=(x_3,y_3)$.

72. A tetrahedron has vertices P, Q, R, and S. Without drawing the tetrahedron, how can one decide whether a point T is inside the tetrahedron?

Just as the complex numbers form a mathematical system on the plane, the **quaternions** form a mathematical system on 4-dimensional space. The elements are of the form $r+a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$, where r, a, b, and c are real numbers, with the rules $\mathbf{i}^2=\mathbf{j}^2=\mathbf{k}^2=-1$ and $\mathbf{ij}=\mathbf{k}$. Quaternion multiplication is not commutative. Otherwise, the quaternions obey the usual rules of arithmetic. They are used in computer graphics to rotate objects.

73. Assuming that $\mathbf{i}^2=\mathbf{j}^2=\mathbf{k}^2=-1$ and $\mathbf{ij}=\mathbf{k}$, show that $\mathbf{ji}=-\mathbf{k}$, $\mathbf{jk}=\mathbf{i}$, $\mathbf{kj}=-\mathbf{i}$, $\mathbf{ki}=\mathbf{j}$, and $\mathbf{ik}=-\mathbf{j}$.

74. Quaternions of the form $a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$ can be viewed as vectors in our usual 3-dimensional space. Let $\mathbf{A}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}$ and $\mathbf{B}=b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k}$ and let $\mathbf{A}\star\mathbf{B}$ denote their product as quaternions. Show that $\mathbf{A}\star\mathbf{B}=-\mathbf{A}\cdot\mathbf{B}+\mathbf{A}\times\mathbf{B}$.

---

75. An industrial hopper is shaped as shown in Figure 14.4.10. Its top and bottom are squares of different sizes. The angle between the plane ABD and the plane BDC is 70°. The angle between the plane ABD and the plane ABC is 80°. What is the angle between plane ABC and plane BCD?

NOTE: The angle is needed during the fabrication of the hopper, since the planes ABC and BCD are made from a single piece of sheet metal bent along the edge BC.

CONTRIBUTED BY: Melvyn Kopald Stein

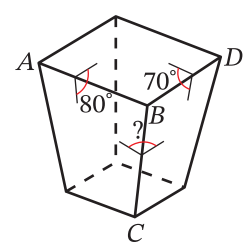


Figure 14.4.10

76. The three vertices of a triangle T of area 1 are P_1, P_2, and P_3. The origin is O. Let $\mathbf{P}_1=\overrightarrow{OP_1}$, $\mathbf{P}_2=\overrightarrow{OP_2}$, and $\mathbf{P}_3=\overrightarrow{OP_3}$. For each point P in triangle T the vector $\mathbf{P}=\overrightarrow{OP}$ can be represented in the form $w_1\mathbf{P}_1+w_2\mathbf{P}_2+w_3\mathbf{P}_3$, where $w_1+w_2+w_3=1$ and each w_i is nonnegative. Triangle T consists of three triangles that share the vertex P, as shown in Figure 14.4.11.

Their areas are A_1, A_2, and A_3.

(a) Using the fact that the area of T is 1, show that

$$\frac{1}{2}\left|\mathbf{P}_1\times\mathbf{P}_2+\mathbf{P}_2\times\mathbf{P}_3+\mathbf{P}_3\times\mathbf{P}_1\right|=1.$$

(b) Show that if $\mathbf{P}=w_1\mathbf{P}_1+w_2\mathbf{P}_2+w_3\mathbf{P}_3$, then $A_3=w_3$.

(c) Show that the points P for which w_3 is fixed lie on a line parallel to the edge P_1P_2.

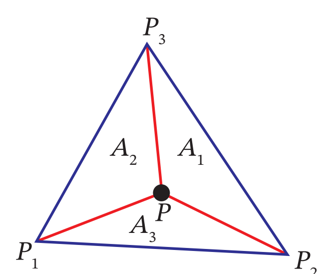


Figure 14.4.11

14.S Chapter Summary

Tables 14.S.1 and 14.S.2 summarize the chapter. Assume $\mathbf{A} = a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}$, $\mathbf{B} = b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k}$, and $\mathbf{C} = c_1\mathbf{i}+c_2\mathbf{j}+c_3\mathbf{k}$. For plane vectors disregard the third component.

Symbol	Name	Comment	Formula
$\mathbf{A}$	Vector	both direction and magnitude	$a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}$ or $\langle a_1, a_2, a_3\rangle$
$\lvert\mathbf{A}\rvert$	Length	also called magnitude or norm	$\sqrt{a_1^2+a_2^2+a_3^2}$
$-\mathbf{A}$	Negative, or opposite, of **A**	points in direction opposite of **A**	$-a_1\mathbf{i}-a_2\mathbf{j}-a_3\mathbf{k}$ or $\langle -a_1, -a_2, -a_3\rangle$
$\mathbf{A}+\mathbf{B}$	Sum of **A** and **B**	place the tail of **B** at the head of **A**	$(a_1+b_1)\mathbf{i}+(a_2+b_2)\mathbf{j}+(a_3+b_3)\mathbf{k}$
$\mathbf{A}-\mathbf{B}$	Difference of **A** and **B**	add $-\mathbf{B}$ to **A**	$(a_1-b_1)\mathbf{i}+(a_2-b_2)\mathbf{j}+(a_3-b_3)\mathbf{k}$
$c\mathbf{A}$	Scalar multiple of **A**	parallel to **A**, opposite direction if c is negative, $\lvert c\rvert$ times as long as **A**	$ca_1\mathbf{i}+ca_2\mathbf{j}+ca_3\mathbf{k}$
$\mathbf{A}\cdot\mathbf{B}$	Dot or scalar product	positive when angle between **A** and **B** is acute; zero when right angle; negative when obtuse	$\lvert\mathbf{A}\rvert\lvert\mathbf{B}\rvert\cos(\mathbf{A},\mathbf{B}) = a_1b_1+a_2b_2+a_3b_3$
$\mathbf{A}\times\mathbf{B}$	Cross or vector product	magnitude: area of parallelogram spanned by **A** and **B**, $\lvert\mathbf{A}\rvert\lvert\mathbf{B}\rvert\sin(\mathbf{A},\mathbf{B})$ direction: perpendicular to **A** and **B**, chosen by right-hand rule	$\det\begin{pmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\end{pmatrix}$
$\mathbf{proj_B}(\mathbf{A})$	(Vector) Projection of **A** on **B**	$\mathbf{A}-\mathbf{proj_B}(\mathbf{A})$ is perpendicular to **B**	$(\mathbf{A}\cdot\mathbf{u})\mathbf{u}$, where $\mathbf{u}=\mathbf{B}/\lvert\mathbf{B}\rvert$
$\mathrm{comp}_{\mathbf{B}}\mathbf{A}$	(Scalar) Component of **A** on **B**	can be positive, negative, or zero	$\mathbf{A}\cdot\mathbf{u}$, where $\mathbf{u}=\mathbf{B}/\lvert\mathbf{B}\rvert$
$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$	Scalar triple product	$\pm$ volume of parallelepiped spanned by **A**, **B**, and **C**	$\det\begin{pmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{pmatrix}$
$\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$	Vector triple product	memory device: 1. first write ____**B** - ____**C**. 2. then fill in the dot products.	$(\mathbf{A}\cdot\mathbf{C})\mathbf{B}-(\mathbf{A}\cdot\mathbf{B})\mathbf{C}$

Table 14.S.1 Basic combinations of vectors.

The following seven mathematical ideas are involved in a wide variety of applications.

1. The angle θ between **A** and **B** satisfies $\cos(\theta) = \dfrac{\mathbf{A}\cdot\mathbf{B}}{\lvert\mathbf{A}\rvert\lvert\mathbf{B}\rvert}$ $(0 \le \theta \le \pi)$.
2. Vectors **A** and **B** are perpendicular when $\mathbf{A}\cdot\mathbf{B} = 0$. They are parallel when $\mathbf{A}\times\mathbf{B} = \mathbf{0}$.
3. The equation of the plane through (x_0, y_0, z_0) perpendicular to **A** is $\mathbf{A}\cdot\langle x-x_0, y-y_0, z-z_0\rangle = 0$.
4. The distance from the point (p, q) to the line $Ax+By+C=0$ is $\dfrac{\lvert Ap+Bq+C\rvert}{\sqrt{A^2+B^2}}$.

Dot Product	Cross Product
$\mathbf{A}\cdot\mathbf{B}$	$\mathbf{A}\times\mathbf{B}$
$\mathbf{A}\cdot\mathbf{B}=\mathbf{B}\cdot\mathbf{A}$	$\mathbf{A}\times\mathbf{B}=-\mathbf{B}\times\mathbf{A}$
$\lvert\mathbf{A}\cdot\mathbf{B}\rvert=\lvert\mathbf{A}\rvert\lvert\mathbf{B}\rvert\lvert\cos(\theta)\rvert$	$\lvert\mathbf{A}\times\mathbf{B}\rvert=\lvert\mathbf{A}\rvert\lvert\mathbf{B}\rvert\sin(\theta)$
test for perpendicular vectors: $\mathbf{A}\cdot\mathbf{B}=0$	test for parallel vectors: $\mathbf{A}\times\mathbf{B}=\mathbf{0}$
formula in components involves a_ib_i (same indices)	formula in components involves a_ib_j (unequal indices)

Table 14.S.2 Comparison of the dot product and vector product.

5. The distance from the plane $Ax+By+Cz+D=0$ to the point (p,q,r) is $\dfrac{|Ap+Bq+Cr+D|}{\sqrt{A^2+B^2+C^2}}$.
6. The line through $P_0=(x_0,y_0,z_0)$ parallel to $\mathbf{A}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}$ is given parametrically as $\overrightarrow{OP}=\overrightarrow{OP_0}+t\mathbf{A}$ or, component-wise,

$$\begin{aligned} x &= x_0+a_1t \\ y &= y_0+a_2t \\ z &= z_0+a_3t \end{aligned}$$

7. The direction cosines of $\mathbf{A}$ are the numbers $\cos(\alpha)$, $\cos(\beta)$, and $\cos(\gamma)$ where α, β, and γ are the direction angles between $\mathbf{A}$ and $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, respectively. An important fact is $\cos^2(\alpha)+\cos^2(\beta)+\cos^2(\gamma)=1$.

EXERCISES for Section 14.S

1. Find a vector perpendicular to the plane determined by $(1,2,1)$, $(2,1,-3)$, and $(0,1,5)$.

2. Find a vector perpendicular to the plane determined by $(1,3,-1)$, $(2,1,1)$, and $(1,3,4)$.

3. Find a vector perpendicular to the line through $(3,6,1)$ and $(2,7,2)$ and to the line through $(2,1,4)$ and $(1,-2,3)$.

4. Find a vector perpendicular to the line through $(1,2,1)$ and $(4,1,0)$ and also to the line through $(3,5,2)$ and $(2,6,-3)$.

5. How far apart are the lines whose parametric forms are $2\mathbf{i}+4\mathbf{j}+\mathbf{k}+t(\mathbf{i}+\mathbf{j}+\mathbf{k})$ and $\mathbf{i}+3\mathbf{j}+2\mathbf{k}+s(2\mathbf{i}-\mathbf{j}-\mathbf{k})$?

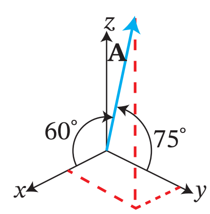


Figure 14.S.1

6. Find the direction cosines of the vector $\mathbf{A}$ shown in Figure 14.S.1.

7. Give two ways to determine whether $\mathbf{A}=\langle a_1,a_2,a_3\rangle$ and $\mathbf{B}=\langle b_1,b_2,b_3\rangle$ are parallel.

8. What is the ratio of the flows across the two sticks (labeled L) in Figure 14.S.2(a) and (b)?

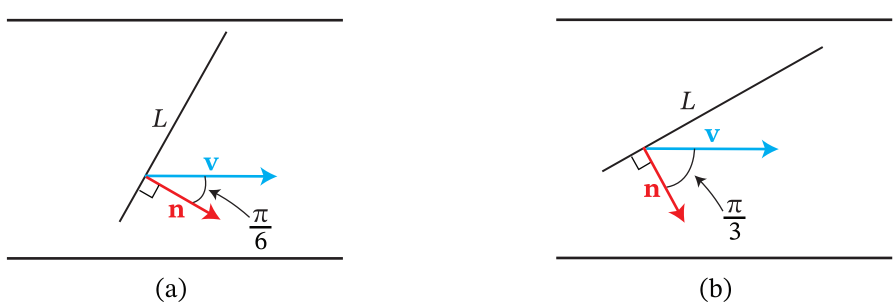


Figure 14.S.2

9. Find the point on the line through $(1,2,1)$ and $(2,-1,3)$ that is closest to the line that goes through $(3,0,3)$ and is parallel to the vector $\mathbf{i}+2\mathbf{j}+5\mathbf{k}$.

In Exercises 10 and 11, find the distance from the point to the line.

10. The point $(0,0)$ to the line $3x+4y-10=0$.

11. The point $\left(\frac{3}{2},\frac{2}{3}\right)$ to the line $2x-y+5=0$.

In Exercises 12 and 13 find a normal and a unit normal to the given planes.

12. $2x-3y+4z+11=0$

13. $z=2x-3y+4$

---

14. Is the line through $(1,1,1)$ and $(3,5,7)$ perpendicular to the plane $x+2y+3z+4=0$?

---

15. (a) Identify, geometrically, the point on the curve $y=\sin(x)$, $0\le x\le\pi$, that is nearest the line $y=\frac{x}{2}+2$.
(b) Find the coordinates of the point identified in (a), and the minimum distance to the line.

16. (a) Identify, geometrically, the point on the curve $y=\sin(x)$, $0\le x\le\pi$, that is nearest the line $y=2x+4$.
(b) Find the coordinates of the point identified in (a), and the minimum distance to the line.

---

17. Find the angle between the planes $x-y-z-1=0$ and $x+y+z+2=0$.

18. A line segment has projections of lengths a, b, and c on the coordinates axes. What, if anything, can be said about its length?

19. Suppose that the direction angles of a vector are equal. What can they be? Draw the cases.

20. What point on the line through $(1,2,5)$ and $(3,1,1)$ is closest to $(2,-1,5)$?

21. Three points $P_1=(x_1,y_1,z_1)$, $P_2=(x_2,y_2,z_2)$, and $P_3=(x_3,y_3,z_3)$ are the vertices of a triangle.
(a) What is the area of the triangle? (b) What is the area of the projection of that triangle on the xy-plane?

22. Explain how to decide whether the line through P and Q is parallel to the plane $Ax+By+Cz+D=0$.

23. Find where the line through $(1,1)$ and $(2,3)$ meets the line $x+2y+3=0$.

24. (a) Give an example of a vector perpendicular to the plane $2x+3y-z+4=0$.
(b) Give an example of a vector parallel to it.

25. **SAM:** Just because $\mathbf{i}\times\mathbf{j}$ obeys the right-hand rule that doesn't mean $\mathbf{A}\times\mathbf{B}$ does in general. I'm not convinced.
JANE: Oh, but it does settle the general case.
SAM: How so?
JANE: Slowly move and alter $\mathbf{i}$ and $\mathbf{j}$ so they become $\mathbf{A}$ and $\mathbf{B}$, never letting either one become $\mathbf{0}$ or letting them be parallel. If it's the right hand rule at the start, it can't shift to the left hand rule.
SAM: Why not?
JANE: Think about it.

Explain what Jane is thinking.

26. Assume that the planes $A_1x+B_1y+C_1z+D_1=0$ and $A_2x+B_2y+C_2z+D_2=0$ meet in a line L.
(a) Explain how to find a vector parallel to L.
(b) Explain how to find a point on L.
(c) Find parametric equations for the line that is the intersection of the planes $2x-y+3z+4=0$ and $3x+2y+5z+2=0$.

27. (a) How far is the point P from the line through Q and R?
(b) How far is $(2,1,3)$ from the line through $(1,5,2)$ and $(2,3,4)$?

28. (a) Explain how to decide whether the points $P_1=(x_1,y_1,z_1)$, $P_2=(x_2,y_2,z_2)$, $P_3=(x_3,y_3,z_3)$ and $P_4=(x_4,y_4,z_4)$ lie in a plane.
(b) Do the points $(1,2,3)$, $(4,1,-5)$, $(2,1,6)$, and $(3,5,3)$ lie in a plane?

29. What is the angle between the line through $(1,2,1)$ and $(-1,3,0)$ and the plane $x+y-2z=0$?

30. Explain why the projection of a circle is an ellipse.

31. Figure 14.S.3 shows a tetrahedron $OABC$ with edges of lengths 4, 5, and 6. Assume B is on the x-axis, C is on the y-axis, and A is on the z-axis.
(a) Find the coordinates of A, B, and C.
(b) Find the volume of the tetrahedron.
(c) Find the area of triangle ABC.
(d) Find the distance from O to the plane in which triangle ABC lies.
(e) Find the cosine of angle ABC.

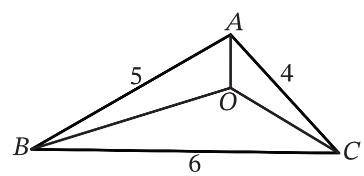


Figure 14.S.3

32. Let f be a differentiable function and L a line that does not meet the graph of f. Let P_0 be the point on the graph that is nearest the line. Assume that f is differentiable at P_0, in particular, that P_0 is not an endpoint of the domain of f. (a) Using calculus, show that the tangent at P_0 is parallel to L. (b) Why is this result expected?

33. Review the Folium of Descartes in Exercise 38 in Section 9.3. Show that its part in the fourth quadrant has the line $x+y+1=0$ as a (tilted) asymptote.

34. A convex set in a plane not parallel to any of the three coordinate planes has area A. The areas of the projections on the coordinate planes are 3, 4, and 5, respectively. What is A?

Application 14.S.1: ***Vectors and Computer Graphics***

Programmers in computer graphics use the dot and cross products for various purposes: finding a vector perpendicular to a polygon in a plane, determining shading and brightness, removing hidden objects, and more.

Exercises 35 to 45 are all related to topics in computer graphics.

35. Let P_1, P_2, and P_3 be the vertices of a triangle in the xy-plane, with the order P_1, P_2, and P_3 counterclockwise. P is a point inside the triangle (not on its border). Let $\mathbf{r}_i=\overrightarrow{PP_i}$, $i=1, 2, 3$. The vector $\frac{1}{2}(\mathbf{r}_1\times\mathbf{r}_2+\mathbf{r}_2\times\mathbf{r}_3+\mathbf{r}_3\times\mathbf{r}_1)$ has the form $A\mathbf{k}$ for a scalar A. Show that A is the area of the triangle.

36. Like Exercise 35, except that P is on the border of the triangle but not a vertex.

37. Like Exercise 35, except that P is a vertex of the triangle.

38. Like Exercise 35, except that P is in the xy-plane is outside the triangle.

The point of Exercises 35 to 38 is that this result is completely independent of the choice of P. So, for simplicity, it is perfectly fine to assume P is the origin. Exercises 39 to 41 show that this result is not limited to convex quadrilaterals: it applies to any polygon in the xy-plane.

39. Assume P_1, P_2, P_3, and P_4 are the vertices of a convex quadrilateral in the xy-plane, with the points in this order arranged counterclockwise, and P is any point in the xy-plane. Show that

$$\frac{1}{2}(\mathbf{r}_1 \times \mathbf{r}_2 + \mathbf{r}_2 \times \mathbf{r}_3 + \mathbf{r}_3 \times \mathbf{r}_4 + \mathbf{r}_4 \times \mathbf{r}_1) = A\mathbf{k}$$

where A is the area of the quadrilateral. (Cut the quadrilateral into two triangles.)

40. Generalize Exercise 39 to any quadrilateral, even concave.

41. Generalize Exercise 40 to any polygon in the xy-plane.

---

42. Light parallel to the vector $\mathbf{A}$ projects the line segment whose ends are P_1 and P_2 onto the plane containing the noncollinear points Q_1, Q_2, and Q_3. How long is the shadow of the segment on that plane?

43. Two points P and Q lie on a line L that has positive slope and P is to the left of Q. If R is a point not on L, how can $\overrightarrow{PQ} \times \overrightarrow{PR}$ be used to determine whether R is above or below L?

44. Sunlight in the direction $\mathbf{j}$ strikes the curve $y = x^2$. (Assume the curve is constructed from a homogeneous material, so light scatters in all directions equally.)

(a) Where is the curve the brightest?

(b) How does that brightness compare with the brightness at $(1, 1)$?

(c) at $(3, 9)$?

45. Lines L_1 and L_2 in space are not parallel and do not intersect. How can one decide which is in front of the other when viewed from point P or whether neither is blocked by the other because P is "between" them?

Calculus is Everywhere # 18

Space Flight: The Gravitational Slingshot

In a slingshot or gravitational assist a spacecraft picks up speed as it passes near a planet. For instance, New Horizons, launched on January 19, 2006, enjoyed a gravitational assist as it passed by Jupiter, February 27, 2007 on its long journey to Pluto. Its speed increased from 42,000 to more than 50,000 miles per hour (mph). It reached Pluto on July 14, 2015, instead of 2018.

Here on Earth baseball offers a simple analogy. A study of a batter's effectiveness compares the speed of the bat as it strikes the ball with the speed of the ball as it rebounds from the bat, the exit speed. The exit speed typically exceeds the bat speed by about 35 mph. As we will see, the batter plays the role of gravity.

Still staying on Earth, we examine in detail the case when a perfectly elastic ball is thrown at the vertical front windshield of an approaching train.

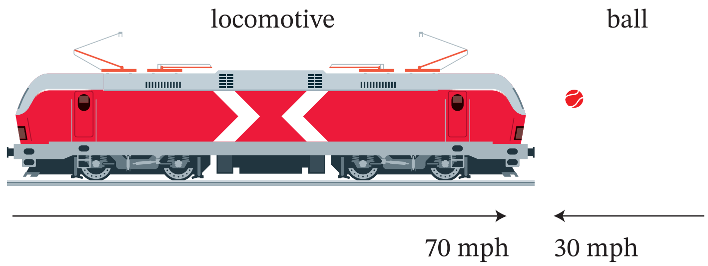


Figure C.18.1

A playful lad throws a perfectly elastic tiny ball at 30 mph directly at a train approaching him at 70 miles per hour, as shown in Figure C.18.1. The train engineer sees the ball coming toward her at $70+30 = 100$ mph. The ball hits the windshield and, because the ball is perfectly elastic, the driver sees it bounce off at 100 mph in the opposite direction. Because the train is moving in the same direction as the ball, the ball is moving through the air at $100+70 = 170$ mph as it returns to the boy. The ball has gained 140 mph, twice the speed of the train.

Instead of a train, think of a planet whose velocity relative to the solar system is represented by the vector $\mathbf{P}$. A spacecraft, moving in the opposite direction with the velocity $\mathbf{v}$ relative to the solar system, comes close to the planet.

An observer on the planet sees the spacecraft approaching with velocity $-\mathbf{P}+\mathbf{v}$. The spacecraft swings around the planet as gravity controls its orbit and sends it off in the opposite direction. Whatever speed it gained as it arrived, it loses as it exits. Its velocity vector when it exits is $-(-\mathbf{P}+\mathbf{v}) = \mathbf{P}-\mathbf{v}$ as viewed by the observer on the planet. Since the planet is moving through the solar system with velocity vector $\mathbf{P}$, the spacecraft is now moving through the solar system with velocity $\mathbf{P}+(\mathbf{P}-\mathbf{v}) = 2\mathbf{P}-\mathbf{v}$. See Figure C.18.2.

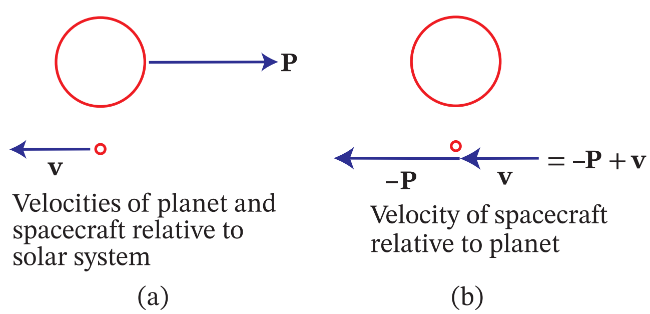


Figure C.18.2

The direction of the spacecraft as it arrives may not be exactly opposite the direction of the planet. To treat the general case assume that $\mathbf{P} = p\mathbf{i}$, where p is positive and $\mathbf{v}$ makes an angle θ, $0 \le \theta \le \pi/2$, with $-\mathbf{i}$, as shown in Figure C.18.3(a). Let $v = |\mathbf{v}|$ be the speed of the spacecraft relative to the solar system. We will assume that the spacecraft's speed (relative to the planet) as it exits is the same as its speed relative to the planet on its arrival. Figure C.18.3(b) shows the arrival vector, $\mathbf{v}-\mathbf{P}$, and the exit vector, $\mathbf{E}$. The y-components of $\mathbf{E}$ and $\mathbf{v}-\mathbf{P}$ are the same, but the x-component of $\mathbf{E}$ is the negative of the x-component of $\mathbf{v}-\mathbf{P}$.

Figure C.18.3(c) shows the arrival vector relative to the solar system, $\mathbf{v} = -v\cos(\theta)\mathbf{i}+v\sin(\theta)\mathbf{j}$.

Relative to the planet we have:

$$\begin{array}{lrcl} \text{Arrival Vector:} & \mathbf{v}-\mathbf{P} & = & -p\mathbf{i}+(-v\cos(\theta)\mathbf{i}+v\sin(\theta)\mathbf{j}) \\ \text{Exit Vector:} & \mathbf{E}_p & = & p\mathbf{i}+v\cos(\theta)\mathbf{i}+v\sin(\theta)\mathbf{j} \end{array}$$

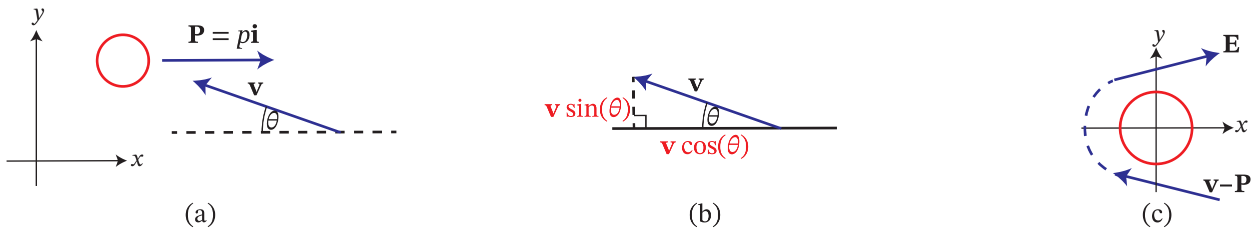


Figure C.18.3

The exit vector, $\mathbf{E}_s$ relative to the solar system is therefore

$$\mathbf{E}_s = (2p + v\cos(\theta))\mathbf{i} + v\sin(\theta)\mathbf{j}.$$

The magnitude of $\mathbf{E}_s$ is

$$\sqrt{(2p + v\cos(\theta))^2 + (v\sin(\theta))^2} = \sqrt{v^2 + 4pv\cos(\theta) + 4p^2}.$$

When $\theta = 0$, we have the train and ball or the planet and spacecraft in Figure C.18.2. Then $\cos(\theta) = 1$ and $|\mathbf{E}_s| = \sqrt{v^2 + 4pv + 4p^2} = v + 2p$, in agreement with our observations.

The scientists controlling a slingshot carry out more extensive calculations, which take into consideration the masses of the spacecraft and the planet, and involve an integration while the spacecraft is near the planet. "Near" for the slingshot around Jupiter means 1.4 million miles. If the spacecraft gets too close, the atmosphere slows down or destroys the craft. The diameter of Jupiter is 86,000 miles.

Historical Note: A Breakthrough for Space Exploration

The gravity assist was proposed by Michael Minovitch in 1963 when he was a graduate student at UCLA. Before then it was felt that to send a spacecraft to the outer solar system and beyond would require launch vehicles with nuclear reactors to achieve the necessary thrust.

Chapter 15

Derivatives and Integrals of Vector-Valued Functions

In Sections 9.3 to 9.6 we studied parametric curves in the plane. Using calculus we saw how to compute arc length, speed, and curvature. We defined curvature as the rate at which an angle changes as a function of arc length.

In this chapter we examine curves in the plane or in space. Of particular interest in Section 15.1 will be velocity and acceleration. For a particle moving along a straight line, say the x-axis, these are dx/dt and d^2x/dt^2. For a particle moving in space, velocity and acceleration involve both magnitude and direction. How should we calculate them?

How can we define curvature for a curve that does not lie in a plane? While arc length still makes sense, there is no angle to differentiate with respect to arc length. In Section 15.2, with the aid of vectors we will be able to define curvature for curves that do not necessarily lie in a plane. Curvature, in turn, will be of use in analyzing acceleration of an object moving along a curve.

In Section 15.3 we introduce integrals over curves and relate them, in Section 15.4 to work, flow of a fluid along the curve or across it, and to the radian measure of angles.

While we could answer these questions using the component notation for a parameterization $\langle x(t), y(t)\rangle$ or $\langle x(t), y(t), z(t)\rangle$, we prefer to use vector notation, where a curve is denoted by one letter. We will sometimes resort to the component notation to carry out computations or a proof.

15.1 The Derivative of a Vector-Valued Function: Velocity and Acceleration

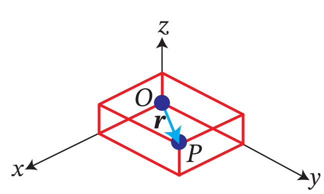


Figure 15.1.1

For motion on a horizontal line the derivative of position with respect to time is sufficient to describe the motion of the particle. If it is positive, the particle is moving to the right. If it is negative, the particle is moving to the left. The speed is the absolute value of the derivative. For motion in the plane or in space we need the derivative of a vector-valued function. This section introduces the calculus of a vector-valued function and applies it to motion along a curve in a plane or in space.

Defining the Derivative

Assume that a curve in the plane is parameterized as $\langle x(t), y(t)\rangle$ or, in space, by $\langle x(t), y(t), z(t)\rangle$. Let $P = P(t)$ be the point corresponding to t, which we may think of as time, though it can be any parameter, such as arc length.

The **position vector**, $\mathbf{r} = \mathbf{r}(t)$, has its tail at the origin O and its tip at P. Then $\mathbf{r} = \overrightarrow{OP}$, as shown in Figure 15.1.1.

We will assume that $\mathbf{r}(t)$ is continuous, in that each of its components is continuous. The vector-valued function $\mathbf{r}(t)$ is said to approach the vector $\mathbf{L}$ as t approaches a if $\lim_{t\to a} |\mathbf{r}(t)-\mathbf{L}| = 0$. We write $\lim_{t\to a} \mathbf{r}(t) = \mathbf{L}$. As Exercises 37 and 38 show, this is equivalent to the assertion that each of the scalar components of $\mathbf{r}(t)$ has a limit and $\mathbf{L} = \left\langle \lim_{t\to a} x(t), \lim_{t\to a} y(t), \lim_{t\to a} z(t)\right\rangle$.

Figure 15.1.2 shows this geometrically. As t approaches a, $\mathbf{r}(t) - \mathbf{r}(a)$ gets shorter as it approaches the zero vector $\mathbf{0}$.

We will say that $\mathbf{r}(t)$ is differentiable at $t = a$ if each of its components is differentiable at $t = a$. Then the derivative of $\mathbf{r}(t)$ is defined as the vector.

$$\langle x'(a), y'(a), z'(a)\rangle.$$

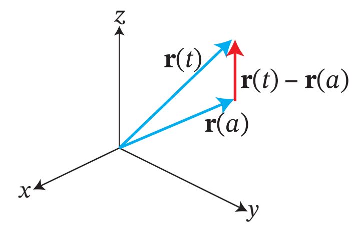


Figure 15.1.2

In vector notation,

$$\mathbf{r}'(a) = \lim_{t\to a} \frac{\mathbf{r}(t) - \mathbf{r}(a)}{t-a} \quad \text{or} \quad \mathbf{r}'(a) = \lim_{\Delta t\to 0} \frac{\mathbf{r}(a+\Delta t) - \mathbf{r}(a)}{\Delta t}.$$

and, if $\Delta\mathbf{r} = \mathbf{r}(a+\Delta t) - \mathbf{r}(t)$, $\mathbf{r}'(a) = \lim_{\Delta t\to 0} \Delta\mathbf{r}/\Delta t$. When t is near a (or Δt is near 0) the vector in the numerator will be short. It is divided by $t - a$ (or Δt), which is small, so the quotient could be a vector of any size.

Some Derivative Formulas

We state some useful identities:

Theorem 15.1.1: Derivative Properties of Vector-Valued Functions

If $\mathbf{r}$ and $\mathbf{s}$ are differentiable vector-valued functions, and f is a differentiable scalar function, then

$(\mathbf{r}+\mathbf{s})' = \mathbf{r}' + \mathbf{s}'$	*derivative of a sum*
$(f\mathbf{r})' = f'\mathbf{r} + f\mathbf{r}'$	*product rule (derivative of a product))*
$(\mathbf{r}\times\mathbf{s})' = \mathbf{r}'\times\mathbf{s} + \mathbf{r}\times\mathbf{s}'$	*derivative of a cross product*
$(\mathbf{r}\cdot\mathbf{s})' = \mathbf{r}'\cdot\mathbf{s} + \mathbf{r}\cdot\mathbf{s}'$	*derivative of a dot product*
$(\mathbf{r}(f(t)))' = \mathbf{r}'(f(t))f'(t)$	*chain rule (derivative of a composition).*

The proofs are straightforward. A key step in each one (except the chain rule) is to distribute a limit over an appropriate sum or product. That is, assuming that $\lim_{t\to a}\mathbf{r}(t)$ and $\lim_{t\to a}\mathbf{r}(t)$ both exist, then

$$\lim_{t\to a}(\mathbf{r}(t)+\mathbf{s}(t)) = \lim_{t\to a}\mathbf{r}(t) + \lim_{t\to a}\mathbf{s}(t)$$
$$\lim_{t\to a}(\mathbf{r}(t)\cdot\mathbf{s}(t)) = \lim_{t\to a}\mathbf{r}(t) \cdot \lim_{t\to a}\mathbf{s}(t)$$
$$\lim_{t\to a}(\mathbf{r}(t)\times\mathbf{s}(t)) = \lim_{t\to a}\mathbf{r}(t) \times \lim_{t\to a}\mathbf{s}(t)$$

These results, which are not surprising, can be proven by expanding each operation componentwise.

We prove only one of the derivative formulas, but do so in two different ways: component and vector notation.

Proof of One Part of Theorem 15.1.1: Derivative of a Dot Product

Two proofs of the derivative rule for the dot product of two vector-valued functions will be provided: one using components and another using vectors.

First, a proof using components. For convenience, assume $\mathbf{r}(t)$ and $\mathbf{s}(t)$ are vectors in the xy-plane: $\mathbf{r}(t) = \langle x(t), y(t)\rangle$ and $\mathbf{s}(t) = \langle u(t), v(t)\rangle$.

$$\begin{aligned}(\mathbf{r}\cdot\mathbf{s})' = (x(t)u(t) + y(t)v(t))' \quad &= x'u + xu' + y'v + yv' \\ &= (x'u + y'v) + (xu' + yv') = \mathbf{r}'\cdot\mathbf{s} + \mathbf{r}\cdot\mathbf{s}'.\end{aligned}$$

And, second, the proof using vectors. This proof starts with the definition of the derivative of a vector-valued function, and then uses the definitions $\Delta\mathbf{r} = \mathbf{r}(a+\Delta t) - \mathbf{r}(a)$ and $\Delta\mathbf{s} = \mathbf{s}(a+\Delta t) - \mathbf{s}(a)$:

$$\begin{aligned}
(\mathbf{r}\cdot\mathbf{s})' &= \lim_{\Delta t\to 0} \frac{\mathbf{r}(a+\Delta t)\cdot\mathbf{s}(a+\Delta t) - \mathbf{r}(a)\cdot\mathbf{s}(a)}{\Delta t} \\
&= \lim_{\Delta t\to 0} \frac{(\mathbf{r}(a)+\Delta\mathbf{r})\cdot(\mathbf{s}(a)+\Delta\mathbf{s}) - \mathbf{r}(a)\cdot\mathbf{s}(a)}{\Delta t} \\
&= \lim_{\Delta t\to 0} \frac{\mathbf{r}(a)\cdot\mathbf{s}(a) + \Delta\mathbf{r}\cdot\mathbf{s}(a) + \mathbf{r}(a)\cdot\Delta\mathbf{s} + \Delta\mathbf{r}\cdot\Delta\mathbf{s} - \mathbf{r}(a)\cdot\mathbf{s}(a)}{\Delta t} \\
&= \lim_{\Delta t\to 0} \left(\frac{\Delta\mathbf{r}}{\Delta t}\cdot\mathbf{s}(a) + \mathbf{r}(a)\cdot\frac{\Delta\mathbf{s}}{\Delta t} + \Delta\mathbf{r}\cdot\frac{\Delta\mathbf{s}}{\Delta t}\right) \\
&= \mathbf{r}'(a)\cdot\mathbf{s}(a) + \mathbf{r}(a)\cdot\mathbf{s}'(a) + \mathbf{0}\cdot\mathbf{s}'(a) \\
&= \mathbf{r}'(a)\cdot\mathbf{s}(a) + \mathbf{r}(a)\cdot\mathbf{s}'(a).
\end{aligned}$$

This resembles the proof for the derivative of the product in Section 4.3. ●

EXAMPLE 1. At time t, a particle has the position vector $\mathbf{r}(t) = 3\cos(2\pi t)\mathbf{i} + 3\sin(2\pi t)\mathbf{j} + 5t\mathbf{k}$. Describe its path.

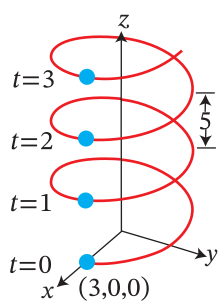


Figure 15.1.3

SOLUTION At time t the particle is at the point (x, y, z) with

$$x = 3\cos(2\pi t), \quad y = 3\sin(2\pi t), \quad \text{and} \quad z = 5t.$$

Because $x^2 + y^2 = (3\cos(2\pi t))^2 + (3\sin(2\pi t))^2 = 9$, the particle is always directly above or below a point on the circle $x^2 + y^2 = 9$.

As t increases, $z = 5t$ increases.

The path is thus a spiral sketched in Figure 15.1.3. When t increases by 1, the angle $2\pi t$ increases by 2π, and the particle goes around the spiral once. The path is called a **helix**. ◆

The Meaning of $\mathbf{r}'$ and $\mathbf{r}''$

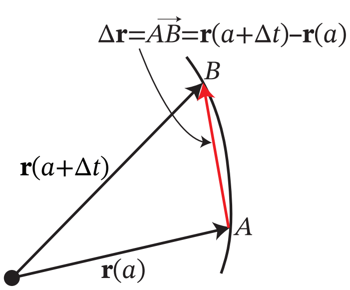


Figure 15.1.4

The vector $\mathbf{r}'(a)$ is the limit of $(\mathbf{r}(a+\Delta t) - \mathbf{r}(a))/\Delta t$ as $\Delta t \to 0$. The numerator $\mathbf{r}(a+\Delta t) - \mathbf{r}(a) = \Delta\mathbf{r}$ is shown in Figure 15.1.4.

Since $\Delta\mathbf{r}$ coincides with a chord, it points almost along the tangent line at the head of $\mathbf{r}(a)$ when Δt is small. Dividing a vector by a scalar produces a parallel vector. The position vector is $\mathbf{r}(t)$, so $(\mathbf{r}(a+\Delta t) - \mathbf{r}(a))/\Delta t$ approximates a vector tangent to the curve at a. We conclude that

$$\mathbf{r}'(a) = \lim_{\Delta t\to 0} \frac{\mathbf{r}(a+\Delta t) - \mathbf{r}(a)}{\Delta t}$$

is a vector tangent to the curve at $\mathbf{r}(a)$. That is the geometric meaning of the direction of the derivative $\mathbf{r}'$.

To see what $\mathbf{r}'$ means when t is time, we compute its length. Since $\mathbf{r}'(t) = \langle x'(t), y'(t), z'(t)\rangle$, its length is

$$\sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2}.$$

This is the natural extension to three dimensions of the speed on a plane curve in Section 9.4.

We can also see that the magnitude of $\mathbf{r}'(t)$ is the speed by using vector language. For small Δt, the vector $\Delta\mathbf{r}$ lies on a short chord of the curve and its length is close to the length of arc swept out during that short interval of time. (See Figure 15.1.4.) Thus the magnitude of $\Delta\mathbf{r}/\Delta t$ approximates the speed.

Since $\mathbf{r}'(t)$ points in the direction of motion and its length is the speed, we call $\mathbf{r}'(t)$ the **velocity vector**. Note that velocity is a vector, while speed is a scalar. Velocity carries much more information than speed; it also tells the direction of the motion.

Definition: *Velocity, Speed, and Acceleration*

The **velocity** is the instantaneous rate of change of the position: $\mathbf{v}(t) = \mathbf{r}'(t)$.
The **speed** is the length of the velocity vector: $s(t) = |\mathbf{v}(t)| = |\mathbf{r}'(t)|$.
The **acceleration** is the derivative of velocity: $\mathbf{a}(t) = \mathbf{v}'(t) = \frac{d\mathbf{v}}{dt} = \mathbf{r}''(t) = \frac{d^2\mathbf{r}}{dt^2}$.

EXAMPLE 2. Let $\mathbf{r}(t) = \langle t, t^3 \rangle$.

(a) Draw and label **r**, **v**, and **a** at $t = 1$. Base the vectors **a** and **a** at the head of **r**.
(b) Draw and label **r** and **r** at $t = 1.1$. Base the vector **a** at the head of **r**.

SOLUTION

(a) $\mathbf{r}(t) = \langle t, t^3 \rangle$, $\mathbf{v}(t) = \langle 1, 3t^2 \rangle$, and $\mathbf{a} = \langle 0, 6t \rangle$. So $\mathbf{r}(1) = \langle 1, 1 \rangle$, $\mathbf{v}(1) = \langle 1, 3 \rangle$ and $\mathbf{a}(1) = \langle 0, 6 \rangle$. We show these in Figure 15.1.5(a).

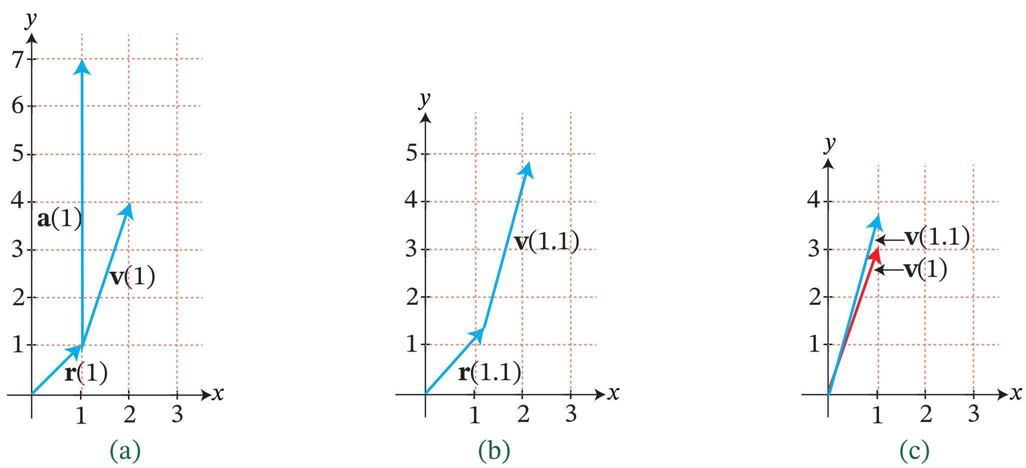


Figure 15.1.5

(b) Before we compute and graph $\mathbf{v}(1.1)$, let us predict how it may change from $\mathbf{v}(1)$. The acceleration vector represents a force representing the rate of change of the velocity vector at an instant in time. Since $\mathbf{a}(1)$ is almost in the direction of $\mathbf{v}(1)$, the particle is speeding up. That is, $\mathbf{v}(1.1)$ should be longer than $\mathbf{v}(1)$.

Also, because the acceleration vector at $t = 1$ points upward, the velocity vector will rotate counterclockwise. So the direction of $\mathbf{v}(1.1)$ should be a bit counterclockwise from that of $\mathbf{v}(1)$. To check, we compute $\mathbf{v}(1.1) = \langle 1.1, 3(1.1)^2 \rangle = \langle 1.1, 3.63 \rangle$.

The vectors $\mathbf{r}(1.1)$ and $\mathbf{v}(1.1)$ are graphed in Figure 15.1.5(b). The velocity vector is, as expected, longer than $\mathbf{v}(1) = \langle 1, 3 \rangle$ since $\sqrt{(1.1)^2 + (3.63)^2}$ is larger than $\sqrt{1 + 3^2}$. Figure 15.1.5(c) shows that it is turned a bit counterclockwise, as expected. Its tail is placed at the head of

$$\mathbf{r}(1.1) = \langle 1.1, 1.331 \rangle = 1.1\mathbf{i} + 1.331\mathbf{j}.$$

◆

EXAMPLE 3. Find the speed at time t of the particle described in Example 1.

SOLUTION From Example 1, the position is $\mathbf{r}(t) = \langle 3\cos(2\pi t), 3\sin(2\pi t), 5t \rangle$. So the velocity is

$$\mathbf{v}(t) = \mathbf{r}'(t) = \langle -6\pi \sin(2\pi t), 6\pi \cos(2\pi t), 5 \rangle$$

and the speed is

$$\begin{aligned}
|\mathbf{r}'(t)| &= \sqrt{(-6\pi\sin 2\pi t)^2 + (6\pi\cos 2\pi t)^2 + 5^2} \\
&= \sqrt{36\pi^2(\sin^2 2\pi t + \cos^2 2\pi t) + 25} \\
&= \sqrt{36\pi^2 + 25}.
\end{aligned}$$

The particle travels at a constant speed along its helical path. In t units of time it travels the distance $\sqrt{36\pi^2+25}\ t$.

The velocity vector is not constant because its direction always changes. However, its length remains constant, which means that its speed is constant. ◆

EXAMPLE 4. Sketch the path of a particle whose position vector at time $t \geq 0$ is $\mathbf{r}(t) = \cos(t^2)\mathbf{i} + \sin(t^2)\mathbf{j}$. Find its speed at time t.

SOLUTION Because $|\mathbf{r}(t)| = \sqrt{\cos^2(t^2) + \sin^2(t^2)} = 1$ the path of the particle is on the circle of radius 1 and center $(0,0)$. Its speed is

$$\begin{aligned}
|\mathbf{v}(t)| = |\mathbf{r}'(t)| &= |-2t\sin(t^2)\mathbf{i} + 2t\cos(t^2)\mathbf{j}| \\
&= \sqrt{(-2t\sin(t^2))^2 + (2t\cos(t^2))^2} \\
&= 2t\sqrt{\sin^2(t^2) + \cos^2(t^2)} \\
&= 2t.
\end{aligned}$$

The particle travels faster and faster as it keeps retracing the unit circle. ◆

EXAMPLE 5. If the acceleration vector and velocity vector are always perpendicular, show that the speed is constant.

SOLUTION The speed is $|\mathbf{v}|$. We will show that the square of the speed, $|\mathbf{v}|^2$, is constant by showing that its derivative with respect to time is zero. Since $|\mathbf{v}|^2 = \mathbf{v}\cdot\mathbf{v}$, we have

$$\frac{d}{dt}\left(|\mathbf{v}|^2\right) = \frac{d}{dt}(\mathbf{v}\cdot\mathbf{v}) = \mathbf{v}'\cdot\mathbf{v} + \mathbf{v}\cdot\mathbf{v}' = 2\,\mathbf{v}\cdot\mathbf{v}' = 2\,\mathbf{v}\cdot\mathbf{a}.$$

Because $\mathbf{a}$ is perpendicular to $\mathbf{v}$ we know that $\mathbf{v}\cdot\mathbf{a} = 0$.

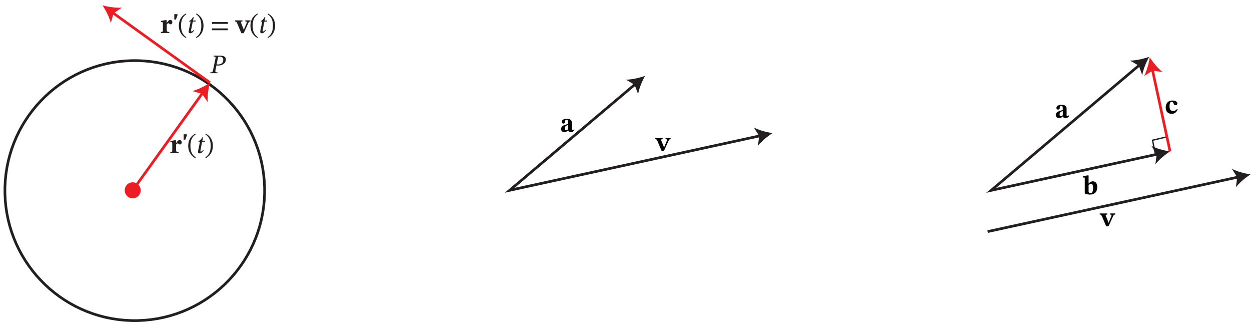
Thus $\mathbf{v}\cdot\mathbf{v}$ is constant, so the speed is constant. ◆

The calculation in Example 5, with $\mathbf{v}$ and $\mathbf{a}$ replaced by $\mathbf{r}$ and $\mathbf{v} = \mathbf{r}'$, implies that if $\mathbf{r}(t)$ is always perpendicular to $\mathbf{r}'(t)$, then the length of $\mathbf{r}(t)$ is constant. The converse is also true: If the length of $\mathbf{r}(t)$ is constant, then $\mathbf{r}'(t)$ is perpendicular to $\mathbf{r}(t)$.

This is not surprising. If $\mathbf{r}(t)$ is constant, then $\mathbf{r}(t)$ lies on a sphere of radius $|\mathbf{r}(t)|$. A tangent to the curve at P is tangent to the sphere. The tangent to a sphere is perpendicular to its radius, as indicated in Figure 15.1.6(a). Consequently, $\mathbf{r}'(t)$ is perpendicular to $\mathbf{r}(t)$.

EXAMPLE 6. Is the particle shown in Figure 15.1.6(b) speeding up or slowing down? Is its direction turning clockwise or counterclockwise?

SOLUTION Represent $\mathbf{a}$ as the sum of two vectors, $\mathbf{b}$ parallel to $\mathbf{v}$, and $\mathbf{c}$ perpendicular to $\mathbf{v}$, as shown in Figure 15.1.6(c). Since $\mathbf{b}$ is in the same direction as $\mathbf{v}$, the particle is speeding up. The direction of $\mathbf{c}$ indicates that the direction of $\mathbf{v}$ is turning counterclockwise. ◆



(a) (b) (c)

Figure 15.1.6

Summary

Instead of parameterizing a curve by displaying its components $\langle x(t), y(t)\rangle$ or $\langle x(t), y(t), z(t)\rangle$, we introduced the position vector $\overrightarrow{OP} = \mathbf{r}(t)$. If $\mathbf{r}(t)$ describes the position of a moving particle at time t, then $\mathbf{r}'(t)$ is the velocity of the particle and $|\mathbf{r}'(t)|$ is its speed. The acceleration $\mathbf{a}(t)$ is the second derivative of $\mathbf{r}(t)$: $\mathbf{a} = \mathbf{r}''$. It is proportional to the force operating on the particle.

We showed that if $\mathbf{r}(t)$ and $\mathbf{r}'(t)$ are perpendicular, then the length of $\mathbf{r}(t)$, $|\mathbf{r}(t)|$, is constant. The converse holds: If $\mathbf{r}(t)$ has constant length, then $\mathbf{r}'(t)$ is perpendicular to $\mathbf{r}(t)$, and $\mathbf{r}(t)\cdot\mathbf{r}'(t) = 0$.

EXERCISES for Section 15.1

1. At time t a particle has the position vector $\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j}$.
(a) Compute and draw, on one set of axes, $\mathbf{r}(1)$, $\mathbf{r}(2)$, and $\mathbf{r}(3)$. (b) Show that the path is a parabola.

2. At time t a particle has the position vector $\mathbf{r}(t) = (2t+1)\mathbf{i} + 4t\mathbf{j}$.
(a) Compute and draw, on one set of axes, $\mathbf{r}(0)$, $\mathbf{r}(1)$, and $\mathbf{r}(2)$. (b) Show that the path is a straight line.

3. Let $\mathbf{r}(t) = 2t\mathbf{i} + t^2\mathbf{j}$.
- (a) Compute and draw, on one set of axes, $\mathbf{r}(1.1)$, $\mathbf{r}(1)$, and their difference $\Delta\mathbf{r} = \mathbf{r}(1.1) - \mathbf{r}(1)$.
- (b) Compute $\dfrac{\Delta\mathbf{r}}{0.1}$, then update the graph in (a) to include this vector.
- (c) Compute $\mathbf{r}'(1)$, then update the graph in (b) to include this vector.

4. Let $\mathbf{r}(t) = 3t\mathbf{i} + t^2\mathbf{j}$.
- (a) Compute and draw, on one set of axes, $\mathbf{r}(2.01)$, $\mathbf{r}(2)$, and their difference $\Delta\mathbf{r} = \mathbf{r}(2.01) - \mathbf{r}(2)$.
- (b) Compute and draw $\dfrac{\Delta\mathbf{r}}{0.01}$, then update the graph in (a) to include this vector.
- (c) Compute and draw $\mathbf{r}'(2)$, then update the graph in (b) to include this vector.

5. Let $\mathbf{r}(t) = 32t\mathbf{i} - 16t^2\mathbf{j}$.
- (a) Draw $\mathbf{r}(1)$ and $\mathbf{r}(2)$.
- (b) Sketch the path.
- (c) Compute and draw $\mathbf{v}(0)$, $\mathbf{v}(1)$, and $\mathbf{v}(2)$, with the tail of each vector at the head of the corresponding position vector.

6. At time $t \geq 0$ a particle is at the point $x = 2t$, $y = 4t^2$.
- (a) What is the position vector $\mathbf{r}(t)$ at time t?
- (b) Sketch the path.
- (c) How fast is the particle moving when $t = 1$?
- (d) Draw $\mathbf{v}(1)$ with its tail at the head of $\mathbf{r}(1)$.

7. The path of a particle moving in the plane is given by $\mathbf{r}(t)$. If $\mathbf{r}(1) = 2.3\mathbf{i} + 4.1\mathbf{j}$ and $\mathbf{r}(1.2) = 2.31\mathbf{i} + 4.05\mathbf{j}$,
- (a) estimate how far the particle moves during the time interval $[1, 1.2]$,
- (b) estimate the slope of the tangent vector at $\mathbf{r}(1)$,
- (c) estimate the velocity vector $\mathbf{r}'(1)$, and
- (d) estimate the speed of the particle at time $t = 1$.

8. The path of a particle moving in space is given by $\mathbf{r}(t)$. If $\mathbf{r}(2) = 1.7\mathbf{i} + 3.6\mathbf{j} + 8\mathbf{k}$ and $\mathbf{r}(2.01) = 1.73\mathbf{i} + 3.59\mathbf{j} + 8.02\mathbf{k}$,
- (a) estimate how far the particle moves during the time interval $[2, 2.01]$,
- (b) estimate the velocity vector $\mathbf{r}'(2)$, and
- (c) estimate the speed of the particle at time $t = 1$.

In Exercises 9 to 12 compute the velocity vector and the speed.

9. $\mathbf{r}(t) = \cos(3t)\mathbf{i} + \sin(3t)\mathbf{j} + 6t\mathbf{k}$

10. $\mathbf{r}(t) = 3\cos(5t)\mathbf{i} + 2\sin(5t)\mathbf{j} + t^2\mathbf{k}$

11. $\mathbf{r}(t) = \ln(1 + t^2)\mathbf{i} + e^{3t}\mathbf{j} + \dfrac{\tan(t)}{1 + 2t}\mathbf{k}$

12. $\mathbf{r}(t) = \sec^2 3t\mathbf{i} + \sqrt{1 + t^2}\mathbf{j}$

---

13. At time t the position vector of a particle is $\mathbf{r}(t) = 2\cos(4\pi t)\mathbf{i} + 2\sin(4\pi t)\mathbf{j} + t\mathbf{k}$.
(a) Sketch its path. (b) Find its speed. (c) Find a unit tangent vector to its path at time t.

In Exercises 14 to 21 the figure shows a velocity vector and an acceleration vector. Decide, when possible, whether (a) the particle is speeding up, slowing down, or neither, (b) the velocity vector is turning clockwise, counterclockwise, or neither.

14. Figure 15.1.7(a) **15.** Figure 15.1.7(b) **16.** Figure 15.1.7(c) **17.** Figure 15.1.7(d)

18. Figure 15.1.7(e) **19.** Figure 15.1.7(f) **20.** Figure 15.1.7(g) **21.** Figure 15.1.7(h)

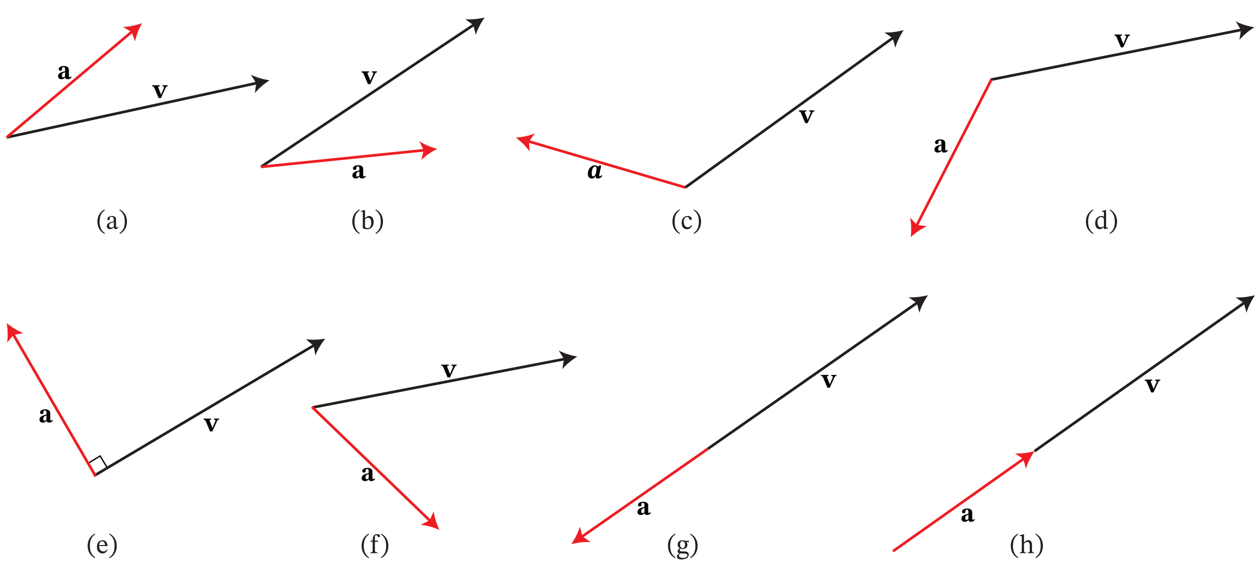


Figure 15.1.7

22. At time t a particle is at $(4t, 16t^2)$.
(a) Show that the particle moves on the curve $y = x^2$.
(b) Draw $\mathbf{r}(t)$ and $\mathbf{v}(t)$ for $t = 0, \frac{1}{4}, \frac{1}{2}$.
(c) What happens to $|\mathbf{v}(t)|$ and the direction of $\mathbf{v}(t)$ for large t?

23. At time $t \geq 1$ a particle is at the point $(x, y) = (t, t^{-1})$.
(a) Draw the path of the particle.
(b) Draw $\mathbf{r}(1)$, $\mathbf{r}(2)$, and $\mathbf{r}(3)$.
(c) Draw $\mathbf{v}(1)$, $\mathbf{v}(2)$ and $\mathbf{v}(3)$.
(d) As t increases, what happens to $\frac{dx}{dt}$, $\frac{dy}{dt}$, $|\mathbf{v}|$, and $\mathbf{v}$?

24. At time t a particle is at $\left(2\cos\left(t^2\right), \sin\left(t^2\right)\right)$.
(a) Show that it moves on an ellipse.
(b) Compute $\mathbf{v}(t)$.
(c) How does $|\mathbf{v}(t)|$ behave for large t? What does this say about the particle?

25. An electron travels at constant speed clockwise in a circle of radius 100 feet 200 times a second. At time $t = 0$ it is at $(100, 0)$. (a) Compute $\mathbf{r}(t)$ and $\mathbf{v}(t)$. (b) Draw $\mathbf{r}(0)$, $\mathbf{r}\left(\frac{1}{800}\right)$, $\mathbf{v}(0)$, and $\mathbf{v}\left(\frac{1}{800}\right)$. (c) How do $|\mathbf{r}(t)|$ and $|\mathbf{v}(t)|$ behave as t increases?

26. A ball is thrown up at an initial speed of 200 feet per second and at an angle of 60° from the horizontal. At time t it is at $\left(100t, 100\sqrt{3}t - 16t^2\right)$. Compute and draw $\mathbf{r}(t)$ and $\mathbf{v}(t)$ (a) when $t = 0$, (b) when the ball reaches its maximum height, and (c) when the ball strikes the ground.

27. A particle moves in a circular orbit of radius a. At time t its position vector is $\mathbf{r}(t) = a\cos(2\pi t)\mathbf{i} + a\sin(2\pi t)\mathbf{j}$.
(a) Draw its position vector when $t = 0$ and when $t = \frac{1}{4}$.
(b) Draw its velocity when $t = 0$ and when $t = \frac{1}{4}$.
(c) Show that its velocity vector is always perpendicular to its position vector.

28. In Example 5 the speed is constant.
(a) Is the acceleration vector constant also? (b) Does the acceleration vector have constant length?

29. Use a computer or graphing calculator to graph $\mathbf{r} = \mathbf{r}(t) = (2\cos(t) + \cos(3t))\mathbf{i} + (3\sin(t) + \sin(3t))\mathbf{j}$, $0 \leq t \leq 2\pi$.

30. If $\mathbf{r}(t)$ is the position vector, $\mathbf{v}$ the velocity vector, and $\mathbf{a}$ the acceleration vector, show that $\frac{d}{dt}(\mathbf{r} \times \mathbf{v}) = \mathbf{r} \times \mathbf{a}$.

31. Let $\mathbf{r}(t) = t^2\mathbf{i} + t^3\mathbf{j}$. (a) Sketch the vector $\Delta\mathbf{r} = \mathbf{r}(1.1) - \mathbf{r}(1)$. (b) Sketch the vector $\frac{\Delta\mathbf{r}}{\Delta t}$ for $\Delta t = 0.1$. (c) Sketch $\mathbf{r}'(1)$. (d) Find $\left|\frac{\Delta\mathbf{r}}{\Delta t} - \mathbf{r}'(1)\right|$ for $\Delta t = 0.1$.

32. A rock is thrown up at an angle θ from the horizontal and at a speed v_0.
(a) Show that $\mathbf{r}(t) = (v_0\cos(\theta))t\mathbf{i} + ((v_0\sin(\theta))t - 16t^2)\mathbf{j}$. (b) Show that the horizontal distance that the rock travels by the time it returns to its initial height is the same whether the angle is θ or $\frac{\pi}{2} - \theta$. (c) What value of θ maximizes the horizontal distance traveled?

LOOK FAMILIAR? Exercise 32 is similar to Exercise 24 in Section 9.3. The only difference is the use of vectors.

33. Instead of t, use the arc length s along the path as a parameter, so $\mathbf{r} = \mathbf{r}(s)$.
(a) Sketch $\Delta\mathbf{r}$ and the arc of length Δs. Why is it reasonable that $\left|\frac{\Delta\mathbf{r}}{\Delta s}\right|$ is near 1 when Δs is small?
(b) Show that $\frac{d\mathbf{r}}{ds}$ is a unit vector.

34. A particle at time $t = 0$ is at the point (x_0, y_0, z_0). It moves on the line through that point in the direction of the unit vector $\mathbf{u} = \cos(\alpha)\mathbf{i} + \cos(\beta)\mathbf{j} + \cos(\gamma)\mathbf{k}$. It travels at the constant speed of 3 feet per second.
(a) Give a formula for its position vector $\mathbf{r} = \mathbf{r}(t)$. (b) Find its velocity vector $\mathbf{v} = \mathbf{r}'(t)$.

35.
(a) Solve Example 5 by writing the speed as $\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}$ and differentiating.
(b) Which way do you prefer? The vector method in Example 5 or the component method in (a)?

36. The force of a magnetic field on a moving electron is always perpendicular to the electron's velocity vector, $\mathbf{v}$. What does this imply about $\mathbf{v}$?

37. Show that if $\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle$ and $\lim_{t\to a} |\mathbf{r}(t) - \langle p, q, r\rangle| = 0$, then $\lim_{t\to a} x(t) = p$, $\lim_{t\to a} y(t) = q$, and $\lim_{t\to a} z(t) = r$.

38. Show the converse of the preceding exercise, namely that if the scalar components have limits, so does $\mathbf{r}(t)$.

39. At time t the position vector of a particle is $\mathbf{r}(t) = t\cos(2\pi t)\mathbf{i} + t\sin(2\pi t)\mathbf{j} + t\mathbf{k}$. Sketch its path.

40. A spaceship is on the path $\mathbf{r}(t) = t^2\mathbf{i} + 3t\mathbf{j} + 4t^3\mathbf{k}$. At time $t = 1$ it shuts off its rockets and coasts along the tangent line to the curve at that point. (a) Where is it at time $t > 1$? (b) Does it pass through the point $(9, 15, 50)$? (c) If not, how close does it get to that point? At what time?

41. A particle traveling on the curve $\mathbf{r}(t) = \ln(t)\mathbf{i} + \cos(3t)\mathbf{j}$, $t \geq 1$, leaves the curve when $t = 2$ and travels on the xy-plane along the tangent to the curve at $\mathbf{r}(2)$. Where is it when $t = 3$?

42. Drawing a picture of $\mathbf{r}(t)$, $\mathbf{r}(t + \Delta t)$, and $\Delta\mathbf{r}$, explain why $\left|\frac{\Delta\mathbf{r}}{\Delta t}\right|$ is an estimate of the speed of a particle moving on the curve $\mathbf{r}(t)$.

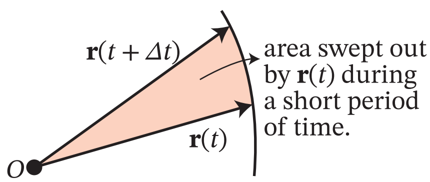


Figure 15.1.8

43. The moment a ball is dropped straight down from a tall tree, you shoot an arrow directly at it. Assume that there is no air resistance (and no obstacles). Show that the arrow will hit the ball, assuming that the ball does not hit the ground first.
(a) Solve using the formulas in Exercise 32.
(b) Solve with a maximum of intuition and a minimum of computation.

44. (a) At time t a particle has the position vector $\mathbf{r}(t)$. Show that for small Δt the area swept out by the position vector is approximately $\frac{1}{2}|\mathbf{r}(t) \times \mathbf{v}(t)|\Delta t$. (See Figure 15.1.8.)
(b) Assume the curve in (a) is parameterized over the interval $[a, b]$. Show that $\frac{1}{2}\int_a^b |\mathbf{r} \times \mathbf{v}|\, dt$ is the area swept out by the curve in (a).
(c) Must the curve lie in a plane for the formula in (b) to hold?

In Exercises 45 to 51 $\mathbf{v}(t)$ is the velocity vector at time t for a moving particle and $\mathbf{r}(0)$ is the particle's position at time $t = 0$. Find $\mathbf{r}(t)$, the position vector of the particle at time t.

45. $\mathbf{v}(t) = \sin^2(3t)\mathbf{i} + \frac{t}{3t^2+1}\mathbf{j}$, $\mathbf{r}(0) = \mathbf{j}$

46. $\mathbf{v}(t) = \frac{t}{t^2+t+1}\mathbf{i} + \tan^{-1}(3t)\mathbf{j}$, $\mathbf{r}(0) = \mathbf{i} + \mathbf{j}$

47. $\mathbf{v}(t) = \frac{t^3}{t^4+1}\mathbf{i} + \ln(t+1)\mathbf{j}$, $\mathbf{r}(0) = \mathbf{0}$

48. $\mathbf{v}(t) = e^{2t}\sin(3t)\mathbf{i} + \frac{t^3}{3t+2}\mathbf{j}$, $\mathbf{r}(0) = \mathbf{i} + 3\mathbf{j}$

49. $\mathbf{v}(t) = \frac{t}{(t+1)(t+2)(t+3)}\mathbf{i} + \frac{t^2}{(t+2)^3}\mathbf{j}$, $\mathbf{r}(0) = \mathbf{i} - \mathbf{j}$

50. $\mathbf{v}(t) = \frac{(\ln(t+1))^3}{t+1}\mathbf{i} + \frac{1}{\sqrt{1-4t^2}}\mathbf{j} + \sec^2(3t)\mathbf{k}$, $\mathbf{r}(0) = \mathbf{i} + \mathbf{j} + \mathbf{k}$

51. $\mathbf{v}(t) = t^3e^{-t}\mathbf{i} + (1+t)(2+t)\mathbf{j}$, $\mathbf{r}(0) = 2\mathbf{i} - \mathbf{j}$

15.2 Curvature and Components of Acceleration

In Section 9.6 we defined the curvature of a plane curve as the absolute value of the derivative $d\phi/ds$, where ϕ is the angle the tangent makes with the x-axis and s is the arc length. This definition does not work for a curve that does not lie in a plane. (Why not?) In this section we use vectors to define curvature for curves in space and then use curvature to analyze the acceleration vector.

To prepare for this discussion it is necessary to extend the definition of curvature of a plane curve (see Section 10.4) to a space curve: $\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$.

Definition: ***Integral Definition of Arc Length of a Space Curve***

The arc length of a space curve $\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$ for $a \le t \le b$ is

$$\text{Arc length} = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt = \int_a^b \left|\mathbf{r}'(t)\right|\, dt.$$

The arc length function $s(t)$ gives the arc length of the curve from its initial point to its position at time t is

$$s(t) = \int_a^t \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt = \int_a^t \left|\frac{\mathbf{r}(t)}{dt}\right|\, dt. \tag{15.2.1}$$

It is important to note that $\dfrac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} = \left|\dfrac{\mathbf{r}(t)}{dt}\right|$. That is, the rate of change of arc length is speed.

Definition of Curvature (in Space)

A particle whose position vector at the time t is $\mathbf{r}(t)$ has velocity $\mathbf{v}(t) = \mathbf{r}'(t)$. When $\mathbf{v}(t)$ is not the zero vector, the unit vector in the direction of $\mathbf{v}(t)$ is

$$\mathbf{T}(t) = \frac{\mathbf{v}(t)}{|\mathbf{v}|} = \frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|},$$

or

$$\mathbf{T} = \frac{\mathbf{v}}{|\mathbf{v}|} = \frac{\mathbf{r}'}{|\mathbf{r}'|} \qquad (\text{assuming } \mathbf{v} \neq \mathbf{0})$$

The **unit tangent vector**, $\mathbf{T}$, records the direction of motion.

As the particle moves along the curve the direction of $\mathbf{T}$ changes most rapidly where the curve is curviest.

Definition: ***Curvature of a (Space) Curve***

The **curvature** of a curve, in the plane or in space, is the length of the rate of change of the unit tangent vector with respect to arc length:

$$\text{Curvature} = \kappa = \left|\frac{d\mathbf{T}}{ds}\right|$$

where s denotes the arc length of a curve, measured from a fixed starting point.

We check in Example 1 that the definition of curvature agrees with the definition for curvature for plane curves in Section 9.6.

EXAMPLE 1. Show that the definition of curvature as $\left|\frac{d\mathbf{T}}{ds}\right|$ agrees with the one for plane curves: $\kappa = \left|\frac{d\phi}{ds}\right|$

SOLUTION As Figure 15.2.1 shows, ϕ is the angle that $\mathbf{T}$ makes with the x-axis. Since $\mathbf{T}$ is a unit vector, $\mathbf{T} = \cos(\phi)\mathbf{i} + \sin(\phi)\mathbf{j}$. Thus

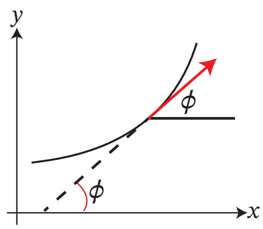


Figure 15.2.1

$$\begin{aligned}
\kappa &= \left|\frac{d\mathbf{T}}{ds}\right| && \text{(definition of } \kappa \text{ for a space curve)}\\
&= \left|\frac{d(\cos(\phi)\mathbf{i} + \sin(\phi)\mathbf{j})}{ds}\right| && \text{(formula for } \mathbf{T} \text{ as a function of } \phi \text{)}\\
&= \left|\frac{d(\cos(\phi)\mathbf{i} + \sin(\phi)\mathbf{j})}{d\phi}\frac{d\phi}{ds}\right| && \text{(chain rule)}\\
&= \left|(-\sin(\phi)\mathbf{i} + \cos(\phi)\mathbf{j})\frac{d\phi}{ds}\right| && \text{(differentiation)}\\
&= \left|-\sin(\phi)\mathbf{i} + \cos(\phi)\mathbf{j}\right|\left|\frac{d\phi}{ds}\right| && \text{(property of lengths of vectors: } |c\mathbf{v}| = |c||\mathbf{v}| \text{)}\\
&= \left|\frac{d\phi}{ds}\right| && \text{(} -\sin(\phi)\mathbf{i} + \cos(\phi)\mathbf{j} \text{ is a unit vector).}
\end{aligned}$$

Consequently, the definitions of curvature of a space curve and for a plane curve (in Section 9.6) are consistent. ♦

We also define, when $\kappa \neq 0$, the **radius of curvature** as the reciprocal of κ.

EXAMPLE 2. Compute the curvature and radius of curvature of the helix $\mathbf{r}(t) = \cos(t)\mathbf{i} + \sin(t)\mathbf{j} + 3t\mathbf{k}$.

SOLUTION To find $\mathbf{T}$ we compute $\mathbf{v} = -\sin(t)\mathbf{i} + \cos(t)\mathbf{j} + 3\mathbf{k}$ and $|\mathbf{v}| = \sqrt{(-\sin(t))^2 + (\cos(t))^2 + 3^2} = \sqrt{10}$. Thus

$$\mathbf{T} = \frac{1}{\sqrt{10}}(-\sin(t)\mathbf{i} + \cos(t)\mathbf{j} + 3\mathbf{k}).$$

By the chain rule, $(d\mathbf{T}/ds)(ds/dt) = d\mathbf{T}/dt$. Therefore, recalling that $ds/dt = |\mathbf{v}|$, the curvature can be computed from its definition:

$$\kappa = \left|\frac{d\mathbf{T}}{ds}\right| = \frac{|d\mathbf{T}/dt|}{|ds/dt|} = \frac{|d\mathbf{T}/dt|}{|\mathbf{v}|} = \frac{\left|\frac{1}{\sqrt{10}}\left(-\cos(t)\mathbf{i} - \sin(t)\mathbf{j}\right)\right|}{\sqrt{10}} = \frac{1}{10}.$$

The curvature is $1/10$ and the radius of curvature is 10. For any helix, the curvature and radius of curvature are both constant (and reciprocals). ♦

The Unit Normal Vector, N

Since $\mathbf{T}(t)$ has constant length,

$$\frac{d}{ds}(|\mathbf{T}|^2) = \frac{d}{ds}(\mathbf{T}\cdot\mathbf{T}) = 2\mathbf{T}\cdot\frac{d\mathbf{T}}{ds} = 0,$$

so $d\mathbf{T}/ds$ is perpendicular to $\mathbf{T}$. By considering small Δs and $\Delta\mathbf{T}$, as in Figure 15.2.2(a), we see that $d\mathbf{T}/ds$ points in the direction in which $\mathbf{T}$ is turning.

Since the length of $d\mathbf{T}/ds$ is the curvature κ, we may write

$$\frac{d\mathbf{T}}{ds} = \kappa\mathbf{N}$$

where $\mathbf{N}$ is a unit vector called the **principal normal**. (Note that $\mathbf{N}$ is not uniquely defined when $\kappa = 0$.) When κ is not zero, it is positive, and $d\mathbf{T}/ds$ and $\mathbf{N}$ point in the same direction. The vectors $\mathbf{T}$ and $\mathbf{N}$, if placed with their tails at a point P on the curve, are perpendicular: $\mathbf{T}\cdot\mathbf{N} = 0$. In particular, they determine a plane. The part of the curve near P stays close to the plane that contains both $\mathbf{T}$ and $\mathbf{N}$. (See Figure 15.2.2(b).)

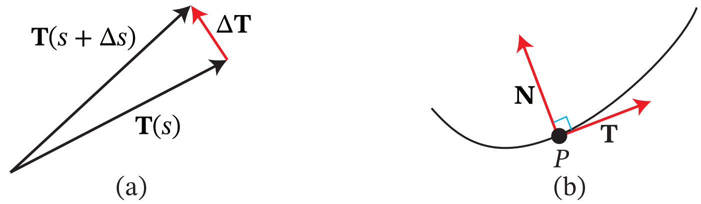


Figure 15.2.2

The Acceleration Vector and the Principal Unit Vectors T and N

The acceleration vector, $\mathbf{a}$, is defined as the second derivative (with respect to time) of the position vector, $\mathbf{r}$. We will show that $\mathbf{a}$ is parallel to the plane determined by $\mathbf{T}$ and $\mathbf{N}$, so $\mathbf{a}$ can be written in the form $c_1\mathbf{T} + c_2\mathbf{N}$, where c_1 and c_2 are scalars.

Since $\mathbf{a} = d\mathbf{v}/dt$, we express $\mathbf{v}$ in terms of $\mathbf{T}$ and $\mathbf{N}$. By the definition of $\mathbf{T}$, $\mathbf{v} = v\mathbf{T}$ where $v = |\mathbf{v}|$, is the speed. $\mathbf{N}$ is not needed to express the velocity vector $\mathbf{v}$.

Thus

$$\begin{aligned}
\mathbf{a} = \frac{d\mathbf{v}}{dt} &= \frac{d(v\mathbf{T})}{dt} \\
&= \frac{dv}{dt}\mathbf{T} + v\frac{d\mathbf{T}}{dt} && (\text{ product rule }) \\
&= \frac{d^2s}{dt^2}\mathbf{T} + v\frac{d\mathbf{T}}{ds}\frac{ds}{dt} && \left(v = \frac{ds}{dt} \text{ and chain rule } \right) \\
&= \frac{d^2s}{dt^2}\mathbf{T} + v^2\frac{d\mathbf{T}}{ds}.
\end{aligned}$$

Replacing $d\mathbf{T}/ds$ with $\kappa\mathbf{N}$, we find an expression for the acceleration vector in terms of the unit vectors $\mathbf{T}$ and $\mathbf{N}$, and the arc length, speed, and curvature of the curve.

Formula 15.2.1: *Expression for Acceleration in Terms of* T, N, *and* κ

The acceleration vector for a particle moving along a curve can be written as

$$\mathbf{a} = \frac{d^2s}{dt^2}\mathbf{T} + v^2\kappa\mathbf{N}$$

where $\mathbf{T}$ and $\mathbf{N}$ are the unit tangent and normal vectors, respectively, s denotes arc length, v is speed, and κ is the curvature.

If κ is not 0, so the radius of curvature is finite, then we have an expression for the acceleration vector that replaces the curvature with the radius of curvature.

Formula 15.2.2: *Expression for Acceleration in Terms of* T, N, *and Radius of Curvature*

The acceleration vector for a particle moving along a curve at a point with nonzero curvature is

$$\mathbf{a} = \frac{d^2s}{dt^2}\mathbf{T} + \frac{v^2}{r}\mathbf{N}$$

where $\mathbf{T}$ and $\mathbf{N}$ are the unit tangent and normal vectors, respectively, s denotes arc length, v is speed, and $r = 1/\kappa$ is the radius of curvature.

The **tangential component of acceleration**, $\mathbf{a}\cdot\mathbf{T} = d^2s/dt^2$, is positive if the particle is speeding up and is negative if it is slowing down. The **normal component of acceleration** is $\mathbf{a}\cdot\mathbf{N} = v^2/r$, which is always positive, provided $\kappa \neq 0$ (in which case $\mathbf{N}$ is not uniquely defined).

Figure 15.2.3 shows how **a** may look relative to **T** and **N**. In both cases **T** turns in the direction of **N**. In Figure 15.2.3 that means that **T** is turning counterclockwise.

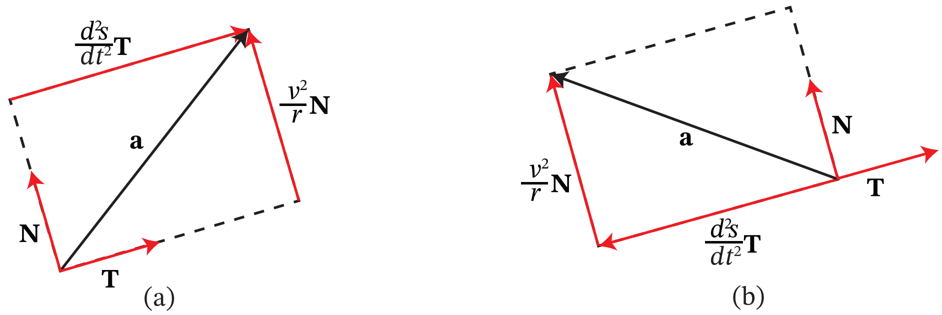


Figure 15.2.3

Computing Curvature

We can compute the curvature directly from its definition. There is also a shorter formula for κ. To develop it we compute

$$\mathbf{T}\times\mathbf{a} = \mathbf{T}\times\left(\frac{d^2s}{dt^2}\mathbf{T} + v^2\kappa\mathbf{N}\right). \tag{15.2.2}$$

We do this for two reasons. First, $\mathbf{T}\times\mathbf{T} = 0$. Second, $|\mathbf{T}\times\mathbf{N}| = 1$, since **T** and **N** are perpendicular unit vectors and therefore span a square of area 1. By (15.2.2), we then have $\mathbf{T}\times\mathbf{a} = \kappa v^2(\mathbf{T}\times\mathbf{N})$. Thus $|\mathbf{T}\times\mathbf{a}| = \kappa v^2$. Because $\mathbf{T} = \mathbf{v}/v$, we have $|\mathbf{v}\times\mathbf{a}|/v = \kappa v^2$ and thus we arrive at a third formula for the curvature:

Formula 15.2.3: *Expression of Curvature in Terms of Speed, Velocity, and Acceleration*

At a point on a curve with velocity **v**, speed v, and acceleration **a**, the curvature is

$$\kappa = \frac{|\mathbf{v}\times\mathbf{a}|}{v^3}. \tag{15.2.3}$$

NOTE: More formulas for κ are found in Exercises 21, 22, and 25.

EXAMPLE 3. Use (15.2.3) to compute the curvature of the helix $\mathbf{r}(t) = \cos(t)\mathbf{i} + \sin(t)\mathbf{j} + 3t\mathbf{k}$.

SOLUTION We begin by computing **v**, v and **a**. From the position, we find the velocity $\mathbf{v} = \mathbf{r}' = -\sin(t)\mathbf{i} + \cos(t)\mathbf{j} + 3\mathbf{k}$, and then the speed $v = |\mathbf{v}| = \sqrt{(-\sin(t))^2 + (\cos(t))^2 + 3^2} = \sqrt{10}$ and the acceleration $\mathbf{a} = d\mathbf{v}/dt = -\cos(t)\mathbf{i} - \sin(t)\mathbf{j}$. Next we compute $\mathbf{v}\times\mathbf{a}$:

$$\begin{aligned}\mathbf{v}\times\mathbf{a} = \det\begin{pmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ -\sin(t) & \cos(t) & 3\\ -\cos(t) & -\sin(t) & 0\end{pmatrix} &= 3\sin(t)\mathbf{i} - 3\cos(t)\mathbf{j} + (\sin^2(t)+\cos^2(t))\mathbf{k}\\ &= 3\sin(t)\mathbf{i} - 3\cos(t)\mathbf{j} + \mathbf{k}.\end{aligned}$$

Thus

Compare with Example 2.

$$\begin{aligned}\kappa = \frac{|\mathbf{v}\times\mathbf{a}|}{v^3} &= \frac{|3\sin(t)\mathbf{i} - 3\cos(t)\mathbf{j} + \mathbf{k}|}{(\sqrt{10})^3}\\ &= \frac{\sqrt{(3\sin(t))^2 + (-3\cos(t))^2 + 1^2}}{(\sqrt{10})^3}\\ &= \frac{\sqrt{10}}{(\sqrt{10})^3} = \frac{1}{10}.\end{aligned}$$

♦

Observation 15.2.1: *Curvature without Arc Length*

Though curvature is defined as a derivative with respect to arc length s, there are two reasons s is rarely used in computations. First, we seldom can obtain a formula for the arc length. Second, if the curve is described in terms of a parameter t, such as time or angle, then we can use the chain rule to express $d\mathbf{T}/ds$ as the directly calculatable

$$\frac{d\mathbf{T}/dt}{ds/dt}.$$

The Meaning of the Components of a

If no force acts on a moving particle it would move in a line at a constant speed. But if there is a force $\mathbf{F}$, then, according to Newton's Laws, it is related to the acceleration vector $\mathbf{a}$ by $\mathbf{F} = m\mathbf{a}$.

If $\mathbf{F}$ is parallel to $\mathbf{T}$, the particle of mass m moves in a line with an acceleration $dv/dt = d^2s/dt^2$. So we would expect $\mathbf{a}$ to equal $d^2s/dt^2\mathbf{T}$.

If $\mathbf{F}$ is perpendicular to $\mathbf{T}$, it would not change the particle's speed, but it would push it away from a straight path.

And, when $\mathbf{F}$ is neither parallel nor perpendicular to $\mathbf{T}$, the force changes both the trajectory and the speed of the object, as shown in Figure 15.2.4.

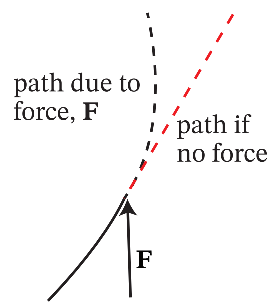


Figure 15.2.4

If you spin a pail of water at the end of a rope you can feel the force. It is proportional to the square of the speed and inversely proportional to the length of the rope. Driving a car around a sharp curve too fast can cause it to skid because the friction of the tires against the road cannot supply the necessary force, whose magnitude is proportional to the speed squared divided by the radius of the turn, to prevent skidding.

The Third Unit Vector, B

The vector $\mathbf{T} \times \mathbf{N}$ has length 1 and is perpendicular to both $\mathbf{T}$ and $\mathbf{N}$. We may think of it as a normal to the plane parallel to $\mathbf{T}$ and $\mathbf{N}$ through a point P on the curve. The unit vector $\mathbf{T} \times \mathbf{N}$ is denoted $\mathbf{B}$ and is called the **binormal**. It is shown in Figure 15.2.5

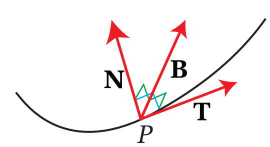


Figure 15.2.5

The three vectors, $\mathbf{T}$, $\mathbf{N}$, and $\mathbf{B}$, may change direction as P moves along the curve. However, they remain a rigid frame, where $\mathbf{T}$ indicates the direction of motion, $\mathbf{N}$ the direction of turning, and $\mathbf{B}$ the tilt of the **osculating plane**, the plane that contains P and the vectors $\mathbf{T}$ and $\mathbf{N}$.

Summary

We defined the curvature of a curve in space (or in the xy-plane) using vectors. The definition agrees with the definition of curvature for curves in the xy-plane given in Section 9.6. The curvature, or its reciprocal, the radius of the curvature, appears in the normal component of the acceleration vector.

The section concluded with the definition of the binormal, $\mathbf{B} = \mathbf{T} \times \mathbf{N}$, which indicates the tilt of the plane determined by $\mathbf{T}$ and $\mathbf{N}$.

EXERCISES for Section 15.2

In Exercises 1 to 4, $\mathbf{v}$ denotes the velocity and $\mathbf{a}$ denotes the acceleration. Evaluate the dot product.

1. $\mathbf{v} \cdot \mathbf{T}$ **2.** $\mathbf{a} \cdot \mathbf{T}$ **3.** $\mathbf{v} \cdot \mathbf{N}$ **4.** $\mathbf{a} \cdot \mathbf{N}$

5. (a) Why is $\mathbf{T}\times\mathbf{N}$ a unit vector? (b) Why is $\mathbf{N}$ perpendicular to $\mathbf{T}$?

In Exercises 6 and 7, $\mathbf{v}$ and $\mathbf{a}$ are given at an instant. Find the (a) curvature , (b) radius of curvature, and (c) $\dfrac{d^2s}{dt^2}$.

6. $\mathbf{v}=2\mathbf{i}+3\mathbf{j}+4\mathbf{k}$, $\mathbf{a}=\mathbf{i}-\mathbf{j}+\mathbf{k}$

7. $\mathbf{v}=\mathbf{i}+\mathbf{j}+\mathbf{k}$, $\mathbf{a}=-\mathbf{i}+\mathbf{j}+\mathbf{k}$

In Exercises 8 and 9 compute the curvature using $\kappa=\dfrac{1}{v}\left|\dfrac{d\mathbf{T}}{dt}\right|$.

8. $\mathbf{r}(t)=t\mathbf{i}+t^2\mathbf{j}+t^3\mathbf{k}$

9. $\mathbf{r}(t)=3\cos(2t)\mathbf{i}+3\sin(2t)\mathbf{j}+4t\mathbf{k}$

In Exercises 10 and 11, compute the curvature using the speed, velocity, and acceleration, that is, using $\kappa=\dfrac{|\mathbf{v}\times\mathbf{a}|}{v^3}$.

10. $\mathbf{r}(t)=t\mathbf{i}+t^2\mathbf{j}+t^3\mathbf{k}$

11. $\mathbf{r}(t)=3\cos(2t)\mathbf{i}+3\sin(2t)\mathbf{j}+4t\mathbf{k}$

---

12. To emphasize the value of the vector approach, compute $\dfrac{d|\mathbf{v}|}{dt}$ in two ways.

(a) Differentiate $|\mathbf{v}|^2=\mathbf{v}\cdot\mathbf{v}$ to conclude that $\dfrac{d|\mathbf{v}|}{dt}=\dfrac{\mathbf{v}\cdot\mathbf{a}}{|\mathbf{v}|}$.

(b) Derive the result starting with $|\mathbf{v}|=\sqrt{\left(\dfrac{dx}{dt}\right)^2+\left(\dfrac{dy}{dt}\right)^2+\left(\dfrac{dz}{dt}\right)^2}$.

13. A particle moves on the helix described by $\mathbf{r}(t)=3\cos(at)\mathbf{i}+3\sin(at)\mathbf{j}+bt\mathbf{k}$, where a and b are constants.

(a) Compute its curvature.
(b) As $b\to\infty$ what happens to the curvature?
(c) Why is the answer to (b) reasonable?
(d) As $a\to\infty$, what happens to the curvature?
(e) Why is the answer to (d) reasonable?

14. Show that for $\mathbf{r}(x)=x\mathbf{i}+f(x)\mathbf{j}+0\mathbf{k}$ the formula $\dfrac{|\mathbf{v}\times\mathbf{a}|}{v^3}$ gives the formula in Section 9.6 for curvature of $y=f(x)$.

15. Show that $\dfrac{d\mathbf{r}}{ds}$ is a unit vector,

(a) by drawing $\mathbf{r}(s+\Delta s)$ and $\mathbf{r}(s)$ and considering $\dfrac{\mathbf{r}(s+\Delta s)-\mathbf{r}(s)}{\Delta s}$ and (b) by writing it as $\dfrac{d\mathbf{r}/dt}{ds/dt}$.

16. Express the area of the parallelogram spanned by $\mathbf{v}$ and $\mathbf{a}$ in terms of the curvature and speed.

17. If a particle reaches a maximum speed at time t_0, must $\dfrac{d^2s}{dt^2}$ be 0 at t_0? Must $\dfrac{d^2\mathbf{r}}{dt^2}$ be $\mathbf{0}$ at t_0? Assume the time interval is $(-\infty,\infty)$.

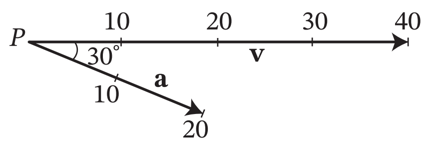
18. Let $\mathbf{r}(t)$ denote the position vector and s the arc length.

(a) Is $\dfrac{d\mathbf{r}}{dt}$ parallel to $\dfrac{d\mathbf{r}}{ds}$? Explain. (b) Is $\dfrac{d^2\mathbf{r}}{dt^2}$ parallel to $\dfrac{d^2\mathbf{r}}{ds^2}$? Explain.

In Exercises 19 and 20 the Figure 15.2.6 shows the velocity and acceleration vectors at a point P on a curve. Find (a) v, (b) $\dfrac{d^2s}{dt^2}$, and (c) κv^2. Then (d) find r, the radius of the curvature, (e) draw the osculating circle, and (f) using the osculating circle, draw an approximation of a short piece of the path near P.

19. Figure 15.2.6(a)

20. Figure 15.2.6(b)



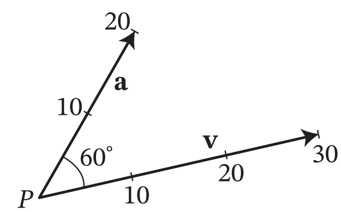


Figure 15.2.6

---

21. **Jane:** After doing Exercises 19 and 20, I have a simpler way to get a formula for curvature. Just look at the right triangle whose hypotenuse has length $|\mathbf{a}|$ and one leg is the component of $\mathbf{a}$ along $\mathbf{v}$. Then, by trigonometry,

$$\kappa v^2 = |\mathbf{a}||\sin(\mathbf{a},\mathbf{v})|. \tag{15.2.4}$$

All that's left is getting $\sin(\mathbf{a},\mathbf{v})$ out and $\cos(\mathbf{a},\mathbf{v})$ in because we know how to express $\cos(\mathbf{a},\mathbf{v})$ in terms of a dot product. Squaring (15.2.4) gives $\kappa^2 v^4 = |\mathbf{a}|^2\left(1-\cos^2(\mathbf{a},\mathbf{v})\right)$. If you use the fact that $\cos(\mathbf{a},\mathbf{v}) = \dfrac{\mathbf{a}\cdot\mathbf{v}}{|\mathbf{a}|v}$ and a little algebra, you get

$$\kappa^2 = \frac{(\mathbf{v}\cdot\mathbf{v})(\mathbf{a}\cdot\mathbf{a}) - (\mathbf{a}\cdot\mathbf{v})^2}{v^6}. \tag{15.2.5}$$

My way is simpler than using the cross product. I guess the authors don't understand trigonometry.

(a) Fill in the missing steps.

(b) Check that Jane's formula agrees with (15.2.3).

22. **Sam:** You used trigonometry. I can do it with just the Pythagorean Theorem. Look at the triangle with hypotenuse $|\mathbf{a}|$. Its legs have lengths $\left|\dfrac{d^2s}{dt^2}\right|$ and κv^2. So $|\mathbf{a}|^2 = \left(\dfrac{d^2s}{dt^2}\right)^2 + (\kappa v^2)^2$. Solve this for κ.

Jane: But you have to express everything in vectors. We're in the chapter on vectors.

Sam: O.K. First $|\mathbf{a}|^2 = \mathbf{a}\cdot\mathbf{a}$ and $v^2 = |\mathbf{v}|^2 = \mathbf{v}\cdot\mathbf{v}$.

Jane: But $\dfrac{d^2s}{dt^2}$?

Sam: That's $\dfrac{dv}{dt}$. So I differentiate both sides of $v^2 = \mathbf{v}\cdot\mathbf{v}$, getting $2v\dfrac{dv}{dt} = 2\mathbf{v}\cdot\mathbf{a}$. So $\dfrac{dv}{dt} = \dfrac{\mathbf{v}\cdot\mathbf{a}}{v}$. So $\left(\dfrac{dv}{dt}\right)^2 = \dfrac{(\mathbf{v}\cdot\mathbf{a})^2}{v^2}$. So $\mathbf{a}\cdot\mathbf{a} = \dfrac{(\mathbf{v}\cdot\mathbf{a})^2}{v^2} + \kappa^2(v^2)^2$. Then solve for κ^2. I get the same result that you got in Exercise 21. It seems quite straightforward. The authors should have used my formula.

Jane: There were so many "so's" that I got lost.

Show that the formula for curvature that Sam obtained agrees with Jane's formula in Exercise 21.

23. **Jane:** I still don't like any of the ways the authors got the formula for curvature. I'm sure they didn't need to drag in the components of the acceleration vector. It's not elegant.

Sam: They're trying to save space. Calculus books are too long.

Jane: My way is neat and short: just calculate $\left|\dfrac{d\mathbf{T}}{ds}\right| = \dfrac{|d\mathbf{T}/dt|}{|\mathbf{v}|}$. To begin I write $\mathbf{T}$ as $\mathbf{v}/|\mathbf{v}|$. Then I differentiate the quotient $\mathbf{v}/|\mathbf{v}|$ (with respect to t). Along the way I'll need $d|\mathbf{v}|/dt$, but I get that by differentiating $|\mathbf{v}|^2 = \mathbf{v}\cdot\mathbf{v}$ (also with respect to t). That will give me

$$\frac{d\mathbf{T}}{ds} = \frac{v^2\mathbf{a} - (\mathbf{a}\cdot\mathbf{v})\mathbf{v}}{v^4}. \tag{15.2.6}$$

SAM: That's a nice formula but there's no cross product.

JANE: If you like cross products, then use (15.2.6) to find $d\mathbf{T}/ds \cdot d\mathbf{T}/ds$ and call on that identity that appeared when getting the length of the cross product $|\mathbf{A}\times\mathbf{B}|$ (see (14.3.9) in Section 14.3). I'll let you fill in the steps. I don't want to deprive you of a little fun.

Fill in the missing steps.

---

24. Using (15.2.6), obtain the formula in (15.2.5) for κ^2.

25. Here is another way to find a formula for curvature. Using the right triangle whose hypotenuse is $|\mathbf{a}|$ and whose legs are parallel to $\mathbf{T}$ and $\mathbf{N}$, show that $\kappa^2 v^4 = \ddot{x}^2 + \ddot{y}^2 + \ddot{z}^2 - \ddot{s}^2$.

NOTATION: Two dots over a variable denotes its second derivative with respect to t.

26. Assume that you are prone to car sickness on curvy roads. Which matters more, $\left|\frac{d\mathbf{T}}{ds}\right|$ where s is arc length or $\left|\frac{d\mathbf{T}}{dt}\right|$ where t is time? Describe the difference in the two quantities.

27. Let $\mathbf{r} = \mathbf{r}(s)$, where s is arc length. Show that the curvature is $\kappa = \left|\frac{d^2\mathbf{r}}{ds^2}\right|$.

The Frenet formulas are

$$\frac{d\mathbf{T}}{ds} = \kappa\mathbf{N}, \qquad \frac{d\mathbf{B}}{ds} = -\tau\mathbf{N}, \qquad \text{and} \qquad \frac{d\mathbf{N}}{ds} = -\kappa\mathbf{T} + \tau\mathbf{B}.$$

Here κ is curvature and τ is torsion, the measure of the tendency of the plane through $\mathbf{T}$ and $\mathbf{N}$ to turn. We already have the first Frenet formula, while Exercises 28 and 29 develop the formulas for $\frac{d\mathbf{B}}{ds}$ and $\frac{d\mathbf{N}}{ds}$.

28.
(a) Why is $\frac{d\mathbf{B}}{ds}$ perpendicular to $\mathbf{B}$?
(b) Why are there scalars p and q such that $\frac{d\mathbf{B}}{ds} = p\mathbf{T} + q\mathbf{N}$?
(c) Using the fact that $\mathbf{B}$ and $\mathbf{T}$ are always perpendicular show that $(p\mathbf{T} + q\mathbf{N})\cdot\mathbf{T} = 0$.
(d) From (c) show that $p = 0$. Thus $\frac{d\mathbf{B}}{ds} = q\mathbf{N}$. The scalar function q is usually denoted $-\tau$. Thus $\frac{d\mathbf{B}}{ds} = -\tau\mathbf{N}$.

NOTE: τ is the Greek letter "tau"

29.
(a) Why are there scalars c and d such that $\frac{d\mathbf{N}}{ds} = c\mathbf{T} + d\mathbf{B}$?
(b) Using the fact that $\mathbf{B}$ and $\mathbf{N}$ are always perpendicular, show that $-\tau\mathbf{N}\cdot\mathbf{N} + \mathbf{B}\cdot(c\mathbf{T} + d\mathbf{B}) = 0$.
(c) From (b) show that $d = \tau$.
(d) Similarly, starting with $\mathbf{T}\cdot\mathbf{N} = 0$, show that $c = -\kappa$. Thus $\frac{d\mathbf{N}}{ds} = -\kappa\mathbf{T} + \tau\mathbf{N}$.

---

30. A pail of water is being swung at the end of a rope. The amount of rope is slowly increased until the radius of the circle the pail sweeps out doubles. Assume the angular velocity remains constant. Does the force of your pull remain the same? Increase? Decrease? Explain.

31. In Example 1 we used calculus to show that for a plane curve $\left|\frac{d\mathbf{T}}{s}\right| = \left|\frac{d\phi}{ds}\right|$, when ϕ is the angle that $\mathbf{T}$ makes with the x-axis. This suggests that for small values of Δs, $|\Delta\phi| = |\phi(s+\Delta s) - \phi(s)|$ is a good approximation of $|\Delta\mathbf{T}| = |\mathbf{T}(s+\Delta s) - \mathbf{T}(s)|$.

(a) Draw $\mathbf{T}(s+\Delta s)$ and $\mathbf{T}(s)$ with their tails at the origin.

(b) Using the diagram, show why $|\Delta\mathbf{T}|$ and $|\Delta\phi|$ should be close to each other in the sense that $|\Delta\mathbf{T}/\Delta\phi|$ would be near 1.

32. Show that a curve that has a constant curvature $\kappa = 0$ is part of a line.

NOTE: Sam's first thought, "Oh, it's a curve with infinite radius of curvature, so it must be a line" is not an acceptable response.

33. Express $d\mathbf{T}/ds$ in terms of the curvature and $\mathbf{N}$.

34. This exercise concerns curves situated on the surface $\mathscr{S}$ of a ball of radius a.

(a) Show that there are curves on $\mathscr{S}$ that have large curvature.

(b) Exhibit a curve whose curvature is as small as $1/a$.

(c) Show that there are no curves with curvature smaller than $1/a$.

CONTRIBUTED BY: G. D. Chakerian

15.3 Line Integrals and Conservative Vector Fields

In Section 6.2, we defined the integral of $f(x)$ over an interval $[a,b]$ as the limit of sums of the form $\sum_{i=1}^{n} f(c_i)\Delta x_i$. Ideas similar to the ones used in this section to define the definite integral over an interval will be used to define two different types of integrals over curves. In the next section we apply them to work, fluid flow, and geometry.

The Integral with Respect to Arc Length s

Let $\mathbf{r}(t)$ be the position vector corresponding to a parameter value t in $[a,b]$. Assume that $\mathbf{r}(t)$ sweeps out a curve C with a continuous unit tangent vector $\mathbf{T}(t)$. Let f be a scalar-valued function defined on C. We will define the integral of f over C with respect to arc length.

Partition $[a,b]$ by $t_0 = a, t_1, \dots, t_n = b$ and let $\mathbf{r}(t_0) = \overrightarrow{OP_0}$, $\mathbf{r}(t_1) = \overrightarrow{OP_1}$, $\dots$, $\mathbf{r}(t_n) = \overrightarrow{OP_n}$ be the position vectors as shown in Figure 15.3.1. The points $P_0, P_1, \dots, P_n$ break the curve into n shorter curves of lengths $\Delta s_1, \Delta s_2, \dots, \Delta s_n$. Form the Riemann sum

$$\sum_{i=1}^{n} f(P_i)\Delta s_i.$$

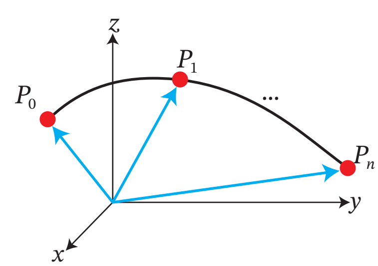


Figure 15.3.1

The limit of sums of this form as all the lengths Δs_i are chosen smaller and smaller is denoted $\int_C f(P)\,ds$. That is,

$$\int_C f(P)\,ds = \lim_{\text{all } \Delta s \to 0} \sum_{i=1}^{n} f(P_i)\Delta s_i.$$

The limit does not depend on the parameterization and so it does not depend on the direction in which the curve is swept out. To compute the integral when the curve is parameterized by t we can use

$$\int_C f(P)\,ds = \int_a^b f(P)\left|\frac{ds}{dt}\right| dt.$$

EXAMPLE 1. A fence is built as a semicircle of radius a with center at the origin. The height of the fence is $\sin^2(\theta)$, where θ is the angle made with the positive x-axis, as in Figure 15.3.2(a)). What is the area of one side of the fence?

SOLUTION Let $f(P)$ be the height of the fence at $P=(a,\theta)$ in polar coordinates. Then $f(a,\theta)=\sin^2(\theta)$. The parameter θ ranges from 0 to π. Let $s=a\theta$ be the arc length subtended by the angle θ, as in Figure 15.3.2(b).

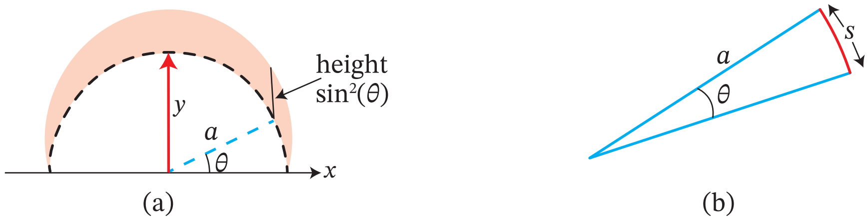


Figure 15.3.2

Then $ds=a\,d\theta$ and we have

$$\text{Area}=\int_C \sin^2(\theta)\,ds=\int_0^{\pi}\sin^2(\theta)a\,d\theta=\frac{a}{2}\int_0^{\pi}(1-\cos(2\theta))\,d\theta=\frac{a}{2}\left(\theta-\frac{1}{2}\sin(2\theta)\right)\Big|_0^{\pi}=\frac{\pi a}{2}.$$

◆

The Integrals with Respect to x, y, and z

Let C be a parameterized curve and let f be a scalar function defined on C. Divide the interval $[a,b]$ into n sections by $t_0=a, t_1,\cdots,t_n=b$. For convenience, take the sections to be of equal lengths.

Let $\mathbf{r}(t_i)=\langle x(t_i), y(t_i), z(t_i)\rangle$. Instead of considering the arc length Δs_i of each short interval we consider instead the change in the x-coordinate, $x_i-x_{i-1}=\Delta x_i$. It can be positive or negative. We have the following definition:

Definition: *Integral Over a Curve (with Respect to x)*

The **integral of f over the curve C with respect to** x is the limit of sums

$$\int_C f\,dx,=\sum_{i=1}^{n} f(x(t_i),y(t_i),z(t_i))\Delta x_i$$

as all Δx_i approach 0. Other notations for this integral include $\int_C f(x,y,z)\,dx$ and $\int_C f(P)\,dx$.

The integrals $\int_C f(P)\,dy$ and $\int_C f(P)\,dz$ are defined similarly.

Each of $\int_C f(P)\,dx$, $\int_C f(P)\,dy$, and $\int_C f(P)\,dz$ is called a **line integral** of f over the curve C. Another line integral is the line integral with respect to arc length: $\int_C f(P)\,ds$. While it would be more natural to call them curve integrals, tradition dictates that they be known as line integrals.

To compute a line integral such as $\int_C f(P)\,dx$ over a parameterized curve C, use the parameterization to express the differential dx as $(dx/dt)\,dt$. Then,

$$\int_C f(P)\,dx=\int_a^b f(x(t),y(t),z(t))(dx/dt)\,dt.$$

In the same way,

$$\int_C f(P)\,ds=\int_a^b f(x(t),y(t),z(t))\sqrt{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2+\left(\frac{dz}{dt}\right)^2}\,dt.$$

A more general line integral is a sum of three types,

$$\int_C \big(P(x,y,z)\ dx + Q(x,y,z)\ dy + R(x,y,z)\ dz\big). \tag{15.3.1}$$

The integrand for this line integral, $P dx + Q dy + R dz$, is sometimes referred to as a **differential form**. This terminology will be encountered again in Chapter 18. If we define the vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ and write the parameterization of the curve C as $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, then the differential form can also be written as the dot product of $\mathbf{F}$ and $d\mathbf{r} = dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k}$: $P dx + Q dy + R dz = \mathbf{F}\cdot d\mathbf{r}$.

In contrast to an integral with respect to arc length, the value of $\int_C f(P)\ dx$ depends on the orientation in which the curve is swept out. If we reverse the orientation, the expression $x_i - x_{i-1}$ changes sign. For instance, if x is an increasing function of the parameter in one parameterization, then $\Delta x_i = x_i - x_{i-1}$ is positive. But in the reverse orientation x is a decreasing function of the parameter, so $\Delta x_i = x_i - x_{i-1}$ is negative.

Observation 15.3.1: *Importance of Orientation*

Denote by $-C$ the curve C swept out in the opposite orientation. Then

$$\int_{-C} f(P)\ ds = \int_C f(P)\ ds. \qquad \text{but} \qquad \int_{-C} f(P)\ dx = -\int_C f(P)\ dx.$$

When evaluating line integrals $\int_C f(P)\ dx$, $\int_C f(P)\ dy$, $\int_C f(P)\ dz$, hence $\int_C \mathbf{F}(P)\cdot d\mathbf{r}$, it is necessary to pay attention to the orientation of C.

Observation 15.3.2: *Importance of Parameterization*

While the orientation of a curve makes a difference for some line integrals, the value of any line integral does not depend on the specific parameterization of the curve that is used. Some parameterizations might lead to definite integrals that are easier to evaluate than other parameterizations. In such situations it is comforting to know that the value of the line integral can be found by evaluating the easier definite integral.

A **closed curve** is a curve that starts and ends at the same point. If the curve does not intersect itself except perhaps at its endpoints, we call the curve **simple**. These are independent: a curve can be neither closed nor simple, closed but not simple, simple but not closed, or both simple and closed. (See Figure 15.3.3, which concerns curves described by a vector-valued function $\mathbf{G}(t)$ with parameter t in an interval $[a,b]$. In each example, the initial and terminal points are $\mathbf{G}(a)$ and $\mathbf{G}(b)$, respectively.)

NOTATION: When C is a closed curve we will usually use the notation $\oint_C \mathbf{F}\cdot\ d\mathbf{r}$ for a line integral over C. (Note the small circle added to the integral sign.)

Orienting a curve orients each section that is part of that curve. For instance, orienting the whole closed curve in Figure 15.3.4 orients the two curves C_1 and C_2 in that figure. When a simple closed curve C lies in the xy-plane, and orientation is not mentioned, it is assumed that the curve is traversed counter-clockwise (when viewed from a position above the xy-plane).

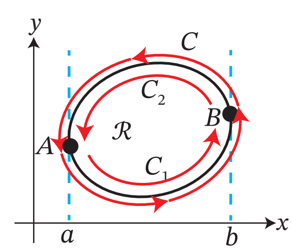


Figure 15.3.4

Just as ordinary integrals $\int_a^c f(x)\ dx$ and $\int_c^b f(x)\ dx$ can be added, with the sum $\int_a^b f(x)\ dx$, so can two line integrals if the end of one curve is the beginning of the other. For instance, the curve C_1 in Figure 15.3.4 ends at B, the initial point of C_2. Thus $\int_{C_1} y\ dx + \int_{C_2} y\ dx$ equals $\oint_C y\ dx$, where C is the entire closed curve surrounding $\mathcal{R}$.

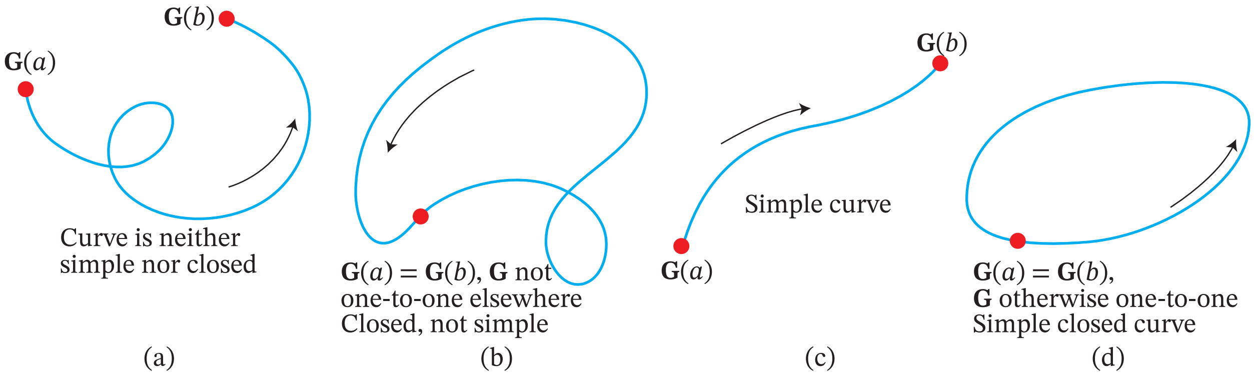


Figure 15.3.3

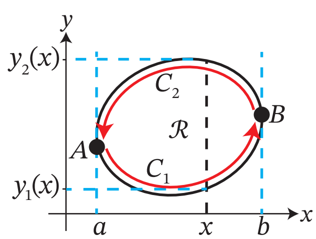


Figure 15.3.5

EXAMPLE 2. The closed curve C, composed of C_1 and C_2, is oriented counterclockwise and bounds a region $\mathscr{R}$, as shown in Figure 15.3.5. Show that the area of $\mathscr{R}$ is given by $-\oint_C y\,dx$.

SOLUTION Each line parallel to the y-axis that meets $[a,b]$ intersects C_2 in a point $(x, y_2(x))$ and C_1 in a point $(x, y_1(x))$. (See Figure 15.3.5.) Because the area of $\mathscr{R}$ equals "the integral of cross-sectional length," we have

$$\text{Area of } \mathscr{R} = \int_a^b (y_2 - y_1)\,dx = \int_a^b y_2\,dx - \int_a^b y_1\,dx. \tag{15.3.2}$$

Next we translate the two integrals in (15.3.2) into the language of line integrals.

Because a is less than b, $\int_a^b y_1\,dx$ is oriented from left to right. Thus $\int_a^b y_1\,dx = \int_{C_1} y\,dx$. The curve C_2 is oriented in the opposite direction, from right to left. Thus $\int_a^b y_2\,dx = \int_{-C_2} y\,dx = -\int_{C_2} y\,dx$. When these two expressions are inserted into (15.3.2), the result is

$$\text{Area of } \mathscr{R} = \int_a^b y_2\,dx - \int_a^b y_1\,dx = -\int_{C_2} y\,dx - \int_{C_1} y\,dx = -\left(\int_{C_2} y\,dx + \int_{C_1} y\,dx\right) = -\oint_C y\,dx.$$

◆

NOTE: If C is the curve in Example 2, then $\oint_C x\,dy$ equals the area inside C. (See Exercise 2.)

For curves in the xy-plane, the most general line integral would be

$$\int_C \big(P(x,y,z)\,dx + Q(x,y,z)\,dy\big). \tag{15.3.3}$$

The expression (15.3.3) can be expressed in the compact language of vectors, as we will now show.

Vector Fields

A **vector field** assigns a vector to each point in some region of space (or the plane). The function that assigns to each point the vector that describes the direction and speed of the wind is an example of a vector field. The use of field instead of function is in deference to physicists and engineers, who speak of magnetic field and electric field, which are also examples of vector fields.

A function that assigns a scalar (real number) to each point in a region in space (or the plane) is called a **scalar field**. The function that assigns the temperature at a point in space is a scalar function as is the function that gives the density of an object at a point.

A typical vector field in space is $\mathbf{F}(x,y,z) = P(x,y,z)\mathbf{i} + Q(x,y,z)\mathbf{j} + R(x,y,z)\mathbf{k}$ where $P(x,y,z)$, $Q(x,y,z)$, and $R(x,y,z)$ are scalar fields. A vector field $\mathbf{F}$ in the plane can be described by two scalar fields with $\mathbf{F}(x,y) = P(x,y)\mathbf{i} + Q(x,y)\mathbf{j}$.

To take advantage of vector notation, we write $d\mathbf{r} = dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k}$. Then (15.3.1) can be written in vector notation as

$$\int_C \mathbf{F}(x,y,z)\cdot d\mathbf{r}, \qquad \int_C \mathbf{F}(\mathbf{r})\cdot d\mathbf{r}, \quad \text{or} \quad \int_C \mathbf{F}\cdot d\mathbf{r}.$$

For computation we may write it as

$$\int_a^b \mathbf{F}(x(t),y(t),z(t))\cdot \frac{d\mathbf{r}}{dt}\,dt \qquad \text{or} \qquad \int_a^b \mathbf{F}\cdot\frac{d\mathbf{r}}{dt}dt.$$

Line integrals in the plane are expressed in the same way.

Another standard notation uses the unit vector $\mathbf{T} = d\mathbf{r}/ds$. Writing $d\mathbf{r}$ as $\mathbf{T}\,ds$ we can rewrite $\int_C \mathbf{F}\cdot d\mathbf{r}$ as $\int_C \mathbf{F}\cdot\mathbf{T}\,ds$. The integrand clearly depends on the orientation of the curve because reversing orientation changes $\mathbf{T}$ to $-\mathbf{T}$.

Conservative Vector Fields

The next example shows that different paths with the same initial point and terminal point may yield different integrals.

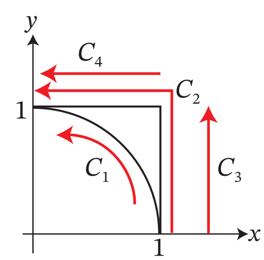


Figure 15.3.6

EXAMPLE 3. Let C_1 be the path from $(1,0)$ to $(0,1)$ along the unit circle with center at the origin. Let C_2 be the path that starts at $(1,0)$, goes to $(1,1)$ on the line $x=1$, and then to $(0,1)$ on the line $y=1$. Compute $\int_{C_1} xy\,dx$ and $\int_{C_2} xy\,dx$.

SOLUTION Figure 15.3.6 shows the two paths C_1 and C_2, together with two more curves, C_3 and C_4, that together make up C_2.

Observation 15.3.3: *What Can Be Said about the Signs of These Line Integrals?*

In $\int_{C_1} xy\,dx$, x and y are both positive and dx is negative, so this line integral should be negative.

It is a little more difficult to analyze $\int_{C_2} xy\,dx$, except that $dx=0$ on the vertical portion of C_2, so this makes no contribution to the overall integral. And, because the horizontal segment is traversed right-to-left, $x>0$, $y=1$, and $dx<0$ and this portion will have a negative contribution to the overall line integral.

To compute $\int_{C_1} xy\,dx$, we parameterize the circle with the angle θ in $[0,\pi/2]$: $x=\cos(\theta)$, $y=\sin(\theta)$, and $dx = (dx/d\theta)d\theta = -\sin(\theta)\,d\theta$. Therefore, with the aid of the substitution $u=\sin(\theta)$, with $du=\cos(\theta)\,d\theta$,

$$\int_{C_1} xy\,dx = \int_0^{\pi/2} (\cos(\theta))(\sin(\theta))(-\sin(\theta))\,d\theta = -\int_0^{\pi/2}\sin^2(\theta)\cos(\theta)\,d\theta = -\left.\frac{\sin^3(\theta)}{3}\right|_0^{\pi/2} = -\frac{1}{3}.$$

Next, we turn our attention to calculating $\int_{C_2} xy\,dx$. It is very natural to break C_2 into two curves: C_3 from $(1,0)$ to $(1,1)$ and C_4 from $(1,1)$ to $(0,1)$. (See Figure 15.3.6.)

Then, on C_3, $x = 1$ and $dx = 0$, so $\int_{C_3} xy\,dx = 0$.

And, on C_4, $y = 1$ and x begins at 1 and ends at 0. A parameterization of C_4 is $x = 1 - t$, $y = 1$ for $0 \le t \le 1$. Then

$$\int_{C_4} xy\,dx = \int_0^1 (1-t)(1)(-dt) = \int_0^1 (t-1)\,dt = \left(\frac{t^2}{2} - t\right)\Big|_0^1 = -\frac{1}{2}.$$

Observation 15.3.4: *Alternate Evaluation of $\int_{C_4} xy\,dx$*

There is an alternate way to evaluate the line integral over C_4 that needs to be discussed. On C_4 we could have used the parameter x itself, which starts at 1 and decreases to 0. Then we would have $\int_{C_4} xy\,dx = \int_1^0 x\,dx = x^2/2\big|_1^0 = -1/2$. It is reassuring that the value agrees with the initial evaluation of this line integral.

Since C_2 is made up of C_3 followed by C_4, we have $\int_{C_2} xy\,dx = \int_{C_3} xy\,dx + \int_{C_4} xy\,dx = 0 + (-1/2) = -1/2$.

The line integrals $\int_{C_1} xy\,dx$ and $\int_{C_2} xy\,dx$ are not equal even though they start at the same point $(1,0)$, end at the same point $(0,1)$, and have the same integrand. ♦

As Example 3 shows, $\int_C xy\,dx$ is not determined by the end points of the curve C. However, some line integrals are. The next example presents a case where the line integral depends only on the endpoints.

EXAMPLE 4. Compute $\displaystyle\int_C \frac{x\,dx + y\,dy}{x^2+y^2}$ on the paths C_1 and C_2 in Example 3.

SOLUTION On the circular path C_1 we use θ as a parameter and have

$$\int_{C_1} \frac{x\,dx + y\,dy}{x^2+y^2} = \int_0^{\pi/2} \frac{(\cos(\theta))(-\sin(\theta)\,d\theta) + \sin(\theta)(\cos(\theta))\,d\theta}{\cos^2(\theta)+\sin^2(\theta)} = \int_0^{\pi/2} \frac{0}{1}\,d\theta = 0.$$

Next we compute the integral on C_2 by adding the contributions on C_3 and C_4. The vertical segment from $(1,0)$ to $(1,1)$ is C_3. There $x = 1$, so $dx = 0$. Therefore, using y as the parameter, we find that

$$\begin{aligned}\int_{C_3} \frac{x\,dx + y\,dy}{x^2+y^2} &= \int_{C_3} \frac{1\cdot 0 + y\,dy}{1+y^2} = \int_{C_3} \frac{y}{1+y^2}\,dy \\ &= \int_0^1 \frac{y}{1+y^2}\,dy = \frac{1}{2}\ln(1+y^2)\Big|_0^1 = \frac{1}{2}\ln(2).\end{aligned}$$

On the horizontal segment C_4, from $(1,1)$ to $(0,1)$, we use x as the parameter starting at $x = 1$ and $y = 1$, so $dy = 0$ and we have

$$\begin{aligned}\int_{C_4} \frac{x\,dx + y\,dy}{x^2+y^2} &= \int_1^0 \frac{x\,dx}{x^2+1} \\ &= \frac{1}{2}\ln(x^2+1)\Big|_1^0 = -\frac{1}{2}\ln(2).\end{aligned}$$

Thus $\int_{C_2} (x\,dx + y\,dy)/(x^2+y^2) = -\frac{1}{2}\ln(2) + \frac{1}{2}\ln(2) = 0$. This is the same value as the integral over C_1. ♦

In Section 18.1 we will develop a simple test that will tell us that $\int_C (x\,dx + y\,dy)/(x^2+y^2)$ depends only on the end points the curve C, provided C does not pass through the origin. That is, if C_1 and C_2 are any two curves from point A to point B that do not pass through the origin, then

$$\int_{C_1} \frac{x\,dx + y\,dy}{x^2+y^2} = \int_{C_2} \frac{x\,dx + y\,dy}{x^2+y^2}.$$

A differential form $P(x,y,z)\;dx + Q(x,y,z)\;dy + R(x,y,z)\;dz$ is called a **conservative form** if its line integrals depend only on the endpoints of the curves over which the integration takes place. Likewise, the vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ is called a **conservative field** when $\int_C \mathbf{F}\cdot d\mathbf{r}$ depends only on the end points of the curve C. In Sections 18.1 and 18.2 we will develop a criterion for identifying conservative fields and see that conservative fields are easier to work with in applications.

Summary

We defined two different types of line integrals in space. The first, $\int_C f\;ds$, is the line integral over a curve C of the function f with respect to arc length, s. The second, in its most general form, is

$$\int_C (P(x,y,z)\;dx + Q(x,y,z)\;dy + R(x,y,z)\;dz) = \int_C (P\;dx + Q\;dy + R\;dz) = \int_C \mathbf{F}\cdot d\mathbf{r}$$

where $\mathbf{F}(x,y,z) = P(x,y,z)\mathbf{i} + Q(x,y,z)\mathbf{j} + R(x,y,z)\mathbf{k}$ and $\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$, $a \le t \le b$, is a parameterization of the curve C. Both types of line integrals come together when it is recalled that $\mathbf{T} = d\mathbf{r}/ds$ so that $\int_C \mathbf{F}\cdot d\mathbf{r} = \int_C \mathbf{F}\cdot\mathbf{T}\;ds$.

EXERCISES for Section 15.3

Exercises 1 and 2 are a continuation of Example 2. In particular, the curve C is the upper semicircle of radius a and center at the origin.

1. Following the approach used in Example 2, show that if curve C were oriented clockwise, then $\oint_C y\;dx$ would equal the area inside C.

2. Let C be oriented counterclockwise. Show why $\oint_C x\;dy$ equals the area inside C.

---

3. Show that the area within a convex curve C is $\frac{1}{2}\oint_C (x\;dy - y\;dx)$ if C is oriented counterclockwise.

4. Compute $\int_C xy\;dx$ on the straight-line path from $(1,0)$ to $(0,0)$, and from there to $(0,1)$. See Example 3.

5. If $\mathbf{F}(P)$ is perpendicular to the curve C at every point P on C, what is $\int_C \mathbf{F}\cdot d\mathbf{r}$?

6. If t represents time and $\mathbf{r}(t)$ describes a curve C, what is the meaning of $\int_C \mathbf{T}\cdot d\mathbf{r}$?

7. Let a and b be positive numbers. Let C be the curve bounding the rectangle with vertices $(0,0)$, $(a,0)$, (a,b), and $(0,b)$. By calculating $\oint_C x\;dy$ with C oriented counterclockwise, confirm the result of Exercise 2. That is, check that the line integral over the closed curve C equals the area of the rectangle.

8. Let a and b be positive numbers. Let C be the curve bounding the triangle with vertices $(0,0)$, $(a,0)$, and $(0,b)$. By calculating $\oint_C y\;dx$ with C oriented clockwise, show that the integral equals the area of the triangle.

9. Let C be the circle of radius a with center at the origin. By calculating $\oint_C x\;dy$ counterclockwise, check that the integral equals the area of the circle.

10. Let $\mathbf{F}(x,y,z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} = \mathbf{r}$. Let C be a curve starting at (x_0, y_0, z_0) and ending at (x_1, y_1, z_1). Calculate $\int_C \mathbf{F}\cdot d\mathbf{r}$ by rewriting it as $\int_a^b (\mathbf{F}\cdot\mathbf{r}')\, dt$. Note that $\int_C \mathbf{F}\cdot d\mathbf{r}$ depends only on the endpoints of C.

In Exercises 11 to 14, sketch the curve and label its start and finish.

11. $\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j}$, t in $[0,1]$

12. $\mathbf{r}(t) = (1-t)\mathbf{i} + (1-t)^2\mathbf{j}$, t in $[0,1]$

13. $\mathbf{r}(t) = (2t+1)\mathbf{i} + 3t\mathbf{j}$, t in $[0,2]$

14. $\mathbf{r}(t) = 4\cos t\mathbf{i} + 5\sin t\mathbf{j}$, t in $[0,1]$

In Exercises 15 to 18, parameterize the curve with the indicated orientation.

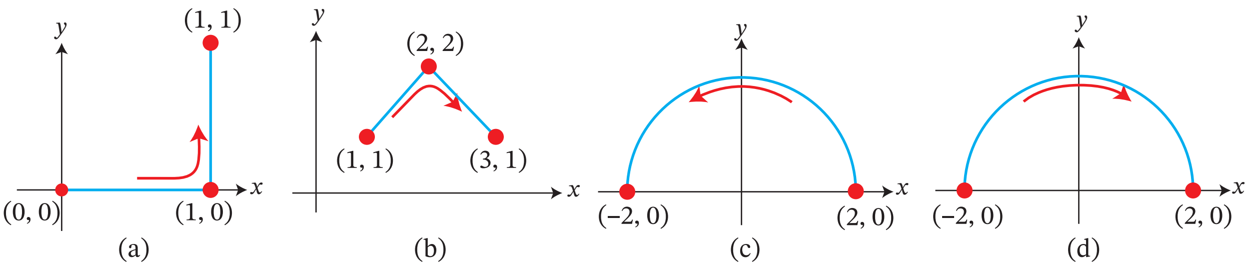


Figure 15.3.7

15. Figure 15.3.7(a)

16. Figure 15.3.7(b)

17. Figure 15.3.7(c)

18. Figure 15.3.7(d)

In Exercises 19 to 22, evaluate

19. $\int_C xy\, dx$, where C is the straight line from $(1,1)$ to $(3,3)$.

20. $\int_C x^2\, dy$, where C is the straight line from $(2,0)$ to $(2,5)$.

21. $\int_C x^2\, dy$, where C is the straight line from $(3,2)$ to $(7,2)$.

22. $\int_C (xy\, dx + x^2\, dy)$, where C is the straight line from $(1,0)$ to $(0,1)$.

In Exercises 23 to 26 evaluate the integral with minimum effort. C is a counterclockwise curve bounding a region of area 5.

23. $\oint_C 3y\, dx$

24. $\oint_C (2y\, dx + 6x\, dy)$

25. $\oint_C (2x\, dx + (x+y)\, dy)$

26. $\oint_C ((x+2y+3)\, dx + (2x-3y+4)\, dy)$

In Exercise 10, the value of the line integral depends only on the endpoints, not on the path that joins them. Exercises 27 and 28 are examples where the path matters.

27. Evaluate $\int_C (xy\,dx + 7\,dy)$ on

(a) the straight path from $(1,1)$ to $(2,4)$ and (b) the path from $(1,1)$ to $(2,4)$ that lies on the parabola $y = x^2$.

28. Evaluate $\int_C x\,dy$ on

(a) the straight path from $(0,0)$ to $(\pi/2,1)$ and (b) the path from $(0,0)$ to $(\pi/2,1)$ that lies on the curve $y = \sin(x)$.

In Exercises 29 and 30, the values of some line integrals are given for curves oriented as shown. Use this information to find $\oint_C f\,dy$. (Note the orientations.)

29. Figure 15.3.8(a), where $\oint_{C_1} f(P)\,dy = 3$, $\oint_{C_2} f(P)\,dy = 5$, and $\oint_{C_3} f(P)\,dy = 4$.

30. Figure 15.3.8(b), where $\oint_{C_1} f(P)\,dy = 1$, $\oint_{C_2} f(P)\,dy = 2$, $\oint_{C_3} f(P)\,dy = 3$, and $\oint_{C_4} f(P)\,dy = 4$.

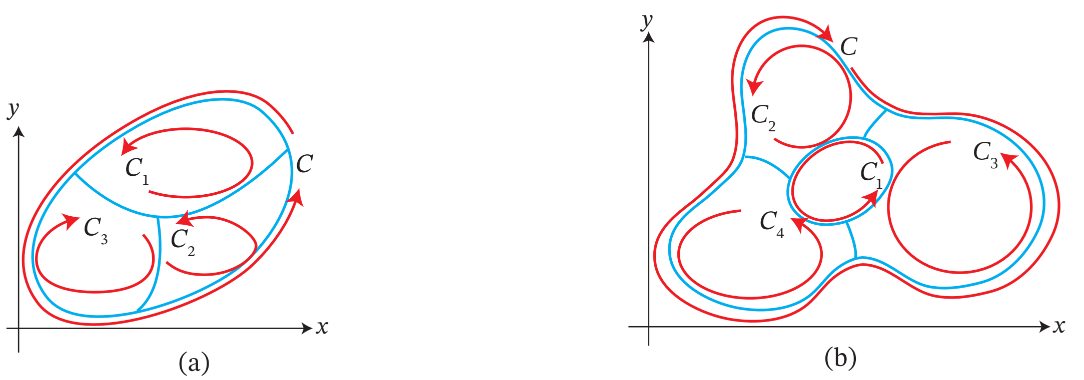


Figure 15.3.8

---

31. Assume the closed curve C is the boundary of the region $\mathscr{R}$, which is broken into regions $\mathscr{R}_i$, $1 \le i \le n$, with each $\mathscr{R}_i$ bounded by a closed curve C_i. Let $\mathbf{F}$ be a vector field on $\mathscr{R}$. If all the curves are swept out counterclockwise, show that $\oint_C \mathbf{F}\cdot d\mathbf{r} = \sum_{i=1}^{n} \oint_{C_i} \mathbf{F}\cdot d\mathbf{r}$.

32. Show that $\mathbf{F} = \frac{x}{x^2+y^2}\mathbf{i}$ is not a conservative vector field by calculating the line integral $\int_C \frac{x\,dx}{x^2+y^2}$ on two paths joining $(1,0)$ to $(1,1)$ for which the integrals are not equal.

33. Let k be a constant. Show that $\oint_C k\,dy = 0$.

34. In Example 2 it was shown that for the counterclockwise-oriented simple closed curve C bounding a region $\mathscr{R}$ that the area of $\mathscr{R}$ can be found as $-\oint_C y\,dx$. Show that the area of $\mathscr{R}$ also equals $\oint_C x\,dy$.

35. Using Exercise 34 and Example 2, show that the area of $\mathscr{R}$ equals $\frac{1}{2}\oint (x\,dy - y\,dx)$.

36. (a) Explain, in words (not equations or formulas), the relationship between a line integral integral over a curve C with respect to x, y, and/or z and the same line integral over $-C$, that is, the curve with reversed orientation.

(b) How does the answer in (a) change for a line integral with respect to arc length?

37. **SAM:** I've a better solution to Example 2.

JANE: You always know better than the authors.

SAM: For simplicity I'll assume $\mathscr{R}$ is in the first quadrant, so I can avoid some negative signs.

JANE: Go ahead.

SAM: I'll start with $\oint\limits_C y\,dx$, where C is oriented counterclockwise. I have

$$\oint\limits_C y\,dx = \int\limits_{C_1} y\,dx + \int\limits_{C_2} y\,dx = \int\limits_a^b y_1(x)\,dx - \int\limits_a^b y_2(x)\,dx.$$

Now, $\int\limits_a^b y_1(x)\,dx$ equals the area of the region below C_1 and above $[a,b]$; call this area A_1. Similarly, $\int\limits_a^b y_2(x)\,dx$ equals the area of the region below C_2 and above $[a,b]$. Call this area A_2. Thus I have $\oint\limits_C y\,dx = A_1 - A_2$. You can easily see that $A_1 - A_2$ is the negative of the area of $\mathscr{R}$ so $\oint\limits_C y\,dx =$ −Area of $\mathscr{R}$. It follows with no more sweat that the area of $\mathscr{R}$ equals $-\oint\limits_C y\,dx$, where C is oriented counterclockwise.

Is Sam correct? Explain.

38. Let $\mathbf{r} = \mathbf{r}(t)$ describe a curve C in the plane or in space. What is the geometric interpretation of $\frac{1}{2}\int\limits_C |\mathbf{r}\times\mathbf{T}|\,ds$?

NOTE: This line integral appears in the derivation of Kepler's area law, in Exercises 1 to 3 in CIE 22 ("Newton's Law Implies Kepler's Three Laws") at the end of this chapter.

15.4 Four Applications of Line Integrals

In the last section we defined line integrals and showed that $\oint_C y\,dx$ and $\oint_C x\,dy$ in the plane are related to the area of the region bounded by the closed curve C. We also introduced the notion of a conservative vector field. In this section we show how line integrals occur in the study of work, fluid flow, and of the angle subtended by a plane curve.

Work Along a Curve

A force $\mathbf{F}$, such as gravity, remains constant (in direction and magnitude) and acts on a particle as it moves in a straight line from A to B. (There could be other forcing acting on the particle.) The work accomplished by $\mathbf{F}$ is defined as $\mathbf{F}\cdot\mathbf{R}$, where $\mathbf{R} = \overrightarrow{AB}$:

$$\text{work} = \mathbf{F}\cdot\mathbf{R}.$$

This is the product of the scalar component of $\mathbf{F}$ in the direction of $\mathbf{R}$ and the distance the particle moves. (See Figure 15.4.1)

What if **F** varies and acts on the particle as it moves along a curve that is not straight? (See Figure 15.4.2(a).)

Assume the curve, C, is parameterized by $\mathbf{r}(t)$ for t in $[a,b]$. Partition $[a,b]$ by $t_0=a,\ t_1,\ \dots,\ t_n=b$ and let $\mathbf{r}(t_0)=\overrightarrow{OP_0}$, $\mathbf{r}(t_1)=\overrightarrow{OP_1}$, …, $\mathbf{r}(t_n)=\overrightarrow{OP_n}$, be the corresponding position vectors. (See Figure 15.4.2(b).) The points $P_0, P_1, \dots, P_n$ break the curve into n shorter curves. The work done by **F** along C between P_{i-1} and P_i is approximately $\mathbf{F}(P_{i-1})\cdot\Delta\mathbf{r}_{i-1}$ where $\Delta\mathbf{r}_{i-1}=\overrightarrow{P_{i-1}P_i}$. The total work done by **F** along C is approximated by

$$\sum_{i=1}^{n}\mathbf{F}(P_{i-1})\cdot\Delta\mathbf{r}_{i-1}.$$

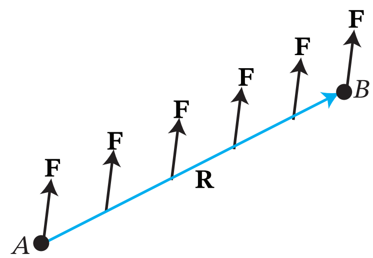


Figure 15.4.1

(a) (b)

Figure 15.4.2

Taking the limit as the largest $|\Delta\mathbf{r}_{i-1}|$ approaches 0 provides the following definition for work as a line integral.

Definition: *Work as a Line Integral*

$$\text{Work done by } \mathbf{F} \text{ along } C = \int_C \mathbf{F}\cdot d\mathbf{r}. \tag{15.4.1}$$

Physicists and engineers commonly use (15.4.1) as a starting point when expressing work.

The vector notation $\mathbf{F}\cdot d\mathbf{r}$ is far more suggestive than the scalar notation $P\,dx+Q\,dy$. It says that work is the dot product of force and displacement. That implies that only the component of the force in the direction of motion accomplishes work.

EXAMPLE 1. How much work is accomplished by the force $\mathbf{F}(x,y)=xy\mathbf{i}+y\mathbf{j}$ in pushing a particle from $(0,0)$ to $(3,9)$ along the parabola $y=x^2$?

SOLUTION Figure 15.4.3 shows the path of the particle. Call it C. Then

$$\text{Work}=\int_C \mathbf{F}\cdot d\mathbf{r}=\int_C (xy\mathbf{i}+y\mathbf{j})\cdot(dx\mathbf{i}+dy\mathbf{j})=\int_C (xy\,dx+y\,dy).$$

To evaluate the line integral, parameterize by x, with x in $[0,3]$. Then $y=x^2$, $dy=2x\,dx$, and

$$\text{Work}=\int_C \mathbf{F}\cdot d\mathbf{r}=\int_C (xy\,dx+y\,dy)=\int_0^3\left(x\cdot x^2\,dx+x^2(2x\,dx)\right)$$

$$=\int_0^3 3x^3\,dx=\frac{3}{4}x^4\Big|_0^3=\frac{243}{4}.$$

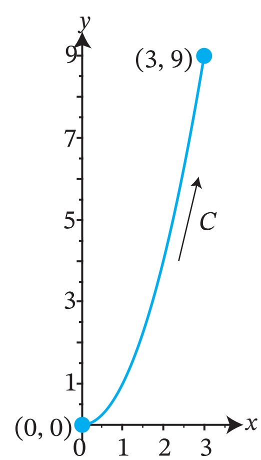


Figure 15.4.3

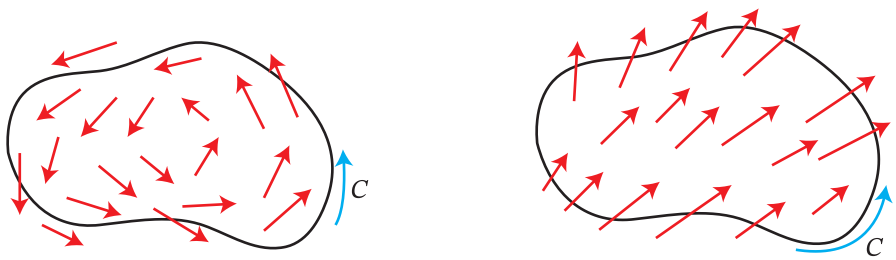
◆

Observation 15.4.1: *Another View of Work*

If we write $d\mathbf{r}$ as $\mathbf{T}\ ds$ then the work integral becomes $\int_C \mathbf{F}\cdot\mathbf{T}\ ds$. This says "Work is the integral of the tangential component of the force."

Circulation of a Fluid

Figures 15.4.4(a) and (b) show a closed curve C and the vectors describing a fluid flow $\mathbf{F}$.



Large circulation along C
(a)

Small circulation along C
(b)

Figure 15.4.4

In Figure 15.4.4(a), C surrounds a whirlpool and there is a tendency for fluid to flow along C rather than across it. In Figure 15.4.4(b) most of the fluid flow is across C rather than parallel to it. The component of $\mathbf{F}$ parallel to the tangent vector determines the tendency of the fluid to flow along C. Because $\mathbf{F}\cdot d\mathbf{r}$ represents flow in the direction of $d\mathbf{r}$, $\oint_C \mathbf{F}\cdot d\mathbf{r}$ represents the tendency of the fluid to flow along C. If C is oriented counterclockwise and $\oint_C \mathbf{F}\cdot d\mathbf{r}$ is positive, the flow of $\mathbf{F}$ along C would be counterclockwise as well. If $\oint_C \mathbf{F}\cdot d\mathbf{r}$ is negative, the flow would tend to be clockwise. The line integral $\oint_C \mathbf{F}\cdot d\mathbf{r}$ is called the **circulation** of $\mathbf{F}$ along C.

Definition: *Circulation as a Line Integral*

$$\text{Circulation of a vector field } \mathbf{F} \text{ along a closed curve } C = \oint\limits_C \mathbf{F}\cdot d\mathbf{r}.$$

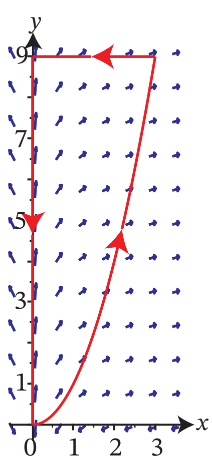


Figure 15.4.5

The integral, $\int_C \mathbf{F}\cdot d\mathbf{r}$, in the definition of circulation also occurs in the study of work and in the study of fluids.

EXAMPLE 2. Find the circulation of the planar flow $\mathbf{F}(x,y) = xy\mathbf{i} + y\mathbf{j}$ around the closed curve that follows $y = x^2$ from $(0,0)$ to $(3,9)$, then horizontally to $(0,9)$, and straight down to $(0,0)$.

SOLUTION The closed curve C comes in three parts: $C = C_1 + C_2 + C_3$ where C_1 is $y = x^2$ for $0 \le x \le 3$, $-C_2$ is $y = 9$, $0 \le x \le 3$, and $-C_3$ is $x = 0$, $0 \le y \le 9$. (Figure 15.4.5 shows the closed curve C in red and the flow is in blue.)

The circulation is

$$\begin{aligned}\oint\limits_C \mathbf{F}\cdot d\mathbf{r} &= \int\limits_{C_1} \mathbf{F}\cdot d\mathbf{r} + \int\limits_{C_2} \mathbf{F}\cdot d\mathbf{r} + \int\limits_{C_3} \mathbf{F}\cdot d\mathbf{r} \\ &= \int\limits_{C_1} \mathbf{F}\cdot d\mathbf{r} - \int\limits_{-C_2} \mathbf{F}\cdot d\mathbf{r} - \int\limits_{-C_3} \mathbf{F}\cdot d\mathbf{r}.\end{aligned}$$

We use $-C_2$ and $-C_3$ because they are easier to parameterize than C_2 and C_3.

By Example 1, $\int_{C_1} \mathbf{F}\cdot d\mathbf{r} = 243/4$. Then, by direct calculation:

$$\int_{-C_2} \mathbf{F}\cdot d\mathbf{r} = \int_0^3 \langle 9x, 9\rangle \cdot \langle dx, 0\rangle = \int_0^3 9x\,dx = \frac{81}{2}$$

and

$$\int_{-C_3} \mathbf{F}\cdot d\mathbf{r} = \int_0^9 \langle 0, y\rangle \cdot \langle 0, dy\rangle = \int_0^9 y\,dy = \frac{81}{2}.$$

The circulation of $\mathbf{F}$ around C is $\oint_C \mathbf{F}\cdot d\mathbf{r} = 243/4 - 81/2 - 81/2 = -81/4$. ♦

Loss or Gain of a Fluid (Flux)

Draw a simple closed curve C on the surface of a fluid that is moving in the xy-plane. In Figure 15.4.6(a), at what rate is fluid escaping or entering the region R whose boundary is C?

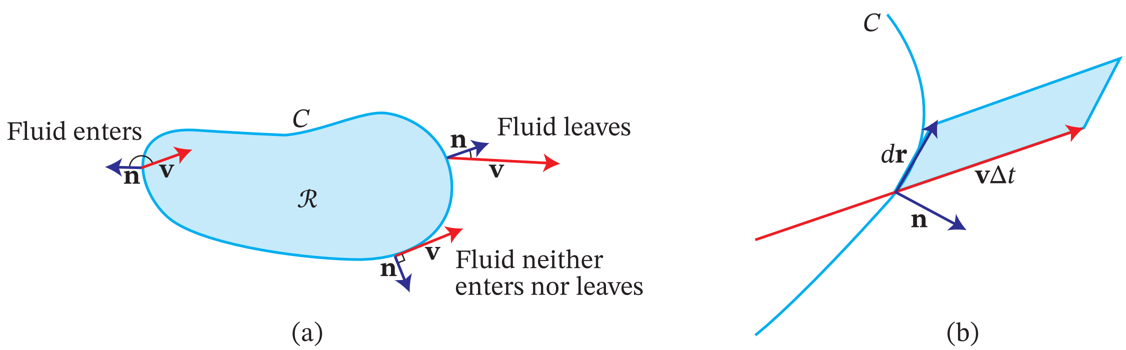


Figure 15.4.6

If the fluid tends to escape, then it is thinning out in R, becoming less dense at some points. If the fluid tends to accumulate, it is becoming denser at some points. (Think of this ideal fluid as a gas rather than a liquid; gases can vary in density while liquids tend to have constant density.)

Since the fluid is escaping or entering R only along its boundary, it suffices to consider the total loss or gain across C. Where $\mathbf{v}$, the fluid velocity, is tangent to C, fluid neither enters nor leaves. Where $\mathbf{v}$ is not tangent to C, fluid is either entering or leaving across C, as indicated in Figure 15.4.6(a). (Remember that simple closed curves are assumed to be oriented counterclockwise, when viewed from above the xy-plane.)

The rate at which fluid crosses C depends not only on its velocity but also on its density, which we denote σ. So the vector field of interest is $\mathbf{F} = \sigma\mathbf{v}$.

The vector $\mathbf{n}$ is a unit vector perpendicular to C and pointing away from the region it bounds. It is called the **exterior unit normal** or **outward unit normal**. We emphasize that this definition of $\mathbf{n}$ applies only when the curve C is closed. If C is not a closed curve, the default orientation of $\mathbf{n}$ is the one that makes an acute angle with $\mathbf{r}$.

To find the total loss or gain of fluid past C, let us look at a short section of C, which we will view as a vector $d\mathbf{r}$. How much fluid crosses $d\mathbf{r}$ in a short interval of time Δt?

During time Δt the fluid moves a distance $|\mathbf{v}|\Delta t$ across $d\mathbf{r}$. The fluid that crosses $d\mathbf{r}$ during the time Δt forms approximately the parallelogram shown in Figure 15.4.6(b).

Its area is the product of its height and its base $|d\mathbf{r}|$. That is,

$$\text{Area of parallelogram} = |\mathbf{proj}_{\mathbf{n}}(\mathbf{v}\Delta t)||d\mathbf{r}| = (\mathbf{v}\Delta t)\cdot\mathbf{n}|d\mathbf{r}|.$$

Since the density of the fluid is σ,

$$\text{Mass in parallelogram} = \sigma(\mathbf{v}\Delta t)\cdot\mathbf{n}|d\mathbf{r}| = (\sigma\mathbf{v})\cdot\mathbf{n}|d\mathbf{r}|\Delta t = \mathbf{F}\cdot\mathbf{n}|d\mathbf{r}|\Delta t.$$

Thus the rate at which fluid crosses $d\mathbf{r}$ per unit time is approximately

$$\frac{\mathbf{F}\cdot\mathbf{n}|d\mathbf{r}|\Delta t}{\Delta t} = \mathbf{F}\cdot\mathbf{n}|d\mathbf{r}|.$$

Since $d\mathbf{r}$ approximates a short piece of the curve, its length $|d\mathbf{r}|$ approximates the arc length ds. Therefore, the rate at which the fluid crosses a short part of C of length ds is approximately

$$\mathbf{F}\cdot\mathbf{n}\,ds.$$

This leads us to make the following definition:

Definition: *Flux Across a Curve as an Integral*

The rate of net loss or gain of fluid inside the region bounded by a closed curve C is called the

$$\text{Flux of the vector field } \mathbf{F} \text{ across the closed curve } C = \oint_C \mathbf{F}\cdot\mathbf{n}\,ds.$$

NOTE: Flux comes from the Latin *fluxus* (flow), from which we also get influx, reflux and fluctuate, but, oddly, not flow, which comes from the Latin *pluere* (to rain).

If the **flux** of $\mathbf{F}$ across C, $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$, is positive, fluid tends to leave R, and the mass of fluid in R decreases. If the flux is negative, fluid tends to enter R, and the mass of fluid in R increases.

Contrast this with $\int \mathbf{F}\cdot\mathbf{T}\,ds$, the integral of the tangential component, which describes circulation and work.

$$\text{Circulation} = \oint_C \mathbf{F}\cdot\mathbf{T}\,ds = \text{integral of tangential component of } \mathbf{F} \text{ along } C$$

$$\text{Flux} = \oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \text{integral of normal component of } \mathbf{F} \text{ across } C.$$

RECALL: $\oint_C \mathbf{F}\cdot\mathbf{T}\,ds$ depends on the direction in which C is traversed, while $\oint_C \mathbf{F}\cdot\mathbf{n}ds$ does not.

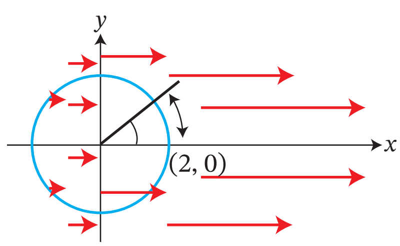


Figure 15.4.7

EXAMPLE 3. Assume the vector field $\mathbf{F} = (2+x)\mathbf{i}$ describes the flow of a fluid in the xy plane. Does the amount of fluid within the circle C of radius 2 and center $(0,0)$ tend to increase or decrease?

SOLUTION Figure 15.4.7 shows the circle and a few of the vectors of $\mathbf{F}$. Since the flow increases as we move to the right, there appears to be more fluid leaving the disk than entering it. We expect the flux $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$ to be positive. To compute $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$, introduce θ as the parameter. Then

$$x = 2\cos(\theta), \qquad y = 2\sin(\theta) \qquad \text{for } 0 \le \theta \le 2\pi.$$

Since the circle has radius 2, $s = 2\theta$ and therefore $ds = 2d\theta$.

The unit normal is parallel to the radius vector $x\mathbf{i} + y\mathbf{j}$. Therefore,

$$\mathbf{n} = \frac{x\mathbf{i}+y\mathbf{j}}{|x\mathbf{i}+y\mathbf{j}|} = \frac{2\cos(\theta)\mathbf{i}+2\sin(\theta)\mathbf{j}}{2} = \cos(\theta)\mathbf{i}+\sin(\theta)\mathbf{j},$$

which leads to

$$\begin{aligned}
\text{Flux} = \oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \int_0^{2\pi} \underbrace{[(2+x)\mathbf{i}\cdot\mathbf{n}]}_{\mathbf{F}\cdot\mathbf{n}}\underbrace{2d\theta}_{ds} &= \int_0^{2\pi} (2+2\cos(\theta))\mathbf{i}\cdot(\cos(\theta)\mathbf{i}+\sin(\theta)\mathbf{j})2\,d\theta \\
= \int_0^{2\pi} (4\cos(\theta)+4\cos^2(\theta))\,d\theta &= \int_0^{2\pi} (4\cos(\theta)+2+2\cos(2\theta))\,d\theta \\
= (4\sin(\theta)+2\theta+\sin(2\theta))\Big|_0^{2\pi} &= 4\pi.
\end{aligned}$$

As expected, the flux is positive since there is a net flow out of the disk. ♦

CHALLENGE: Evaluate the definite integrals in Example 3 in your head – use symmetry.

The Angle Subtended by a Curve

Our fourth illustration of a line integral concerns the angle subtended at a point O by a curve C in the plane. We assume that each ray from O meets C in at most one point. We include this example as background for the solid angle subtended by a surface, which appears in Section 18.4.

The curve C in Figure 15.4.8(a) subtends an angle θ at the point O. We will show that θ can be expressed as a line integral along C. Of course, the angle θ can also be found using the points A, O, and B. There will be instances where this integral arises, and we will want to remember that it represents the angle subtended by a curve.

The idea that θ can be expressed as a line integral will be important in Section 17.7 when this idea is extended from a curve to a surface. There, this concept will be relevant to the theory of gravity and electromagnetism.

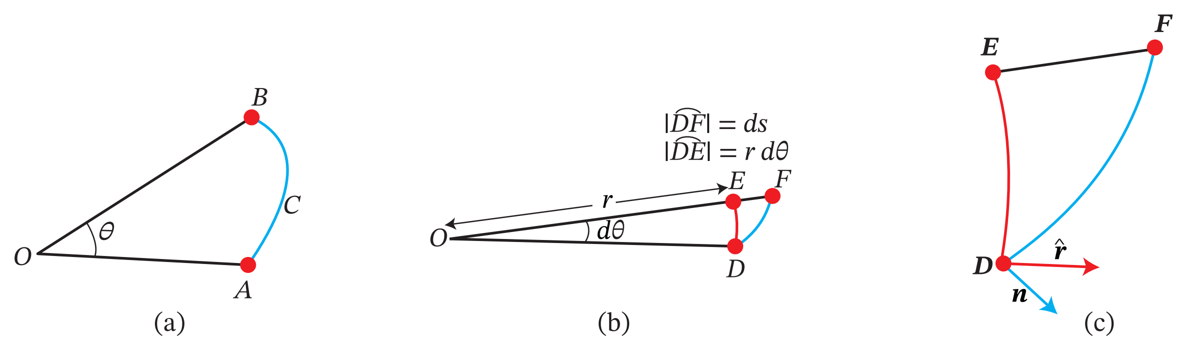


Figure 15.4.8

To express θ in Figure 15.4.8(a) as an integral over C we develop a local estimate, $d\theta$, of the radians subtended by a short section of the curve C, shown as $\widehat{DF}$ in Figure 15.4.8(b) and (c). Segment OD coincides with the vector $\mathbf{r}$ of length r and $\hat{\mathbf{r}}$ is the unit vector in the direction of $\mathbf{r}$. Arc $\widehat{DF}$ is part of the curve, and $\widehat{DE}$ is part of the circle. Because both of these arcs are almost straight, they can be approximated by a line segment of length ds. When all of this is put together, we find

$$|\widehat{DE}| \approx |\widehat{DF}|\ \cos(\hat{\mathbf{r}},\mathbf{n}) = |\widehat{DF}|\ \frac{\hat{\mathbf{r}}\cdot\mathbf{n}}{|\hat{\mathbf{r}}||\mathbf{n}|} = |\widehat{DF}|\,\hat{\mathbf{r}}\cdot\mathbf{n} \approx \hat{\mathbf{r}}\cdot\mathbf{n}\,ds.$$

Recall that $\mathbf{n}$ is the outward pointing unit normal vector to a closed curve C. In cases where C is not closed, the unit normal vector $\mathbf{n}$ is the unit normal vector to C that makes an acute angle with $\mathbf{r}$. (When $\mathbf{r}$ perpendicular to both normal vectors, it is customary to choose $\mathbf{n}$ to be consistent with nearby unit normal vectors; the choice does not matter as $\mathbf{r}\cdot\mathbf{n} = 0$ for either choice.) Thus, $\mathbf{r}\cdot\mathbf{n}$ is positive (or zero) and so

$$d\theta = \frac{|\widehat{DE}|}{r} \approx \frac{\hat{\mathbf{r}}\cdot\mathbf{n}}{r}\,ds.$$

From the local estimate we conclude that

Definition: *Radians as an Integral*

$$\text{Angle } \theta \text{ subtended by arc } C = \int_C \frac{\hat{\mathbf{r}}\cdot\mathbf{n}}{r}\,ds. \qquad (15.4.2)$$

Observation 15.4.2: *Understanding the Angle Subtended by a Curve*

The angle subtended by a curve C is the integral with respect to arc length of the normal component of the vector function $\hat{\mathbf{r}}/|\mathbf{r}|$. That is, it is the flux of the vector field $\hat{\mathbf{r}}/r$ (in the plane).

In the cases considered here the geometry ensures that $\mathbf{r}\cdot\mathbf{n}$ is positive. However, there are occasions where it is natural to have $\mathbf{r}\cdot\mathbf{n}$ negative, that is, the angle between $\mathbf{n}$ and $\mathbf{r}$ is obtuse, as in Exercise 33.

EXAMPLE 4. Verify (15.4.2) for the angle subtended at the origin by the line segment that joins $(1,0)$ and $(1,1)$. Choose $\mathbf{n}$ so that $\mathbf{r}\cdot\mathbf{n}$ is positive.

SOLUTION The subtended angle θ is shown in Figure 15.4.9(a). Obviously, $\theta = \pi/4$.

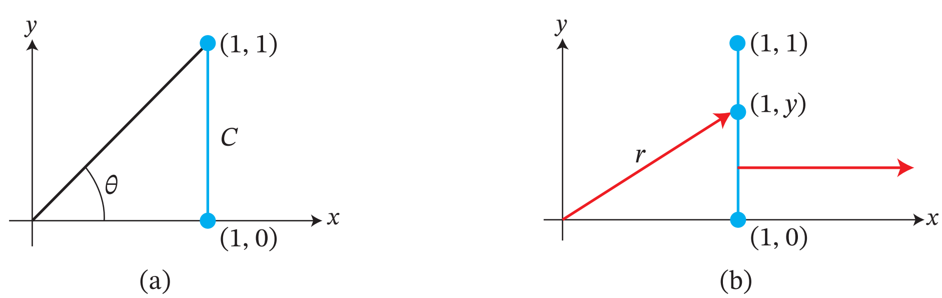


Figure 15.4.9

To confirm our intuition, evaluate the integral in (15.4.2). Figure 15.4.9(b) shows that $\mathbf{n} = \mathbf{i}$ and $\mathbf{r} = \mathbf{i} + y\mathbf{j}$. To parameterize the curve C, use $s = y$:

$$\begin{aligned}\theta &= \int_C \frac{\mathbf{n}\cdot\hat{\mathbf{r}}}{|\mathbf{r}|}\,ds \quad = \int_C \frac{1}{\sqrt{1+y^2}}\mathbf{i}\cdot\left(\frac{\mathbf{i}+y\mathbf{j}}{\sqrt{1+y^2}}\right)ds \\ &= \int_C \frac{1}{1+y^2}\,ds = \int_0^1 \frac{1}{1+y^2}\,dy \\ &= \tan^{-1}(y)\Big|_0^1 \quad = \frac{\pi}{4}.\end{aligned}$$

This agrees with our observation. ♦

Summary

The basic concepts introduced in this section are summarized in Table 15.4.1.

Application	Integral	Description
Work	$\int_C \mathbf{F}\cdot\mathbf{T}\,ds = \int_C \mathbf{F}\cdot d\mathbf{r}$	integral of tangential component of force $\mathbf{F}$ along C
Circulation	$\oint_C \mathbf{F}\cdot\mathbf{T}\,ds$	integral of tangential component of flow $\mathbf{F}$ around closed curve C
Flux	$\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$	integral of normal component of flow $\mathbf{F}$ around closed curve C
Angle Subtended	$\int_C \frac{\hat{\mathbf{r}}\cdot\mathbf{n}}{r}\,ds$	integral of normal component of $\hat{\mathbf{r}}/r$ along C

Table 15.4.1

When the vector field $\mathbf{F}$ is in the xy-plane, $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$, the integral of the tangential component of $\mathbf{F}$ along C is $\int_C \mathbf{F}\cdot\mathbf{T}\,ds = \int_C (P\,dx + Q\,dy)$.

EXERCISES for Section 15.4

In Exercises 1 to 4 decide whether the work accomplished by the vector field in moving a particle along the curve from A to B is positive, negative, or zero.

1. Figure 15.4.10(a) **2.** Figure 15.4.10(b) **3.** Figure 15.4.10(c) **4.** Figure 15.4.10(d)

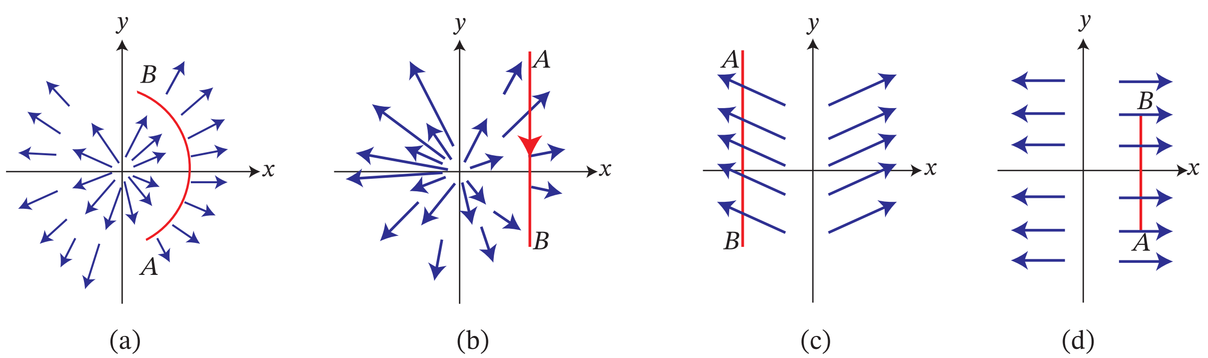


Figure 15.4.10

In Exercises 5 to 8 decide whether fluid is tending to leave, enter, or neither.

5. Figure 15.4.11(a) **6.** Figure 15.4.11(b) **7.** Figure 15.4.11(c) **8.** Figure 15.4.11(d)

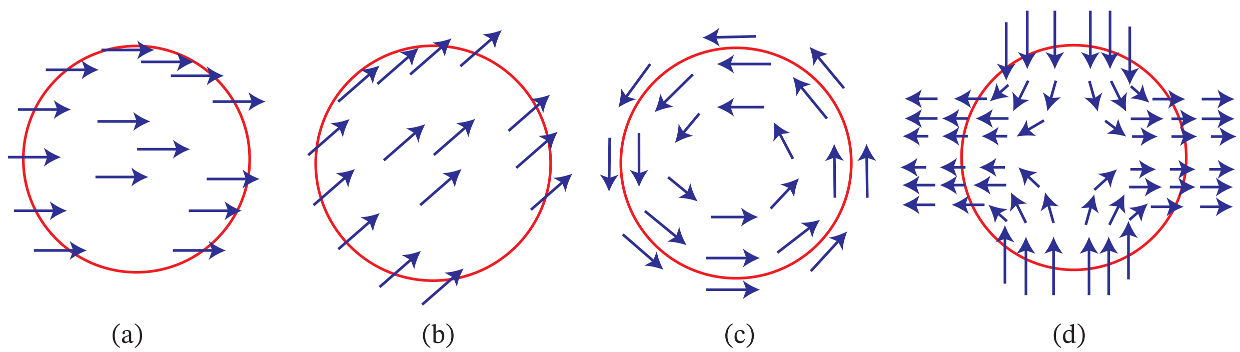


Figure 15.4.11

In Exercises 9 to 12 compute the work accomplished by the force $\mathbf{F} = x^2 y\mathbf{i} + y\mathbf{j}$ along the curve.

9. From $(0,0)$ to $(2,4)$ along the parabola $y = x^2$. **10.** From $(0,0)$ to $(2,4)$ along the line $y = 2x$.

11. From $(0,0)$ to $(2,4)$ along the path in Figure 15.4.12(a). **12.** From $(0,0)$ to $(2,4)$ along the path in Figure 15.4.12(b).

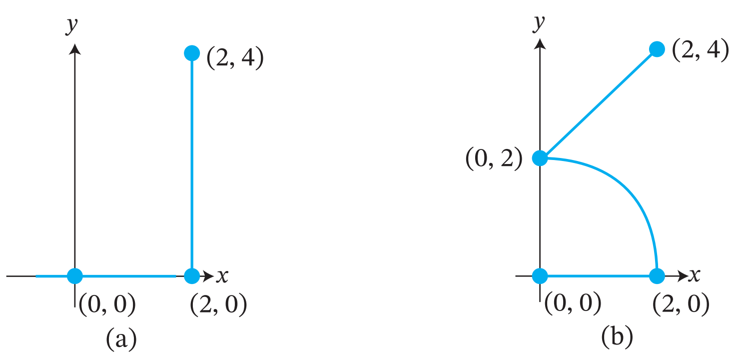


Figure 15.4.12

13. Verify (15.4.2) for the angle subtended at the origin by the line segment that joins $(2,0)$ to $(2,3)$.

14. Verify (15.4.2) for the angle subtended at the origin by the line segment that joins $(1,0)$ to $(0,1)$.

15. Let $\mathbf{F} = xy\mathbf{i} + y\mathbf{j}$ and C be the closed curve along $y = x^2$ from $(0,0)$ to $(3,9)$, then horizontally to $(0,9)$, and straight down to $(0,0)$. (a) Draw $\mathbf{F}$ at a few points on each part of C. (b) Use (a) to determine if the flow of $\mathbf{F}$ along C is clockwise or counterclockwise. (c) Does this agree with the result in Example 2?

16. Find the work done by the force $-3\mathbf{j}$ in moving a particle from $(0,3)$ to $(3,0)$ along (a) The circle of radius 3 with center at the origin. (b) The straight path from $(0,3)$ to $(3,0)$. (c) The answers to (a) and (b) are the same. Will they by the same for all curves from $(0,3)$ to $(3,0)$?

17. Figure 15.4.13(a) shows some vectors for a radially symmetric vector field $\mathbf{F}$ and a closed curve C. Use them to estimate (a) the circulation of $\mathbf{F}$ along the boundary curve C and (b) the flux of $\mathbf{F}$ across C.

NOTE: Since no formula for **F** is given, there is a range of correct answers.

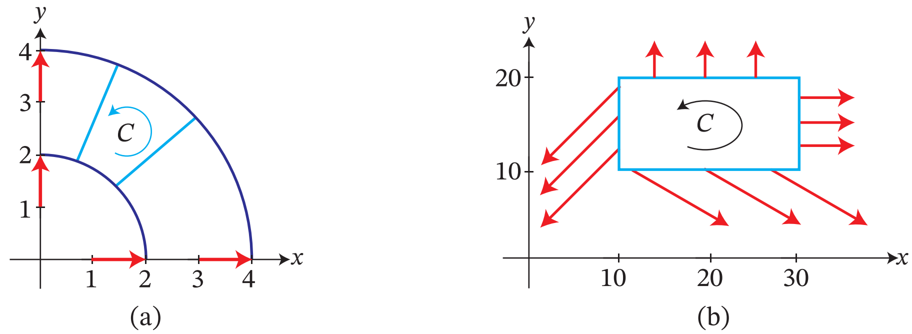


Figure 15.4.13

18. Repeat Exercise 17 for the (continuous) vector field represented in Figure 15.4.13(b).

19. The gravitational force **F** of Earth, which is located at the origin $(0,0)$ of a rectangular coordinate system, on a particle at (x,y) is

$$\frac{-x\mathbf{i}}{(\sqrt{x^2+y^2})^3}+\frac{-y\mathbf{j}}{(\sqrt{x^2+y^2})^3}=\frac{-\mathbf{r}}{|\mathbf{r}|^3}=\frac{-\widehat{\mathbf{r}}}{r^2}$$

where $\widehat{\mathbf{r}}=\dfrac{\mathbf{r}}{|\mathbf{r}|}$. Compute the total work done by **F** if the particle goes from $(2,0)$ to $(0,1)$ along (a) along the portion of the ellipse $x=2\cos(t)$, $y=\sin(t)$ in the first quadrant and (b) along the line parameterized as $x=2-2t$, $y=t$.

20. (a) Compute $W(b)$, the work done by the force in Exercise 19 in moving a particle along the straight line from $(1,0)$ to $(b,0)$. (b) What is $\lim\limits_{b\to\infty} W(b)$?

21. Consider the vector field $\mathbf{F}(x,y)$ describing a fluid flow has the value $(x+1)^2\mathbf{i}+y\mathbf{j}$ at the point (x,y) and the curve C is the unit circle with parametric equations $x=\cos(t)$, $y=\sin(t)$, for t in $[0,2\pi]$.

(a) Draw **F** at eight equally spaced points on the circle.
(b) Is fluid tending to leave or enter the region bounded by C?
(c) Compute the net outward flow using a line integral.

22. Repeat Exercise 21 where $\mathbf{F}(x,y)=(2-x)\mathbf{i}+y\mathbf{j}$ and C is the square with vertices $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$.

23. Let $\mathbf{F}(x,y)=\sigma\mathbf{v}$ be fluid flow, and let C be a closed curve in the xy-plane. If $\oint\limits_C \mathbf{F}\cdot d\mathbf{r}$ is positive and C is counterclockwise, does the motion along C tend to be clockwise or counterclockwise?

24. Let $\mathbf{F}(x,y)=\sigma\mathbf{v}$ be fluid flow, and let C be a closed curve in the xy-plane. If $\oint\limits_C \mathbf{F}\cdot\mathbf{n}\,ds$ is positive, is fluid tending to leave the region bounded by C or to enter it?

25. Let C be a closed convex curve that encloses the point O. Let $\mathbf{r}$ be the position vector $\overrightarrow{OP}$ for points P on the curve. Determine the value of $\oint\limits_C \dfrac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r}\,ds$ where $\mathbf{n}$ is the outward unit normal to C.

26. Write in your own words and diagrams why $\int\limits_C \mathbf{F}\cdot d\mathbf{r}$ represents the work done by force **F** along the curve C.

27. Write in your own words and diagrams why $\int_C \mathbf{F}\cdot\mathbf{n}\, ds$ represents the net loss of fluid across C if $\mathbf{F}$ is the fluid flow and $\mathbf{n}$ is a unit external normal to C. Include the definition of $\mathbf{F}$.

28. Explain why $\int_C \mathbf{F}\cdot d\mathbf{r}$ represents the tendency of a fluid to move along C, if $\mathbf{F}$ is the fluid flow.

29. Explain why $\int_C \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r}\, ds$ represents the angle subtended by a curve C at the origin. Assume that each ray from the origin meets C at most once.

30. Let C be a curve in space and C^* its projection on the xy plane. Assume that distinct points of C project onto distinct points of C^*. The line integral $\int_C 1\, ds$ equals the arc length of C. What integral over C equals the arc length of C^*?

31. Sam, Jane, and Sarah are debating a delicate issue.

Sam: Let C be the circle in the xy-plane whose polar equation is $r = 2\cos(\theta)$. It is a unit circle that passes through the origin O. Let $\mathbf{F}$ be the field $\frac{\widehat{\mathbf{r}}}{r}$. What is the flux of $\mathbf{F}$ across C?

Jane: The field blows up at O, so the flux is an improper integral.

Sam: Yes, but if I move C rigidly just a tiny bit so O is inside it, the flux is 2π. So I say the flux across C is 2π.

Sarah: I say it's π. Just draw a figure 8 made of two copies of C joined smoothly to form one curve, as in Figure 15.4.14(a). The flux across the curve is 2π. Each half must have flux π. Since each half looks like C, the flux across C must be π.

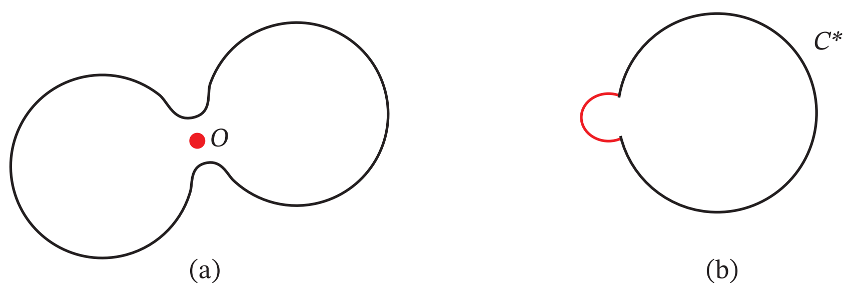


Figure 15.4.14

Settle the issue by:

(a) Evaluating the integral $\oint_C \mathbf{F}\cdot\mathbf{n}\, ds$ by the fundamental theorem of calculus.

(b) Considering the flux across the curve C^* obtained from C by replacing the small part of C near O by a semicircle C, as in Figure 15.4.14(b).

(c) By considering the angle the curve C subtends at O.

32. Assume that $\mathbf{F}(P) = \sigma(P)\mathbf{v}(P)$ represents the flow of a fluid and C is a closed curve that forms the boundary of the region R. Let $Q(t)$ be the total mass of the fluid in R at time t. Express dQ/dt in terms of a line integral.

33. Show that $\oint_C \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r}\, ds = 0$ for any convex curve C in the xy-plane which does not surround the origin.

34. For positive constants a, b, and c, verify each antiderivative formula by showing that the derivative of the right-hand side of the equation is the integrand on the left-hand side.

(a) $\int \frac{x}{ax^2+c}\,dx = \frac{1}{2a}\ln\left(ax^2+c\right)$

(b) $\int x\sqrt{ax+b}\,dx = \frac{2(3ax-2b)}{15a^2}\sqrt{(ax+b)^3}$

(c) $\int \cos^3(ax)\,dx = \frac{1}{a}\sin(ax) - \frac{1}{3a}\sin^3(ax)$

(d) $\int \tan^2(ax)\,dx = \frac{1}{a}\tan(ax) - x$

(e) $\int x\cos(ax)\,dx = \frac{1}{a^2}\cos(ax) + \frac{x}{a}\sin(ax)$

(f) $\int \arctan(ax)\,dx = x\arctan(ax) - \frac{1}{2a}\ln\left(1+a^2x^2\right)$

15.S Chapter Summary

This chapter concerned the derivatives of vector functions and integrals over curves.

Let $\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle$ be the position vector from the origin to a point on a curve. We defined its derivative, $\mathbf{r}'(t)$, in terms of the derivatives of its components. We could just as well define it as

$$\mathbf{r}'(t) = \lim_{\Delta t\to 0}\frac{\mathbf{r}(t+\Delta t)-\mathbf{r}(t)}{\Delta t} = \lim_{\Delta t\to 0}\frac{\Delta\mathbf{r}}{\Delta t}. \qquad (15.S.1)$$

This definition reveals the underlying geometry, as Figure 15.S.1(a) shows. For small Δt, the direction of $\Delta\mathbf{r}$ is almost along the tangent. The length of $\Delta\mathbf{r}$ is almost the same as the scalar length Δs along the curve. Thus, $\Delta\mathbf{r}/\Delta t$ is a vector pointing almost in the direction of motion and with a magnitude approximating the instantaneous speed.

The limit in (15.S.1) is called the derivative of the function $\mathbf{r}(t)$. If we think of t as time, then $\mathbf{r}'$ is called the velocity vector, denoted $\mathbf{v}$. The derivative of $\mathbf{v}$ is the acceleration vector: $\mathbf{v}' = \mathbf{a}$.

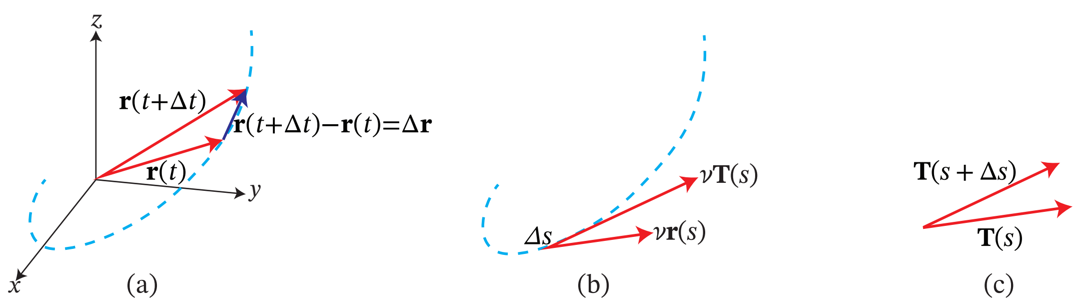


Figure 15.S.1

The vector $\mathbf{T} = \mathbf{r}'/|\mathbf{r}'|$ is a unit tangent vector. The magnitude of its derivative with respect to arc length, s, is the curvature of the path, κ, as suggested by Figure 15.S.1. Though the curve may not lie in a plane, the figure resembles Figure 15.2.5 in Section 15.2. It was shown that curvature equals $|\mathbf{v}\times\mathbf{a}|/|\mathbf{v}|^3$.

The vector $d\mathbf{T}/ds$ is perpendicular to $\mathbf{T}$. (Why?) The unit vector $\mathbf{N} = (d\mathbf{T}/ds)/|d\mathbf{T}/ds|$ is called the principal normal to the curve at the given point. The vector $\mathbf{T}\times\mathbf{N} = \mathbf{B}$ is the third unit vector forming a frame that moves along the curve, with $\mathbf{T}$ and $\mathbf{N}$ indicating the plane in which the curve locally almost lies.

The acceleration vector $\mathbf{a}$, even for space curves, can be expressed relative to $\mathbf{T}$ and $\mathbf{N}$ ($\mathbf{B}$ is not involved):

$$\mathbf{a} = \frac{d^2s}{dt^2}\mathbf{T} + \frac{v^2}{r}\mathbf{N}$$

where $r = 1/\kappa$ is the radius of curvature. The second coefficient shows that the force needed to keep the particle in the path is proportional to the square of the velocity and inversely proportional to the radius of curvature.

This chapter then introduced four integrals involving a curve C:

$$\int_C f(P)\,ds, \qquad \int_C f(P)\,dx, \qquad \int_C f(P)\,dy, \quad \text{and} \quad \int_C f(P)\,dz,$$

whose definitions resemble those in Chapter 6 for definite integrals. The first integral is a line integral with respect to arc length; its value does not depend on the orientation of the curve. The other three types of integral do depend on the orientation of the curve: switching the direction in which the curve is swept out changes the sign of dx, dy, or dz, and thus that of the integral.

For a closed curve taken counterclockwise $\oint_C y\,dx$ is the negative of the area enclosed by the curve. (Why?) On the other hand, $\int_C x\,dy$ taken counterclockwise is the area enclosed.

The most general integral considered was

$$\int_C \left(P(x,y,z)\,dx + Q(x,y,z)\,dy + R(x,y,z)\,dz\right).$$

whose integrand is called a differential form. For $\mathbf{F} = \langle P, Q, R\rangle$ it can be written as $\int_C \mathbf{F}\cdot\,d\mathbf{r}$. However, in proofs or computations we often need the differential form.

If $\int_C \mathbf{F}\cdot d\mathbf{r}$ depends only on the ends of C, $\mathbf{F}$ is called a conservative vector field, which will be important in Chapter 18.

Line integrals were applied to work, circulation, flux, and the angle subtended by a curve (the last in preparation for the solid angle subtended by a surface in Sections 18.7 and 18.4).

EXERCISES for Section 15.S

In Exercises 1 to 6, evaluate $\int_C \mathbf{F}\cdot d\mathbf{r}$ for the vector field $\mathbf{F}$ and curve C.

1. $\mathbf{F}(x,y) = 2x\,\mathbf{i}$ and C is a semicircle, $\mathbf{r}(\theta) = 3\cos(\theta)\,\mathbf{i} + 3\sin(\theta)\,\mathbf{j}$, $0 \le \theta \le \pi$.
2. $\mathbf{F}(x,y) = x^2\,\mathbf{i} + 2xy\,\mathbf{j}$ and C is a line segment, $\mathbf{r}(t) = 2t^2\,\mathbf{i} + 3t^2\,\mathbf{j}$, $1 \le t \le 2$.
3. $\mathbf{F}(x,y,z) = x\,\mathbf{i} + y\,\mathbf{j} + z\,\mathbf{k}$ and C is a helix, $\mathbf{r}(t) = \cos(t)\,\mathbf{i} + \sin(t)\,\mathbf{j} + 3t\,\mathbf{k}$, $0 \le t \le 4\pi$.
4. $\mathbf{F}(x,y,z) = x^2\,\mathbf{i} + xy\,\mathbf{j} + 3\,\mathbf{k}$ and C is a line segment, $\mathbf{r}(t) = 2t\,\mathbf{i} + (3t+1)\,\mathbf{j} + t\,\mathbf{k}$, $1 \le t \le 2$.
5. $\mathbf{F}(\mathbf{r}) = \dfrac{\widehat{\mathbf{r}}}{|\mathbf{r}|^2}$ and C is a line, $\mathbf{r}(t) = 2t\,\mathbf{i} + 3t\,\mathbf{j} + 4t\,\mathbf{k}$, $1 \le t \le 2$.
6. $\mathbf{F}(\mathbf{r}) = \mathbf{r}$ and C is the circle $\mathbf{r}(t) = \cos(\theta)\,\mathbf{i} + \sin(\theta)\,\mathbf{j} + 2\,\mathbf{k}$, $0 \le \theta \le 2\pi$.

---

7. **SAM:** Remember when they defined curvature back in Section 9.6? They used an angle ϕ and then the curvature was defined as the limit of $\dfrac{\Delta\phi}{\Delta s}$.
JANE: Yes.
SAM: Now, they say that definition will not work for space curves because there is no angle ϕ.
JANE: Right.
SAM: Well, I don't need an angle ϕ. Watch how I define curvature for a space curve.
JANE: I can't wait.
SAM: I take a nearby point a distance Δs from the point of interest. The angle between the tangent at the nearby point and the point of interest is small when Δs is. I call it A. Then I define the curvature as the limit of $A/\Delta s$ as Δs goes to 0. In short, I use the same definition as in Chapter 9.
JANE: It looks reasonable, but does it give the same curvature as the book gets?

Does it? If so, why did the authors not use Sam's approach?

8. Figure 15.S.2(a) shows $\mathbf{T}$ and $\mathbf{N}$ for a point P on a curve C. The curve is not shown. Sketch what a short part of C may look like.

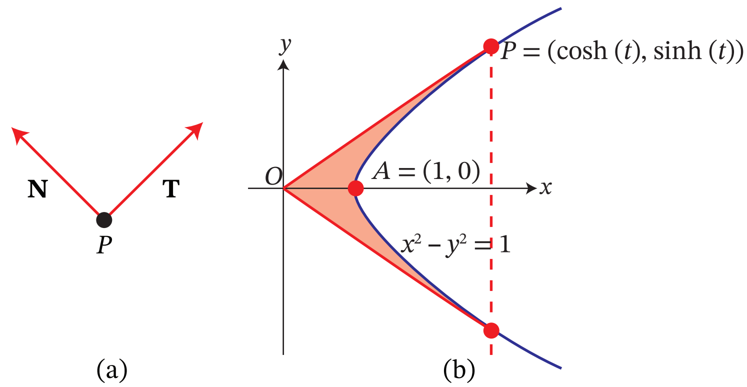


Figure 15.S.2

9. (a) Express the area under the hyperbola $x^2 - y^2 = 1$ and above the interval $[1, \cosh(t)]$ as a line integral.
(b) Evaluate it.
(c) What is the area of the shaded region in Figure 15.S.2(b)?

See also Exercises 64 in Section 6.S and 76 in Section 8.S.

In "Reflections on Reflections" (CIE 4 at the end of Chapter 3) the reflection properties of parabolas and ellipses were developed. Exercises 10 and 11 show an easier way to obtain the results using vectors.

10. A parabola consists of the points P equidistant from a point F and a line L, as in Figure 15.S.3.

Let O be a point on L and let $\mathbf{u}$ be a unit vector perpendicular to L aimed toward P. Let $\mathbf{r} = \overrightarrow{OP}$ and $\mathbf{F} = \overrightarrow{OF}$. We assume the curve is parameterized so that there is a well-defined tangent vector, $\mathbf{r}'$.

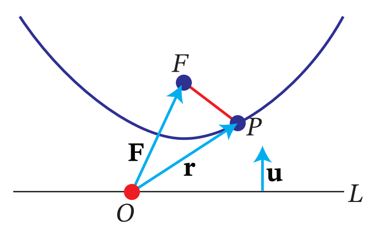


Figure 15.S.3

(a) Show that $|\mathbf{r} - \mathbf{F}| = \mathbf{r} \cdot \mathbf{u}$.
(b) From (a) deduce that $\dfrac{\mathbf{r} - \mathbf{F}}{|\mathbf{r} - \mathbf{F}|} \cdot \mathbf{r}' = \mathbf{r}' \cdot \mathbf{u}$.
(c) From (b) deduce that $|\mathbf{r}'|\cos(\mathbf{r}', \mathbf{r} - \mathbf{F}) = |\mathbf{r}'|\cos(\mathbf{r}', \mathbf{u})$.
(d) From (c) deduce the reflection principle of a parabola.

REFERENCE: The proof of the reflection principle of a parabola presented in Exercise 10 appears in Harley Flanders', "The Optical Properties of the Conics," *American Mathematical Monthly*, 1968, p. 399.

11. This exercise develops the reflection property of an ellipse. Start with its geometric definition as the locus of points such that the sum of their distances from two fixed points is constant. Let $\mathbf{p}$ and $\mathbf{q}$ be the position vectors of the fixed points and $\mathbf{r}$ the position vector for a point P on the ellipse, which is parameterized so we may speak of $\mathbf{r}'$, a tangent vector.
(a) Differentiate both sides of $|\mathbf{r} - \mathbf{p}| + |\mathbf{r} - \mathbf{q}| = c$, a constant.
(b) Let $\mathbf{u}_1$ be the unit vector in the direction of $\mathbf{r} - \mathbf{p}$ and $\mathbf{u}_2$ be the unit vector in the direction of $\mathbf{r} - \mathbf{q}$. Show that $\mathbf{u}_1 \cdot \mathbf{r}' + \mathbf{u}_2 \cdot \mathbf{r}' = 0$.
(c) Show that $\mathbf{u}_1 + \mathbf{u}_2$ is normal to the curve at P.
(d) Show that $\mathbf{u}_1$ and $\mathbf{u}_2$ make equal angles with $\mathbf{u}_1 + \mathbf{u}_2$.
(e) From (d) deduce the reflection property of an ellipse.

In Exercises 12 to 20 evaluate each antiderivative.

12. $\int 108t\,(\ln(t))^2 \; dt$

13. $\int \ln(1+t^2)\, dt$

14. $\int t \arctan(t)\, dt$

15. $\int \frac{\tan(t)+\sin(t)}{\sec(t)}\, dt$

16. $\int \frac{t^4}{t^2+4}\, dt$

17. $\int \frac{2t-4}{t^2+2t+1}\, dt$

18. $\int (t^2+4t+5)^{-1}\, dt$

19. $\int t^2 \cos(t)\, dt$

20. $\int \frac{1}{t^2+4}\, dt$

Calculus is Everywhere # 19

The Suspension Bridge and the Hanging Cable

In a suspension bridge the roadway hangs from a cable, as shown in Figure C.19.1. We will use calculus to find the shape of the cable. We assume that the weight of a section of the roadway is proportional to its length. That is, there is a constant k such that x feet of the roadway weighs kx pounds. We will assume that the cable is weightless. That is justified for it weighs little in comparison to the roadway.

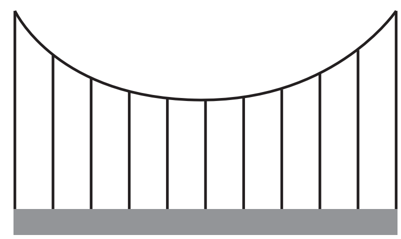


Figure C.19.1

We introduce an xy-coordinate system with origin at the lowest point of the cable, and consider the section of the cable that goes from $(0,0)$ to (x,y), as shown in Figure C.19.2. Three forces act on it. The force at $(0,0)$ is horizontal and pulls the cable to the left. Call its magnitude T. Gravity pulls the cable down with a force whose magnitude is kx, the weight of the roadway beneath it. At the top of the section, at (x,y), the cable above it pulls to the right and upward along the tangent line to the cable.

The section does not move. The horizontal part of the force at (x,y) must have magnitude T and the vertical part of the force must have magnitude kx, also shown in Figure C.19.2.

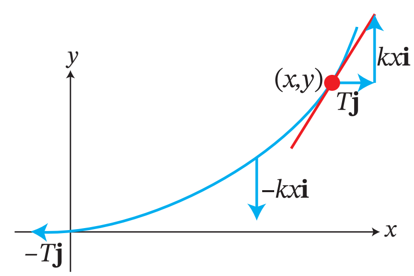


Figure C.19.2

Since the force at the point (x,y) is directed along the tangent line there, we have

$$\frac{dy}{dx} = \frac{kx}{T}.$$

Therefore

$$y = \frac{kx^2}{2T} + C$$

for some constant C. Since $(0,0)$ is on the curve, $C = 0$, and the cable has the equation

$$y = \frac{kx^2}{2T}.$$

The cable forms a parabola.

What if we have the cable but no roadway? That is also the case for a laundry line, or a telephone wire, or a hanging chain. In these cases the downward force is due to the weight of the cable. If s feet of cable weighs ks pounds, reasoning similar to that for the suspension bridge leads to

$$\frac{dy}{dx} = \frac{ks}{T}.$$

Since

$$s = \int_0^x \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$

we have

$$\frac{dy}{dx} = \frac{k}{T}\int_0^x \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx. \tag{C.19.1}$$

If we differentiate both sides of (C.19.1), and use the second part of the fundamental theorem of calculus, we get

$$\frac{d^2y}{dx^2} = \frac{k}{T}\sqrt{1 + \left(\frac{dy}{dx}\right)^2}. \tag{C.19.2}$$

It can be shown (see Exercise 1) that

$$y = \frac{T}{2k}\left(e^{kx/T} + e^{-kx/T}\right) - \frac{T}{k}.$$

The curve is called a **catenary**, after the Latin catena, meaning chain. It may look like a parabola, but it is not. The 630-foot tall Gateway Arch in St. Louis, completed October 28, 1965, is a famous catenary.

EXERCISES for CIE C.19

1. Check that the solution of (C.19.1) that passes through $(0,0)$ with slope 0 is $y = \frac{T}{2k}\left(e^{kx/T} + e^{-kx/T}\right) - \frac{T}{k}$.

Calculus is Everywhere # 20

The Path of the Rear Wheel of a Scooter

When the front wheel of a scooter follows a certain path, what is the path of its rear wheel? This question could be asked for a bicycle or car, but the scooter is more convenient for carrying out experiments. As you read this section we encourage you to do some experiments. Get a scooter (or bike) and implement some of the scenarios that we discuss.

The **tractrix problem** is the special case when the front wheel moves in a straight line. (See Exercise 66 in Section 8.5.) Now, using vectors, we will look at the case when the front wheel sweeps out a circular path.

The Basic Equation

Figure C.20.1 shows the geometry at any instant. Let s denote the arc length of the path swept out by the rear wheel as measured from its starting point. Let a be the length of the wheel base, that is, the distance between the front and rear axles. The vector $\mathbf{r}(s)$ gives the position of the rear wheel and $\mathbf{f}(s)$ gives the position of the front wheel. Because the rear wheel is parallel to $\mathbf{f}(s)-\mathbf{r}(s)$, the unit vector $\mathbf{r}'(s) = d\mathbf{r}/ds$ points directly toward the front wheel (when the scooter is moving forward) or directly away from it (when the scooter is backing up).

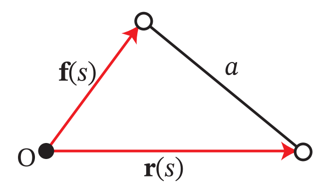


Figure C.20.1

Thus $\mathbf{f}(s) = \mathbf{r}(s) + a\mathbf{r}'(s)$ or $\mathbf{f}(s) = \mathbf{r}(s) - a\mathbf{r}'(s)$. In short, we will write

$$\mathbf{f}(s) = \mathbf{r}(s) \pm a\mathbf{r}'(s).$$

Assume that the front wheel is moving, say, counterclockwise and tracing out a circular path with center O and radius c. Because

$$\mathbf{f}(s)\cdot\mathbf{f}(s) = c^2$$

we have

$$\left(\mathbf{r}(s) \pm a\mathbf{r}'(s)\right)\cdot\left(\mathbf{r}(s) \pm a\mathbf{r}'(s)\right) = c^2.$$

By distributivity of the dot product,

$$\mathbf{r}(s)\cdot\mathbf{r}(s) + a^2\mathbf{r}'(s)\cdot\mathbf{r}'(s) \pm 2a\mathbf{r}(s)\cdot\mathbf{r}'(s) = c^2. \tag{C.20.1}$$

Letting $r(s) = |\mathbf{r}(s)|$, we may rewrite (C.20.1) as

$$(r(s))^2 + a^2 \pm 2a\mathbf{r}(s)\cdot\mathbf{r}'(s) = c^2. \tag{C.20.2}$$

Differentiate $\mathbf{r}(s)\cdot\mathbf{r}(s) = r(s)^2$ to obtain $\mathbf{r}(s)\cdot\mathbf{r}'(s) = r(s)r'(s)$, which changes (C.20.1) to an equation involving the scalar function $r(s)$. If we write $r(s)$ as r and $r'(s)$ as r' we arrive at

$$r^2 + a^2 \pm 2arr' = c^2. \tag{C.20.3}$$

This is the equation we will use to analyze the path of the rear wheel of a scooter.

The Direction of $\mathbf{r}'$

We first find when $\mathbf{r}'$ points towards the front wheel and when it points away from the front wheel.

The movement of the back wheel is determined by the projection of $\mathbf{f}'$ on the line of the scooter, which is the same as $\mathbf{r}'$.

Observation C.20.1: ***The Connection between θ and $\mathbf{r}'$***

When the angle θ between the front wheel and the line of the scooter is obtuse, as in Figure C.20.2(a), $\mathbf{r}'$ points towards the front wheel. When θ is acute, the scooter backs up and $\mathbf{r}'$ points away from the front wheel, as shown in Figure C.20.2(b).

When the direction of $\mathbf{r}'$ shifts from pointing towards the front wheel to pointing away from it, the path of the rear wheel also changes, as shown in Figure C.20.2(c).

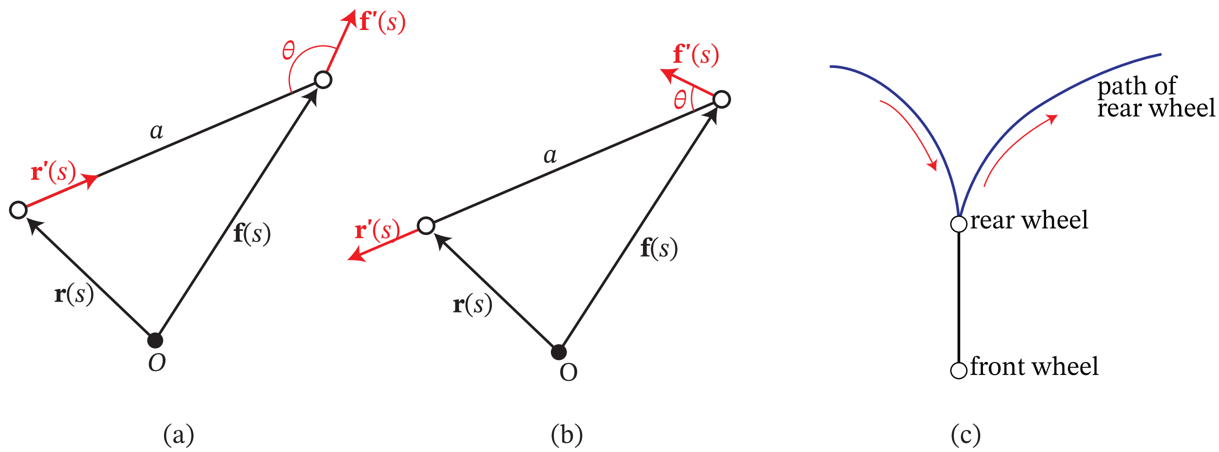


Figure C.20.2

The path of the rear wheel is continuous but the unit tangent vector $\mathbf{r}'$ is not defined where its direction suddenly shifts. The path is said to contain a cusp and the point at which $\mathbf{r}'(s)$ shifts direction by the angle π is the vertex of the cusp.

The Path of the Rear Wheel for a Short Scooter

Assume the wheelbase a is less than the radius of the circle c, θ is obtuse, and r^2 is less than c^2-a^2. Thus, $c^2-a^2-r^2$ is positive. (Exercise 5 shows the significance of c^2-a^2.)

We write $c^2 = a^2+r^2+2rr'a$ in the form

$$\frac{-2rr'}{c^2-a^2-r^2} = \frac{-1}{a}. \tag{C.20.4}$$

Integration of both sides of (C.20.4) with respect to arc length s shows that there is a constant k such that

$$\ln(c^2-a^2-r^2) = \frac{-s}{a}+k,$$

so

$$c^2-a^2-r^2 = e^k e^{-s/a}. \tag{C.20.5}$$

Equation (C.20.5) tells us that r^2 increases but remains less than c^2-a^2, and approaches c^2-a^2 as s increases. Thus the rear wheel traces a spiral path that gets arbitrarily close to the circle of radius $\sqrt{c^2-a^2}$ and center O, as in Figure C.20.3. This figure shows the path of the rear wheel of a scooter with length $a=1$, whose front wheel moves counterclockwise around the circle with radius $c=2$ from the point $(2,0)$ with the line of the scooter at an angle $\theta = 3\pi/4$ with the front wheel. The snapshots are taken when $s=0$, $s=1.25$, $s=2.50$, $s=5.0$, $s=10.0$, and $s=15.0$. Because this is a short scooter ($a<c$), the rear wheel approaches the circle with radius $r=\sqrt{c^2-a^2}=\sqrt{3}$. (Recall that s is the arc length of the rear wheel's path.)

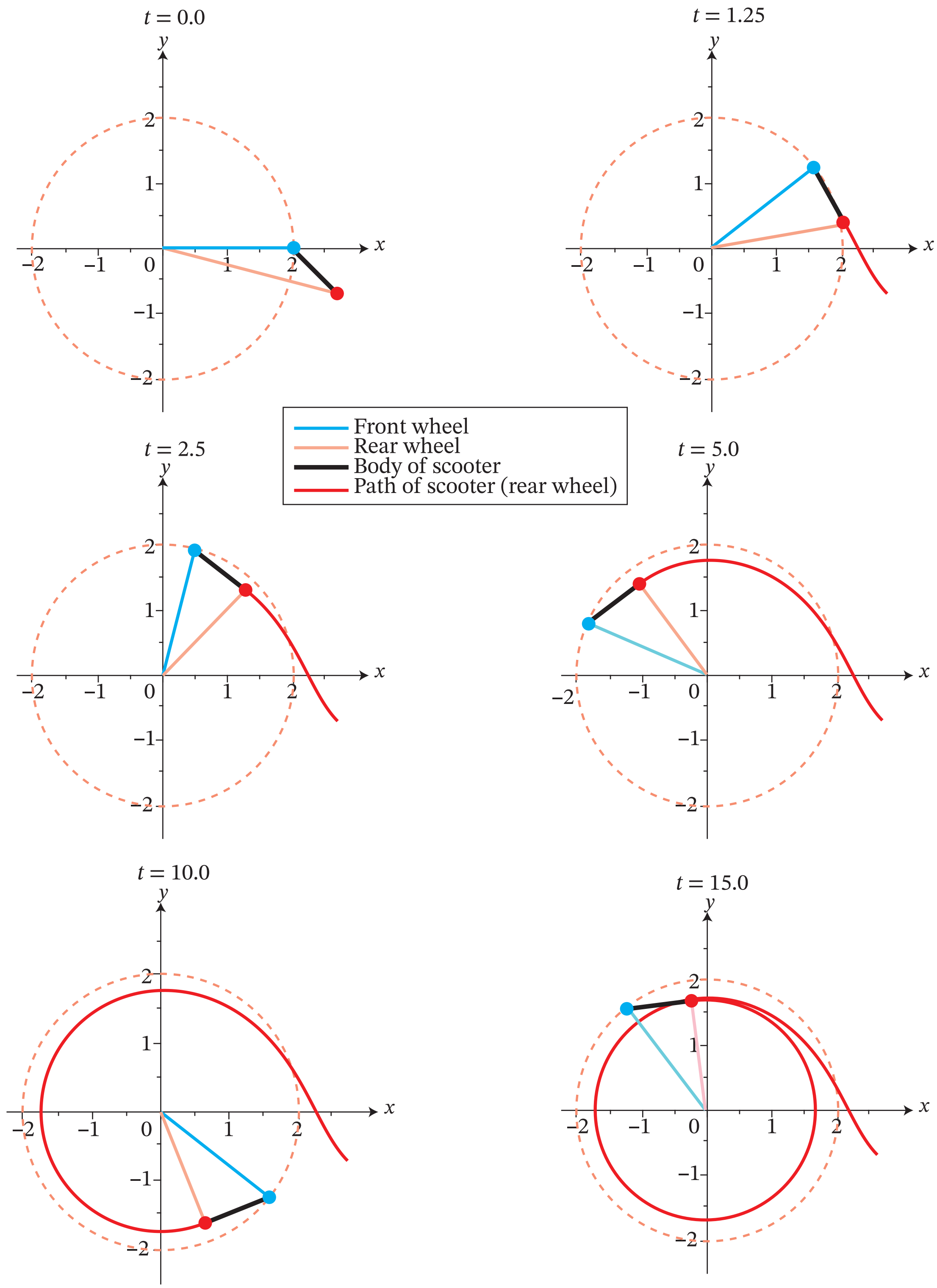

t = 0.0
t = 1.25
t = 2.5
t = 5.0
t = 10.0
t = 15.0
Front wheel
Rear wheel
Body of scooter
Path of scooter (rear wheel)

Figure C.20.3

The Path of the Rear Wheel for a Long Scooter

Assume that the wheelbase is longer than the radius of the circle on which the front wheel moves, that is, $a > c$. Assume that initially the scooter is moving forward, so we again have

$$c^2 = a^2 + r^2 + 2rr'a. \tag{C.20.6}$$

The initial position is indicated in the snapshot of the scooter's configuration at $t = 0$ (see the upper-left frame in Figure C.20.4.

Now $c^2 - a^2 - r^2$ is negative, and we have

$$\frac{2rr'}{a^2 + r^2 - c^2} = \frac{-1}{a},$$

where the denominator on the left-hand side is positive. Thus there is a constant k such that

$$a^2 + r^2 - c^2 = e^k e^{-s/a}. \tag{C.20.7}$$

As s gets arbitrarily large, (C.20.7) implies that r^2 approaches $c^2 - a^2$. But this cannot happen because $c^2 - a^2$ is negative. Our assumption that (C.20.6) holds for all s must be wrong. There must be a cusp and the equation switches to

$$c^2 = a^2 + r^2 - 2arr'.$$

This leads to

$$a^2 + r^2 - c^2 = e^k e^{s/a},$$

which implies that as s increases r becomes arbitrarily large. However, r can never exceed $c + a$. So, another cusp must form.

It can be shown that the cusps occur when $r = a - c$ (assuming $a > c$) and $r = a + c$. At the vertex of a cusp, $\mathbf{r}'$ is not defined; it changes direction by π.

Figure C.20.4 shows the shape of the path of the rear wheel of a scooter with length $a = 4$ whose front wheel moves counterclockwise around the circle with radius $c = 2$ from the point $(2, 0)$ with the line of the scooter at an angle $\theta = \pi$ with the front wheel. The snapshots are taken when $s = 0$, $s = 3$, $s = 9$, $s = 18$, $s = 36$, an $s = 72$. Because the scooter is long ($a > c$), the rear wheel travels along a path that has cusps when $r = c + a$ and $r = |c - a|$. (For $a > 2c$, that path remains outside the circle.)

EXERCISES for CIE C.20

1. When a bus or car (or scooter) turns a corner why does a rear tire sometimes go over the curb even though a front tire does not?

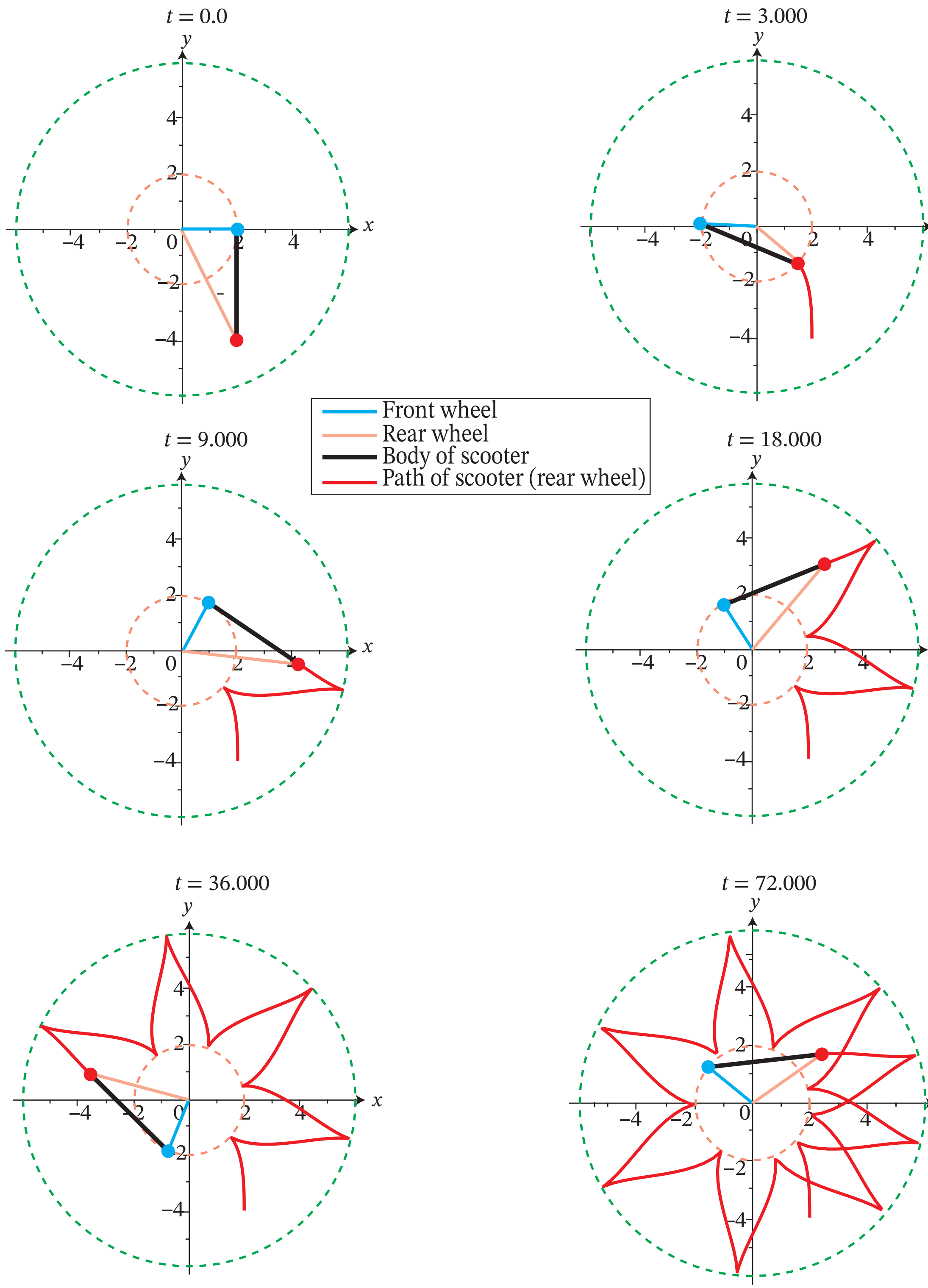
2. It is a belief among many bicyclists that the rear tire of a bicycle wears out more slowly than the front tire. Decide whether the belief is justified. (Assume both tires support the same weight.)

3. (a) Is $\dfrac{d\mathbf{r}}{ds}$ a unit vector? (b) Is $\dfrac{d\mathbf{f}}{ds}$ a unit vector?

4. When the front wheel is held at a fixed angle ϕ, $0 < \phi < \frac{\pi}{2}$, to the body of the scooter both wheels travel on circular paths. Express the radii of these paths in terms of ϕ and a, the wheel base of the scooter.

5. (a) Assume a and c are positive with $c > a$ and that the front wheel moves on a circle of radius c. Show that the rear wheel could remain on a concentric circle of radius $b = \sqrt{c^2 - a^2}$.

(b) Draw the triangle whose sides are a, b, and c and explain why the result in (a) is plausible.



$t = 0.0$
$t = 3.000$
$t = 9.000$
$t = 18.000$
$t = 36.000$
$t = 72.000$
Front wheel
Rear wheel
Body of scooter
Path of scooter (rear wheel)

Figure C.20.4

6. We assumed for a short scooter that initially $r^2 < c^2 - a^2$. Examine the case in which initially $r^2 > c^2 - a^2$. Assume that initially the scooter is not backing up.

7. We assumed in the case of the short scooter that initially $r^2 < c^2 - a^2$ and that the scooter is not backing up. Investigate what happens when we assume that initially $r^2 < c^2 - a^2$ and the scooter is backing up.

(a) Draw such an initial position. (b) Predict what will happen. (c) Carry out the mathematics.

8. Show that if the path of the front wheel of a scooter is a circle and a cusp forms in the path of the rear wheel, the scooter at that moment lies on a line through the center of the circle.

9. For a long scooter, $a > c$, do cusps always form, whatever the initial value of r and θ?

10. Extend the analysis of the scooter to the case when $a = c$.

11. Assume that the scooter is moving forward with the path of the front wheel following a straight line. For convenience, choose that line as the x-axis. The path of the rear wheel is called a **tractrix**. (This case appeared in Section 8.5, Exercise 66.)

(a) Write $\mathbf{r}(s)$ as $x(s)\mathbf{i} + y(s)\mathbf{j}$. Then, show that $y(s) + y'(s)a = 0$.

(b) Deduce that there is a constant k such that $y(s) = ke^{-s/a}$. Thus the distance from the rear wheel to the x-axis decays exponentially as a function of arc length.

12. **Sam:** I can use the scooter to show $1 = 0$.
Jane: Remarkable.
Sam: If I hold the front wheel perpendicular to the scooter base then $\mathbf{f}' \cdot \mathbf{r}' = 0$.
Jane: That makes sense.
Sam: But $\mathbf{f}' \cdot \mathbf{r}' = (\mathbf{r}' + \mathbf{r}''a) \cdot \mathbf{r}' = 1 + a\mathbf{r}' \cdot \mathbf{r}'' = 1 + \frac{a}{2}\frac{d(\mathbf{r}' \cdot \mathbf{r}')}{ds} = 1 + \frac{a}{2}\frac{d(1)}{ds} = 1 + 0 = 1$.
Jane: So?
Sam: So $1 = 0$.
Jane: Something must be wrong.

Is something wrong? If so, fix it.

Calculus is Everywhere # 21

How to Find Planets Around Stars

The approach described in this CIE is not the only way astronomers find planets. Another approach is to measure the dimming of a star when a planet passes in front of the star.

Astronomers have discovered that stars other than the sun have planets circling them. How do they do this, since the planets are too small to be seen? They combine some vector calculus with observations of the star.

Suppose the only objects in a solar system are a star S and a planet P in orbit around it. (If there were additional objects in this solar system, their influence would have to be included in the model.) To describe the situation, we are tempted to choose a coordinate system attached to the star. In that case the star would appear motionless, hence having no acceleration. However, the planet exerts a gravitational force on the star and the equation force = mass · acceleration would be violated. We get around this obstacle as follows.

Let $\mathbf{X}$ be the position vector of the planet P and $\mathbf{Y}$ be the position vector of the star S, relative to our standard coordinates with a fixed point, say Earth, as its origin. A key step in this discussion is the introduction of a second coordinate system with the property that bodies have no acceleration unless they are influenced by an external force. Such a coordinate system is called an **inertial system**.

Let M be the mass of the star and m the mass of planet P. Let $\mathbf{r} = \mathbf{X} - \mathbf{Y}$ be the vector from the star to the planet, as shown in Figure C.21.1.

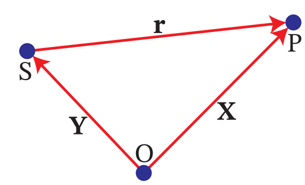


Figure C.21.1

The gravitational pull of the star on the planet is proportional to the product of their masses and the reciprocal of the square of the distance between them:

$$\mathbf{F} = \frac{-GmM\hat{\mathbf{r}}}{r^2} = \frac{-GmM\mathbf{r}}{r^3}.$$

where G is a constant. Equating the force with mass times acceleration, we have

$$m\mathbf{X}'' = \frac{-GmM\mathbf{r}}{r^3}.$$

Thus

$$\mathbf{X}'' = \frac{-GM\mathbf{r}}{r^3}.$$

By calculating the force that the planet exerts on the star, we have

$$\mathbf{Y}'' = \frac{Gm\mathbf{r}}{r^3}.$$

The center of mass of the system consisting of the planet and the star, is denoted as C (see Figure C.21.2):

$$\mathbf{C} = \frac{M\mathbf{Y} + m\mathbf{X}}{M+m}.$$

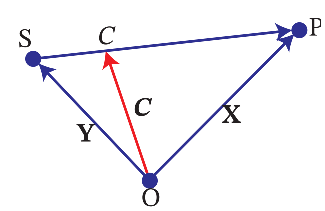


Figure C.21.2

The center of mass is closer to the star than to the planet. For our sun and Earth, the center of mass is 300 miles from the center of the sun.

The acceleration of the center of mass is

$$\mathbf{C}'' = \frac{M\mathbf{Y}'' + m\mathbf{X}''}{M+m} = \frac{1}{M+m}\left(M\left(\frac{Gm\mathbf{r}}{r^3}\right) + m\left(\frac{-GM\mathbf{r}}{r^3}\right)\right) = \mathbf{0}.$$

Because the center of mass has acceleration $\mathbf{0}$, it is moving at a constant velocity relative to the coordinate system. Therefore a coordinate system rigidly attached to the center of mass may also serve as a system in which the laws of physics still hold.

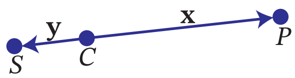


Figure C.21.3

We now describe the position of the star and planet in the new coordinate system. Star S has the vector $\mathbf{y}$ from C to it. Denote the vector from C to the star by $\mathbf{y}$ and the vector from C to the planet by $\mathbf{x}$, as shown in Figure C.21.3. Then the vector from S to P is $\mathbf{r} = \mathbf{x} - \mathbf{y}$.

To obtain a relation between $\mathbf{x}$ and $\mathbf{y}$, we express them in terms of $\mathbf{r}$. (See Figure C.21.3.) We have

$$\mathbf{y} = \mathbf{Y} - \mathbf{C} = \mathbf{Y} - \frac{M\mathbf{Y} + m\mathbf{X}}{M+m} = \frac{m}{M+m}\mathbf{Y} - \frac{m}{M+m}\mathbf{X}.$$

Letting $k = m/M$, a small quantity, we have

$$\mathbf{y} = \frac{k}{1+k}(\mathbf{Y} - \mathbf{X}) = \frac{-k}{1+k}\mathbf{r}. \tag{C.21.1}$$

Since $\mathbf{r} = \mathbf{x} - \mathbf{y}$, it follows that $\mathbf{x} = \mathbf{r} + \mathbf{y}$, hence

$$\mathbf{x} = \mathbf{r} + \left(\frac{-k}{1+k}\right)\mathbf{r} = \frac{1}{1+k}\mathbf{r}. \tag{C.21.2}$$

Combining (C.21.1) and (C.21.2) shows that

Theorem C.21.2: The Key Equation for Discovering Planets

In a planet-star system with center of mass at C, the vector $\mathbf{y}$ from C to the star (with mass M), and the vector $\mathbf{x}$ from C to the planet (with mass m) are related by

$$\mathbf{y} = -k\mathbf{x} \tag{C.21.3}$$

where $k = m/M$ is a small quantity.

Equation (C.21.3) contains a lot of information, including:

1. The star and planet remain on opposite sides of the center of mass, C, and on a straight line through C.
2. The star is always closer to C than the planet is.
3. The orbit of the star is similar in shape to the orbit of the planet, but smaller and reflected through C.
4. If the orbit of the star is periodic so is the orbit of the planet, with the same period.

Equation (C.21.3) is the key to the discovery of planets around stars. The astronomers look for a star that wobbles, which is the sign that it is in orbit around the center of mass of a planet-star couple. The time it takes for the planet to orbit the star is the time it takes for the star to oscillate back and forth once. The reference cited below shows that the star and the planet sweep out elliptical orbits in the second coordinate system (the one relative to C).

Astronomers have found hundreds of stars with planets, some with several planets. A registry of these **exoplanets** discovered through June 2018 can be found at `http://exoplanets.org/`.

REFERENCE: Osserman, R., Kepler's Laws, Newton's Laws, and the search for new planets, *American Mathematical Monthly* **108** (2001), pp. 813–820.

EXERCISES for CIE C.21

The mass of the sun is about 330,000 times that of Earth's mass. The closest Earth gets to the sun is about 91,341,000 miles, and the farthest is about 94,448,000 miles. The diameter of the sun is about 870,000 miles.

1. What are the closest and farthest the center of the sun gets to the center of mass of the sun-Earth system?

2. What would Earth's mass have to be for the center of mass of the sun-Earth system to lie outside the sun?

Calculus is Everywhere # 22

Newton's Law Implies Kepler's Three Laws

After hundreds of pages of computation based on observations by the astronomer Tycho Brahe (1546—1601) in the last thirty years of the sixteenth century, lengthy detours, and lucky guesses, Johannes Kepler (1571–1630) arrived at his three laws of planetary motion:

Theorem C.22.3: Kepler's Three Laws

1. *Every planet travels around the sun in an elliptical orbit such that the sun is situated at one focus (discovered in 1605, published in 1609).*
2. *The velocity of a planet varies so that the line joining the planet to the sun sweeps out equal areas in equal times (discovered 1602, published 1609).*
3. *The square of the time required by a planet for one revolution around the sun is proportional to the cube of its mean distance from the sun (discovered 1618, published 1619).*

Kepler's work shattered the crystal spheres that for 2,000 years had been thought to carry the planets. Before Kepler, astronomers used only circular motion and motion built of circular arcs. Copernicus (1473–1543) used five circles to describe Mars' motion. Ellipses were unwelcome. In 1605 Kepler wrote to a skeptical astronomer:

> You have disparaged my oval orbit . . . . If you are enraged because I cannot take away oval flight how much more you should be enraged by the motions assigned by the ancients, which I did take away . . . . You disdain my oval, a single cart of dung, while you endure the whole stable. (If indeed my oval is a cart of dung.)

The astronomical tables that Kepler based on his Laws, published in 1627, proved to be more accurate than any others, and ellipses gradually gained acceptance. Kepler's Laws stood as mysteries alongside a related question: If there are no crystal spheres, what propels the planets? Bullialdus (1605–1694), a French astronomer and mathematician, suggested in 1645:

> The force with which the sun seizes or pulls the planets, a physical force which serves as hands for it, is sent out in straight lines into all the world's space . . . ; since it is physical it is decreased in greater space; . . . the ratio of this distance is the same as that for light, namely as the reciprocal of the square of the distance.

In 1666, Hooke (1635–1703), more of an experimental scientist than a mathematician, wondered:

> why the planets should move about the sun . . . being not included in any solid orbs . . . nor tied to it . . . by any visible strings . . . . I cannot imagine any other likely cause besides these two: The first may be from an unequal density of the medium . . . ; if we suppose that part of the medium, which is farthest from the centre, or sun, to be more dense outward, than that which is more near, it will follow, that the direct motion will be always deflected inwards, by the easier yielding of the inwards . . . .
>
> But the second cause of inflecting a direct motion into a curve may be from an attractive property of the body placed in the center; whereby it continually endeavors to attract or draw it to itself. For if such a principle be supposed all the phenomena of the planets seem possible to be explained by the common principle of mechanic motions. . . . By this hypothesis, the phenomena of the comets as well as of the planets may be solved.

In 1675, Hooke, in an announcement to the Royal Society, went further:

> All celestial bodies have an attraction towards their own centres, whereby they attract not only their own parts but also other celestial bodies that are within the sphere of their activity .... All bodies that are put into direct simple motion will so continue to move forward in a single line till they are, by some other effectual powers, deflected and bent into a motion describing a circle, ellipse, or some other more compound curve .... These attractive powers are much more powerful in operating by how much the nearer the body wrought upon is to their own centers .... It is a notion which if fully prosecuted as it ought to be, will mightily assist the astronomer to reduce all the celestial motions to a certain rule ....

Trying to interest Newton in the question, Hooke wrote on November 24, 1679: "I shall take it as a great favor if ... you will let me know your thoughts of that of compounding the celestial motion of planets of a direct motion by the tangent and an attractive motion toward the central body." But four days later Newton replied:

> My affection to philosophy [science] being worn out, so that I am almost as little concerned about it as one tradesman used to be about another man's trade or a countryman about learning. I must acknowledge myself averse from spending that time in writing about it which I think I can spend otherwise more to my own content and the good of others ....

In a letter to Newton on January 17, 1680, Hooke returned to the problem of planetary motion:

> It now remains to know the properties of a curved line (not circular ...) made by a central attractive power which makes the velocities of descent from the tangent line or equal straight motion at all distances in a duplicate proportion to the distances reciprocally taken. I doubt not that by your excellent method you will easily find out what that curve must be, and its properties, and suggest a physical reason for this proportion.

Hooke succeeded in drawing Newton back to science, as Newton admitted in his *Philosophiae Naturalis Principia*, usually referred to as the *Principia*, published in 1687: "I am beholden to him only for the diversion he gave me from the other studies to think on these things and for his dogmaticalness in writing as if he had found the motion in the ellipse, which inclined me to try it."

It seems that Newton then obtained a proof, perhaps containing a mistake (the history is not clear), that if the motion is elliptical, the force varies as the inverse-square. In 1684, at the request of the astronomer Halley, Newton provided a correct proof. With Halley's encouragement (and funding), Newton spent the next year and a half writing the *Principia*.

In the *Principia*, which develops the science of mechanics and applies it to celestial motions, Newton began with two laws:

1. Every body continues in its state of rest, or of uniform motion in a straight line, unless it is compelled to change this state by forces impressed upon it.
2. The change of momentum is proportional to the motive force impressed, and is made in the direction of the straight line in which that force is impressed.

To state these in the language of vectors, let $\mathbf{v}$ be the velocity of the body, $\mathbf{F}$ the force, and m the mass of the body. The first law asserts that $\mathbf{v}$ is constant if $\mathbf{F}$ is $\mathbf{0}$. **Momentum** is $m\mathbf{v}$ so the second law asserts that

$$\mathbf{F} = \frac{d}{dt}(m\mathbf{v}).$$

If m is constant, this reduces to

$$\mathbf{F} = m\mathbf{a},$$

where $\mathbf{a}$ is the acceleration vector.

Newton assumed a universal **law of gravitation**, that a particle P exerts an attractive force on any other particle Q, and the direction of the force is from Q toward P. Then assuming that the orbit of a planet moving about the sun is an ellipse with one focus at the sun, he deduced that the force is inversely proportional to the square of the distance between the particles P and Q.

Nowhere in the *Principia* does he deduce from the inverse-square law of gravity that the planets' orbits are ellipses. (Though there are theorems in *Principia* on the basis of which this deduction could have been made.) In the *Principia* he showed that Kepler's second law (concerning areas) was equivalent to the assumption that the force acting on a planet is directed toward the sun. He also deduced Kepler's third law.

Newton's universal law of gravitation asserts that a particle, of mass M, exerts a force on another particle, of mass m, and that the magnitude of the force is proportional to the product of the masses, mM, inversely proportional to the square of the distance between them, and is directed toward the particle with the mass M.

Assume that the sun has mass M and is located at O and that the planet has mass m and is located at P. (See Figure C.22.1.) Let $\mathbf{r} = \overrightarrow{OP}$ and $r = |\mathbf{r}|$. Then the sun exerts a force $\mathbf{F}$ on the planet given by

$$\mathbf{F} = -\frac{GmM}{r^2}\widehat{\mathbf{r}}, \tag{C.22.1}$$

where G is a universal constant and $\widehat{\mathbf{r}} = \mathbf{r}/r$ is the unit vector that points in the direction of $\mathbf{r}$.

Now, $\mathbf{F} = m\mathbf{a}$, where $\mathbf{a}$ is the acceleration vector of the planet. Thus

$$m\mathbf{a} = -\frac{GmM}{r^2}\widehat{\mathbf{r}},$$

from which it follows that

$$\mathbf{a} = -\frac{q\widehat{\mathbf{r}}}{r^2}, \tag{C.22.2}$$

where $q = GM$ is independent of the planet.

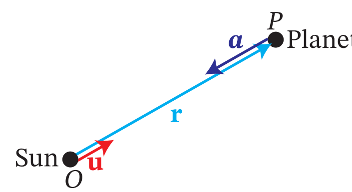


Figure C.22.1

The vectors $\widehat{\mathbf{r}}$, $\mathbf{r}$, and $\mathbf{a}$ are shown in Figure C.22.1.

The following exercises show how to obtain Kepler's laws from the single law of Newton, $\mathbf{a} = -q\widehat{\mathbf{r}}/r^2$.

EXERCISES for CIE C.22

Exercises 1 to 3 obtain Kepler's area law.

1. Let $\mathbf{r}(t)$ be the position vector of a planet at time t. Let $\Delta\mathbf{r} = \mathbf{r}(t+\Delta t) - \mathbf{r}(t)$. Show that for small Δt, $\frac{1}{2}|\mathbf{r}\times\Delta\mathbf{r}|$ approximates the area swept out by the position vector during the small interval of time Δt.

2. From Exercise 1 deduce that $\frac{1}{2}\left|\mathbf{r}\times\frac{d\mathbf{r}}{dt}\right|$ is the rate at which the position vector $\mathbf{r}$ sweeps out area. (See Figure C.22.2.)

3. Define $\mathbf{v} = \frac{d\mathbf{r}}{dt}$. With the aid of (C.22.2), show that $\frac{1}{2}\mathbf{r}\times\mathbf{v}$ is constant.

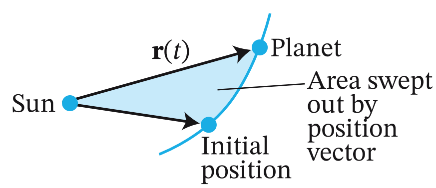


Figure C.22.2

Since $\mathbf{r}\times\mathbf{v}$ is constant, $\frac{1}{2}|\mathbf{r}\times\mathbf{v}|$ is constant. In view of Exercise 2, it follows that the radius vector of a given planet sweeps out area at a constant rate.

To put it another way, the radius vector sweeps out equal areas in equal times. This is Kepler's second law.

Introduce an xyz-coordinate system with origin at the center of mass of the sun and planet such that the unit vector $\mathbf{k}$, which points in the direction of the positive z-axis, has the same direction as the constant vector $\mathbf{r}\times\mathbf{v}$. Assume the center of mass of the sun and planet does not move over time. Thus there is a positive constant h such that

$$\mathbf{r}\times\mathbf{v} = h\mathbf{k}. \tag{C.22.3}$$

Exercises 4 to 13 obtain Kepler's ellipse law. In Exercises 7 to 10 we use the dot notation for differentiation with respect to time, so $\dot{\mathbf{r}} = \mathbf{v}$, $\dot{\mathbf{v}} = \mathbf{a}$, and $\dot{\theta} = \frac{d\theta}{dt}$.

4. Show that h in (C.22.3) is twice the rate at which the position vector of the planet sweeps out area.

5. Show that the planet remains in the plane perpendicular to $\mathbf{k}$ that passes through the sun.

By Exercise 5, and the assumption that the center of mass of the sun and the planet does not move, the orbit of the planet is planar. We may assume that it lies in the xy-plane; as above, we locate the origin of the xy-coordinates at the center of mass of the sun and the planet. Introduce polar coordinates in the plane, with the pole at the center of mass of the sun and the planet and the polar axis along the positive x-axis, as in Figure C.22.3.

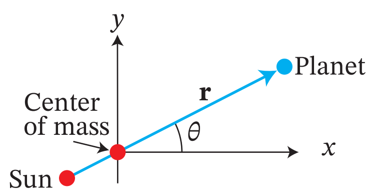


Figure C.22.3

6. (a) Show that during the time interval $[t_0, t]$ the position vector of the planet sweeps out the area $\frac{1}{2}\int_{t_0}^{t} r^2 \frac{d\theta}{dt}\,dt$.

(b) Deduce that the radius vector sweeps out area at the rate $\frac{1}{2}r^2\frac{d\theta}{dt}$.

---

7. Show that $\mathbf{r} \times \mathbf{v} = r^2\dot{\theta}\mathbf{k}$.

8. Show that $\dot{\hat{r}} = \frac{d\hat{\mathbf{r}}}{d\theta}\dot{\theta}$ and is perpendicular to $\hat{\mathbf{r}}$. (Recall that $\hat{\mathbf{r}}$ is defined as $\mathbf{r}/|\mathbf{r}|$.)

9. From $\mathbf{r} = r\hat{\mathbf{r}}$, show that $h\mathbf{k} = r^2(\hat{\mathbf{r}} \times \dot{\hat{r}})$.

10. Using (C.22.2) and Exercise 9, show that $\mathbf{a} \times h\mathbf{k} = q\dot{\hat{r}}$.

11. Deduce from Exercise 10 that $\mathbf{v} \times h\mathbf{k}$ and $q\hat{\mathbf{r}}$ differ by a constant vector.

By Exercise 11, there is a constant vector $\mathbf{C}$ such that $\mathbf{v} \times h\mathbf{k} = q\hat{\mathbf{r}} + \mathbf{C}$. Moreover, the coordinate system can be chosen so that $\mathbf{C}$ points in the direction of the positive x-axis. Then θ will be the counterclockwise angle from $\mathbf{C}$ to $\mathbf{r}$.

12. (a) Show that $(\mathbf{r} \times \mathbf{v}) \cdot h\mathbf{k} = h^2$.

(b) Show that $\mathbf{r} \cdot (\mathbf{v} \times h\mathbf{k}) = rq + \mathbf{r} \cdot \mathbf{C}$.

(c) Combining (a) and (b), deduce that $h^2 = rq + rc\cos(\theta)$, where $c = |\mathbf{C}|$.

It follows from Exercise 12 that the polar equation for the orbit of the planet is given by

$$r(\theta) = \frac{h^2}{q + c\cos(\theta)}. \tag{C.22.4}$$

An equation of the form $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$, where not all coefficients are 0, describes a **conic section**. Its graph may be an ellipse, hyperbola, parabola, or a straight line. Special cases, such as $x^2 + y^2 + 1 = 0$ or $x^2 + y^2 = 0$ describe "degenerate cases" such as the empty set or a single point.

13. By expressing (C.22.4) in rectangular coordinates, show that (C.22.4) describes a conic section.

NOTE: Recall that the general form for a conic section is $Ax^2 + Bxy + Cy^2 + Dx + Ey = F$.

Since the orbit of a planet is bounded and is also a conic section, it must be an ellipse. This establishes Kepler's first law.

Kepler's third law asserts that the square of the time required for a planet to complete one orbit is proportional to the cube of its mean distance from the sun.

For Kepler, **mean distance** meant the average of the shortest distance and the longest distance from the planet to the sun. Let us compute the average for the ellipse with semimajor axis a and semiminor axes b, shown in Figure C.22.4. The sun is at the focus F, which is also the pole of the polar coordinate system we are using. The line through the foci contains the polar axis.

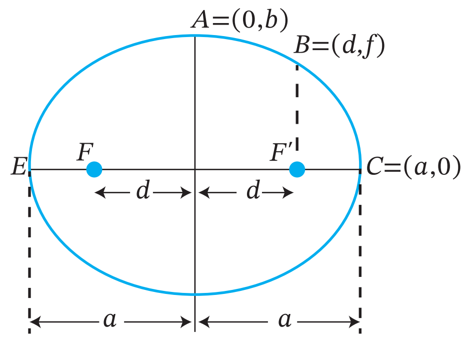


Figure C.22.4

An ellipse is the set of points P such that the sum of the distances from P to the foci F and F' is constant, $2a$. The shortest distance from the planet to the sun is $|\overrightarrow{FE}| = a - d$ and the longest distance is $|\overrightarrow{FC}| = a + d$. Thus Kepler's mean distance is

$$\frac{(a-d)+(a+d)}{2} = a.$$

Let T be the time required by the given planet to complete one orbit. Kepler's third law asserts that T^2 is proportional to a^3. Exercises 14 to 18 establish this by showing that T^2/a^3 is the same for all planets.

14. Using the fact that the area of the ellipse in Figure C.22.4 is πab, show that $Th/2 = \pi ab$, hence that

$$T = \frac{2\pi ab}{h}. \qquad \text{(C.22.5)}$$

The rest of the argument depends only on (C.22.4) and (C.22.5) and the fixed sum of two distances property of an ellipse.

15. Using (C.22.4), show that f in Figure C.22.4 equals $\dfrac{h^2}{q}$.

16. Show that $b^2 = af$:

(a) From $|\overrightarrow{F'A}| + |\overrightarrow{FA}| = 2a$, deduce that $a^2 = b^2 + d^2$.

(b) From $|\overrightarrow{F'B}| + |\overrightarrow{FB}| = 2a$, deduce that $d^2 = a^2 - af$.

(c) From (a) and (b), deduce that $b^2 = af$.

17. From Exercises 15 and 16, deduce that $b^2 = \dfrac{ah^2}{q}$.

18. Combining (C.22.5) and Exercise 17, show that $\dfrac{T^2}{a^3} = \dfrac{4\pi^2}{q}$.

Since $\dfrac{4\pi^2}{q}$ is a constant, the same for all planets, Kepler's third law is established.

Chapter 16

Partial Derivatives

So far we have been concerned mainly with functions whose domains are part or all of a line or curve. This chapter generalizes the derivative to functions whose domains are part or all of a plane or space, called functions of two or three variables. Chapter 17 does the same for definite integrals.

The first seven sections generalize Chapters 1 to 4: picturing functions of two or three variables (Section 16.1), their derivatives (Section 16.2), the chain rule (Section 16.3), more on their derivatives (Section 16.4), the tangent plane to a surface (Section 16.5), and finding extrema (Sections 16.6 and 16.7). In preparation for extending the method of substitution for evaluating definite integrals, the magnification of a function is introduced in Section 16.8. Section 16.9 applies ideas learned earlier in this chapter to obtain some fundamental equations in introductory thermodynamics.

16.1 Picturing a Function of Several Variables

The graph of $y = f(x)$, a function of one variable, x, is a curve in the xy-plane. The graph of a function of two variables, $z = f(x, y)$ is a surface in space. It consists of the points (x, y, z) for which $z = f(x, y)$. For instance, if $z = 2x + 3y$, the graph is the plane $z = 2x + 3y$.

This section describes some ways of picturing a scalar-valued function of two or three variables.

Contour Lines

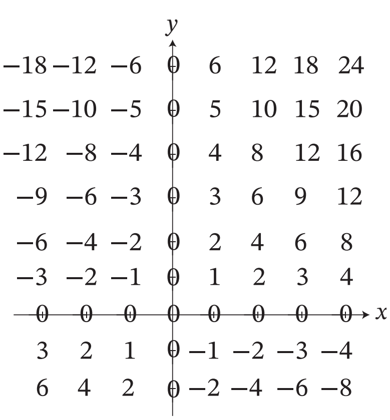


Figure 16.1.1

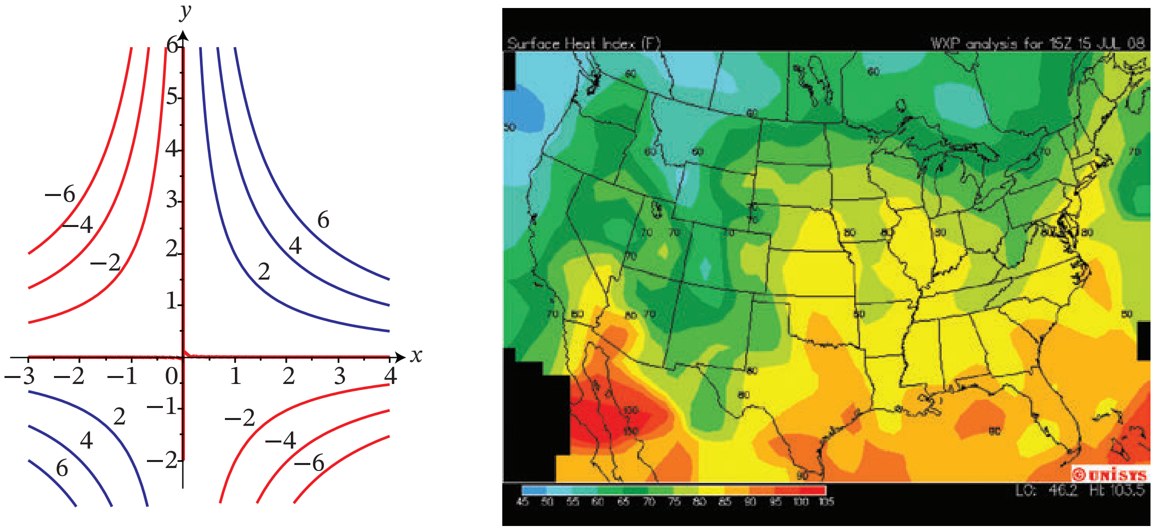
For $z = f(x, y)$, the simplest method is to attach at (x, y) the value of the function there. Figure 16.1.1 illustrates this for $z = xy$. It gives a sense of the function. Its values are positive in the first and third quadrants, negative in the second and fourth. For (x, y) far from the origin near the lines $y = x$ or $y = -x$ the values are large.

Rather than attach the values at points, we could indicate points where the function has a specific fixed value. We could graph for a constant k all the points (x, y) where $f(x, y) = k$, called a **contour** or **level curve**.

For $z = xy$, contours are hyperbolas $xy = k$. In Figure 16.1.2(a) the contours corresponding to $k = 2, 4, 6, 0, -2, -4, -6$ are shown.

Many newspapers publish a daily map showing temperatures using contour lines. Figure 16.1.2(b) is an example; in this case the contour lines are the boundary curves between the differently colored regions.

At a glance you can see where it is hot or cold and in what direction to travel to warm up or cool off.



(a) (b)

Figure 16.1.2

Traces

Another way to see the surface $z = f(x, y)$ is to sketch the intersections of various planes with the surface. They are cross sections that are called **traces**.

Figure 16.1.3 shows a trace created by the plane $z = k$, which is parallel to the xy-coordinate plane. The curve is a copy of the contour $f(x, y) = k$.

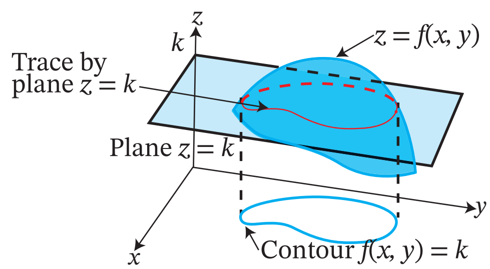


Figure 16.1.3

EXAMPLE 1. Sketch the traces of the surface $z = xy$ with the planes (a) $z = 1$, (b) $x = 1$, (c) $y = x$, (d) $y = -x$, and (e) $x = 0$.

SOLUTION

The surface $z = xy$ can be viewed as made up of lines, of parabolas, or of hyperbolas.

(a) The trace with the plane $z = 1$ is shown in Figure 16.1.4(a). For points (x, y, z) on this trace, $xy = 1$, a hyperbola. It is the contour line $xy = 1$ in the xy-plane raised by one unit, as shown in Figure 16.1.4(a)

(b) The trace in the plane $x = 1$ satisfies the equation $z = 1 \cdot y = y$. It is a straight line, shown in Figure 16.1.4(b)

(c) The trace in the plane $y = x$ satisfies the equation $z = x^2$. It is the parabola shown in Figure 16.1.4(c).

(d) The trace in the plane $y = -x$ satisfies the equation $z = x(-x) = -x^2$. It is an upside-down parabola, shown in Figure 16.1.4(d).

(e) The intersection with the coordinate plane $x = 0$ satisfies the equation $z = 0 \cdot y = 0$. This is the y-axis, shown in Figure 16.1.4(e). ♦

Notice that the surface $z = xy$ looks like a saddle or the pass between two hills, as shown in Figure 16.1.5.

Functions of Three Variables

The graph of $y = f(x)$ consists of points in the xy-plane. The graph of $z = f(x, y)$ consists of points in xyz-space. What if we have a function of three variables, $u = f(x, y, z)$? (For example, the volume of a box with sides x, y, and z can be modeled with the function $f(x, y, z) = xyz$.) We cannot graph the set of points (x, y, z, u) where $u = f(x, y, z)$ since we cannot draw graphs in four dimensions. What we could do is pick a constant k and draw a level surface,

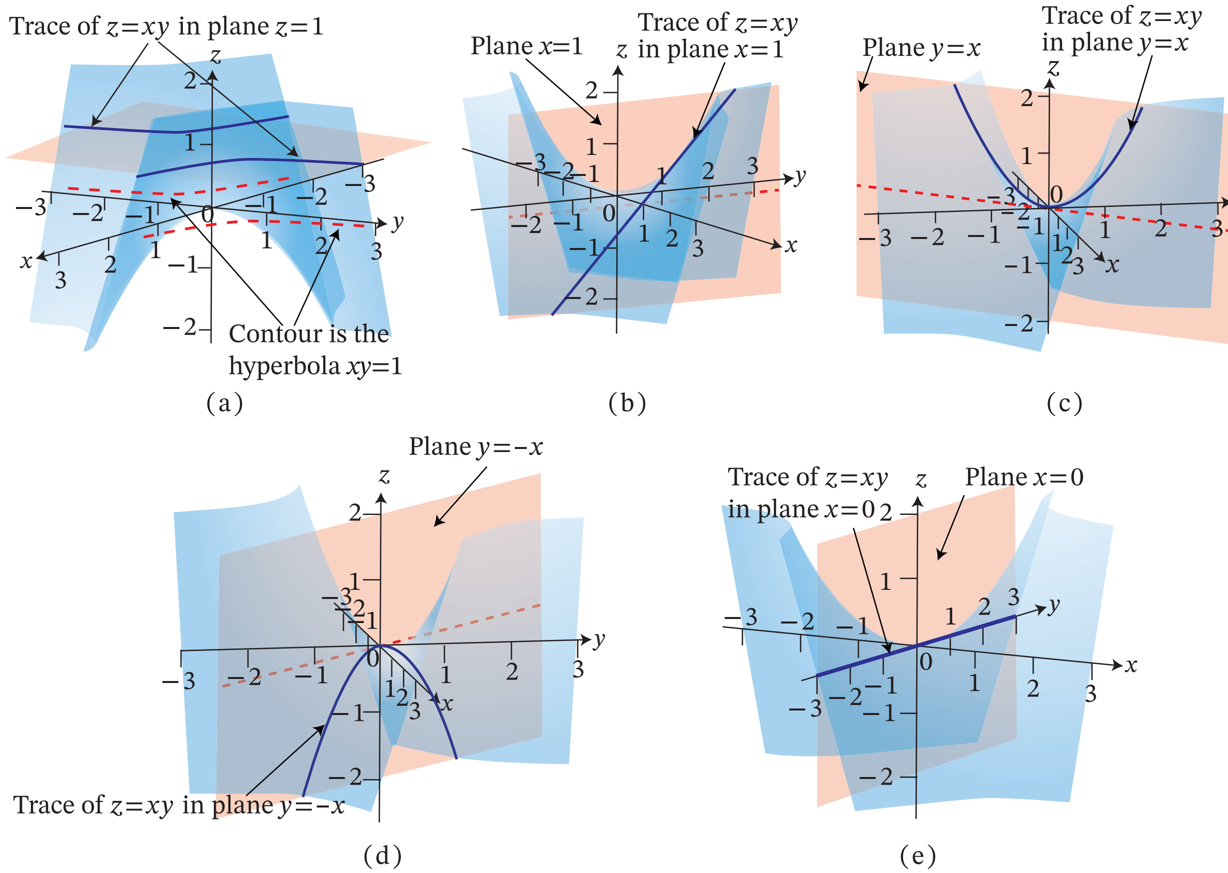


Figure 16.1.4

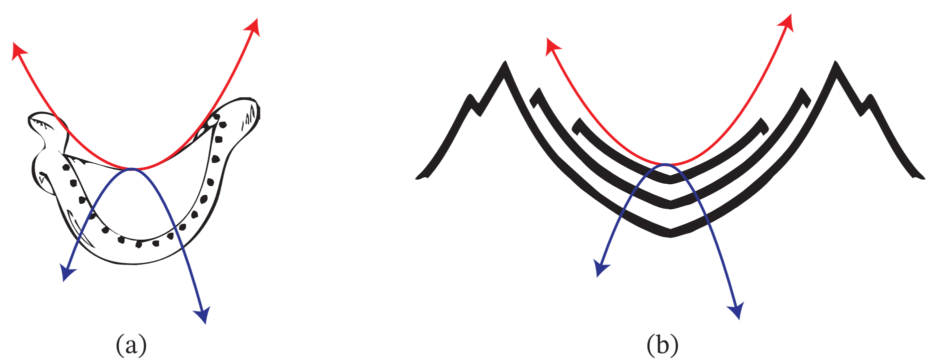


Figure 16.1.5

the set of points where $f(x,y,z)=k$. Varying k may give an idea of the function's behavior, just as varying k in $f(x,y)=k$ yields information about the behavior of a function of two variables.

For example, let $T=f(x,y,z)$ be the temperature (in degrees Fahrenheit) at the point (x,y,z). Then the level surface $68=f(x,y,z)$ consists of all points in space where the temperature is 68°F.

EXAMPLE 2. Describe the level surfaces of $f(x,y,z)=x^2+y^2+z^2$.

SOLUTION For each k we examine $f(x,y,z)=x^2+y^2+z^2=k$. If k is negative, there are no points on the level surface. If $k=0$, there is only one point, the origin $(0,0,0)$. If $k=1$, the equation is $x^2+y^2+z^2=1$, which describes a sphere of radius 1, with center $(0,0,0)$. If k is positive, the level surface $f(x,y,z)=k$ is a sphere of radius $\sqrt{k}$, with center $(0,0,0)$. ◆

Summary

We introduced the idea of a function of two variables $z=f(P)$ defined for points P in a region in the xy-plane. The graph of $z=f(P)$ is usually a surface. It is often more useful to sketch a few of its level curves than to sketch the surface. A level curve in the xy-plane is a copy of a trace by a plane parallel to that plane. At all points (x,y) on a level curve the function has the same value, so it is constant on a level curve.

For functions of three variables $u=f(x,y,z)$, which could also be written as $u=f(P)$, we defined level surfaces $k=f(x,y,z)$ on which f is constant.

Historical Note: History of Contours

The use of contour lines goes back to 1774. Surveyors had collected a large number of the elevations of points on Mount Schiehalli in Scotland in order to estimate its mass and, by its gravitational attraction, the mass of Earth. They asked the mathematician Charles Hutton for help in using the data. Hutton saw that if he connected points on the map that showed the same elevation, the resulting curves, contour lines, suggested the shape of the mountain.

REFERENCE: Bill Bryson, *A Short History of Nearly Everything*, Broadway Books, New York, 2003, p. 57.

EXERCISES for Section 16.1

In Exercises 1 to 10, graph the given function. That is, graph $z=f(x,y)$.

1. $f(x,y)=y$
2. $f(x,y)=x+1$
3. $f(x,y)=3$
4. $f(x,y)=-2$
5. $f(x,y)=x^2$
6. $f(x,y)=y^2$
7. $f(x,y)=x+y+1$
8. $f(x,y)=2x-y+1$
9. $f(x,y)=x^2+2y^2$
10. $f(x,y)=\sqrt{x^2+y^2}$

In Exercises 11 to 14 draw the level curves corresponding to the values −1, 0, 1, and 2 if they are not empty.

11. $f(x,y)=x+y$
12. $f(x,y)=x+2y$
13. $f(x,y)=x^2+2y^2$
14. $f(x,y)=x^2-2y^2$

In Exercises 15 to 18 draw the level curves that pass through the given points.

15. $f(x,y)=x^2+y^2$ through $(1,1)$
16. $f(x,y)=x^2+3y^2$ through $(1,2)$
17. $f(x,y)=x^2-y^2$ through $(3,2)$
18. $f(x,y)=x^2-y^2$ through $(2,3)$

19. (a) Draw the level curves for $f(x, y) = x^2 + y^2$ corresponding to $k = 0, 1, \dots, 9$.
(b) By inspection of the curves in (a), decide where the function changes most rapidly. Explain your answer.

In Exercises 20 to 27, sketch three level curves in the $r\theta$-plane for the corresponding function.

20. $f(r,\theta) = r$
21. $f(r,\theta) = r^2$
22. $f(r,\theta) = r\cos(\theta)$
23. $f(r,\theta) = r\sin(\theta)$
24. $f(r,\theta) = e^{-r}$
25. $f(r,\theta) = \ln(r)$
26. $f(r,\theta) = \dfrac{1}{r}$
27. $f(r,\theta) = \sin(r)$

28. Compare and contrast the level curves drawn in Exercises 20 to 27. How are the level curves similar? How are they different?

29. Let $u = g(x, y, z)$ be a function of three variables. Describe the level surface $g(x, y, z) = 1$ when
(a) $g(x, y, z) = x + y + z$ (b) $g(x, y, z) = x^2 + y^2 + z^2$ (c) $g(x, y, z) = x^2 + y^2 - z^2$ (d) $g(x, y, z) = x^2 - y^2 - z^2$

30. A weather map, Figure 16.1.6, shows level curves of constant barometric pressure (called **isobars**).
(a) Where is the lowest pressure? (b) Where is the highest pressure? (c) Where do you think the wind at ground level is the fastest? Why?

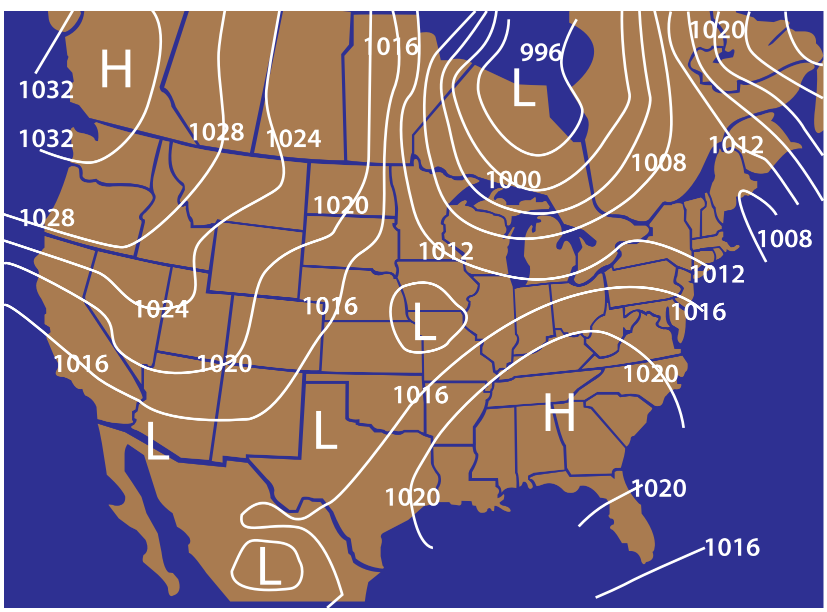


Figure 16.1.6

31. (a) Sketch the surface $z = x^2 + y^2$.
(b) Show that traces by planes parallel to the xz-plane are parabolas.
(c) Show that the parabolas in (b) are congruent. Thus, the surface is made up of identical parabolas.
(d) What kind of curve is a trace in a plane parallel to the xy-plane?
(e) Are the traces identified in (d) congruent?

32. For the surface $z = x^2 + 4y^2$, what type of curve is produced by a trace by a plane parallel to
(a) the xy-plane? (b) the xz-plane? (c) the yz-plane?

33. Let $f(P)$ be the average daily solar radiation at point P, measured in langleys. The level curves corresponding to 350, 400, 450, and 500 langleys are shown in Figure 16.1.7.

(a) What can be said about the ratio between the maximum and minimum solar radiation at points in the United States?

(b) Why are there sharp bends in the level curves in two areas?

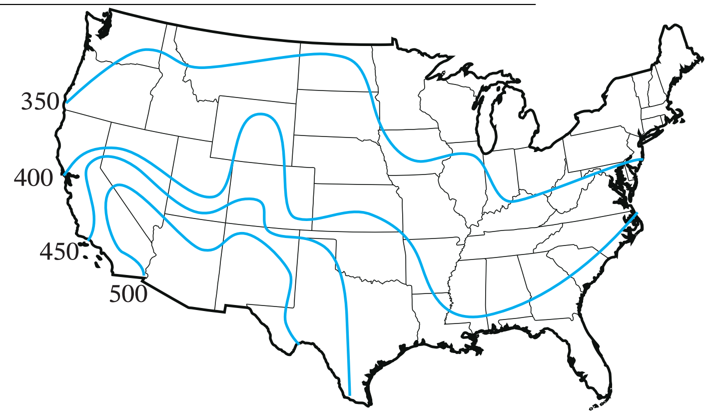


Figure 16.1.7

Historical Note: Samuel Langley

The **langley** (Ly) is a (non-SI) unit of heat transmission commonly used to express the rate of solar radiation received by the earth.

1 Ly = 1 thermochemical calorie per square centimeter = 41,840 J/m^2(joules per square meter).

Franz Linke first introduced the Langley in 1947 as a way to honor Samuel Langley, an American physicist, astronomer, and aviation pioneer. He might be most well-known as the person who almost beat the Wright Brothers to have the first manned powered and controllable flight. Samuel Langley is also the namesake of Langley AFB and NASA's Langley Research Center, both in Hampton, VA, and the neighborhood in McLean, VA where the CIA's headquarters is located.

16.2 Limits, Continuity, and Partial Derivatives

The concepts of limit, continuity, and derivative carry over with similar definitions from functions of one variable to functions of several variables. But there are differences. A differentiable function $f(x)$ has one first derivative, $f'(x)$, but $f(x, y)$ has many first derivatives.

Limits and Continuity of $f(x, y)$

The **domain** of a function $f(x, y)$ is the set of points where it is defined. The domain of $f(x, y) = x + y$ is the entire xy-plane. The domain of $f(x, y) = \sqrt{1 - x^2 - y^2}$ is smaller because for the square root of $1 - x^2 - y^2$ to be defined $1 - x^2 - y^2$ must not be negative. So, $x^2 + y^2 \leq 1$. The domain is the disk bounded by the circle $x^2 + y^2 = 1$, shown in Figure 16.2.1(a).

The domain of $g(x, y) = 1/\sqrt{1 - x^2 - y^2}$ is even smaller. Now $1 - x^2 - y^2$ cannot be 0 or negative. The domain of $1/\sqrt{1 - x^2 - y^2}$ consists of points (x, y) such that $x^2 + y^2 < 1$. It is the (cyan) disk in Figure 16.2.1 (a) without its (blue) boundary.

The function $h(x, y) = 1/(y - x)$ is defined everywhere except on the line $y - x = 0$. Its domain is the xy-plane from which the line $y = x$ is removed. (See Figure 16.2.1(b).)

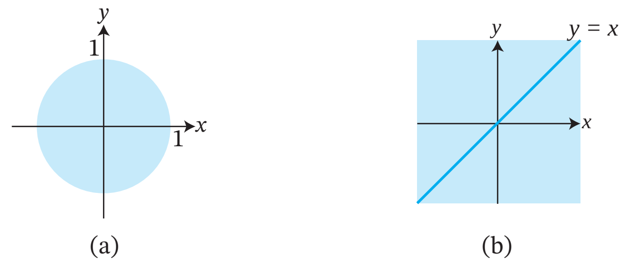


Figure 16.2.1

The domains of functions we look at will either be the entire xy-plane or a region bordered by curves or lines, or perhaps one with a few points (or curves) omitted.

Let P_0 be a point in the domain of a function f. If there is a disk with center P_0 that lies within the domain of f, we call P_0 an **interior** point of the domain. (See Figure 16.2.2(a).) When P_0 is an interior point of the domain of f, we know that $f(P)$ is defined for all points P sufficiently near P_0. A set R is called **open** if each point P of R is an interior point of R. Any disk without its circumference is open. The set of points inside a closed curve but not on it forms an open set.

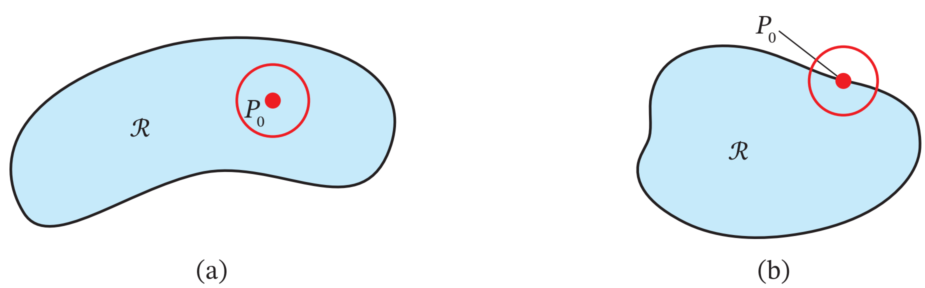


Figure 16.2.2

A point P_0 is on the **boundary** of a set if every disk centered at P_0, no matter how small, contains points in the set and points not in the set. (See Figure 16.2.2(b).) The boundary of the disk $x^2 + y^2 \le 1$ is the circle $x^2 + y^2 = 1$.

A set R is called **closed** if it includes all of its boundary points. Any disk with its circumference is closed. The set of points inside a closed curve, including the boundary curve, forms a closed set.

The entire xy-plane is both closed and open. The empty set is also both closed and open. These are the only sets in the plane with this property. All other sets in the plane are either open or closed or neither.

The domains of the functions f, g, and h defined at the beginning of this section are closed, open, and open, respectively. Returning to the four functions defined at the beginning of this section: the domain of $f(x, y) = x + y$ is the entire xy-plane, which is both open and closed, the domain of $f(x, y) = \sqrt{1 - x^2 - y^2}$ is closed, and the domains of $g(x, y) = 1/\sqrt{1 - x^2 - y^2}$ and $h(x, y) = 1/(y - x)$ are both open.

The definition of the limit of $f(x, y)$ as (x, y) approaches $P_0 = (a, b)$ should not come as a surprise.

Definition: *Limit of* $f(x, y)$ *at* $P_0 = (a, b)$

Let f be a function defined at least at every point in some disk with center $P_0 = (a, b)$, except perhaps at P_0. If there is a number L such that $f(P)$ approaches L whenever P approaches P_0 (through points in the domain of f) we call L the **limit of** $f(P)$ **as** P **approaches** P_0. We write

$$\lim_{P \to P_0} f(P) = L \qquad \text{or} \qquad f(P) \to L \text{ as } P \to P_0 \qquad \text{or} \qquad \lim_{(x,y)\to(a,b)} f(x, y) = L.$$

EXAMPLE 1. Define $f(x,y) = \dfrac{x^3}{x^2+y^2}$. Determine whether $\lim_{(x,y)\to(0,0)} f(x,y)$ exists.

SOLUTION The domain of f is the xy-plane without the origin. When P is near $(0,0)$ both numerator and denominator approach 0, so we have an indeterminate limit.

Because of the presence of x^2+y^2, we introduce polar coordinates, replacing x^2+y^2 by r^2 and x^3 by $r^3\cos^3(\theta)$. The quotient now reads

$$\frac{r^3\cos^3(\theta)}{r^2} = r\cos^3(\theta).$$

Because $r\cos^3(\theta)$ approaches 0 as $r\to 0$, we conclude that $\lim_{P\to(0,0)} f(P) = 0$. ♦

EXAMPLE 2. Let $g(x,y) = \dfrac{x^2-y^2}{x^2+y^2}$. Does $\lim\limits_{(x,y)\to(0,0)} g(x,y)$ exist?

SOLUTION The function is not defined at $(0,0)$. When (x,y) is near $(0,0)$, both the numerator and denominator of $(x^2-y^2)/(x^2+y^2)$ are small numbers. As in Example 1 there are two opposing influences.

We try a few inputs near $(0,0)$. For instance, $(0.01,0)$ is near $(0,0)$ and

$$g(0.01,0) = \frac{(0.01)^2-0^2}{(0.01)^2+0^2} = 1.$$

Also, $(0,0.01)$ is near $(0,0)$ and

$$g(0,0.01) = \frac{0^2-(0.01)^2}{0^2+(0.01)^2} = -1.$$

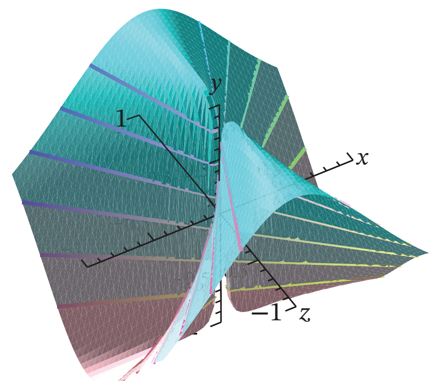


Figure 16.2.3

More generally, for $x\neq 0$, $g(x,0)=1$ while for $y\neq 0$, $g(0,y)=-1$. Since x can be as near 0 as we please and y can be as near 0 as we please, $\lim_{P\to(0,0)} g(P)$ does not exist. Figure 16.2.3 shows one view of the graph of $z=g(x,y)$. ♦

NOTE: The true nature of the graph in Figure 16.2.3 is difficult to show in a single figure. This is due, primarily, to the discontinuity at $(0,0)$. Notice the ridge at height $z=1$ along the y-axis - except for $y=0$ and the valley at height $z=-1$ along the x-axis - except for $x=0$.

If P_0 is not an interior point of the domain of a function f, we modify the definition of limit slightly. Let P_0 be a point on the boundary of the domain of f. If $f(P)\to L$ as P approaches P_0 *through points in the domain of* f, we say that "L is the limit of $f(P)$ as $P\to P_0$." Example 2 is such a case.

Continuity of $f(x,y)$ at $P_0=(a,b)$

The definition of continuity for $f(x)$ in Section 2.4 generalizes to the definition of continuity for $f(x,y)$.

Definition: *Continuity of $f(x,y)$ at $P_0=(a,b)$*

Assume that $f(P)$ is defined throughout some disk with center P_0. Then f is **continuous** at P_0 if and only if $\lim\limits_{P\to P_0} f(P) = f(P_0)$. This means all three of the following statements must be true:

1. $f(P_0)$ is defined (that is, P_0 is in the domain of f),
2. $\lim\limits_{P\to P_0} f(P)$ exists, and
3. $\lim\limits_{P\to P_0} f(P) = f(P_0)$.

Continuity at a point on the boundary of the domain can be defined similarly.

A function $f(P)$ is **continuous** if it is continuous at every point in its domain.

EXAMPLE 3. Determine whether $g(x,y) = \dfrac{x^2-y^2}{x^2+y^2}$ is continuous at $(1,1)$.

SOLUTION This is the function in Example 2. First, $g(1,1)$ is defined and equals 0. Second, $\lim_{(x,y)\to(1,1)}(x^2-y^2)/(x^2+y^2)$ exists and is $0/2=0$. Third, $\lim_{(x,y)\to(1,1)} g(x,y)$ equals $g(1,1)$. Hence, $g(x,y)$ is continuous at $(1,1)$. ◆

In fact, the function of Examples 2 and 3 is continuous at every point (x,y) in its domain. We do not need to worry about the behavior of $g(x,y)$ when (x,y) is near $(0,0)$ because $(0,0)$ is not in the domain. Since $g(x,y)$ is continuous at every point in its domain, it is a continuous function.

The Two Partial Derivatives of $f(x,y)$

Let (a,b) be a point in the domain of $f(x,y)$. The trace on the surface $z=f(x,y)$ by a plane through (a,b) and parallel to the z-axis is a curve, as shown in Figure 16.2.4(a).

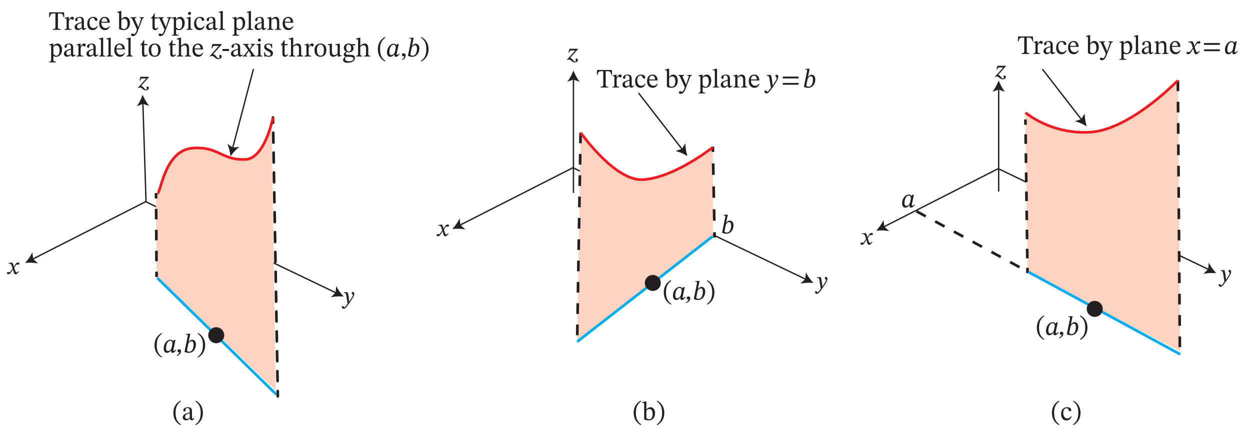


Figure 16.2.4

If f is well behaved then at the point $P=(a,b,f(a,b))$ the trace has a slope. It depends on the plane through (a,b). In this section we consider only the planes parallel to the coordinate planes $y=0$ and $x=0$, shown in Figures 16.2.4(b) and (c). In the next section we treat the general case.

For $f(x,y)=x^2y^3$, if we hold y constant and differentiate with respect to x, we obtain $2xy^3$. This will be called the partial derivative of x^2y^3 with respect to x. We could hold x fixed instead and find the derivative of x^2y^3 with respect to y, that is, $3x^2y^2$. This derivative will be called the partial derivative of x^2y^3 with respect to y. Now we give a precise definition of partial derivatives and see what they mean in terms of slope and rate of change.

Definition: *Partial derivatives.*

Assume that the domain of $f(x,y)$ includes the interior of some disk with center (a,b).

If
$$\lim_{\Delta x\to 0}\frac{f(a+\Delta x,b)-f(a,b)}{\Delta x}$$
exists, it is called the **partial derivative of** f **with respect to** x **at** (a,b).

Similarly, if
$$\lim_{\Delta y\to 0}\frac{f(a,b+\Delta y)-f(a,b)}{\Delta y}$$
exists, it is called the **partial derivative of** f **with respect to** y **at** (a,b).

There are several notations for the partial derivatives of $z=f(x,y)$ with respect to x:
$$\frac{\partial z}{\partial x}, \frac{\partial f}{\partial x}, f_x, f_1, D_x f, \text{ or } z_x$$

and with respect to y:

$$\frac{\partial z}{\partial y}, \frac{\partial f}{\partial y}, f_y, f_2, D_y f, \text{ or } z_y.$$

The symbol $\partial f/\partial x$ may be viewed as the rate at which the function $f(x,y)$ changes when x varies and y is kept fixed; $\partial f/\partial y$ records the rate at which the function $f(x,y)$ changes when y varies and x is kept fixed.

The value of $\partial f/\partial x$ at (a,b) is denoted

$$\frac{\partial f}{\partial x}(a,b) \qquad \text{or} \qquad \left.\frac{\partial f}{\partial x}\right|_{(a,b)} \qquad \text{or} \qquad D_x f(a,b)$$

In the middle of a sentence, we will sometimes write it as $f_x(a,b)$ or $\partial f/\partial x(a,b)$.

The partial derivative $\frac{\partial f}{\partial y}(a,b)$ is the slope of the tangent line to the trace $z = f(a,y)$ at the point (a,b).

EXAMPLE 4. If $f(x,y) = \sin(x^2 y)$, find (a) $\dfrac{\partial f}{\partial x}$, (b) $\dfrac{\partial f}{\partial y}$, and (c) $\dfrac{\partial f}{\partial y}$ at $\left(1, \dfrac{\pi}{4}\right)$.

SOLUTION

(a) To find $D_x(\sin(x^2 y))$, keep y constant and differentiate with respect to x:

$$\begin{aligned} \frac{\partial}{\partial x}\sin(x^2 y) &= \cos(x^2 y)\frac{\partial}{\partial x}(x^2 y) && \text{(chain rule)} \\ &= \cos(x^2 y)(2xy) && \text{(}y\text{ is constant)} \\ &= 2xy\cos(x^2 y). \end{aligned}$$

(b) To find $D_y(\sin(x^2 y))$, keep x constant and differentiate with respect to y:

$$\begin{aligned} \frac{\partial}{\partial y}\sin(x^2 y) &= \cos(x^2 y)\frac{\partial}{\partial y}(x^2 y) && \text{(chain rule)} \\ &= \cos(x^2 y)(x^2) && \text{(}x\text{ is constant)} \\ &= x^2\cos(x^2 y). \end{aligned}$$

(c) From (b), $f_y(1,\pi/4) = 1^2\cos\left(1^2 \cdot \pi/4\right) = \sqrt{2}/2$. ◆

As Example 4 shows, because partial derivatives are really ordinary derivatives, the procedures for computing derivatives of a function $f(x)$ of a single variable carry over to functions of two variables.

Higher-Order Partial Derivatives

Just as there are derivatives of derivatives there are partial derivatives of partial derivatives. For instance, when $z = 2x + 5x^4 y^7$, we have

$$\frac{\partial z}{\partial x} = 2 + 20x^3 y^7 \qquad \text{and} \qquad \frac{\partial z}{\partial y} = 35x^4 y^6.$$

Then we may compute partial derivatives of $\partial z/\partial x$ and $\partial z/\partial y$:

$$\begin{aligned} \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right) &= 60x^2 y^7 & \qquad \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right) &= 140x^3 y^6 \\ \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) &= 140x^3 y^6 & \qquad \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right) &= 210x^4 y^5. \end{aligned}$$

There are four partial derivatives of the second order,

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right), \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right), \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right), \text{ and } \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right).$$

These can also be denoted, in the same order, as

$$\frac{\partial^2 z}{\partial x^2}, \frac{\partial^2 z}{\partial y \partial x}, \frac{\partial^2 z}{\partial y^2}, \text{ and } \frac{\partial^2 z}{\partial x \partial y}.$$

To compute $\frac{\partial^2 z}{\partial x \partial y}$, we first differentiate with respect to y, then with respect to x. And, to compute $\frac{\partial^2 z}{\partial y \partial x}$, first differentiate with respect to x, then with respect to y. In both cases, we differentiate from right to left in the order that the variables occur in the "denominator".

Recall that the partial derivative $\frac{\partial f}{\partial x}$ is also denoted f_x and $\frac{\partial f}{\partial y}$ is denoted f_y. Similarly, the second-order partial derivative $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial (f_y)}{\partial x} = (f_y)_x$ is denoted f_{yx}. When using subscript notation, we differentiate from left to right, first f_y, then $(f_y)_x$. That is, $f_{yx} = (f_y)_x$, $f_{yy} = (f_y)_y$, and $f_{xy} = (f_x)_y$. In both notations the mixed partial is computed in the order that resembles its definition with the parentheses removed:

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) \quad \text{and} \quad f_{xy} = (f_x)_y$$

The subscript notation, f_{yx}, is generally preferred in the middle of a sentence.

are the two mixed second-order partial derivatives of f.

In the computations just done, the two mixed partials z_{xy} and z_{yx} are equal. This is not a coincidence. For functions commonly encountered, the two mixed partials are equal. This equality is used in Section 16.8 and in Chapter 18.

Exercise 56 has a function for which the two mixed particles are not equal.

EXAMPLE 5. Compute $\frac{\partial^2 z}{\partial x^2} = z_{xx}$, $\frac{\partial^2 z}{\partial y \partial x} = z_{xy}$, $\frac{\partial^2 z}{\partial x \partial y} = z_{yx}$, and $\frac{\partial^2 z}{\partial y^2} = z_{yy}$ for $z = y\cos(xy)$.

SOLUTION The first partial derivatives are

$$\frac{\partial z}{\partial x} = y\left(-\sin(xy)\frac{\partial}{\partial x}(xy)\right) = y(-\sin(xy)y) = -y^2\sin(xy)$$

and

$$\frac{\partial z}{\partial y} = \frac{\partial}{\partial y}(y)\cdot\cos(xy) + y\cdot\frac{\partial}{\partial y}\cos(xy) = \cos(xy) + y(-\sin(xy)x) = \cos(xy) - xy\sin(xy).$$

Now we can compute the four second-order derivatives.

$$\begin{aligned}
\frac{\partial^2 z}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial}{\partial x}(-y^2\sin(xy)) = -y^3\cos(xy), \\
\frac{\partial^2 z}{\partial y \partial x} &= \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial}{\partial y}(-y^2\sin(xy)) = -2y\sin(xy) - x^2y\cos(xy), \\
\frac{\partial^2 z}{\partial x \partial y} &= \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial}{\partial x}(\cos(xy) - xy\sin(xy)) \\
&= \frac{\partial}{\partial x}(\cos(xy)) - y\frac{\partial}{\partial x}(x\sin(xy)) \\
&= -y\sin(xy) - y(\sin(xy) + xy\cos(xy)) \\
&= -y\sin(xy) - y\sin(xy) - xy^2\cos(xy) = -2y\sin(xy) - xy^2\cos(xy),
\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2 z}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial}{\partial y}(\cos(xy) - yx\sin(xy))\\ &= (-x\sin(xy)) + \big(-x\sin(xy) - yx(x\cos(xy))\big)\\ &= -x\sin(xy) - x\sin(xy) - x^2 y\cos(xy)\\ &= -2x\sin(xy) - x^2 y\cos(xy).\end{aligned}$$

While the computations of the two mixed partials are different, the results are the same, as is expected. ◆

In view of the importance of the equation $f_{xy} = f_{yx}$, we state it as a theorem.

Theorem 16.2.1: Equality of Mixed Partial Derivatives

Assume that $f(x, y)$ is defined in some disk centered at (a, b). If f_x and f_y exist in the disk and f_{xy} is continuous at (a, b), then $f_{yx}(a, b)$ exists and equals $f_{xy}(a, b)$.

This is not obvious. Why should the rate at which the slope in the y-direction changes with respect to x be the same as the rate at which the slope in the x-direction changes with respect to y? A proof is outlined in Exercise 21 in Section 16.S.

Differentiating Under the Integral Sign

Let $g(y) = \int_a^b f(x, y)\,dx$, an integral that depends on y. The following theorem expresses the derivative of g in terms of a partial derivative of f. It provides a general condition that allows us to differentiate under the integral sign.

Theorem 16.2.2: Differentiating Under the Integral Sign

Assume that $f(x, y)$ is defined in the rectangle whose vertices are (a, a), (a, b), (b, b), and (b, a). Assume also that $\partial f/\partial y$ is continuous there. Define $g(y) = \int_a^b f(x, y)\,dx$. Then

$$g'(y) = \int_a^b \frac{\partial f}{\partial y}(x, y)\,dx.$$

While an example is not a substitute for a proof, Example 6 suggests that shows this theorem is, at least, plausible.

EXAMPLE 6. Verify Theorem 16.2.2 when $g(y) = \displaystyle\int_1^2 e^{xy}\,dx$.

SOLUTION To evaluate $\int_1^2 D_y\left(e^{xy}\right)\,dx$ we find

$$\begin{aligned}\int_1^2 \frac{\partial}{\partial y}e^{xy}\,dx &= \int_1^2 xe^{xy}\,dx && (\text{ compute } D_y(e^{xy})\text{ })\\ &= \left(\frac{xe^{xy}}{y} - \frac{e^{xy}}{y^2}\right)\Bigg|_{x=1}^{x=2} && (\text{ integration by parts: } u = x,\ dv = e^{xy}\,dx\text{ })\\ &= \left(\frac{2e^{2y}}{y} - \frac{e^{2y}}{y^2}\right) - \left(\frac{e^y}{y} - \frac{e^y}{y^2}\right) && (\text{ evaluating at endpoints and subtracting })\\ &= \frac{(2y-1)e^{2y} - (y-1)e^y}{y^2} && (\text{ simplifying }). \qquad (16.2.1)\end{aligned}$$

On the other hand

$$g(y) = \int_1^2 e^{xy}\,dx = \left.\frac{e^{xy}}{y}\right|_{x=1}^{x=2} = \frac{e^{2y}}{y} - \frac{e^{y}}{y}.$$

Then

$$\begin{aligned} g'(y) &= \frac{2ye^{2y} - e^{2y}}{y^2} - \frac{ye^{y} - e^{y}}{y^2} \\ &= \frac{(2y-1)e^{2y} - (y-1)e^{y}}{y^2}. \end{aligned} \tag{16.2.2}$$

The final expressions in (16.2.1) and (16.2.2) are equal, as guaranteed by Theorem 16.2.2. ♦

Functions of More Than Two Variables

A quantity may depend on more than two variables. For instance, the chill factor depends on temperature, humidity, and wind velocity. Also, the temperature T at a point in the atmosphere is a function of the three space coordinates, x, y, and z: $T = f(x,y,z)$.

The definitions and notations of partial derivatives carry over to functions of more than two variables. If $u = f(x,y,z,t)$, there are four first-order partial derivatives. The partial derivative of u with respect to x, holding y, z, and t fixed, is denoted

$$\frac{\partial u}{\partial x}, \ \frac{\partial f}{\partial x}, \text{ or } u_x.$$

Higher-ordered partial derivatives are defined and denoted similarly. Section 16.8 and the "Wave in a Rope" CIE (CIE 23) at the end of Chapter 17 illustrate their use in physics.

To differentiate with respect to a variable, hold all variables constant except the one being differentiated with respect to. Then, differentiate using the usual methods and techniques.

Summary

We defined limits, continuity, and partial derivatives for functions of several variables. These are closely related to the one-variable versions.

A partial derivative with respect to one variable, say x, is found by treating all other variables as constants and applying the standard differentiation rules with respect to x. Higher-order partial derivatives are defined much like higher-order derivatives of a function of a single variable. An important property of higher-order partial derivatives of functions usually met in applications is that the order in which the partial derivatives are calculated does not affect the results. For instance,

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right).$$

EXERCISES for Section 16.2

1. (a) Draw a clear picture that shows the trace of a surface $z = f(x,y)$ by the plane $y = b$.

(b) Draw the line through $(a, b, f(a,b))$ whose slope is $\dfrac{\partial f}{\partial x}$ at (a,b).

2. (a) Draw a clear picture that shows the trace of a surface $z = f(x,y)$ by the plane $x = a$.

(b) Draw the line through $(a, b, f(a,b))$ whose slope is $\dfrac{\partial f}{\partial y}$ at (a,b).

In Exercises 3 to 16 evaluate the limits, if they exist.

3. $\lim_{(x,y)\to(2,3)} \frac{x+y}{x^2+y^2}$

4. $\lim_{(x,y)\to(1,1)} \frac{x^2}{x^2+y^2}$

5. $\lim_{(x,y)\to(0,0)} \frac{x^2}{x^2+y^2}$

6. $\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2}$

7. $\lim_{(x,y)\to(2,3)} x^y$

8. $\lim_{(x,y)\to(0,0)} x^y$

9. $\lim_{(x,y)\to(1,0)} \frac{e^{2xy}}{xy}$

10. $\lim_{(x,y)\to(1,2)} \frac{e^{2xy}}{xy}$

11. $\lim_{(x,y)\to(0,0)} \frac{\sin(x+y)}{x+x^2}$

12. $\lim_{(x,y)\to(0,0)} \frac{\sin(2x)}{\sin(3y)}$

13. $\lim_{(x,y)\to(2,2)} \frac{y^3-8}{x^2-4}$

14. $\lim_{(x,y)\to(3,2)} \frac{e^{2x}-e^{3y}}{4x^2-9y^2}$

15. $\lim_{(x,y)\to(0,0)} (1+xy)^{1/(xy)}$

16. $\lim_{(x,y)\to(0,0)} (1+x)^{1/y}$

In Exercises 17 to 24, describe the domain of the function.

17. $f(x,y) = \frac{1}{x+y}$

18. $f(x,y) = \frac{1}{x^2+2y^2}$

19. $f(x,y) = \frac{1}{9-x^2-y^2}$

20. $f(x,y) = \sqrt{49-x^2-y^2}$

21. $f(x,y) = \ln(x+2y)$

22. $f(x,y) = \ln(4-x^2+y^2)$

23. $f(x,y) = \frac{1}{(9-x^2-y^2)^{1/3}}$

24. $f(x,y) = \frac{1}{\sqrt{x^2+y^2-25}}$

In Exercises 25 to 30, R consists of all points (x,y) that satisfy the condition. Find the boundary of R.

25. $x^2+y^2 \le 1$

26. $x^2+y^2 < 1$

27. $1/(x^2+y^2)$ is defined

28. $1/(x+y)$ is defined

29. $y < x^2$

30. $y \le x$

In Exercises 31 to 36 compute $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$ and $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)$.

31. $f(x,y) = e^{3x^2y}$

32. $f(x,y) = \sqrt{x^2+3y^2}$

33. $f(x,y) = \ln(2x+3y)$ (Assume $2x+3y>0$.)

34. $f(x,y) = \arctan\left(\sqrt{xy^3}\right)$ (Assume $x, y \ge 0$.)

35. $f(x,y) = \frac{y}{x}$ (Assume $x \ne 0$.)

36. $f(x,y) = \frac{\sin(x+2y)}{x}$ (Assume $x \ne 0$.)

In Exercises 37 to 44 compute $\frac{\partial^2 f}{\partial x^2}$ and $\frac{\partial^2 f}{\partial y^2}$.

37. $f(x,y) = \ln\left(\sin^2(x)\cos^3(xy)\right)$

38. $f(x,y) = \exp\left(x^3\right)$

39. $f(x,y) = \tan\left(3x^2y^3\right)$

40. $f(x,y) = \frac{x^3}{y^2}$

41. $f(x,y) = 3x^2y^3$

42. $f(x,y) = \arctan\left(\frac{y}{x}\right)$

43. $f(x,y) = e^{x^2+y^2}$

44. $f(x,y) = \ln\left(y^2+y^4\right)$

45. Let $T(x,y,z) = \frac{1}{\sqrt{x^2+y^2+z^2}}$, if (x,y,z) is not the origin $(0,0,0)$. Show that $\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} = 0$.

NOTE: This equation arises in the theory of heat, as we will see in CIE 25 at the end of Chapter 18.

46. Solve Example 3 by using polar coordinates to express the function.

Check that differentiating under the integral sign gives correct results for the functions in Exercises 47 to 50.

47. $g(y)=\int_a^b x^m y^n\,dx$ (Assume m, $n>1$.)

48. $g(y)=\int_a^b \sin(xy)\,dx$

49. $g(y)=\int_a^b x^y\,dx$ (Assume a, b, $y>1$.)

50. $g(y)=\int_a^b \frac{dx}{xy}$

---

51. View $\int_a^b f(x,y)\,dx$ as a function of a, b, and y, say, $g(a,b,y)$. Find (a) $\frac{\partial g}{\partial b}$, (b) $\frac{\partial g}{\partial a}$, and (c) $\frac{\partial g}{\partial y}$.

52. Assume that $f(1,2)$ is 3, $\frac{\partial f}{\partial x}$ at $(1,2)$ is 2, and $\frac{\partial f}{\partial y}$ at $(1,2)$ is 1.2.
Estimate (a) $f(1,2.01)$, (b) $f(0.98,2)$, and (c) $f(1.02,2.03)$. Explain your reasoning in each case.

53. Develop a convincing, but not necessarily rigorous, argument justifying differentiation under the integral sign, that is, $\frac{d}{dy}\int_a^b f(x,y)\,dx=\int_a^b \frac{\partial f}{\partial y}(x,y)\,dx$.

54. Assume that the domain of f is the xy-plane and $\frac{\partial f}{\partial x}=0$ everywhere. (a) Give an example of a nonconstant function for which $\frac{\partial f}{\partial x}=0$. (b) What is the most general function for which $\frac{\partial f}{\partial x}=0$ everywhere?

55. Find all functions f defined throughout the xy-plane for which both $\frac{\partial f}{\partial x}=0$ and $\frac{\partial f}{\partial y}=0$ everywhere. Explain.

56. This exercise concerns a function $f(x,y)$ whose mixed partial derivatives at $(0,0)$ are not equal.

(a) Let $g(x,y)=\frac{x^2-y^2}{x^2+y^2}$ for $(x,y)\neq(0,0)$. Show that $\lim_{k\to 0}(\lim_{h\to 0} g(h,k))=-1$ but $\lim_{h\to 0}(\lim_{k\to 0} g(h,k))=1$.

(b) Let $f(x,y)=xyg(x,y)$ for (x,y) not $(0,0)$ and $f(0,0)=0$. Show that $f(x,y)=0$ if x or y is 0.

(c) Show that $f_{xy}(0,0)=\lim_{k\to 0}\frac{f_x(0,k)-f_x(0,0)}{k}$.

(d) Show that $f_{xy}(0,0)=\lim_{k\to 0}\left(\lim_{h\to 0}\frac{f(h,k)-f(0,k)-f(h,0)+f(0,0)}{hk}\right)$.

(e) Show that $f_{xy}(0,0)=-1$.

(f) Similarly, show that $f_{yx}(0,0)=1$.

(g) Show that in polar coordinates the value of f at (r,θ) is $\frac{1}{4}r^2\sin(4\theta)$

16.3 Change and the Chain Rule

For a function of one variable, $f(x)$, the change in its value as the input changes from a to $a+\Delta x$ is approximately $f'(a)\,\Delta x$. In this section we estimate the change in $f(x,y)$ as (x,y) moves from (a,b) to $(a+\Delta x,b+\Delta y)$.

This change is used to obtain the chain rule for functions of several variables.

Estimating the Change Δf

Let $z=f(x,y)$ be a function of two variables with continuous partial derivatives at least throughout a disk centered at the point (a,b). We will express $\Delta f=f(a+\Delta x,b+\Delta y)-f(a,b)$ in terms of f_x and f_y.

We first observe that, in Figure 16.3.1, Δf is the the length of the segment AC: $\Delta f=|AC|$. The same figure suggests we can view $\Delta f=|AC|$ as being obtained in two steps. First, there is the change from $f(a,b)$ to $f(a+\Delta x,y)$

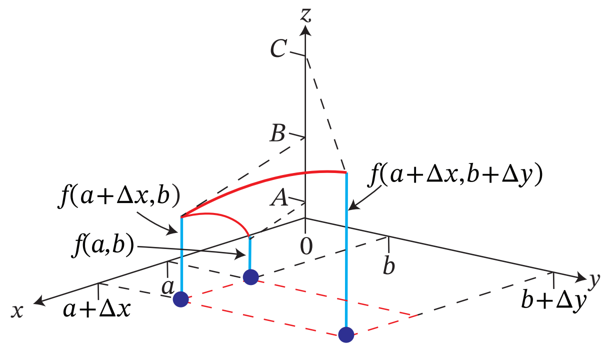


Figure 16.3.1

as x goes from a to $a+\Delta x$, that is, $f(a+\Delta x, b) - f(a,b) = |AB|$. Second, there is the change $|BC|$ from $f(a+\Delta x, b)$ to $f(a+\Delta x, b+\Delta y)$, as y changes from b to $b+\Delta y$. That is,

$$\begin{aligned}\Delta f &= f(a+\Delta x, b+\Delta y) - f(a,b) = |AC| = |AB| + |BC| \\ &= \big(f(a+\Delta x, b) - f(a,b)\big) + \big(f(a+\Delta x, b+\Delta y) - f(a+\Delta x, b)\big).\end{aligned} \tag{16.3.1}$$

By the mean value theorem, there is a number c_1 between a and $a+\Delta x$ such that

$$|AB| = f(a+\Delta x, b) - f(a,b) = \frac{\partial f}{\partial x}(c_1, b)\,\Delta x. \tag{16.3.2}$$

The mean value theorem also applies to the second bracketed expression in (16.3.1). This time it says there is a number c_2 between b and $b+\Delta y$ such that

$$|BC| = f(a+\Delta x, b+\Delta y) - f(a+\Delta x, b) = \frac{\partial f}{\partial y}(a+\Delta x, c_2)\,\Delta y. \tag{16.3.3}$$

Combining (16.3.1), (16.3.2), and (16.3.3) produces

$$\Delta f = \frac{\partial f}{\partial x}(c_1, b)\,\Delta x + \frac{\partial f}{\partial y}(a+\Delta x, c_2)\,\Delta y. \tag{16.3.4}$$

When both Δx and Δy are small, (c_1, b) and $(a+\Delta x, c_2)$ are near (a,b). If we assume that the partial derivatives f_x and f_y are continuous at (a,b), then

$$\frac{\partial f}{\partial x}(c_1, b) = \frac{\partial f}{\partial x}(a,b) + \epsilon_1 \quad \text{and} \quad \frac{\partial f}{\partial y}(a+\Delta x, c_2) = \frac{\partial f}{\partial y}(a,b) + \epsilon_2, \tag{16.3.5}$$

where ϵ_1 and ϵ_2 approach 0 as Δx and Δy approach 0. Combining (16.3.4) and (16.3.5) gives the key to estimating the change in the function f. We state this important result as a theorem.

Theorem 16.3.1: Differential of $f(x,y)$

Let f have continuous first-order partial derivatives f_x and f_y for all points within some disk with center at the point (a,b). Then Δf, the change $f(a+\Delta x, b+\Delta y) - f(a,b)$, can be written

$$\Delta f = \frac{\partial f}{\partial x}(a,b)\,\Delta x + \frac{\partial f}{\partial y}(a,b)\,\Delta y + \epsilon_1\,\Delta x + \epsilon_2\,\Delta y, \tag{16.3.6}$$

where ϵ_1 and ϵ_2 approach 0 as Δx and Δy approach 0. (Both ϵ_1 and ϵ_2 are functions of a, b, Δx, and Δy.)

Equation (16.3.6) is the core of this section. The term $f_x(a,b)\Delta x$ estimates the change due to the change in the x-coordinate and $f_y(a,b)\Delta y$ estimates the change due to the change in the y-coordinate.

We call $f(x,y)$ **differentiable** at (a,b) if (16.3.6) holds. If f_x and f_y exist in a disk around (a,b) and are continuous at (a,b), then f is differentiable at (a,b).

Since ϵ_1 and ϵ_2 in (16.3.6) approach 0 as Δx and Δy approach 0,

$$\Delta f \approx \frac{\partial f}{\partial x}(a,b)\,\Delta x + \frac{\partial f}{\partial y}(a,b)\,\Delta y, \tag{16.3.7}$$

The approximation (16.3.7) gives us a way to estimate Δf when Δx and Δy are small.

EXAMPLE 1. Estimate $(2.1)^2(0.95)^3$.

SOLUTION Let $f(x,y) = x^2y^3$. Because $f(2,1) = 2^2 1^3 = 4$ all that remains is to estimate $f(2.1, 0.95)$. Equation (16.3.7) is used to estimate the differential $\Delta f = f(2.1,0.95) - f(2,1)$. The first-order partial derivatives of f are

$$\frac{\partial(x^2y^3)}{\partial x} = 2xy^3 \quad \text{and} \quad \frac{\partial(x^2y^3)}{\partial y} = 3x^2y^2,$$

which, when evaluated at the point $(2,1)$, lead to

$$\frac{\partial f}{\partial x}(2,1) = 4 \quad \text{and} \quad \frac{\partial f}{\partial y}(2,1) = 12.$$

Since $\Delta x = 2.1 - 2 = 0.1$ and $\Delta y = 0.95 - 1 = -0.05$,

$$\Delta f \approx \frac{\partial f}{\partial x}(2,1)\,\Delta x + \frac{\partial f}{\partial y}(2,1)\,\Delta y, = 4(0.1) + 12(-0.05) = 0.4 - 0.6 = -0.2.$$

Thus $(2.1)^2(0.95)^3$ is approximately $4 + (-0.2) = 3.8$. The exact value is 3.78102375. ♦

The Chain Rule

We begin with two special cases of the chain rule for functions of more than one variable. Afterward we will state the chain rule for functions of any number of variables.

In the first special case, Theorem 16.3.2, $z = f(x,y)$ where x and y are functions of t. The second special case, Theorem 16.3.4, is more general, where $z = f(x,y)$ with x and y functions of two variables, t and u.

Theorem 16.3.2: Chain Rule, Special Case I

Let $z = f(x,y)$ have continuous partial derivatives f_x and f_y, and let $x = x(t)$ and $y = y(t)$ be differentiable functions of t. Then $z = f(x(t), y(t))$ is a differentiable function of t and

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}. \tag{16.3.8}$$

Proof of Theorem 16.3.2

A change Δt causes changes Δx and Δy, which cause a change Δz in z.

By definition,

$$\frac{dz}{dt} = \lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t}.$$

According to Theorem 16.3.1,

$$\Delta z = \frac{\partial f}{\partial x}(x,y)\,\Delta x + \frac{\partial f}{\partial y}(x,y)\,\Delta y + \epsilon_1\,\Delta x + \epsilon_2\,\Delta y,$$

where $\epsilon_1 \to 0$ and $\epsilon_2 \to 0$ as Δx and Δy approach 0. (x and y are fixed.) Thus

$$\frac{\Delta z}{\Delta t} = \frac{\partial f}{\partial x}(x,y)\frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y}(x,y)\frac{\Delta y}{\Delta t} + \epsilon_1\frac{\Delta x}{\Delta t} + \epsilon_2\frac{\Delta y}{\Delta t},$$

and

$$\frac{dz}{dt} = \lim_{\Delta t\to 0}\frac{\Delta z}{\Delta t} = \frac{\partial f}{\partial x}(x,y)\frac{dx}{dt} + \frac{\partial f}{\partial y}(x,y)\frac{dy}{dt} + 0\frac{dx}{dt} + 0\frac{dy}{dt}.$$

This proves the theorem. •

The two summands on the right-hand side of (16.3.8) remind us of the chain rule for functions of one variable. Why is there a "+" in (16.3.8)? It first appears in (16.3.4) and can be traced back to Figure 16.3.1.

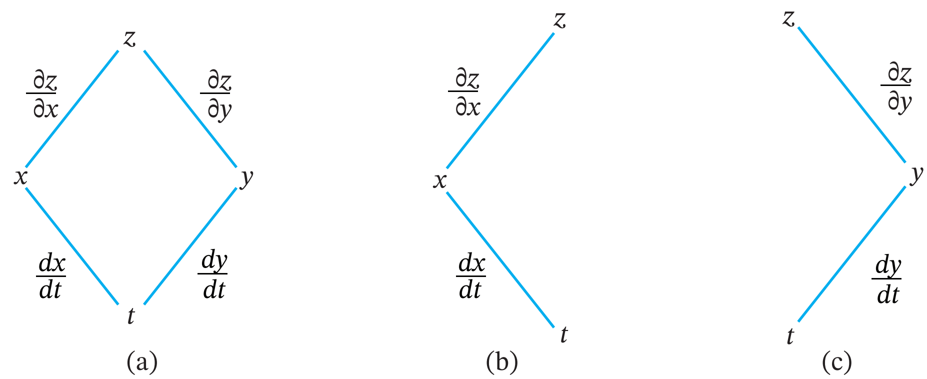


Figure 16.3.2

The diagram in Figure 16.3.2(a) helps in using this case of the chain rule. There are two paths from the top variable z down to the bottom variable t. Label each edge with a partial derivative (or derivative). For each path there is a summand in the chain rule. The left-hand path (see Figure 16.3.2(b)) gives the summand

$$\frac{\partial z}{\partial x}\frac{dx}{dt}.$$

The right-hand path (see Figure 16.3.2(c)) gives the summand

$$\frac{\partial z}{\partial y}\frac{dy}{dt}.$$

And, the instantaneous rate of change of z with respect to t, dz/dt, is their sum.

The total number of terms in a partial derivative of a "top" variable with respect to a "bottom" variable is the number of paths from the top one to the bottom one. Each path produces one summand, which consists of the product of the factors that appear in the path. This simple fact provides a quick way to check if any terms have been omitted from a partial derivative.

EXAMPLE 2. Let $z = x^2y^3$, $x = 3t^2$, and $y = \frac{t}{3}$. Find $\frac{dz}{dt}$ when $t = 1$.

SOLUTION To apply Theorem 16.3.2 compute z_x, z_y, dx/dt, and dy/dt:

$$\frac{\partial z}{\partial x} = 2xy^3, \quad \frac{\partial z}{\partial y} = 3x^2y^2, \quad \frac{dx}{dt} = 6t, \text{ and } \frac{dy}{dt} = \frac{1}{3}.$$

By Theorem 16.3.2,

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = 2xy^3\cdot 6t + 3x^2y^2\cdot\frac{1}{3}.$$

When $t = 1$, x is 3, y is 1/3, and

$$\frac{dz}{dt} = 2\cdot 3\left(\frac{1}{3}\right)^3 6\cdot 1 + 3\cdot 3^2\left(\frac{1}{3}\right)^2\frac{1}{3} = \frac{36}{27} + \frac{27}{27} = \frac{7}{3}.$$

◆

Observation 16.3.3: *Alternate Solution to Example 2*

In Example 2, dz/dt can be found without using Theorem 16.3.2 by writing z in terms of t:

$$z = x^2y^3 = (3t^2)^2\left(\frac{t}{3}\right)^3 = \frac{t^7}{3}.$$

Then $dz/dt = 7t^6/3$. When $t = 1$, this gives $dz/dt = 7/3$, as before.

EXAMPLE 3. The temperature at the point (x, y) on a window is $T(x, y)$. A bug wandering on the window is at the point $(x(t), y(t))$ at time t. How fast does the bug observe that the temperature of the glass changes as he crawls about?

SOLUTION The bug is asking us to find dT/dt. The chain rule (16.3.8) asserts that

$$\frac{dT}{dt} = \frac{\partial T}{\partial x}\frac{dx}{dt} + \frac{\partial T}{\partial y}\frac{dy}{dt}.$$

The bug can influence this rate by crawling faster or slower, changing dx/dt and dy/dt. He may want to know the direction he should choose in order to cool off or warm up as quickly as possible. We will be able to tell him how to do this in the next section. ◆

The proof of the next chain rule is almost identical to the proof of Theorem 16.3.2. (See Exercise 24.)

Theorem 16.3.4: Chain Rule, Special Case II

Let $z = f(x, y)$ have continuous first-order partial derivatives, f_x and f_y. Let $x = x(t, u)$ and $y = y(t, u)$ also have continuous first-order partial derivatives $\partial x/\partial t$, $\partial x/\partial u$, $\partial y/\partial t$, and $\partial y/\partial u$. Then, the two first-order partial derivatives of $z = f(x(t, u), y(t, u))$ are

$$\frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} \quad \textit{and} \quad \frac{\partial z}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u}.$$

The dependencies among the variables are recorded in Figure 16.3.3(a).

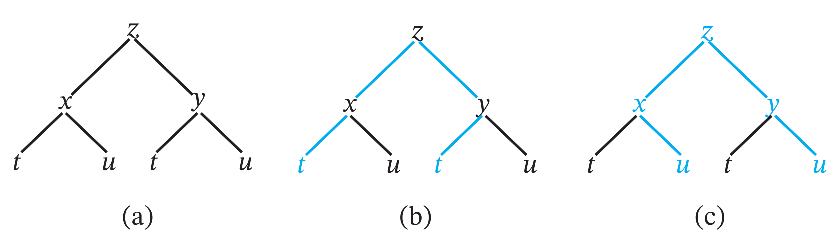


Figure 16.3.3

To find z_t, draw paths from z down to t. Label their edges by the appropriate partial derivative, as shown in cyan in Figure 16.3.3(b). Finding z_u is shown in Figure 16.3.3(c).

Each path from the top variable to the bottom variable contributes a summand in the chain rule. The only difference between Figure 16.3.2(a) and Figure 16.3.3(b) is that ordinary derivatives dx/dt and dy/dt appear in Figure 16.3.2(a), while partial derivatives x_t and y_t appear in Figure 16.3.3(b).

In Theorem 16.3.2 there are two middle variables and one bottom variable. In Theorem 16.3.4 there are two middle variables and two bottom variables. The chain rule holds for any number of middle variables and any number of bottom variables. There may be three middle variables and, say, four bottom variables. Then there are three summands for each of four first-order partial derivatives.

In the next example there is one middle variable and two bottom variables.

EXAMPLE 4. Let $z = f(u)$ be a function of a single variable. Let $u = 2x + 3y$. Then z is a composite function of x and y. Show that

$$2\frac{\partial z}{\partial y} = 3\frac{\partial z}{\partial x}. \tag{16.3.9}$$

SOLUTION We will evaluate z_x and z_y by the chain rule and then show that (16.3.9) is true.

To find z_x we use paths from z down to x. (See Figure 16.3.4.) There is only one middle variable so there is only one such path. Since $u = 2x + 3y$, $u_x = 2$ and

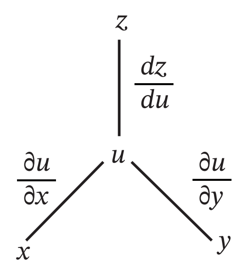


Figure 16.3.4

$$\frac{\partial z}{\partial x} = \frac{dz}{du}\frac{\partial u}{\partial x} = \frac{dz}{du}\cdot 2 = 2\frac{dz}{du}.$$

NOTE: One derivative is ordinary, while the other is partial.

Next we find z_y. Again, there is only one summand. Since $u = 2x + 3y$, $u_y = 3$. Consequently

$$\frac{\partial z}{\partial y} = \frac{dz}{du}\frac{\partial u}{\partial y} = \frac{dz}{du}\cdot 3 = 3\frac{dz}{du}.$$

From these results, $3z_x = 3(2z_u) = 6z_u$ and $2z_y = 2(3z_u) = 6z_u$ we see that $3z_x = 2z_y$. ♦

An Important Use of the Chain Rule

There is a difference between Example 2 and Example 4. In the first, we were dealing with explicitly given functions. As pointed out in Observation 16.3.3, we did not need to use the chain rule to find the derivative, dz/dt. In Example 4, we were dealing with a general type of function formed in a certain way: We showed that (16.3.9) holds for every differentiable function $f(u)$. No matter what $f(u)$ we choose, if $u = 2x + 3y$, we know that $2z_y = 3z_x$.

Example 4 shows why the chain rule is important. It enables us to make general statements about the partial derivatives of an infinite number of functions, all of which are formed the same way. The next example illustrates this use again.

We begin with a brief introduction to that example. In 1746, when working to model the motion of a vibrating string, D'Alembert proposed the partial differential equation:

$$\frac{\partial^2 y}{\partial t^2} = k^2\frac{\partial^2 y}{\partial x^2}. \tag{16.3.10}$$

This **wave equation** created a great deal of excitement, especially since d'Alembert showed that any twice differentiable function of the form

The wave equation also arises in the study of sound or light.

$$y(x,t) = g(x + kt) + h(x - kt)$$

is a solution of (16.3.10). Here k is a constant, which can be assumed to be positive.

To confirm that d'Alembert was right, we first check his claim for $g(x + kt)$. The check for $h(x - kt)$ is similar.

EXAMPLE 5. Show that any function $y(x,t) = g(x + kt)$, where k is a constant and g is a twice differentiable function, satisfies the partial differential equation (16.3.10).

SOLUTION To find the partial derivatives y_{xx} and y_{tt} express $y = g(x + kt)$ as a composition of functions:

$$y = g(u) \qquad \text{where} \qquad u = x + kt.$$

Note that g is a function of one variable. Figure 16.3.5 shows the dependencies among the variables.

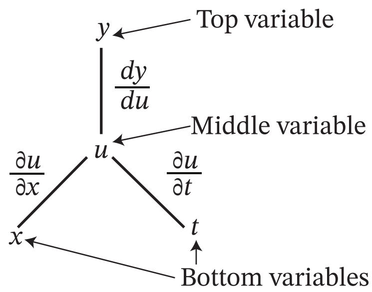


Figure 16.3.5

We will compute y_{xx} and y_{tt} in terms of derivatives of g and then check whether (16.3.10) holds. First, to compute y_{xx}, note that there is only one path from y down to x. (See Figure 16.3.5.)

$$\frac{\partial y}{\partial x} = \frac{dy}{du}\frac{\partial u}{\partial x} = \frac{dy}{du}\cdot 1 = \frac{dy}{du}. \tag{16.3.11}$$

In (16.3.11), dy/du is viewed as a function of x and t; that is, u is replaced by $x + kt$. Next,

$$\frac{\partial^2 y}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial y}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{dy}{du}\right).$$

Now, dy/du, viewed as a function of x and t, may be expressed as a composite function. Letting $w = dy/du$, we have

$$w = f(u), \quad \text{where} \quad u = x + kt.$$

Therefore, since there is only one path from w down to x,

$$\frac{\partial^2 y}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial y}{\partial x}\right) = \frac{\partial w}{\partial x} = \frac{dw}{du}\cdot\frac{\partial u}{\partial x} = \frac{d}{du}\left(\frac{dy}{du}\right)\frac{\partial u}{\partial x}.$$

Thus

$$\frac{\partial^2 y}{\partial x^2} = \frac{d^2 y}{du^2}. \tag{16.3.12}$$

Then we express y_{tt} in terms of d^2y/du^2 as follows. First,

$$\frac{\partial y}{\partial t} = \frac{dy}{du}\frac{\partial u}{\partial t} = \frac{dy}{du}\cdot k = k\frac{dy}{du}.$$

Hence:

$$\frac{\partial^2 y}{\partial t^2} = \frac{\partial}{\partial t}\left(\frac{\partial y}{\partial t}\right) = \frac{\partial}{\partial t}\left(k\frac{dy}{du}\right) = k\frac{d}{du}\left(\frac{dy}{du}\right)\cdot\frac{\partial u}{\partial t} = k\frac{d^2 y}{du^2}\cdot k$$

and we have

$$\frac{\partial^2 y}{\partial t^2} = k^2\frac{d^2 y}{du^2}. \tag{16.3.13}$$

Comparing (16.3.12) and (16.3.13) shows that

$$\frac{\partial^2 y}{\partial t^2} = k^2\frac{\partial^2 y}{\partial x^2}.$$

◆

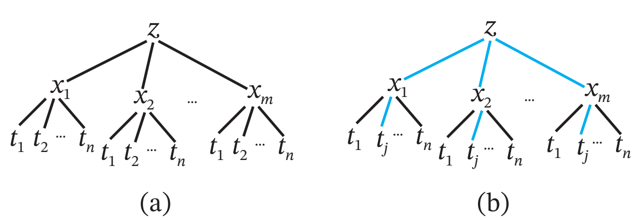


Figure 16.3.6

If z is a function of $x_1, x_2, \ldots, x_m$ and each x_i is a function of $t_1, t_2, \ldots, t_n$, then there are n partial derivatives $\partial z/\partial t_j$, $j = 1, 2, \ldots, n$. Each is a sum of m products of the form $(\partial z/\partial x_i)(\partial x_i/\partial t_j)$. To organize the calculation, first make a roster as shown in Figure 16.3.6(a). To compute $\partial z/\partial t_j$, list all paths from z down to t_j, as shown in Figure 16.3.6(b). Each path that starts at z and goes to t_j contributes a summand for $\partial z/\partial t_j$. The result is

$$\frac{\partial z}{\partial t_j} = \frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial t_j} + \frac{\partial z}{\partial x_2}\frac{\partial x_2}{\partial t_j} + \cdots + \frac{\partial z}{\partial x_m}\frac{\partial x_m}{\partial t_j}.$$

Summary

The section opened by showing that, under suitable assumptions on $z = f(x, y)$,

$$\Delta f = f(a+\Delta x, b+\Delta y) - f(a,b) = \frac{\partial f}{\partial x}(a,b)\,\Delta x + \frac{\partial f}{\partial y}(a,b)\,\Delta y + \epsilon_1\,\Delta x + \epsilon_2\,\Delta y, \qquad (16.3.14)$$

where ϵ_1 and ϵ_2 approach 0 as Δx and Δy approach 0. This gave a way to estimate Δf, namely

$$\Delta f \approx \frac{\partial f}{\partial x}(a,b)\,\Delta x + \frac{\partial f}{\partial y}(a,b)\,\Delta y.$$

The change in z is due to both the change in x and the change in y. Equation (16.3.14) generalizes to any number of variables and also is the basis for the various chain rules for partial derivatives.

EXERCISES for Section 16.3

In Exercises 1 to 4 verify the chain rule, special case I (Theorem 16.3.2), by computing $\frac{dz}{dt}$ two ways: (a) with the chain rule and (b) without the chain rule, by writing z as a function of t.

1. $z = x^2y^3$, $x = t^2$, $y = t^3$

2. $z = xe^y$, $x = t$, $y = 1+3t$

3. $z = \cos(xy^2)$, $x = e^{2t}$, $y = \sec(3t)$

4. $z = \ln(x+3y)$, $x = t^2$, $y = \tan(3t)$.

In Exercises 5 and 6 verify the chain rule, special case II (Theorem 16.3.4), by computing $\frac{\partial z}{\partial t}$ two ways: (a) with the chain rule and (b) without the chain rule, by writing z as a function of t and u.

5. $z = x^2y$, $x = 3t+4u$, $y = 5t-u$

6. $z = \sin(x+3y)$, $x = \sqrt{\frac{t}{u}}$, $y = \sqrt{t}+\sqrt{u}$

7. Assume that $z = f(x_1, x_2, x_3, x_4, x_5)$ and that x_i is a function of t_1, t_2, t_3.

(a) List all variables, showing top, middle, and bottom variables.

(b) Draw the paths involved in expressing $\frac{\partial z}{\partial t_3}$ in terms of the chain rule.

(c) Express $\frac{\partial z}{\partial t_3}$ in terms of the sum of products of partial derivatives.

(d) When computing $\frac{\partial z}{\partial x_2}$, which variables are constant?

(e) When computing $\frac{\partial z}{\partial t_3}$, which variables are constant?

8. If $z = f(g(t_1, t_2, t_3), h(t_1, t_2, t_3))$ (a) How many middle variables are there? (b) How many bottom variables? (c) What does the chain rule say about $\frac{\partial z}{\partial t_3}$? Include a diagram showing the paths.

9. Let $z = f(x(t), y(t))$. Find $\frac{dz}{dt}$ if $z_x = 4$, $z_y = 3$, $\frac{dx}{dt} = 4$, and $\frac{dy}{dt} = 1$.

10. Let $z = f(x(t), y(t))$. Find $\frac{dz}{dt}$ if $z_x = 3$, $z_y = 2$, $\frac{dx}{dt} = 4$, and $\frac{dy}{dt} = -3$.

11. Let $z = f(x, y)$, $x = u+v$, and $y = u-v$.
(a) Show that $(z_x)^2 - (z_y)^2 = (z_u)(z_v)$. Include diagrams. (b) Verify (a) when $f(x, y) = x^2 + 2y^3$.

12. Let $z = f(x,y)$, $x = u^2 - v^2$, and $y = v^2 - u^2$. (a) Show that $u\frac{\partial z}{\partial v} + v\frac{\partial z}{\partial u} = 0$. Include diagrams. (b) Verify (a) when $f(x,y) = \sin(x+2y)$.

13. Let $z = f(t-u, -t+u)$.
(a) Show that $\frac{\partial z}{\partial t} + \frac{\partial z}{\partial u} = 0$. Include diagrams. (b) Verify (a) when $f(x,y) = x^2 y$

14. Let $w = f(x-y, y-z, z-x)$.
(a) Show that $\frac{\partial w}{\partial x} + \frac{\partial w}{\partial y} + \frac{\partial w}{\partial z} = 0$. Include diagrams. (b) Verify (a) when $f(s,t,u) = s^2 + t^2 - u$.

15. Let $z = f(u,v)$, where $u = ax + by$, $v = cx + dy$, and a, b, c, d are constants. Assume that f has second-order derivatives and f_{uv} is continuous. Show that

(a) $\frac{\partial^2 z}{\partial x^2} = a^2\frac{\partial^2 f}{\partial u^2} + 2ac\frac{\partial^2 f}{\partial u \partial v} + c^2\frac{\partial^2 f}{\partial v^2}$

(b) $\frac{\partial^2 z}{\partial y^2} = b^2\frac{\partial^2 f}{\partial u^2} + 2bd\frac{\partial^2 f}{\partial u \partial v} + d^2\frac{\partial^2 f}{\partial v^2}$

(c) $\frac{\partial^2 z}{\partial x \partial y} = ab\frac{\partial^2 f}{\partial u^2} + (ad+bc)\frac{\partial^2 f}{\partial u \partial v} + cd\frac{\partial^2 f}{\partial v^2}$.

16. Let a, b, and c be constants and consider the partial differential equation $a\frac{\partial^2 z}{\partial x^2} + b\frac{\partial^2 z}{\partial x \partial y} + c\frac{\partial^2 z}{\partial y^2} = 0$. Suppose it has a solution $z = f(y + mx)$, where m is a constant. Assume that f is twice differentiable. Show that either (i) $am^2 + bm + c$ must be 0 or (ii) f is a linear function.

17. (a) Show that the partial differential equation $\frac{\partial^2 z}{\partial x^2} - 2\frac{\partial^2 z}{\partial x \partial y} + \frac{\partial^2 z}{\partial y^2} = 0$ is satisfied by any function of the form $z = f(x+y)$, where f is a twice differentiable function. (b) Verify (a) when $z = (x+y)^3$.

18. (a) Suppose f and g are twice differentiable. Show that any function of the form $z = f(x+y) + e^y g(x-y)$ satisfies $\frac{\partial^2 z}{\partial x^2} - \frac{\partial^2 z}{\partial y^2} - \frac{\partial z}{\partial x} + \frac{\partial z}{\partial y} = 0$. (b) Check (a) for $z = (x+y)^2 + e^y \sin(x-y)$.

19. Let $z = f(x,y)$ denote the temperature at the point (x,y) in the first quadrant. If polar coordinates are used, then we would write $z = g(r,\theta)$.

(a) Express z_r in terms of z_x and x_y.

(b) Express z_θ in terms of z_x and z_y.

(c) Show that $\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2 = \left(\frac{\partial z}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial z}{\partial \theta}\right)^2$.

(d) What variable is held constant in $\frac{\partial z}{\partial \theta}$?

(e) What variable is held constant in $\frac{\partial z}{\partial x}$?

20. Let $u = f(r)$ and $r = (x^2+y^2+z^2)^{1/2}$. Assume f is twice differentiable. Show that $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \frac{d^2 u}{dr^2} + \frac{2}{r}\frac{du}{dr}$.

21. At what rate is the volume of a rectangular box changing when its width is 3 feet and increasing at the rate of 2 feet per second, its length is 8 feet and decreasing at the rate of 5 feet per second, and its height is 4 feet and increasing at the rate of 2 feet per second?

22. The temperature T at (x,y,z) in space is $f(x,y,z)$ degrees. An astronaut is traveling so that his x- and y-coordinates increase at the rate of 4 miles per second and his z-coordinate decreases at the rate of 3 miles per second. Compute the rate $\frac{dT}{dt}$ at which the temperature changes at a point where $\frac{\partial T}{\partial x} = 4$, $\frac{\partial T}{\partial y} = 7$, and $\frac{\partial T}{\partial z} = 9$.

23. Let $u(x,t)$ be the temperature at point x along a rod at time t. The function u satisfies the one-dimensional heat equation for a constant k: $\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2}$.

(a) Show that $u(x,t) = e^{kt}g(x)$ satisfies the heat equation if $g(x)$ is any function such that $g''(x) = g(x)$.

(b) Show that if $g(x) = 3e^{-x} + 4e^{x}$, then $g''(x) = g(x)$.

24. We proved Theorem 16.3.2 when there are two middle variables and one bottom variable. Prove Theorem 16.3.4 when there are two middle variables and two bottom variables.

25. To prove the general chain rule when there are three middle variables, we need an analog of Theorem 16.3.1 concerning Δf when f is a function of three variables.

(a) Let $w = f(x,y,x)$ be a function of three variables. Show that

$$\begin{aligned}\Delta f &= f(x+\Delta x, y+\Delta y, z+\Delta z) - f(x,y,z)\\ &= \big(f(x+\Delta x, y+\Delta y, z+\Delta z) - f(x+\Delta x, y+\Delta y, z)\big) + \big(f(x+\Delta x, y+\Delta y, z) - f(x+\Delta x, y, z)\big)\\ &\quad + \big(f(x+\Delta x, y, z) - f(x,y,z)\big).\end{aligned}$$

(b) Using (a) show that $\Delta f = \frac{\partial f}{\partial x}(x,y,z)\,\Delta x + \frac{\partial f}{\partial y}(x,y,z)\,\Delta y + \frac{\partial f}{\partial z}(x,y,z)\,\Delta z + \epsilon_1\,\Delta x + \epsilon_2\,\Delta y + \epsilon_3\,\Delta z$, where ϵ_1, ϵ_2, $\epsilon_3 \to 0$ as Δx, Δy, $\Delta z \to 0$.

(c) Obtain the general chain rule for three middle variables and any number of bottom variables.

26. Let $z = f(x,y)$, where f has second-order derivatives, f_{xy} is continuous, and $x = r\cos(\theta)$ and $y = r\sin(\theta)$. Show that $\frac{\partial^2 z}{\partial r^2} = \cos^2(\theta)\frac{\partial^2 f}{\partial x^2} + 2\cos(\theta)\sin(\theta)\frac{\partial^2 f}{\partial x\partial y} + \sin^2(\theta)\frac{\partial^2 f}{\partial y^2}$.

27. Let $u = f(x,y)$, where $x = r\cos(\theta)$ and $y = r\sin(\theta)$. Verify the following equation, which appears in electromagnetic theory: $\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$.

28. Let u be a function of x and y, where u has second-order derivatives, u_{xy} is continuous, and x and y are both functions of s and t. Show that $\frac{\partial^2 u}{\partial s^2} = \frac{\partial^2 u}{\partial x^2}\left(\frac{\partial x}{\partial s}\right)^2 + 2\frac{\partial^2 u}{\partial x\partial y}\frac{\partial x}{\partial s}\frac{\partial y}{\partial s} + \frac{\partial^2 u}{\partial y^2}\left(\frac{\partial y}{\partial s}\right)^2 + \frac{\partial u}{\partial x}\frac{\partial^2 x}{\partial s^2} + \frac{\partial u}{\partial y}\frac{\partial^2 y}{\partial s^2}$.

29. Let (r,θ) be polar coordinates for the point (x,y) given in rectangular coordinates.

(a) From $r = \sqrt{x^2+y^2}$, show that $\frac{\partial r}{\partial x} = \cos(\theta)$.

(b) From $r = \frac{x}{\cos(\theta)}$, show that $\frac{\partial r}{\partial x} = \frac{1}{\cos(\theta)}$.

(c) Explain why (a) and (b) are not contradictory.

30. In developing (16.3.6), we used the path that started at (a,b), went to $(a+\Delta x, b)$, and ended at $(a+\Delta x, b+\Delta y)$. Could we have used the path from (a,b), through $(a, b+\Delta y)$, to $(a+\Delta x, b+\Delta y)$ instead? If not, explain why. If so write out the argument, using the path.

In Exercises 31 to 35 concern **homogeneous functions**. A function $f(x,y)$ is homogeneous of degree r if $f(kx,ky) = k^r f(x,y)$ for all $k > 0$.

31. Show that the following functions are homogeneous, and find the degree r.

(a) $f(x,y) = x^2(\ln x - \ln y)$ (b) $f(x,y) = \frac{1}{\sqrt{x^2+y^2}}$ (c) $f(x,y) = \sin\left(\frac{y}{x}\right)$.

32. Show that if f is homogeneous of degree r, then $xf_x + yf_y = rf$. This result is known as **Euler's theorem**.

33. Verify that the following functions are homogeneous of degree 1 and that they satisfy the conclusion of Euler's theorem (with $r = 1$): $f(x,y) = x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y}$. (a) $f(x,y) = 3x + 4y$ (b) $f(x,y) = x^3y^{-2}$ (c) $f(x,y) = xe^{x/y}$

34. Verify Euler's theorem for the functions in Exercise 31. See Exercise 32.

35. Show that if f is homogeneous of degree r, then $\dfrac{\partial f}{\partial x}$ is homogeneous of degree $r-1$. See Exercise 31.

16.4 Directional Derivatives and the Gradient

In this section we generalize the notion of a partial derivative to that of a directional derivative. Then we introduce a vector, called the gradient, to provide a formula for the directional derivative. The gradient will have other uses later in this chapter and in Chapter 18.

Directional Derivatives

If $z = f(x, y)$, the partial derivative $\partial f/\partial x = f_x$ tells how rapidly z changes as we move parallel to the x-axis with increasing x. Similarly, $\partial f/\partial y = f_y$ tells how fast z changes as we move parallel to the y-axis with increasing y. How rapidly does z change when we move the input point in any fixed direction in the xy-plane? The answer is given by the directional derivative.

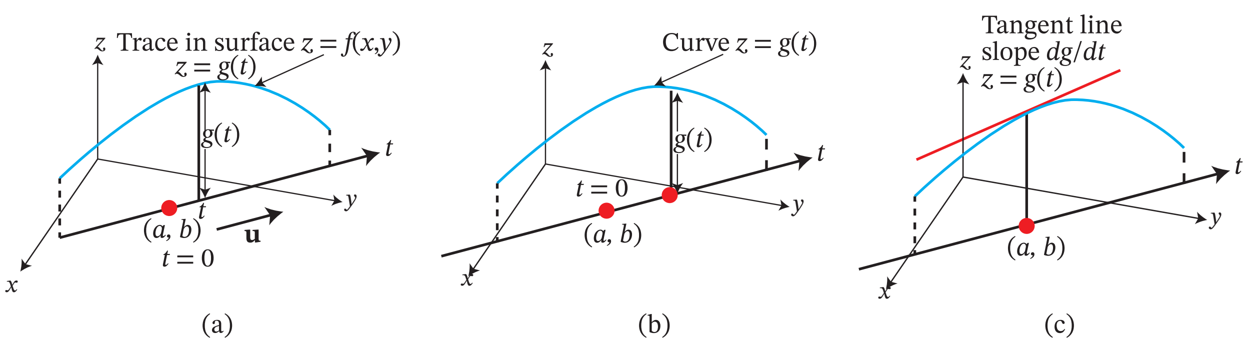


Figure 16.4.1

Let $z = f(x, y)$, (a, b) be a point, and $\mathbf{u}$ be a unit vector in the xy-plane. Draw a line through (a, b) parallel to $\mathbf{u}$. Call it the t-axis and let its positive part point in the direction of $\mathbf{u}$. Place the 0 of the t-axis at (a, b). (See Figure 16.4.1(a).) A value of t determines a point (x, y) on the t-axis and thus a value of z. Along the t-axis, z is a function of t, $z = g(t)$. (See Figure 16.4.1(b).) The derivative dg/dt, evaluated at $t = 0$, is called the **directional derivative** of $z = f(x, y)$ at (a, b) in the direction $\mathbf{u}$. It is denoted $D_{\mathbf{u}}f(a, b)$. The directional derivative is the slope of the tangent line to the curve $z = g(t)$ at $t = 0$. (See Figure 16.4.1(c).)

When $\mathbf{u} = \mathbf{i}$ we obtain the directional derivative $D_{\mathbf{i}}f$, which is f_x. When $\mathbf{u} = \mathbf{j}$ we obtain $D_{\mathbf{j}}f$, which is f_y.

The directional derivative generalizes the partial derivatives f_x and f_y, giving the rate of change of $z = f(x, y)$ in any direction in the xy-plane, not just in the directions indicated by $\mathbf{i}$ and $\mathbf{j}$.

The following theorem provides a definition of directional derivatives.

Theorem 16.4.1: Directional Derivatives

If $f(x, y)$ has continuous partial derivatives f_x and f_y, then the directional derivative of f at (a, b) in the direction of $\mathbf{u} = \cos(\theta)\mathbf{i} + \sin(\theta)\mathbf{j}$, where θ is the counterclockwise angle between $\mathbf{u}$ and $\mathbf{i}$, is

$$D_{\mathbf{u}}f = \frac{\partial f}{\partial x}(a, b)\cos(\theta) + \frac{\partial f}{\partial y}(a, b)\sin(\theta). \tag{16.4.1}$$

Proof of Theorem 16.4.1
The points on the line through (a,b) that makes an angle θ with the x-axis are $(a+t\cos(\theta), b+t\sin(\theta))$. Thus the directional derivative of f at (a,b) in the direction **u** is the derivative of

$$g(t) = f(a+t\cos(\theta), b+t\sin(\theta))$$

when $t=0$. (See Figure 16.4.2.)

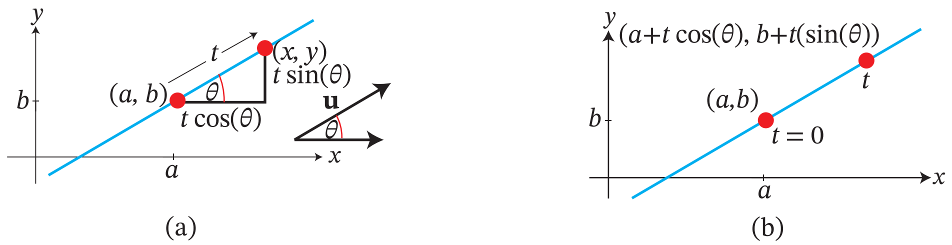


Figure 16.4.2

Because g is a composite function, $g(t) = f(x,y)$ where $x = a+t\cos(\theta)$ and $y = b+t\sin(\theta)$, the chain rule tells us that

$$g'(t) = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

From $dx/dt = \cos(\theta)$ and $dy/dt = \sin(\theta)$ it follows that

$$g'(0) = \frac{\partial f}{\partial x}(a,b)\cos\theta + \frac{\partial f}{\partial y}(a,b)\sin\theta,$$

and the theorem is proved. •

When $\theta = 0$, that is, when $\mathbf{u} = \mathbf{i}$, (16.4.1) becomes

$$\begin{aligned} D_{\mathbf{i}}f(a,b) &= \frac{\partial f}{\partial x}(a,b)\cos(0) + \frac{\partial f}{\partial y}(a,b)\sin(0) \\ &= \frac{\partial f}{\partial x}(a,b)(1) + \frac{\partial f}{\partial y}(a,b)(0) \\ &= \frac{\partial f}{\partial x}(a,b). \end{aligned}$$

When $\theta = \pi$, that is when $\mathbf{u} = -\mathbf{i}$, (16.4.1) becomes

$$\begin{aligned} D_{-\mathbf{i}}f(a,b) &= \frac{\partial f}{\partial x}(a,b)\cos(\pi) + \frac{\partial f}{\partial y}(a,b)\sin(\pi) \\ &= \frac{\partial f}{\partial x}(a,b)(-1) + \frac{\partial f}{\partial y}(a,b)(0) \\ &= -\frac{\partial f}{\partial x}(a,b). \end{aligned}$$

NOTE: If g increases in one direction then it decreases (by the same amount) in the opposite direction.

When $\theta = \pi/2$, that is when $\mathbf{u} = \mathbf{j}$, (16.4.1) asserts that the directional derivative is

$$\begin{aligned} D_{\mathbf{j}}f(a,b) &= \frac{\partial f}{\partial x}(a,b)\cos\left(\frac{\pi}{2}\right) + \frac{\partial f}{\partial y}(a,b)\sin\left(\frac{\pi}{2}\right) \\ &= \frac{\partial f}{\partial x}(a,b)(0) + \frac{\partial f}{\partial y}(a,b)(1) \\ &= \frac{\partial f}{\partial y}(a,b), \end{aligned}$$

also as expected.

EXAMPLE 1. Compute the derivative of $f(x, y) = x^2 y^3$ at $(1,2)$ in the direction given by the angle $\frac{\pi}{3}$. Interpret the result if f describes a temperature distribution.

SOLUTION First of all $f_x = 2xy^3$ and $f_y = 3x^2y^2$. Therefore $f_x(1,2) = 16$ and $f_y(1,2) = 12$. Also, because $\cos(\pi/3) = 1/2$ and $\sin(\pi/3) = \sqrt{3}/2$, the derivative of f in the direction $\mathbf{u} = \cos(\pi/3)\mathbf{i} + \sin(\pi/3)\mathbf{j} = (\mathbf{i} + \sqrt{3}\mathbf{j})/2$ is

$$D_{\mathbf{u}} f(1,2) = 16\left(\frac{1}{2}\right) + 12\left(\frac{\sqrt{3}}{2}\right) = 8 + 6\sqrt{3} \approx 18.4.$$

If x^2y^3 is the temperature in degrees at (x, y), where x and y are measured in centimeters, then the rate at which the temperature changes at $(1,2)$ in the direction given by $\theta = \pi/3$ is approximately 18.4 degrees per centimeter. ♦

The Gradient

Equation (16.4.1) resembles a dot product. To exploit this similarity, we introduce the vector whose scalar components are $f_x(a, b)$ and $f_y(a, b)$.

Definition: *The gradient of* $f(x, y)$

The **gradient** of f at (a, b) is the vector

$$\nabla f(a, b) = \frac{\partial f}{\partial x}(a, b)\mathbf{i} + \frac{\partial f}{\partial y}(a, b)\mathbf{j}.$$

∇f is called both "grad f" and "del f"; the latter because ∇ is an upside-down delta.

NOTATION: The del symbol, ∇, is in boldface because the gradient of f is a vector.

Let $f(x, y) = x^2 + y^2$. We compute and draw ∇f at a few points, listed in Table 16.4.1. Figure 16.4.3 shows ∇f at the same points, with the tail of the vector placed where ∇f is evaluated.

(x, y)	$\frac{\partial f}{\partial x} = 2x$	$\frac{\partial f}{\partial y} = 2y$	∇f
$(1,2)$	2	4	$2\mathbf{i} + 4\mathbf{j}$
$(3,0)$	6	0	$6\mathbf{i}$
$(2,-1)$	4	-2	$4\mathbf{i} - 2\mathbf{j}$

Table 16.4.1

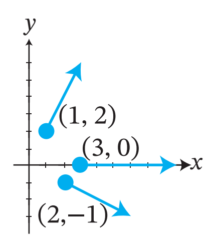


Figure 16.4.3

Theorem 16.4.1 can be restated in a vector form that is more conducive to computing directional derivatives:

Theorem 16.4.2: Directional Derivative, Rephrased

If $z = f(x, y)$ has continuous partial derivatives f_x and f_y in a neighborhood of (a, b), then the directional derivative of f at (a, b) in the direction of the unit vector $\mathbf{u}$ is

$$D_{\mathbf{u}} f = \nabla f(a, b) \cdot \mathbf{u} = (f_x(a, b)\mathbf{i} + f_y(a, b)\mathbf{j}) \cdot \mathbf{u}.$$

To find the directional derivative in the direction of a vector $\mathbf{v}$ that is not a unit vector, use $\mathbf{u} = \mathbf{v}/|\mathbf{v}|$.

The gradient is introduced not merely to provide a short notation for directional derivatives. Its importance is made clear in the next theorem.

A Different View of the Gradient

The gradient vector $\nabla f(a,b)$ provides two pieces of geometric information about a function. First, it points in the direction in the xy-plane in which the function increases most rapidly from the point (a,b). Second, its length, $|\nabla f(a,b)|$, is the largest directional derivative of f at (a,b).

Theorem 16.4.3: Geometric Significance of ∇f

Let $z = f(x,y)$ have continuous partial derivatives f_x and f_y. Let (a,b) be a point where ∇f is not $\mathbf{0}$. Then the largest directional derivative of f at (a,b) is $|\nabla f(a,b)|$, the length of $\nabla f(a,b)$. Moreover, this occurs in the direction

$$\mathbf{u} = \frac{\nabla f(a,b)}{\left|\nabla f(a,b)\right|}.$$

The direction of maximum increase for the function f at the point (a,b) is the gradient vector $\nabla f(a,b)$.

Proof of Theorem 16.4.3

If $\mathbf{u}$ is a unit vector, then, at (a,b), $D_{\mathbf{u}} f = \nabla f \cdot \mathbf{u}$. By the definition of the dot product, where the angle between ∇f and $\mathbf{u}$ is labeled α, $\nabla f \cdot \mathbf{u} = \left|\nabla f\right| |\mathbf{u}| \cos(\nabla f, \mathbf{u})$. Since $|\mathbf{u}| = 1$,

$$D_{\mathbf{u}} f = |\nabla f| \cos(\nabla f, \mathbf{u}). \tag{16.4.2}$$

The largest value of $\cos(\nabla f, \mathbf{u})$ occurs when the angle between ∇f and u is 0. That is, when $\mathbf{u}$ points in the direction of ∇f. (See Figure 16.4.4.) Thus, by (16.4.2), the largest directional derivative of $f(x,y)$ at (a,b) occurs when the direction is that of ∇f at (a,b). For that $\mathbf{u}$, $D_{\mathbf{u}} f = |\nabla f|$. This proves the theorem.

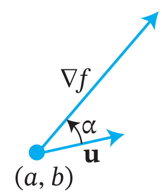


Figure 16.4.4

EXAMPLE 2. What is the largest directional derivative of $f(x,y) = x^2y^3$ at $(2,3)$? In what direction does the maximum directional derivative occur?

SOLUTION By direct calculation, $\nabla f = 2xy^3\mathbf{i} + 3x^2y^2\mathbf{j}$. The gradient vector at $(2,3)$ is $\nabla f(2,3) = 108\mathbf{i} + 108\mathbf{j}$. The maximal directional derivative of x^2y^3 at $(2,3)$ is $|\nabla f(2,3)| = 108\sqrt{2}$. This occurs at the angle $\theta = \pi/4$ relative to the x-axis, that is, for

$$\mathbf{u} = \frac{\nabla f(2,3)}{|\nabla f(2,3)|} = \frac{108\mathbf{i} + 108\mathbf{j}}{108\sqrt{(2)}} = \frac{\sqrt{2}}{2}\mathbf{i} + \frac{\sqrt{2}}{2}\mathbf{j}.$$

If $f(x,y)$ denotes the temperature at (x,y), the gradient ∇f indicates the direction in which heat flows. It tends to flow toward the cold, which is the mathematical assertion that heat tends to flow in the direction of $-\nabla f$.

The gradient and directional derivative have been interpreted in terms of a temperature distribution in the plane. It is also instructive to interpret them for a hiker on a mountain.

The elevation of a point on the surface of a mountain above the point (x,y) will be denoted by $f(x,y)$. The directional derivative $D_{\mathbf{u}} f$ indicates the rate at which elevation changes per unit change in horizontal distance in the direction of $\mathbf{u}$. The gradient ∇f at (a,b) points in the direction of steepest ascent. The length of ∇f tells the hiker how steep the slope is in that direction. (See Figure 16.4.5.)

Generalization to $f(x,y,z)$

Directional derivatives and gradients can be generalized to functions of three or more variables. The directional derivative of $f(x,y,z)$ in a direction in space indicates the rate of change in that direction.

Let $\mathbf{u}$ be a unit vector in space with direction angles α, β, and γ. Then $\mathbf{u} = \cos(\alpha)\mathbf{i} + \cos(\beta)\mathbf{j} + \cos(\gamma)\mathbf{k}$. We now define the derivative of $f(x,y,z)$ in the direction $\mathbf{u}$.

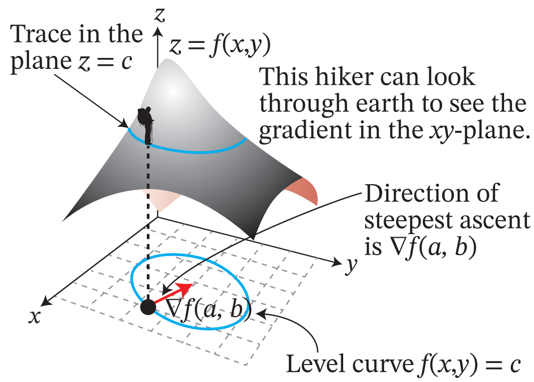


Figure 16.4.5

Definition: *Directional Derivative of* $f(x, y, z)$.

The **directional derivative** of f at (a, b, c) in the direction of the unit vector $\mathbf{u} = \cos(\alpha)\mathbf{i} + \cos(\beta)\mathbf{j} + \cos(\gamma)\mathbf{k}$ is $g'(0)$, where g is defined by $g(t) = f(a + t\cos(\alpha), b + t\cos(\beta), c + t\cos(\gamma))$. The directional derivative of f at (a, b, c) in the unit direction $\mathbf{u}$ is denoted $D_{\mathbf{u}} f(a, b, c)$.

The variable t measures the length along the line through (a, b, c) with direction angles α, β, and γ. Therefore $D_{\mathbf{u}} f$ is a derivative along the t-axis.

The proof of Theorem 16.4.4 for a function of three variables is essentially the same as the proof of this fact for a function of two variables.

Theorem 16.4.4: Directional Derivative of $f(x, y, z)$

If $f(x, y, z)$ has continuous partial derivatives f_x, f_y, and f_z, then the directional derivative of f at (a, b, c) in the direction of the unit vector $\mathbf{u} = \cos(\alpha)\mathbf{i} + \cos(\beta)\mathbf{j} + \cos(\gamma)\mathbf{k}$ is

$$\frac{\partial f}{\partial x}(a, b, c)\cos(\alpha) + \frac{\partial f}{\partial y}(a, b, c)\cos(\beta) + \frac{\partial f}{\partial z}(a, b, c)\cos(\gamma).$$

To write the directional derivative in space as a dot product brings us to the definition of the gradient of a function of three variables.

Definition: *The gradient of* $f(x, y, z)$.

The vector

$$\nabla f(a, b, c) = \frac{\partial f}{\partial x}(a, b, c)\mathbf{i} + \frac{\partial f}{\partial y}(a, b, c)\mathbf{j} + \frac{\partial f}{\partial z}(a, b, c)\mathbf{k}$$

is the **gradient** of f at (a, b, c).

The definition of the gradient and Theorem 16.4.4 provide a definition of directional derivatives as a dot product.

Definition: *Directional Derivative as a Dot Product*

The derivative of $f(x, y, z)$ in the direction of the unit vector $\mathbf{u}$ equals the dot product of the gradient of f and $\mathbf{u}$:

$$D_{\mathbf{u}} f = \nabla f \cdot \mathbf{u}.$$

As for a function of two variables, ∇f evaluated at (a,b,c) points in the direction $\mathbf{u}$ that produces the largest directional derivative at (a,b,c) and $|\nabla f(a,b,c)|$ is the largest directional derivative there. As for two variables, the key step in the proof of this theorem is writing $\nabla f \cdot \mathbf{u} = |\nabla f||\mathbf{u}|\cos(\nabla f, \mathbf{u})$ and remembering that $\mathbf{u}$ is a unit vector.

EXAMPLE 3. The temperature at the point (x,y,z) in a solid piece of metal is given by $f(x,y,z) = e^{2x+y+3z}$ degrees. In what direction at the point $(0,0,0)$ does the temperature increase most rapidly?

SOLUTION Because

$$\frac{\partial f}{\partial x} = 2e^{2x+y+3z}, \quad \frac{\partial f}{\partial y} = e^{2x+y+3z}, \quad \frac{\partial f}{\partial z} = 3e^{2x+y+3z},$$

the gradient vector is

$$\nabla f = 2e^{2x+y+3z}\mathbf{i} + e^{2x+y+3z}\mathbf{j} + 3e^{2x+y+3z}\mathbf{k}.$$

At $(0,0,0)$, $\nabla f(0,0,0) = 2\mathbf{i}+\mathbf{j}+3\mathbf{k}$. The largest rate of increase is

$$|\nabla f(0,0,0)| = |2\mathbf{i}+\mathbf{j}+3\mathbf{k}| = \sqrt{2^2+1^2+3^2} = \sqrt{14} \text{ degrees per unit length.}$$

The direction of most rapid increase in temperature is that given by the unit vector

$$\begin{aligned}\mathbf{u} &= \frac{\nabla f(0,0,0)}{|\nabla f(0,0,0)|} \\ &= \frac{1}{\sqrt{14}}(2\mathbf{i}+\mathbf{j}+3\mathbf{k}).\end{aligned}$$

◆

The previous example can also be interpreted in the context of the initial definition of a directional derivative in space. Construct a coordinate system so that the line through $(0,0,0)$ parallel to $2\mathbf{i}+\mathbf{j}+3\mathbf{k}$ is the t-axis, with $t=0$ at the origin and the positive part in the direction of $2\mathbf{i}+\mathbf{j}+3\mathbf{k}$, then $df/dt = \sqrt{14}$ at $t=0$.

Historical Note: The Origins of Gradient and ∇

The gradient was denoted Δ by Hamilton in 1846. By 1870 it was denoted ∇, an upside-down delta, and therefore called "atled." In 1871 Maxwell wrote, "The quantity ∇P is a vector. I venture, with much diffidence, to call it the slope of P." Slope is no longer used in this context, having been replaced by gradient, which comes from grade, the slope of a road or surface. The name "del" first appeared in print in 1901, in *Vector Analysis: A Text-Book for the Use of Students of Mathematics and Physics: Founded Upon the Lectures of J.W. Gibbs*, by J.W. Gibbs and E.B. Wilson.

Summary

We defined the derivative of $f(x,y)$ at (a,b) in the direction of the unit vector $\mathbf{u}$ in the xy-plane and the derivative of $f(x,y,z)$ at (a,b,c) in the direction of the unit vector $\mathbf{u}$ in space. Both are denoted $D_{\mathbf{u}}f$. Then we introduced the gradient vector ∇f in terms of its components and obtained the general formula

$$D_{\mathbf{u}}(P)f = \nabla f(P)\cdot\mathbf{u}.$$

By examining this formula we saw that the length and direction of $\nabla f(P)$ at a given point P are significant: $\nabla f(P)$ points in the (unit) direction $\mathbf{u}$ that maximizes $D_{\mathbf{u}}f(P)$ at the given point and $|\nabla f(P)|$ is the maximum directional derivative of f at the given point.

An alternate definition of the gradient vector is that it is that vector whose dot product with a unit vector $\mathbf{u}$ gives the directional derivative in the direction $\mathbf{u}$.

EXERCISES for Section 16.4

REMINDER: As usual, all functions in these exercises are assumed to have continuous partial derivatives.

1. In what direction from (a, b) does a function decrease most rapidly? What is the maximum rate of decrease?

2. Explain in words, using no symbols, the meaning of $D_{\mathbf{u}} f$.

In Exercises 3 and 4 compute the directional derivative of $x^4 y^5$ at $(1,1)$ in the given direction.

3. (a) $\mathbf{i}$, (b) $-\mathbf{i}$, and (c) $\cos\left(\frac{\pi}{4}\right)\mathbf{i} + \sin\left(\frac{\pi}{4}\right)\mathbf{j}$.

4. (a) $\mathbf{j}$, (b) $-\mathbf{j}$, and (c) $\cos\left(\frac{\pi}{3}\right)\mathbf{i} + \sin\left(\frac{\pi}{3}\right)\mathbf{j}$.

In Exercises 5 and 6 compute the directional derivative of $x^2 y^3$ in the given direction.

NOTE: If the given direction is not a unit vector, remember to construct a unit vector with the same direction.

5. (a) $\mathbf{j}$, (b) $\mathbf{k}$, and (c) $-\mathbf{i}$.

6. (a) $\mathbf{i} + \mathbf{j} + \mathbf{k}$, (b) $2\mathbf{i} - \mathbf{j}$, and (c) $2\mathbf{i} - 3\mathbf{j}$.

---

7. Assume that at the point $(2,3)$, $\frac{\partial f}{\partial x} = 4$ and $\frac{\partial f}{\partial y} = 5$.

(a) Draw ∇f at $(2,3)$.
(b) What is the maximal directional derivative of f at $(2,3)$?
(c) For which $\mathbf{u}$ is $D_{\mathbf{u}} f$ at $(2,3)$ maximal? (Write $\mathbf{u}$ in the form $x\mathbf{i} + y\mathbf{j}$.)

8. Assume that at the point $(1,1)$, $\frac{\partial f}{\partial x} = 3$ and $\frac{\partial f}{\partial y} = -3$.

(a) Draw ∇f at $(1,1)$.
(b) What is the maximal directional derivative of f at $(1,1)$?
(c) For which $\mathbf{u}$ is $D_{\mathbf{u}} f$ at $(1,1)$ maximal? (Write $\mathbf{u}$ in the form $x\mathbf{i} + y\mathbf{j}$.)

In Exercises 9 and 10 compute and draw ∇f at the given point.

9. $f(x, y) = x^2 y$ at (a) $(2,5)$ and (b) $(3,1)$.

10. $f(x, y) = \dfrac{1}{\sqrt{x^2 + y^2}}$ at (a) $(1,2)$ and (b) $(3,0)$.

---

11. If the maximal directional derivative of f at (a, b) is 5, what is the minimal directional derivative there? Explain.

12. For a function $f(x, y)$ at (a, b) is there always a direction in which the directional derivative is 0? Explain.

13. If $\frac{\partial f}{\partial x}(a, b) = 2$ and $\frac{\partial f}{\partial y}(a, b) = 3$, in what direction should a directional derivative at (a, b) be computed in order that it be (a) 0? (b) as large as possible? (c) as small as possible?

14. If, at the point (a, b, c), $\frac{\partial f}{\partial x} = 2$, $\frac{\partial f}{\partial y} = 3$, and $\frac{\partial f}{\partial z} = 4$, what is the largest directional derivative of f at (a, b, c)?

15. Assume that $f(1,2) = 2$ and $f(0.99, 2.01) = 1.98$.

(a) Which directional derivatives $D_{\mathbf{u}} f$ at $(1,2)$ can be estimated? (Give $\mathbf{u}$.)
(b) Estimate the directional derivatives in (a).

16. Assume that $f(1,1,1)=3$ and $f(1.1,1.2,1.1)=3.1$.
  (a) Which directional derivatives $D_{\mathbf{u}}f$ at $(1,1,1)$ can be estimated? (Give $\mathbf{u}$.)
  (b) Estimate the directional derivatives in (a).

17. When moving east on the xy-plane, the temperature increases at the rate of $0.02°$ per centimeter. When moving north, the temperature decreases at the rate of $-0.03°$ per centimeter.
  (a) At what rate does the temperature change when going south?
  (b) At what rate does the temperature change when moving $30°$ north of east?
  (c) In what direction should one move to make the temperature change as small as possible? (That is, moving in what direction stays on the same level curve of the temperature?)

18. The temperature increases at the rate of $2°$ per kilometer towards the east and decreases at the rate of $2°$ per kilometer towards the north. In what direction does the temperature
  (a) increase most rapidly? (b) decrease most rapidly? (c) change as little as possible?

19. In the direction $\mathbf{i}$, the temperature increases at the rate of $0.03°$ per centimeter. In the direction $\mathbf{j}$, the temperature decreases at the rate of $0.02°$ per centimeter. In the direction $\mathbf{k}$ the temperature increases at the rate of $0.05°$ per centimeter. Does the temperature tend to increase, decrease, or not change in the direction $\langle 2,5,1\rangle$?

20. Assume that $f(1,2)=3$ and that the directional derivative of f at $(1,2)$ in the direction of the nonunit vector $\mathbf{i}+\mathbf{j}$ is 0.7. Estimate $f(1.1,2.1)$.

21. Assume that $f(1,1,2)=4$ and that the directional derivative of f at $(1,1,2)$ in the direction of the vector from $(1,1,2)$ to $(1.01,1.02,1.99)$ is 3. Estimate $f(0.99,0.98,2.01)$.

In Exercises 22 to 27 find the directional derivative of the function in the given direction and the maximum directional derivative.

22. xyz^2 at $(1,0,1)$, $\mathbf{i}+\mathbf{j}+\mathbf{k}$

23. x^3yz at $(2,1,-1)$, $2\mathbf{i}-\mathbf{k}$

24. $e^{xy\sin(z)}$ at $\left(1,1,\frac{\pi}{4}\right)$, $\mathbf{i}+\mathbf{j}+3\mathbf{k}$

25. $\arctan\left(\sqrt{x^2+y+z}\right)$ at $(1,1,1)$, $-\mathbf{i}$

26. $\ln(1+xyz)$ at $(2,3,1)$, $-\mathbf{i}+\mathbf{j}$

27. $x^x y e^{z^2}$ at $(1,1,0)$, $\mathbf{i}-\mathbf{j}+\mathbf{k}$

---

28. Let $f(x,y,z)=2x+3y+z$.
  (a) Compute ∇f at (i) $(0,0,0)$, and (ii) $(1,1,1)$.
  (b) Draw ∇f for the two points in (a), putting the tail at the point.

29. Let $f(x,y,z)=x^2+y^2+z^2$.
  (a) Compute ∇f at (i) $(2,0,0)$, (ii) $(0,2,0)$, and (iii) $(0,0,2)$.
  (b) Draw ∇f for the three points in (a), putting the tail at the point.

30. Let $T(x,y,z)$ be the temperature at the point (x,y,z). Assume that ∇T at $(1,1,1)$ is $2\mathbf{i}+3\mathbf{j}+4\mathbf{k}$.
  (a) Find $D_{\mathbf{u}}T$ at $(1,1,1)$ if $\mathbf{u}$ is in the direction of $\mathbf{i}-\mathbf{j}+2\mathbf{k}$.
  (b) Estimate the change in temperature moving a distance 0.2 in the direction $\mathbf{i}-\mathbf{j}+2\mathbf{k}$ from the point $(1,1,1)$.
  (c) Find three unit vectors $\mathbf{u}$ such that $D_{\mathbf{u}}T=0$ at $(1,1,1)$.

31. Let $f(x,y)=\dfrac{1}{\sqrt{x^2+y^2}}$ and $\mathbf{r}=\langle x,y\rangle$. Show that (a) $\nabla f=\dfrac{-\mathbf{r}}{|\mathbf{r}|^3}$ and (b) $|\nabla f|=\dfrac{1}{|\mathbf{r}|^2}$.

32. Let $g(x,y,z)=\dfrac{1}{\sqrt{x^2+y^2+z^2}}$ and $\mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$. Express (a) ∇g in terms of $\mathbf{r}$ and (b) $|\nabla g|$ in terms of $\mathbf{r}$.

The functions f and g in Exercises 31 and 32 are not defined at the origin.

33. Let $f(x,y) = x^2+y^2$. Prove that if (a,b) is a point on the curve $x^2+y^2=9$, then ∇f computed at (a,b) is perpendicular to the tangent line to the curve at (a,b).

34. Let $f(x,y,z)$ be the temperature at (x,y,z). Let $P=(a,b,c)$ and Q be a point near (a,b,c). Show that $\nabla f \cdot \overrightarrow{PQ}$ is a good estimate of the change in temperature from P to Q.

35. If $f(P)$ is the electric potential at the point P, then the electric field $\mathbf{E}$ at P is given by $\dfrac{-1}{c^2}\nabla f$, where c is a constant. Calculate $\mathbf{E}$ if $f(x,y) = \sin(\alpha x)\cos(\beta y)$, where α and β are constants.

36. The equality $\dfrac{\partial^2 f}{\partial x \partial y} = \dfrac{\partial^2 f}{\partial y \partial x}$ can be written as $D_{\mathbf{i}}\left(D_{\mathbf{j}} f\right) = D_{\mathbf{j}}\left(D_{\mathbf{i}} f\right)$. Show that $D_{\mathbf{u}_2}\left(D_{\mathbf{u}_1} f\right) = D_{\mathbf{u}_1}\left(D_{\mathbf{u}_2} f\right)$ for any unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$.

REMEMBER: f is assumed to have continuous first- and second-order partial derivatives.

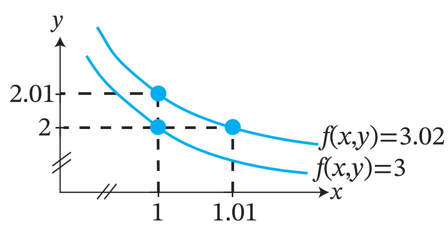


Figure 16.4.6

37. Figure 16.4.6 shows two level curves of $f(x,y)$ near $(1,2)$, namely $f(x,y)=3$ and $f(x,y)=3.02$. Use it to (a) estimate $D_{\mathbf{i}} f$ at $(1,2)$, (b) estimate $D_{\mathbf{j}} f$ at $(1,2)$, and (c) draw ∇f at $(1,2)$.

38. Assume that ∇f at (a,b) is not $\mathbf{0}$. Show that there are two unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ such that the directional derivatives of f at (a,b) in the directions of $\mathbf{u}_1$ and $\mathbf{u}_2$ are 0.

39. Assume that ∇f at (a,b,c) is not $\mathbf{0}$. How many unit vectors $\mathbf{u}$ are there such that $D_{\mathbf{u}} f(a,b,c) = 0$? Explain.

40. Let $f(x,y)$ be the temperature at (x,y). Assume that ∇f at $(1,1)$ is $2\mathbf{i}+3\mathbf{j}$. A particle moves northwest at the rate of 3 centimeters per second. Let $g(t)$ be the temperature at the point where the particle is at time t seconds. Then $\dfrac{dg}{dt}$ is the rate at which temperature changes on the particle's journey (degrees per second.) Find $\dfrac{dg}{dt}$ when the particle is at $(1,1)$.

41. Assume that f is defined throughout the xy-plane. If $f(x,y) = g(y)$, then $\dfrac{\partial f}{\partial x} = 0$. Is the converse true? That is, if $\dfrac{\partial f}{\partial x} = 0$, is there a function g of one variable such that $f(x,y) = g(y)$?

42. Let f and g be scalar functions defined throughout the xy-plane. Assume they have the same gradient, $\nabla f = \nabla g$ at all points. Must $f = g$? Is there any relation between f and g?

43. Let $f(x,y) = xy$.

(a) Draw the level curve $xy = 4$ carefully.

(b) Compute ∇f at three points on the level curve and draw it with its tail at the point where it is evaluated.

(c) What angle does ∇f seem to make with the curve at the point where it is evaluated?

(d) Prove that the angle is what you think it is.

44. Prove, without the aid of vectors, that the maximum value of $g(\theta) = \dfrac{\partial f}{\partial x}(a,b)\cos(\theta) + \dfrac{\partial f}{\partial y}(a,b)\sin(\theta)$ is $\sqrt{\left(\dfrac{\partial f}{\partial x}(a,b)\right)^2 + \left(\dfrac{\partial f}{\partial y}(a,b)\right)^2}$.

NOTE: This is the first part of Theorem 16.4.3.

45. **SAM:** They said that if you know $\nabla f \cdot \mathbf{u}$ for every unit vector $\mathbf{u}$, then you know ∇f.
JANE: What's wrong with that?
SAM: What if there's another unit vector $\mathbf{A}$ such that $\mathbf{A}\cdot\mathbf{u} = D_{\mathbf{u}} f$ for every vector $\mathbf{u}$?
JANE: So?
SAM: Then you don't pinpoint ∇f by its dot products with all unit vectors.
JANE: Maybe they have to rewrite this section?
SAM: If so, it's their own fault. Sixteen chapters have taught me to question authority.

Is Sam right? Must part of this section be rewritten?

16.5 Normal Vectors and Tangent Planes

In this section we find a normal vector to a curve whose equation is $f(x, y) = k$ and to a surface whose equation is $f(x, y, z) = k$. We also find the tangent plane at a point on a surface.

Normal Vectors to a Curve in the xy-Plane

We saw in Section 14.4 how to find a normal vector to a curve when the curve is given parametrically. Now we will learn how to find a normal vector when the curve is given as a level curve, that is as the graph of $f(x, y) = k$. We assume that functions are well behaved, meaning curves have continuous tangent vectors and functions have continuous partial derivatives.

Theorem 16.5.1: Normal Vector to a Level Curve

The gradient ∇f at (a, b) is a normal vector to the level curve of f passing through (a, b).

Proof of Theorem 16.5.1

Let $\mathbf{G}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}$ be a parameterization of the level curve of f that passes through $(a, b) = \mathbf{G}(t_0)$. On it $f(x, y)$ is a constant and has value $f(a, b)$. The tangent vector to the curve at (a, b) is $\mathbf{G}'(t_0) = x'(t_0)\mathbf{i} + y'(t_0)\mathbf{j}$ and the gradient of f at (a, b) is $\nabla f(a, b) = f_x(a, b)\mathbf{i} + f_y(a, b)\mathbf{j}$. We wish to show that

$$\nabla f \cdot \mathbf{G}'(t_0) = 0.$$

We will do this by showing that

$$\frac{\partial f}{\partial x}(a, b)\frac{dx}{dt}(t_0) + \frac{\partial f}{\partial y}(a, b)\frac{dy}{dt}(t_0) = 0. \tag{16.5.1}$$

The left-hand side of (16.5.1) has the form of a chain rule. To exploit this let

$$u(t) = f(x(t), y(t)).$$

Because f has the same value at every point on a level curve, $u(t) = f(a, b)$. Thus $u(t)$ is a constant function, which implies $du/dt = 0$.

Because u can be viewed as a function of x and y, where x and y are functions of t, the chain rule gives

$$\frac{du}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Since $du/dt = 0$, (16.5.1) follows. Hence ∇f evaluated at (a, b) is a normal vector to the level curve of f that passes through (a, b).

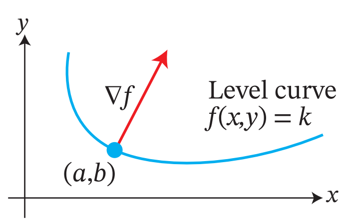


Figure 16.5.1

What does the theorem say about a weather map that shows the barometric pressure? On level curves the pressures are equal. The gradient ∇f points in the direction in which the pressure increases most rapidly. Consequently $-\nabla f$ points in the direction the pressure is decreasing most rapidly. Since the wind tends to go from high pressure to low pressure, we can think of $-\nabla f$ as representing the wind. (Earth's rotation also affects the wind.) Figure 16.5.1 shows a level curve and gradient. The gradient is perpendicular to the level curve. As we saw in Section 16.4, the gradient points in the direction in which the function increases most rapidly.

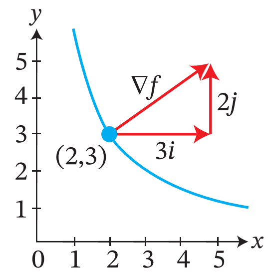


Figure 16.5.2

EXAMPLE 1. Find and draw a normal vector to the hyperbola $xy = 6$ at the point $(2,3)$.

SOLUTION Let $f(x, y) = xy$. Then $f_x = y$ and $f_y = x$, so

$$\nabla f = y\mathbf{i} + x\mathbf{j}$$

and

$$\nabla f(2,3) = 3\mathbf{i} + 2\mathbf{j}.$$

This level curve (cyan) $xy = 6$ and its gradient at $(2,3)$ are shown in Figure 16.5.2. ♦

EXAMPLE 2. Find an equation of the tangent line to the ellipse $x^2 + 3y^2 = 7$ at the point $(2,1)$.

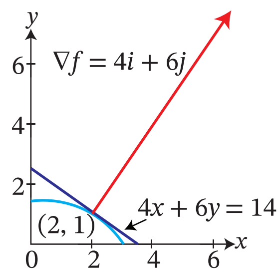


Figure 16.5.3

SOLUTION As we saw in Section 14.4, we may write the equation of a line in the plane if we know a point on it and a vector normal to it. We know that $(2,1)$ lies on the tangent line. We use the gradient to produce a normal vector to the curve, and to the tangent line, at the point $(2,1)$.

The ellipse $x^2 + 3y^2 = 7$ is a level curve of $f(x, y) = x^2 + 3y^2$. Since $f_x = 2x$ and $f_y = 6y$, $\nabla f = 2x\mathbf{i} + 6y\mathbf{j}$. In particular

$$\nabla f(2,1) = 4\mathbf{i} + 6\mathbf{j}.$$

Hence the tangent line at $(2,1)$ has the equation

$$4(x-2) + 6(y-1) = 0 \qquad \text{or} \qquad 4x + 6y = 14.$$

The level curve (cyan), normal vector (red), and tangent line (blue) are shown in Figure 16.5.3. ♦

Normal Vectors to a Surface

We can construct a vector perpendicular to a surface $f(x, y, z) = k$ at a point $P = (a, b, c)$ as we constructed a vector perpendicular to a curve. The gradient vector ∇f evaluated at (a, b, c) is perpendicular to the surface $f(x, y, z) = k$. The proof of this is similar to the proof of Theorem 16.5.1 for normal vectors to a level curve.

To define what is meant by a vector being perpendicular to a surface it is important to remember that a smooth curve is a curve with a unit normal vector that is continuous. (For example, a circle or ellipse is a smooth curve but a square or any polygon is not — because the unit normal is discontinuous at each corner.)

Definition: *Normal Vector to a Surface*

A vector is perpendicular to a surface at (a, b, c) on the surface if it is perpendicular to each smooth curve on the surface through (a, b, c). Such a vector is called a **normal vector** to the surface.

The next theorem provides a way to find normal vectors to a level surface $f(x, y, z) = k$.

Theorem 16.5.2: Normal Vectors to a Level Surface

The gradient ∇f at (a, b, c) is a normal to the level surface of f passing through (a, b, c).

The proof consists of showing that ∇f at (a,b,c) is perpendicular to each curve in the level surface of f at (a,b,c). It differs from the proof of Theorem 16.5.1 only in that the vectors have three components instead of two.

A check of this theorem is to see whether it is correct when the level surfaces are planes. Let $f(x,y,z) = Ax + By + Cz + D$. The plane $Ax + By + Cz + D = 0$ is the level surface $f(x,y,z) = 0$. According to the theorem, ∇f is perpendicular to it. Because $f_x = A$, $f_y = B$, and $f_z = C$,

$$\nabla f = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k} = A\mathbf{i} + B\mathbf{j} + C\mathbf{k}.$$

This agrees with the fact that $A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$ is normal to the plane, as we saw by vector algebra in Section 14.4.

EXAMPLE 3. Find a normal vector to the ellipsoid $x^2 + \frac{y^2}{4} + \frac{z^2}{9} = 3$ at $(1,2,3)$.

SOLUTION The ellipsoid is a level surface of $f(x,y,z) = x^2 + y^2/4 + z^2/9$, which has gradient $\nabla f = 2x\mathbf{i} + (y/2)\mathbf{j} + (2z/9)\mathbf{k}$. A normal vector to the ellipsoid at the point $(1,2,3)$, is the gradient vector

$$\nabla f(1,2,3) = 2\mathbf{i} + \mathbf{j} + \frac{2}{3}\mathbf{k}.$$

◆

Tangent Planes to a Surface

Now that we can find a normal to a surface we can define a tangent plane at a point on the surface.

Definition: *Tangent Plane to a Surface*

Let (a,b,c) be a point on a surface. The tangent plane to the surface at (a,b,c) is the plane through (a,b,c) that is perpendicular to a normal vector to the surface at (a,b,c).

The next theorem provides a way to find the tangent plane to a level surface $f(x,y,z) = k$. This result is a direct consequence of the previous definition of a tangent plane and the fact that the gradient ∇f at (a,b,c) is a normal vector to the level surface of f passing through (a,b,c) (see Theorem 16.5.2).

Theorem 16.5.3: Tangent Plane to a Level Surface

Let (a,b,c) be a point on a level surface of f at which ∇f is not $\mathbf{0}$. The tangent plane there is the plane through (a,b,c) that is perpendicular to ∇f evaluated at (a,b,c):

$$\nabla f(a,b,c) \cdot \langle x-a, y-b, z-c\rangle = 0.$$

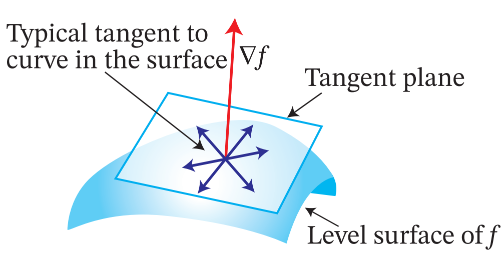


Figure 16.5.4

The tangent plane at (a,b,c) is the plane that best approximates the surface near (a,b,c). It consists of all the tangent lines to curves in the surface that pass through (a,b,c). See Figure 16.5.4.

Recall that an equation of the plane through (x_0, y_0, z_0) perpendicular to $\langle A, B, C\rangle$ is $A(x-x_0)+B(y-y_0)+C(z-z_0)=0$. Then, an equation of the tangent plane to the surface $f(x, y, z) = k$ at (a, b, c) is

$$\frac{\partial f}{\partial x}(a,b,c)(x-a)+\frac{\partial f}{\partial y}(a,b,c)(y-b)+\frac{\partial f}{\partial z}(a,b,c)(z-c)=0.$$

EXAMPLE 4. Find an equation of the tangent plane to the ellipsoid $x^2+\frac{y^2}{4}+\frac{z^2}{9}=3$ at $(1,2,3)$.

SOLUTION By Example 3, $2\mathbf{i}+\mathbf{j}+(2/3)\mathbf{k}$ is normal to the surface at the point $(1,2,3)$. The tangent plane consequently is described by the equation

$$2(x-1)+1(y-2)+\frac{2}{3}(z-3)=0.$$

This equation can be simplified to $2x+y+2z/3=6$ or, clearing fractions, $6x+3y+2z=18$. ♦

Normal Vectors and Tangent Planes to $z = f(x, y)$

A surface may be described explicitly in the form $z = f(x, y)$ rather than implicitly in the form $f(x, y, z) = k$. The techniques we developed enable us to find the normal and tangent plane for $z = f(x, y)$ as well.

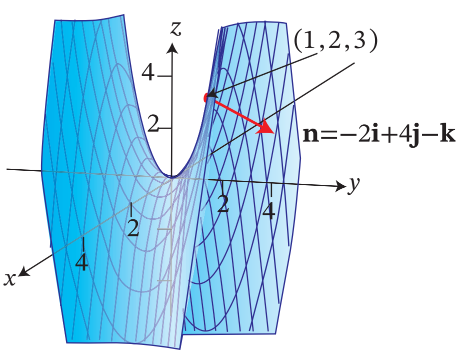


Figure 16.5.5

If we write $z = f(x, y)$ as $f(x, y)-z=0$ and let $g(x, y, z)$ be $f(x, y)-z$, then the surface $f(x, y)-z=0$ is the level surface $g(x, y, z)=0$. There is no need to memorize a formula for a vector normal to the surface $z = f(x, y)$, as illustrated in the next example.

EXAMPLE 5. Find a vector perpendicular to $z = y^2-x^2$ at $(1,2,3)$.

SOLUTION Write $z = y^2-x^2$ as $-x^2+y^2-z=0$, which is a level surface of $g(x, y, z)=-x^2+y^2-z$. Hence $\nabla g=-2x\mathbf{i}+2y\mathbf{j}-\mathbf{k}$ and therefore $\nabla g(1,2,3)=-2\mathbf{i}+4\mathbf{j}-\mathbf{k}$ is perpendicular to the surface $z = y^2-x^2$ at the point $(1,2,3)$.
The surface looks like a saddle near the origin. The surface (cyan) and the normal vector (red) are shown in Figure 16.5.5. ♦

Estimates by Tangent Planes

For a function of one variable, $y = f(x)$, the tangent line at $(a, f(a))$ closely approximates the graph of $y = f(x)$ near $(a, f(a))$. The equation of the tangent line, $y = f(a)+f'(a)(x-a)$, gives us a linear approximation of $f(x)$. (See Section 5.5.)

We can use the tangent plane to the surface $z = f(x, y)$ similarly. To find the equation of the plane tangent at $(a, b, f(a,b))$, we first write the equation of the surface as

$$g(x, y, z) = f(x, y)-z=0.$$

Then $\nabla g(a,b)$ is a normal to the surface at $(a, b, f(a,b))$. Because

$$\nabla g(a,b)=\frac{\partial f}{\partial x}(a,b)\mathbf{i}+\frac{\partial f}{\partial y(a,b)}\mathbf{j}-\mathbf{k},$$

the equation of the tangent plane at $(a, b, f(a,b))$ is

$$\frac{\partial f}{\partial x}(a,b)(x-a)+\frac{\partial f}{\partial y}(a,b)(y-b)-(z-f(a,b))=0.$$

We can write this as

$$z = f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b). \tag{16.5.2}$$

Letting $\Delta x = x - a$ and $\Delta y = y - b$, (16.5.2) becomes

$$z = f(a,b) + \frac{\partial f}{\partial x}(a,b)\Delta x + \frac{\partial f}{\partial y}(a,b)\Delta y.$$

From this we learn that the change of the z-coordinate on the tangent plane, as the x-coordinate changes from a to $a+\Delta x$ and the y-coordinate changes from b to $b+\Delta y$, is exactly

$$\frac{\partial f}{\partial x}(a,b)\Delta x + \frac{\partial f}{\partial y}(a,b)\Delta y.$$

By (16.3.7) in Section 16.3, this is an estimate of the change Δf in the function f as its argument changes from (a,b) to $(a+\Delta x, b+\Delta y)$. This is another way of saying that the tangent plane to $z = f(x,y)$ at $(a,b,f(a,b))$ is close to the surface. See Figure 16.5.6.

The expression $f_x(a,b)\,dx + f_y(a,b)\,dy$ is called the **differential** of f at (a,b). For small values of dx and dy it is a good estimate of $f(a+dx, b+dy) - f(a,b)$.

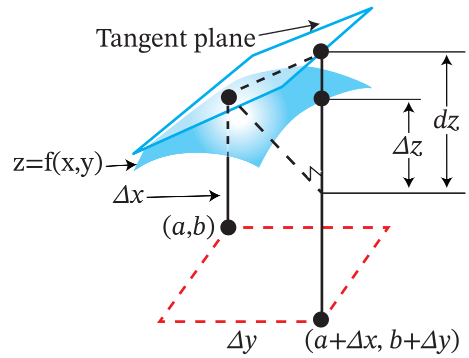


Figure 16.5.6

EXAMPLE 6. Let $z = f(x,y) = x^2 y$. Let $\Delta z = f(1.01, 2.02) - f(1,2)$ and let $dz = \frac{\partial f}{\partial x}(1,2)\cdot 0.01 + \frac{\partial f}{\partial y}(1,2)\cdot 0.02$. Compute Δz and dz.

SOLUTION First,

$$\Delta z = f(1.01, 2.02) - f(1,2) = (1.01)^2(2.02) - 1^2 2 = 2.060602 - 2 = 0.060602.$$

Since $f_x = 2xy$ and $f_y = x^2$, at the point $(1,2)$ $f_x(1,2) = 4$ and $f_y(1,2) = 1$. Hence,

$$dz = \frac{\partial f}{\partial x}(1,2)\cdot 0.01 + \frac{\partial f}{\partial y}(1,2)\cdot 0.02 = (4)(0.01) + (1)(0.02) = 0.06.$$

The differential of f at $(1,2)$ produces a good approximation of Δz at $(1,2)$. ♦

Summary

The following table summarizes what we know about the normal and tangent vectors to a curve in the plane, or to a surface in space.

Function	Level Curve / Surface	Normal Vector	Tangent Line / Plane
$f(x,y)$	$f(x,y) = k$ (level curve)	$\nabla f = f_x\mathbf{i} + f_y\mathbf{j}$	$f_x(a,b)(x-a) + f_y(a,b)(y-b) = 0$ (tangent line at (a,b))
$f(x,y,z)$	$f(x,y,z) = k$ (level surface)	$\nabla f = f_x\mathbf{i} + f_y\mathbf{j} + f_z\mathbf{k}$	$f_x(a,b,c)(x-a)$ $+ f_y(a,b,c)(y-b)$ $+ f_z(a,b,c)(z-c) = 0$ (tangent plane at (a,b,c))

Table 16.5.1

To find a normal vector or tangent plane to a surface given as $z = f(x,y)$, treat the surface as a level surface of $f(x,y) - z$, namely $f(x,y) - z = 0$.

Recall that we have previously shown, in Section 16.3, that the differential approximation of Δf is the change in f along the tangent plane.

EXERCISES for Section 16.5

The angle between two surfaces that pass through (a, b, c) is defined as the angle between the lines through (a, b, c) that are perpendicular to the surfaces at the point (a, b, c). This angle is always chosen to be either acute or a right angle. Use this definition in Exercises 1 to 3.

1. (a) Show that $(1, 1, 2)$ lies on the surfaces $xyz = 2$ and $x^3yz^2 = 4$.
(b) Find the angle between the surfaces at $(1, 1, 2)$.

2. (a) Show that $(1, 2, 3)$ lies on the plane $2x + 3y - z = 5$ and the sphere $x^2 + y^2 + z^2 = 14$.
(b) Find the angle between them at the point $(1, 2, 3)$.

3. (a) Show that the surfaces $z = x^2y^3$ and $z = 2xy$ pass through $(2, 1, 4)$.
(b) At what angle do they cross at that point?

---

4. Let $z = f(x, y)$ describe a surface. Assume that at $(3, 5)$, $z = 7$, $\frac{\partial z}{\partial x} = 2$, and $\frac{\partial z}{\partial y} = 3$. (a) Find a normal to the surface at $(3, 5, 7)$. (b) Find two vectors that are tangent to it at $(3, 5, 7)$. (c) Estimate $f(3.02, 4.99)$.

5. Let $T(x, y, z)$ be the temperature at the point (x, y, z), where ∇T is not $\mathbf{0}$. A level surface $T(x, y, z) = k$ is called an **isotherm**. Show that if you are at the point (a, b, c) and wish to move in the direction in which the temperature changes most rapidly, you would move in a direction perpendicular to the isotherm that passes through (a, b, c).

6. Write a short essay on the chain rule. Include a description of how it was used to show that $D_{\mathbf{u}}f = \nabla f \cdot \mathbf{u}$ and that ∇f is a normal to a level surface of f.

7. Suppose you are at (a, b, c) on the level surface $f(x, y, z) = k$. There $\nabla f = 2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}$.
(a) If $\mathbf{u}$ is tangent to the surface at (a, b, c), what would $D_{\mathbf{u}}f$ equal? (There are infinitely many $\mathbf{u}$s.)
(b) If $\mathbf{u}$ is normal to the level surface at (a, b, c), what would $D_{\mathbf{u}}f$ equal? (There are two $\mathbf{u}$s.)

8. We have found a way to find a normal and a tangent plane to a surface. How would you find a tangent line to a surface? Illustrate your method by finding a line that is tangent to $z = xy$ at $(2, 3, 6)$.

9. (a) Draw three level curves of $f(x, y) = xy$. Include the curve through $(1, 1)$ as one of them.
(b) Draw three level curves of $g(x, y) = x^2 - y^2$. Include the curve through $(1, 1)$ as one of them.
(c) Prove that a level curve of f intersects a level curve of g at a right angle.
(d) If we think of f as air pressure, how may we interpret the level curves of g?

10. (a) Draw a level curve for $2x^2 + y^2$.
(b) Draw a level curve for $\frac{y^2}{x}$.
(c) Prove that a level curve of $2x^2 + y^2$ crosses a level curve of $\frac{y^2}{x}$ at a right angle.

11. Two surfaces $f(x, y, z) = 0$ and $g(x, y, z) = 0$ pass through (a, b, c). Their intersection is a curve. How would you find a tangent vector to it at (a, b, c)?

12. Prove Theorem 16.5.3.

13. The surfaces $x^2yz = 1$ and $xy + yz + zx = 3$ pass through $(1, 1, 1)$. Their tangent planes meet in a line. Find parametric equations for it.

14. The map in Figure 16.5.7 shows isobars, level curves of the pressure $p(x, y)$. At which of the labeled points on the map is the gradient of p, ∇p, the longest? In what direction does it point? In which direction (approximately) would the wind vector point?

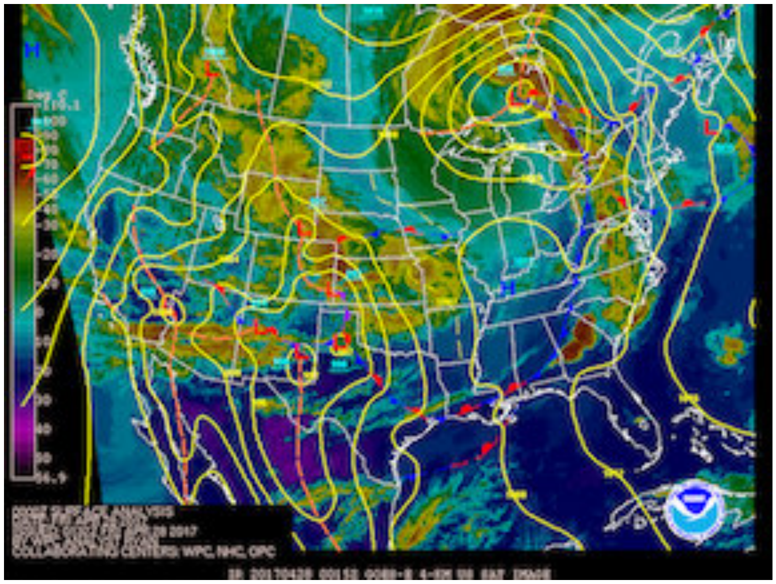


Figure 16.5.7

15. The surface $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}+\dfrac{z^2}{c^2}=1$ is called an ellipsoid. If $a^2=b^2=c^2$ it is a sphere. Show that if a^2, b^2, and c^2 are distinct, then there are exactly six normal vectors to the ellipsoid that pass through the origin.

16. If $f(x)$ is defined for all x and its derivative is always 0, it is constant. Assume $f(x, y)$ is defined at all points (x, y) and its gradient is always $\mathbf{0}$. Must $f(x, y) = C$ for some constant C?

17. Let $\mathscr{S}$ be a surface with equation $f(x, y, z) = 0$, such that each ray from the origin O meets $\mathscr{S}$ at exactly one point. Assume that at each point P in $\mathscr{S}$, the tangent plane at P is perpendicular to the radial vector $\mathbf{r} = \overrightarrow{OP}$. Show that $\mathscr{S}$ is a sphere.

18. Assume that $f(x, y, z)$ and $g(x, y, z)$ have the property that ∇f and ∇g, evaluated at the same point, are always parallel with the same nonzero scale factor throughout the domains of f and g. Show that they have the same level surfaces.

19. How far is the point $(2,1,3)$ from the tangent plane to $z = xy$ at $(3,4,12)$?

20. How far is the tangent plane to $xyz = 1$ at $(1,1,1)$ from the point $(2,1,3)$?

21. Find three vectors that are tangent to the surface $x^2 + 3y^2 + 4z^2 = 8$ at $(1,1,1)$.

In Exercises 22 to 23 check the equations by differentiation.

22. $\displaystyle\int \cos^3(ax)\,dx = \frac{1}{a}\sin(ax) - \frac{1}{3a}\sin^3(ax)$

23. $\displaystyle\int (p^2-x^2)^{3/2}\,dx = \frac{x}{4}(p^2-x^2)^{3/2} + \frac{3p^2x}{8}\sqrt{p^2-x^2} + \frac{3p^4}{8}\arcsin\left(\frac{x}{p}\right)$

In Exercises 24 and 25 evaluate each antiderivative.

24. $\displaystyle\int \frac{x+2}{\sqrt{4x^2+9}}\,dx$

25. $\displaystyle\int \frac{x+2}{4x^2+9}\,dx$

16.6 Critical Points and Extrema

For a function of one variable, $y = f(x)$, the first and second derivatives were of use in searching for relative extrema. We looked for critical numbers, that is, solutions of the equation $f'(x) = 0$ and checked the value of $f''(x)$ there. If $f''(x)$ was positive, a critical number gave a relative minimum. If $f''(x)$ was negative, a critical number gave a relative maximum. If $f''(x)$ was 0, then anything might happen: a minimum, maximum, or neither; other tests are needed to classify the critical point.

This section extends the idea of a critical point to functions $f(x, y)$ of two variables and shows how to use the second-order partial derivatives f_{xx}, f_{yy}, and f_{xy} to see whether the critical point provides a relative maximum, relative minimum, or neither.

Extrema of $f(x, y)$

The number M is called the **maximum** (or **global maximum**) of f over a set R in the plane if it is the largest value of $f(x, y)$ for (x, y) in R. A **relative maximum** (or **local maximum**) of f occurs at a point (a, b) in R if there is a disk around (a, b) such that $f(a, b) \geq f(x, y)$ for all points (x, y) in the disk. **Minimum** and **relative** (or **local) minimum** are defined similarly.

Consider the surface $z = f(x, y)$ above a point (a, b) where a relative maximum of f occurs. This implies that f is defined for all points in some disk around (a, b). In addition, assume f possesses partial derivatives at (a, b). Assume, for convenience, that the values of f are positive. Let L_1 be the line $y = b$ in the xy-plane; let L_2 be the line $x = a$ in the xy-plane. (See Figure 16.6.1.)

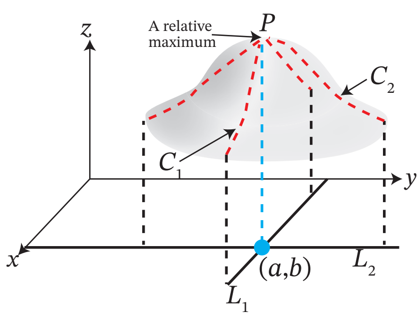


Figure 16.6.1

Let C_1 be the curve on the surface above the line L_1, C_2 be the curve on the surface above the line L_2, and let P be the point on the surface above (a, b).

Since f has a relative maximum at (a, b), no point on the surface near P is higher than P. Thus P is a highest point on C_1 and on C_2 (for points near P). So both curves have horizontal tangents at P. That is, at (a, b) both partial derivatives of f must be 0:

$$\frac{\partial f}{\partial x}(a, b) = 0 \qquad \text{and} \qquad \frac{\partial f}{\partial y}(a, b) = 0.$$

This conclusion is summarized in the following theorem.

Theorem 16.6.1: Relative Extremum of $f(x, y)$

Let f be defined on a region that includes the point (a, b) and all points in some disk whose center is (a, b). If f has a relative maximum (or relative minimum) at (a, b) and f_x and f_y exist at (a, b), then

$$\frac{\partial f}{\partial x}(a, b) = 0 = \frac{\partial f}{\partial y}(a, b),$$

which says that ∇f is $\mathbf{0}$ at a relative extremum.

Points (a, b) where both partial derivatives f_x and f_y are 0 are clearly of importance. They are analogous to the critical points of a function of one variable.

Definition: *Critical Point*

If $f_x(a, b) = 0$ and $f_y(a, b) = 0$, the point (a, b) is a **critical point** of the function $f(x, y)$.

One might expect that if (a,b) is a critical point of f and the two second partial derivatives f_{xx} and f_{yy} are both positive at (a,b), then f necessarily has a relative minimum at (a,b). The next example shows that this is false.

EXAMPLE 1. Find the critical points of $f(x,y) = x^2 + 3xy + y^2$ and determine whether there are extrema there.

SOLUTION Find critical points by setting both f_x and f_y equal to 0. This leads to $2x + 3y = 0$ and $3x + 2y = 0$, whose only solution is $(x,y) = (0,0)$. The function has one critical point, $(0,0)$.

Now look at the graph of f for (x,y) near $(0,0)$. Because $f(x,0) = x^2 + 3\cdot x\cdot 0 + 0^2 = x^2$, considered as a function of x, the function has a minimum at the origin. (See the solid red parabola in Figure 16.6.2(a).)

On the y-axis, the function reduces to $f(0,y) = y^2$, whose graph is another parabola with a minimum at the origin. (See the dashed red parabola in Figure 16.6.2(a).) Also, $f_{xx} = 2$ and $f_{yy} = 2$, so both are positive at $(0,0)$.

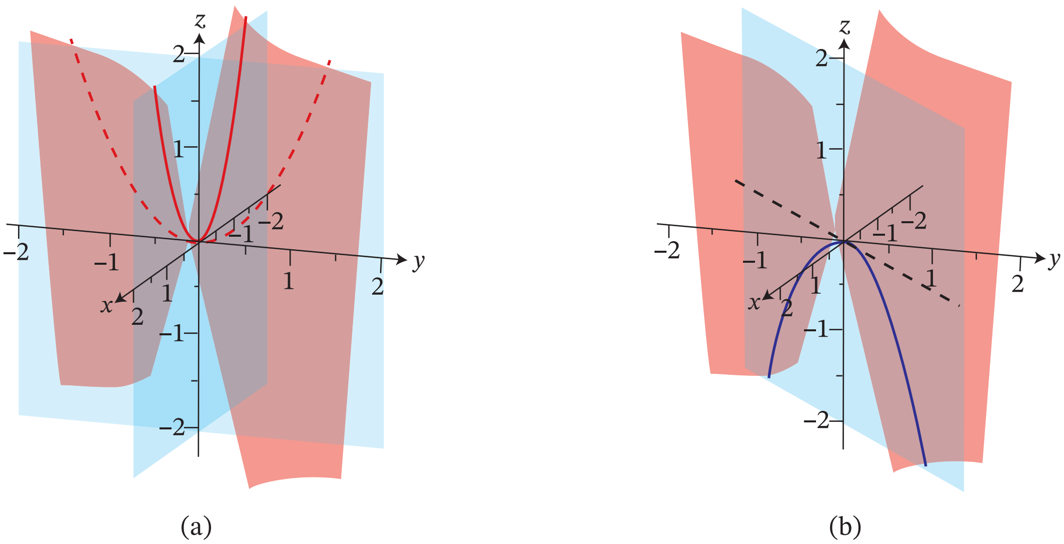


Figure 16.6.2

We might think that f has a relative minimum at $(0,0)$. However, on the line $y = -x$,

$$f(x,y) = f(x,-x) = x^2 + 3x(-x) + (-x)^2 = -x^2.$$

On that line the function assumes nonpositive values, and its graph is a parabola opening downward, with a maximum at $(0,0)$. (See the solid blue parabola in Figure 16.6.2(b).)

Thus $f(x,y)$ has neither a relative maximum nor minimum at the origin. Its graph resembles a saddle. ◆

Example 1 shows that to determine whether a critical point of $f(x,y)$ provides an extremum, it is not enough to look at f_{xx} and f_{yy}. The criteria are more complicated and involve the mixed partial derivative f_{xy} as well. Exercise 59 outlines a proof of Theorem 16.6.2. At the end of this section we give a proof when $f(x,y)$ is a polynomial of the form $Ax^2 + Bxy + Cy^2$, where A, B and C are constants.

Theorem 16.6.2: Second-Partial-Derivative Test for Extrema of $f(x,y)$

*Let (a,b) be a critical point of the function $f(x,y)$. Assume that the partial derivatives f_x, f_y, f_{xx}, f_{xy}, and f_{yy} are continuous at and near (a,b). The **discriminant** of f at (a,b) is*

$$\begin{aligned} D(a,b) &= \frac{\partial^2 f}{\partial x^2}(a,b)\frac{\partial^2 f}{\partial y^2}(a,b) - \left(\frac{\partial^2 f}{\partial x \partial y}(a,b)\right)^2 \\ &= f_{xx}(a,b) f_{yy}(a,b) - \left(f_{xy}(a,b)\right)^2. \end{aligned}$$

Case 1: *If $D(a,b) > 0$ and $f_{xx}(a,b) > 0$, then f has a relative minimum at (a,b).*

Case 2: *If $D(a,b) > 0$ and $f_{xx}(a,b) < 0$, then f has a relative maximum at (a,b).*

Case 3: *If $D(a,b) < 0$, then f has a* ***saddle point*** *at (a,b).*
(That is, the critical point (a,b) is neither a relative minimum nor a relative maximum of f.)

NOTE: *If $D(a,b) = 0$ anything can happen at the critical point (a,b): there may be a relative minimum, a relative maximum, or a saddle point. (See Exercise 43.)*

To see what the theorem says, consider Case 1, the test for a relative minimum. The hypotheses are that $f_x(a,b) = f_y(a,b) = 0$, $f_{xx}(a,b) > 0$, and $D > 0$. This means

$$\frac{\partial^2 f}{\partial x^2}(a,b)\frac{\partial^2 f}{\partial y^2}(a,b) - \left(\frac{\partial^2 f}{\partial x \partial y}(a,b)\right)^2 > 0,$$

or, equivalently,

$$\left(\frac{\partial^2 f}{\partial x \partial y}(a,b)\right)^2 < \frac{\partial^2 f}{\partial x^2}(a,b)\frac{\partial^2 f}{\partial y^2}(a,b). \tag{16.6.1}$$

Since the square of a real number is never negative and $f_{xx}(a,b)$ is positive, it follows that $f_{yy}(a,b)$ must also be positive.

But (16.6.1) says more. It says that the mixed partial $f_{xy}(a,b)$ must not be too large. This form of the discriminant condition might be easier to remember in some situations.

At a critical point (a,b) where (16.6.1) holds, then the discriminant is positive and the graph of $z = f(x,y)$ has either a relative maximum or a relative minimum. To determine which, just examine the sign of $f_{xx}(a,b)$ or $f_{yy}(a,b)$. (Why must both of these have the same sign?) If either is positive, then (a,b) is a relative minimum. And, if either is negative, then (a,b) is a relative maximum.

EXAMPLE 2. Find all relative extrema of:
(a) $f(x,y) = x^2 + 3xy + y^2$, (b) $g(x,y) = x^2 + 2xy + y^2$, and (c) $h(x,y) = x^2 + xy + y^2$.

SOLUTION

(a) The function $f(x,y) = x^2 + 3xy + y^2$ was studied in Example 1, where it was found that the origin is the only critical point, and it is provides neither a relative maximum nor a relative minimum (hence, (0,0 is a saddle point). We can check this by the discriminant:

$$\frac{\partial^2 f}{\partial x^2}(0,0) = 2, \qquad \frac{\partial^2 f}{\partial x \partial y}(0,0) = 3, \qquad \text{and} \quad \frac{\partial^2 f}{\partial y^2}(0,0) = 2.$$

Hence $D(0,0) == 2\cdot 2 - 3^2 = -5$ is negative. By Theorem 16.6.1, $(0,0)$ is a saddle point, there is neither a relative maximum nor a relative minimum at the origin.

(b) It is straightforward to show that all points on the line $x + y = 0$ are critical points of $g(x,y) = x^2 + 2xy + y^2$. So $(x,-x)$ is a critical point for any number x. Because

$$\frac{\partial^2 g}{\partial x^2}(x,y) = 2, \qquad \frac{\partial^2 g}{\partial x \partial y}(x,y) = 2, \qquad \text{and} \quad \frac{\partial^2 g}{\partial y^2}(x,y) = 2,$$

the discriminant $D(x,y) = 2{\cdot}2 - 2^2 = 0$. Since $D(x,y) = 0$ for every critical point of g, it gives no information. Observe that $g(x,y) = (x+y)^2$ is constant on the line $x + y = c$. (See Figure 16.6.3(a).) Also, $x^2 + 2xy + y^2 = (x+y)^2$ is always greater than or equal to 0. Every point $(x,-x)$ on the line $x + y = 0$ is a relative minimum of $x^2 + 2xy + y^2$.

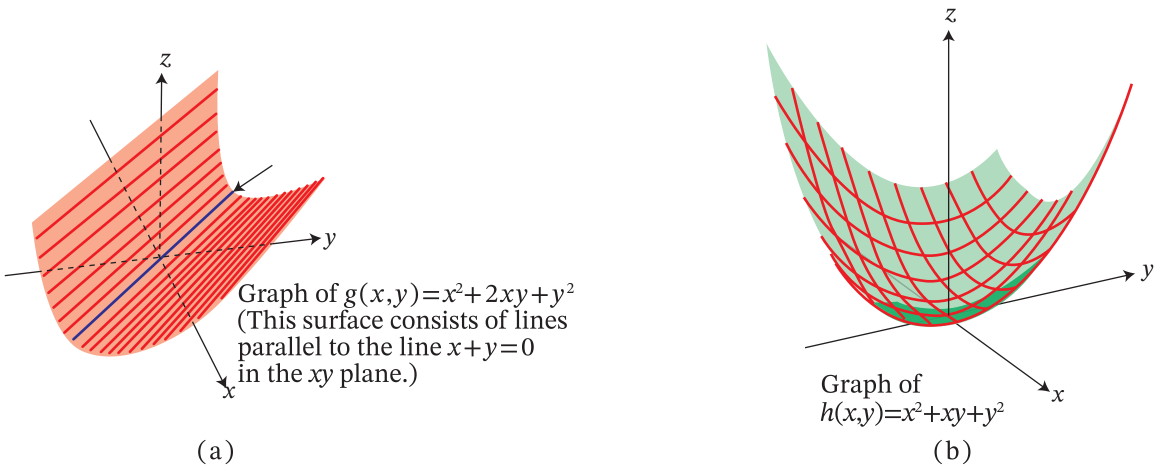


Figure 16.6.3

(c) For $h(x, y) = x^2 + xy + y^2$, it is easy to check that the origin is the only critical point and we have

$$\frac{\partial^2 h}{\partial x^2}(0,0) = 2, \qquad \frac{\partial^2 h}{\partial x \partial y}(0,0) = 1, \qquad \text{and} \qquad \frac{\partial^2 h}{\partial y^2}(0,0) = 2.$$

Then, $D = 2\cdot 2 - 1^2 = 3$ is positive and $h_{xx}(0,0) > 0$. Hence $x^2 + xy + y^2$ has a relative minimum at the origin. The graph of h, shown in Figure 16.6.3(b), clearly shows the relative minimum at the origin. ♦

EXAMPLE 3. Examine $f(x, y) = x + y + \dfrac{1}{xy}$ for global and relative extrema.

SOLUTION When x and y are both large positive numbers or small positive numbers, then $f(x, y)$ may be arbitrarily large. There is therefore no global maximum. By allowing x and y to be negative numbers of large absolute values, we see that there is also no global minimum.

Local extrema will occur at critical points. We have

$$\frac{\partial f}{\partial x} = 1 - \frac{1}{x^2 y} \qquad \text{and} \qquad \frac{\partial f}{\partial y} = 1 - \frac{1}{xy^2}.$$

Setting each partial derivative equal to 0 gives

$$\frac{1}{x^2 y} = 1 \qquad \text{and} \qquad \frac{1}{xy^2} = 1. \tag{16.6.2}$$

Hence $x^2 y = xy^2$. Since f is not defined when x or y is 0, we may assume $xy \neq 0$. Dividing both sides of $x^2 y = xy^2$ by xy leaves $x = y$. By (16.6.2), $1/x^3 = 1$ so $x = 1$, and so $y = 1$. Thus there is only one critical point, $(1,1)$.

To find whether it is a relative extremum, use Theorem 16.6.2. We have

$$\frac{\partial^2 f}{\partial x^2} = \frac{2}{x^3 y}, \qquad \frac{\partial^2 f}{\partial x \partial y} = \frac{1}{x^2 y^2}, \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y^2} = \frac{2}{xy^3}.$$

Thus at $(1,1)$,

$$\frac{\partial^2 f}{\partial x^2}(1,1) = 2, \qquad \frac{\partial^2 f}{\partial x \partial y}(1,1) = 1, \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y^2}(1,1) = 2.$$

Therefore, $D(1,1) = 2\cdot 2 - 1^2 = 3 > 0$. Since the discriminant is positive and $f_{xx}(1,1) = 2 > 0$, the point $(1,1)$ provides a relative minimum. ♦

Extrema on a Bounded Region

In Section 4.2, we saw how to find a maximum of a differentiable function $y = f(x)$ on closed interval $[a, b]$. The procedure was:

1. Find x in $[a, b]$ (other than a or b) where $f'(x) = 0$. It is called a critical number. If there are no critical numbers, the maximum occurs at a or b.
2. Evaluate f at each critical number. Also find $f(a)$ and $f(b)$. The maximum of f in $[a, b]$ is the largest of $f(a)$, $f(b)$, and the values of f at critical numbers.

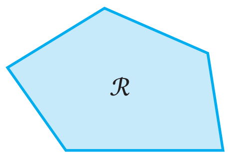


Figure 16.6.4

We can similarly find the maximum of $f(x, y)$ in a region $\mathscr{R}$ in the plane bounded by some polygon or curve. (See Figure 16.6.4.) It is assumed that $\mathscr{R}$ includes its border and is a finite region in the sense that it lies within some disk. (In advanced courses it is proved that a continuous function defined on such a domain has a maximum – and a minimum – value.)

If f has continuous partial derivatives, the procedure for finding a maximum is similar to that for maximizing a function on a closed interval.

1. First find points that are in $\mathscr{R}$ but not on the boundary of $\mathscr{R}$ where both f_x and f_y are 0. They are called **critical points**. (See Note 1, below.)
2. Evaluate f at each critical number. Also find the maximum of f on the boundary. (See Note 2, below.) The maximum of f on $\mathscr{R}$ is the largest value of f on the boundary and at critical points.

NOTES:
1. If there are no critical points in $\mathscr{R}$, then the maximum occurs on the boundary.
2. Finding the maximum of f on the boundary often involves finding the maximum of a one-variable function on a closed interval.
3. A similar procedure is used to find the minimum value of a function on a bounded region.

EXAMPLE 4. Maximize the function $f(x, y) = xy(108 - 2x - 2y) = 108xy - 2x^2y - 2xy^2$ on the triangle $\mathscr{R}$ bounded by the x-axis, the y-axis, and the line $x + y = 54$. (See Figure 16.6.5.)

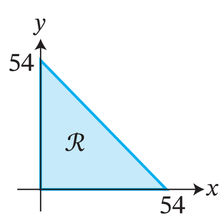


Figure 16.6.5

SOLUTION When x or y is 0 the function has the value 0. When x and y are small positive numbers the function is positive. Thus the maximum cannot occur when either x or y is 0. Similarly, if $x + y = 54$, then $f(x, y) = 2xy(54 - x - y) = 0$, so the maximum must occur at a critical point in the interior of $\mathscr{R}$. To find these critical points we consider

$$\frac{\partial f}{\partial x}(x, y) = 108y - 4xy - 2y^2 = 0$$
$$\frac{\partial f}{\partial y}(x, y) = 108x - 2x^2 - 4xy = 0$$

which give the simultaneous equations

$$2y(54 - 2x - y) = 0,$$
$$2x(54 - x - 2y) = 0.$$

Because neither x nor y is zero, the first equation reduces to $2x + y = 54$ and the second equation reduces to $x + 2y = 54$. To eliminate y, compute two times the first equation minus the second equation: $3x = 54$. So, $x = 18$, and then substituting this value of x into either equation shows that $y = 18$.

The point $(18, 18)$ lies in the interior of $\mathscr{R}$, since it lies above the x-axis, to the right of the y-axis, and below the line $x + y = 54$. The maximum value of the function is $f(18, 18) = 18 \cdot 18(108 - 2 \cdot 18 - 2 \cdot 18) = 11,664$.

Note that our preliminary analysis has already shown that $f(x, y) = 0$ at every point on the boundary of R: on the x-axis, on the y-axis, and on the slanted line $x + y = 54$. The boundary needs no further consideration.

Therefore, the local maximum occurs at the critical point $(18, 18)$ and has the value $f(18, 18) = 11,664$. ♦

EXAMPLE 5. The combined height and girth (distance around) of a package sent through the mail cannot exceed 108 inches. If the package is a rectangular box, how large can its volume be? (The "height" is defined to be the length of the package's longest side.)

SOLUTION We label the height of the box (the longest side) z and the other sides x and y, as in Figure 16.6.6. The volume $V = xyz$ is to be maximized, subject to girth plus length being at most 108. This constraint is expressed as an inequality:

$$2x + 2y + z \leq 108.$$

Since we want the largest box, we will restrict our attention to boxes for which the combined height and girth is as large as possible:

$$2x + 2y + z = 108. \qquad (16.6.3)$$

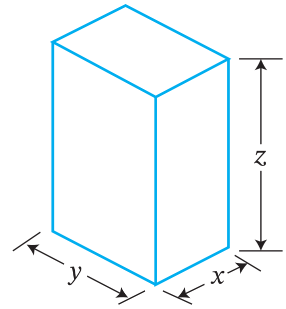


Figure 16.6.6

By (16.6.3), $z = 108 - 2x - 2y$ and the formula for the volume, $V = xyz$, can be expressed as a function of two variables:

$$V = f(x, y) = xy(108 - 2x - 2y).$$

This is to be maximized on the triangle described by $x \geq 0$, $y \geq 0$, $2x + 2y \leq 108$, that is, $x + y \leq 54$.

These are the same objective function and feasible region as in Example 4. Hence, the largest box has $x = y = 18$ and $z = 108 - 2x - 2y = 108 - 2\cdot 18 - 2\cdot 18 = 36$. Its dimensions are 18 inches by 18 inches by 36 inches and its volume is 11,664 cubic inches.

Note that the optimal package has $z = 36$ which is larger than the other two sides: $x = y = 18$. Thus, the requirement that z is the length of the longest side was satisfied without any special consideration. If the postage calculation had been worded slightly differently, say "The length of one edge plus the girth around the other edges shall not exceed 108 inches," the effect would be the same. You would not be able to send a larger box by measuring the girth around the base formed by its largest edges. ◆

EXAMPLE 6. Let $f(x, y) = x^2 + y^2 - 2x - 4y$. Find the maximum and minimum values of $f(x, y)$ on the disk $\mathscr{R}$ of radius 3 and center $(0, 0)$.

SOLUTION Before starting, note that because the function f is continuous and the region $\mathscr{R}$ is closed and bounded, f does have a maximum minimum value on $\mathscr{R}$.

First, find all critical points of f that are inside $\mathscr{R}$. The two equations

$$\frac{\partial f}{\partial x} = 2x - 2 = 0 \qquad \text{and} \qquad \frac{\partial f}{\partial y} = 2y - 4 = 0.$$

have the unique solution $x = 1$ and $y = 2$. The point $(1, 2)$ lies in $\mathscr{R}$ since its distance from the origin is $\sqrt{1^2 + 2^2} = \sqrt{5}$, which is less than 3. At $(1, 2)$ the value of the function is $f(1, 2) = 1^2 + 2^2 - 2(1) - 4(2) = 5 - 2 - 8 = -5$.

Now we find the behavior of f on the boundary, which is a circle of radius 3. We parameterize this circle as

$$x = 3\cos(\theta), \qquad y = 3\sin(\theta) \qquad \text{for } 0 \leq \theta \leq 2\pi.$$

On this circle

$$\begin{aligned} f(x, y) &= x^2 + y^2 - 2x - 4y \\ &= (3\cos(\theta))^2 + (3\sin(\theta))^2 - 2(3\cos(\theta)) - 4(3\sin(\theta)) \\ &= 9\cos^2(\theta) + 9\sin^2(\theta) - 6\cos(\theta) - 12\sin(\theta) \\ &= 9 - 6\cos(\theta) - 12\sin(\theta). \end{aligned}$$

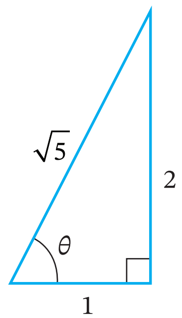


Figure 16.6.7

To find the maximum and minimum of the single-variable function $g(\theta) = 9 - 6\cos(\theta) - 12\sin(\theta)$ for θ in $[0, 2\pi]$, begin by finding $g'(\theta)$: $g'(\theta) = 6\sin\theta - 12\cos\theta$. Setting $g'(\theta) = 0$ gives $0 = 6\sin(\theta) - 12\cos(\theta)$ or

$$\sin(\theta) = 2\cos(\theta). \tag{16.6.4}$$

To solve (16.6.4), divide by $\cos(\theta)$ (Why is $\cos(\theta) \neq 0$?), getting $\sin(\theta)/\cos(\theta) = 2$ or $\tan(\theta) = 2$.

There are two angles θ in $[0, 2\pi]$ such that $\tan(\theta) = 2$. One is in the first quadrant, $\theta = \arctan(2)$, and the other is in the third quadrant, $\theta = \pi + \arctan(2)$. To evaluate $g(\theta) = 9 - 6\cos(\theta) - 12\sin(\theta)$ for these two angles we must compute $\cos(\theta)$ and $\sin(\theta)$. The right triangle in Figure 16.6.7 helps us do this.

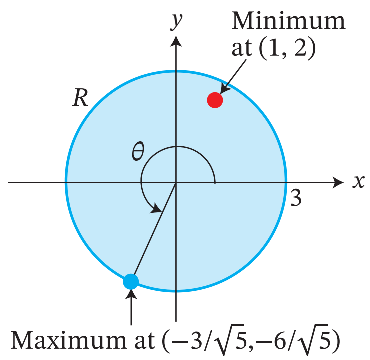


Figure 16.6.8

Figure 16.6.7 shows that for $\theta = \arctan(2)$,

$$\cos(\theta) = \frac{1}{\sqrt{5}} \quad \text{and} \quad \sin(\theta) = \frac{2}{\sqrt{5}}.$$

Then $g(\arctan(2)) = 9 - 6\left(1/\sqrt{5}\right) - 12\left(2/\sqrt{5}\right) = 9 - 30/\sqrt{5} \approx -4.4164$.

When $\theta = \pi + \arctan(2)$,

$$\cos(\theta) = \frac{-1}{\sqrt{5}} \quad \text{and} \quad \sin(\theta) = \frac{-2}{\sqrt{5}}.$$

Then $g(\pi + \arctan(2)) = 9 - 6\left(-1/\sqrt{5}\right) - 12\left(-2/\sqrt{5}\right) = 9 + 30/\sqrt{5} \approx 22.4164$.

At the endpoints of the interval $[0, 2\pi]$, $g(2\pi) = g(0) = 9 - 6(1) - 12(0) = 3$. Thus the maximum of f on the boundary of $\mathscr{R}$ is about 22.4164 and the minimum is about -4.4164.

At the critical point $(1, 2)$ the value of f is $f(1, 2) = -5$. We conclude that the maximum value of f on $\mathscr{R}$ is $f(-1/\sqrt{(}5), -2/\sqrt{5}) = 9 + 30/\sqrt{5} \approx 22.4164$ and the minimum value is $f(1, 2) = -5$. See Figure 16.6.8. ♦

Observation 16.6.3: *Different Types of Boundaries*

The big difference between finding extrema of a function of 1, 2, or 3 variables is this: When $\mathscr{R}$ is an interval, $[a, b]$, its boundary consists of the two points $x = a$ and $x = b$. But, when $\mathscr{R}$ is a region in the xy-plane, its boundary will be a curve. And, when $\mathscr{R}$ is a region in xyz-space, then its boundary will be a surface.

Proof of a Special Case of Theorem 16.6.2

We will prove Theorem 16.6.2 when $f(x, y)$ is a second-degree polynomial: $f(x, y) = Ax^2 + Bxy + Cy^2$, where A, B, and C are constants. (The proof of Case 2 of Theorem 16.6.2 for a general function f is developed in Exercise 59.)

Corollary 16.6.4: Special Case of Theorem 16.6.4

Let $f(x, y) = Ax^2 + Bxy + Cy^2$, where A, B, and C are constants. Then $(0, 0)$ is a critical point of f and the discriminant is $D = 4AC - B^2$. Moreover,

Case 1: *if $D > 0$ and $A > 0$, then f has a relative minimum at $(0, 0)$.*

Case 2: *if $D > 0$ and $A < 0$, then f has a relative maximum at $(0, 0)$.*

Case 3: *if $D < 0$, then f has neither a relative minimum nor a relative maximum at $(0, 0)$.*

NOTE: *As in Theorem 16.6.2, if $D = 0$ then any of the three cases are possible.*

Proof of Corollary 16.6.4

We prove Case 1, leaving Cases 2 and 3 for Exercises 56 and 57, respectively. The situation when $D=0$ is addressed in Exercise 58.

First, the first- and second-order derivatives are: $f_x(x,y)=2Ax+By$, $f_y(x,y)=Bx+2Cy$, $f_{xx}(x,y)=2A$, $f_{xy}(x,y)=B$, and $f_{yy}(x,y)=2C$. The discriminant is $D=f_{xx}f_{yy}-(f_{xy})^2=4AC-B^2$.

Both f_x and f_y are 0 at $(0,0)$. Hence $(0,0)$ is a critical point. (In fact, when $D\neq 0$, $(0,0)$ is the only critical point of f.)

Because $f(0,0)=0$, showing that $(0,0)$ is a relative minimum of f when $D>0$ and $A>0$ requires that we demonstrate that $f(x,y)\geq 0$ for all (x,y) near $(0,0)$.

$$\begin{aligned}
f(x,y) &= Ax^2+Bxy+Cy^2 \\
&= A\left(x^2+\frac{B}{A}xy+\frac{C}{A}y^2\right) && (A>0,\text{ so } A\neq 0) \\
&= A\left(x+\frac{B}{2A}y\right)^2-\frac{B^2}{4A}y^2+Cy^2 && \text{(complete the square)} \\
&= A\left(x+\frac{B}{2A}y\right)^2+\left(C-\frac{B^2}{4A}\right)y^2 && \text{(simplifying)} \\
&= A\left(x+\frac{B}{2A}y\right)^2+\frac{D}{4A}y^2
\end{aligned}$$

For Case 1, we are assuming $A>0$ and $D>0$, so $D/(4A)$ is also positive. So $f(x,y)$ is the sum of two terms that have the same form: a positive coefficient times the square of a number. As such, $f(x,y)$ is never negative. (In fact, $f(x,y)$ is positive for all $(x,y)\neq(0,0)$.)

Thus $f(x,y)\geq f(0,0)=0$ holds for all (x,y), not only for (x,y) near $(0,0)$. Since the critical point $(0,0)$ is a relative minimum of f, the proof of Case 1 of the corollary is complete.

Summary

We defined a critical point of $f(x,y)$ as a point where both partial derivatives f_x and f_y are 0. Even if f_{xx} and f_{yy} are negative there, such a point need not provide a relative maximum. We must also know that $|f_{xy}|$ is not too large.

If $f_{xx}<0$ and $f_{xy}^2<f_{xx}f_{yy}$, then there is a relative maximum at the critical point. (The two inequalities imply $f_{yy}<0$.)

Similar criteria hold for a relative minimum: if $f_{xx}>0$ and $f_{xy}^2<f_{xx}f_{yy}$, then this critical point is a relative minimum.

The critical point is a saddle point when $f_{xy}^2>f_{xx}f_{yy}$.

When $f_{xy}^2=f_{xx}f_{yy}$, the critical point may be a relative maximum, relative minimum, or neither.

We also described how to find extrema of a function defined on a bounded region.

EXERCISES for Section 16.6

In Exercises 1 to 9 use Theorems 16.6.1 and 16.6.2 to determine relative maxima or minima of the given functions.

1. $x^2+3xy+y^2$
2. x^2-y^2
3. $x^2-2xy+2y^2+4x$
4. $x^4+8x^2+y^2-4y$
5. x^2-xy+y^2
6. $x^2+2xy+2y^2+4x$
7. $2x^2+2xy+5y^2+4x$
8. $-4x^2-xy-3y^2$
9. x^3-y^3+3xy

In each of Exercises 10 to 15 let f by a function of x and y such that at (a,b) both f_x and f_y equal 0. Values are specified for f_{xx}, f_{xy}, and f_{yy} at (a,b). Assume that the partial derivatives are continuous. Decide whether, at (a,b), f has a relative maximum, a relative minimum, a saddle point, or there is not enough information to classify the critical point.

10. $f_{xy}=4$, $f_{xx}=2$, $f_{yy}=8$

11. $f_{xy}=-3$, $f_{xx}=2$, $f_{yy}=4$

12. $f_{xy}=3$, $f_{xx}=2$, $f_{yy}=4$

13. $f_{xy}=2$, $f_{xx}=3$, $f_{yy}=4$

14. $f_{xy}=-2$, $f_{xx}=-3$, $f_{yy}=-4$

15. $f_{xy}=-2$, $f_{xx}=3$, $f_{yy}=-4$

In Exercises 22 to 24 find the critical points and the relative extrema of the given functions.

16. $3xy-x^3-y^3$

17. $12xy-x^3-y^3$

18. $6xy-x^2y-xy^2$

19. $\exp(x^3+y^3)$

20. 2^{xy}

21. $3x+xy+x^2y-2y$

22. $x+y-\dfrac{1}{xy}$

23. $\dfrac{4}{x}+\dfrac{2}{y}+xy$

24. $x+y+\dfrac{8}{xy}$

---

25. Find the dimensions of the open rectangular box of volume 1 that has the smallest surface area. Use Theorem 16.6.2 as a check that the critical point provides a minimum.

26. The material for the top and bottom of a rectangular box costs 3 cents per square foot, and that for the sides 2 cents per square foot. What is the least expensive box that has a volume of 1 cubic foot? Use Theorem 16.6.2 as a check that the critical point provides a minimum.

27. UPS ships rectangular packages whose combined length and girth is at most 165 inches (and weigh at most 150 pounds).

(a) What are the dimensions of the rectangular package with the largest volume that it ships?

(b) What are the dimensions of the rectangular package with maximum surface area that UPS ships?

28. Let $P_1=(x_1,y_1)$, $P_2=(x_2,y_2)$, $P_3=(x_3,y_3)$, and $P_4=(x_4,y_4)$. Find the coordinates of the point P that minimizes the sum of the squares of the distances from P to the four points.

29. Find the dimensions of the rectangular box of largest volume whose total surface area is 12 square meters.

30. Three nonnegative numbers x, y, and z have sum 1. (a) How small can $x^2+y^2+z^2$ be? (b) How large can it be?

31. Each year a firm can produce r radios and t television sets at a cost of $2r^2+rt+2t^2$ dollars. It sells a radio for \$600 and a television set for \$900.

(a) What is the profit from the sale of r radios and t television sets?

NOTE: Profit is revenue less cost.

(b) Find the combination of r and t that maximizes profit. Use the discriminant as a check.

32. Find the dimensions of the rectangular box of largest volume that can be inscribed in a sphere of radius 1.

33. For what values of k does $x^2+kxy+3y^2$ have a relative minimum at $(0,0)$?

34. For what values of k does $kx^2+5xy+4y^2$ have a relative minimum at $(0,0)$?

35. Let $f(x,y)=(2x^2+y^2)e^{-x^2-y^2}$.

(a) Find all critical points of f.

(b) Examine the behavior of f when (x,y) is far from the origin.

(c) What is the minimum value of f?

(d) What is the maximum value of f?

36. Find the maximum and minimum values of the function in Exercise 35 (a) on the circle $x^2 + y^2 = 1$ and (b) on the circle $x^2 + y^2 = 4$.

37. Maximize the function $f(x, y) = 3x^2 - 4y^2 + 2xy$ on the square region with vertices $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$.

38. Find the maximum value of $f(x, y) = xy$ on the triangular region whose vertices are $(0,0)$, $(1,0)$, and $(0,1)$.

39. Maximize the function $-x + 3y + 6$ on the quadrilateral region with vertices $(1,1)$, $(4,2)$, $(5,6)$, and $(0,3)$.

40. (a) Show that $z = x^2 - y^2 + 2xy$ has no maximum and no minimum.
(b) Find the minimum and maximum of z if we consider only (x, y) on the circle of radius 1 and center $(0,0)$. That is, all (x, y) such that $x^2 + y^2 = 1$.
(c) Find the minimum and maximum of z if we consider all (x, y) in the disk of radius 1 and center $(0,0)$. That is, all (x, y) such that $x^2 + y^2 \le 1$.

41. Suppose z is a function of x and y with continuous second partial derivatives. If, at (x_0, y_0), $z_x = 0 = z_y$, $z_{xx} = 3$, and $z_{yy} = 12$, for what values of z_{xy} is it certain that z has a relative minimum at (x_0, y_0)?

42. Let $U(x, y, z) = x^{1/2}y^{1/3}z^{1/6}$ be the utility to a consumer of the amounts x, y, and z of three commodities. Their prices are, respectively, 2 dollars, 1 dollar, and 5 dollars, and the consumer has 60 dollars to spend. How much of each product should he buy to maximize the utility?

Exercise 43 shows that if the discriminant D is zero, then any of the three outcomes mentioned in Cases 1, 2, and 3 of Theorem 16.6.2 are possible.

43. (a) Let $f(x, y) = x^2 + 2xy + y^2$. Show that at $(0,0)$ both f_x and f_y are 0, f_{xx} and f_{yy} are positive, $D = 0$, and f has a relative minimum.
(b) Let $f(x, y) = x^2 + 2xy + y^2 - x^4$. Show that at $(0,0)$ both f_x and f_y are 0, f_{xx} and f_{yy} are positive, $D = 0$, and f has neither a relative maximum nor a relative minimum.
(c) Give an example of a function $f(x, y)$ for which $(0,0)$ is a critical point and $D = 0$ there, but f has a relative maximum at $(0,0)$.

44. Let $f(x, y) = ax + by + c$ for nonzero constants a, b, and c. Let R be a polygon in the xy-plane. Show that the maximum and minimum values of $f(x, y)$ on R are assumed at vertices of the polygon.

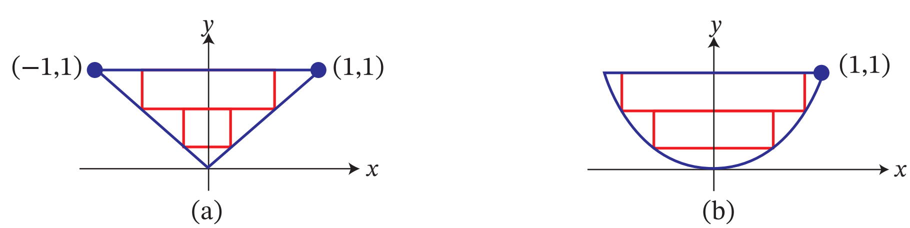


Figure 16.6.9

45. Two rectangles are placed in the triangle whose vertices are $(0,0)$, $(1,1)$, and $(-1,1)$ as shown in Figure 16.6.9(a). Show that they can fill as much as 2/3 of the area of the triangle.

46. Two rectangles are placed in the region bounded by the line $y = 1$ and the parabola $y = x^2$ as shown in Figure 16.6.9(b). How large can their total area be?

47. Let $P_0 = (a, b, c)$ be a point not on the surface $f(x, y, z) = 0$. Let P be a point on the surface nearest P_0. Show that $\overrightarrow{P_0P}$ is perpendicular to the surface at P.

48. Let $f(x, y) = (y - x^2)(y - 2x^2)$.
(a) Show that f has neither a local minimum nor a local maximum at $(0,0)$.
(b) Show that f has a local minimum at $(0,0)$ when considered only on any line through $(0,0)$. (Graph $y = x^2$ and $y = 2x^2$ and show where $f(x, y)$ is positive and where it is negative.)

49. Find (a) the minimum value of xyz, and (b) the maximum value of xyz, for nonnegative real numbers x, y, z such that $x+y+z=1$.

50. (a) Deduce from Exercise 49 that for three nonnegative numbers a, b, and c, $\sqrt[3]{abc} \le (a+b+c)/3$.

This shows that the **geometric mean** of three numbers is not larger than their **arithmetic mean**. See also Exercise 28 in Section 16.7.

(b) Obtain a corresponding result for four numbers.

Exercises 51 and 52 are related.

51. Let (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) be n points in the plane. Statisticians define the **line of regression** as the line that minimizes the sum of the squares of the differences between y_i and the ordinates of the line at x_i. (See Figure 16.6.10.) Let a line in the plane have the equation $y = mx + b$.

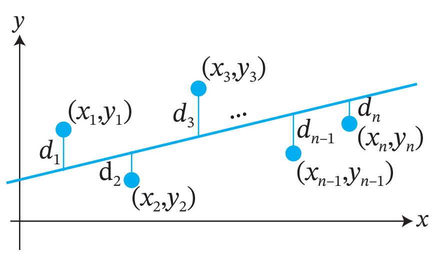


Figure 16.6.10

(a) Show that the line of regression minimizes the sum of the squares of the vertical distance from each point to the line $y = mx + b$, $\sum_{i=1}^{n} \left(y_i - (mx_i + b)\right)^2$, considered as a function of m and b.

(b) Let $f(m,b) = \sum_{i=1}^{n} \left(y_i - (mx_i + b)\right)^2$. Compute f_m and f_b.

(c) Show that when $f_m = 0 = f_b$, we have $m\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i$ and $m\sum_{i=1}^{n} x_i + nb = \sum_{i=1}^{n} y_i$.

(d) When do the simultaneous equations in (c) have a unique solution for m and b?

(e) Find the regression line for the points $(1,1)$, $(2,4)$, and $(3,5)$.

52. If your calculator is programmed to compute lines of regression, find and draw the line of regression for the points $(1,1)$, $(2,1.5)$, $(3,3)$, $(4,2)$ and $(5,3.5)$.

---

53. A surface is called **closed** when it is the boundary of a region R, as a balloon is the boundary of the air within it. A surface is called **smooth** when it has a continuous outward unit normal vector at each point of the surface. Let S be a smooth closed surface bounding a bounded region R. Show that for a point P_0 in R, there are at least two points on S such that $\overrightarrow{P_0P}$ is normal to S. (It is conjectured that if P_0 is the centroid of R, that is, the center of mass of R, then there are at least four points on S such that $\overrightarrow{P_0P}$ is normal to S. The **centroid** is defined in Section 17.8.)

54. Find the point P on the plane $Ax + By + Cz + D = 0$ nearest the point $P_0 = (x_0, y_0, z_0)$, which is not on that plane. (a) Find P by calculus. (b) Find P by using the algebra of vectors. (Why is $\overrightarrow{P_0P}$ perpendicular to the plane?)

Exercise 55 provides another motivation for the definition in Section 12.7 of the Fourier series of a function f defined on the interval $[0, 2\pi]$.

55. For an integer n let $S(x) = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos(kx) + b_k \sin(kx))$. Let $f(x)$ be a continuous function defined on $[-\pi, \pi]$. The definite integral $\int_{-\pi}^{\pi} \left(f(x) - S(x)\right)^2 \, dx$ is a measure of how close $S(x)$ is to $f(x)$ on the interval $[-\pi, \pi]$. The integral can never be negative. (Why?) The smaller the integral, the better S approximates f on $[-\pi, \pi]$. Determine the coefficients a_k and b_k that minimize the integral. Show that these choices for the coefficients mean that S is a partial sum of the Fourier series associated with $f(x)$.

Exercises 56 and 57 complete the proof of Corollary 16.6.4. Exercise 58 extends that result to the case when $D = 0$.

56. Prove Case 2 of Corollary 16.6.4.

57. Prove Case 3 of Corollary 16.6.4.

58. Corollary 16.6.4 does not address the case when $D = 0$. Assume $A = 0$.

(a) Show that the critical point $(0,0)$ can be either a relative minimum or a relative maximum of f.

(b) Show that there is, in fact, an entire line of critical points that can be either a relative minimum or a relative maximum of f.

Exercise 59 outlines the proof of Case 2 of Theorem 16.6.2: if $f_x(a,b) = f_y(a,b) = 0$, $f_{xx}(a,b) > 0$, and $f_{xx}(a,b)f_{yy}(a,b) - (f_{xy})^2(a,b) > 0$, then $f(a,b)$ has a relative minimum at (a,b).

59. Assume that f_{xx}, f_{yy}, and f_{xy} are continuous, and that f_{xx} and $f_{xx}f_{yy} - (f_{xy})^2$ remain positive on some disk R whose center is (a,b) (that is, there is a version of the permanence property for two-dimensional sets). The following steps show that f has a minimum at (a,b) on each line L through (a,b). Let $\mathbf{u} = \cos(\theta)\mathbf{i} + \sin(\theta)\mathbf{j}$ be a unit vector.

(a) Show that $D_{\mathbf{u}} f(a,b) = 0$.

(b) Show that $D_{\mathbf{u}}(D_{\mathbf{u}} f) = f_{xx}\cos^2(\theta) + 2f_{xy}\sin(\theta)\cos(\theta) + f_{yy}\sin^2(\theta)$.

(c) Show that $(f_{xx})D_u(D_u f) = \left(f_{xx}\cos(\theta) + f_{xy}\sin(\theta)\right)^2 + \left(f_{xx}f_{yy} - (f_{xy})^2\right)\sin^2(\theta)$.

(d) Deduce from (c) that f is concave up on the part of each line through (a,b) inside the disk R.

(e) Deduce that the graph of f for points in R lies above its tangent lines at $(a,b,f(a,b))$, so f has a relative minimum at (a,b).

16.7 Lagrange Multipliers

The method of Lagrange multipliers is another way to find maxima or minima of a function that satisfies one or more additional constraints. These problems are called **constrained optimization problems**. The key observation is that a gradient of a function is perpendicular to its level curves (or level surfaces).

Giuseppe Lodovico Lagrangi, 1736–1813, was an Italian mathematician, physicist, and astronomer who later changed his citizenship to France and changed his name to Joseph-Louis Lagrange

The Essence of the Method

We first consider a simple case. Suppose we want to find a maximum or a minimum of $f(x,y)$ for points (x,y) on the line L that has the equation $g(x,y) = k$. See Figure 16.7.1(a).)

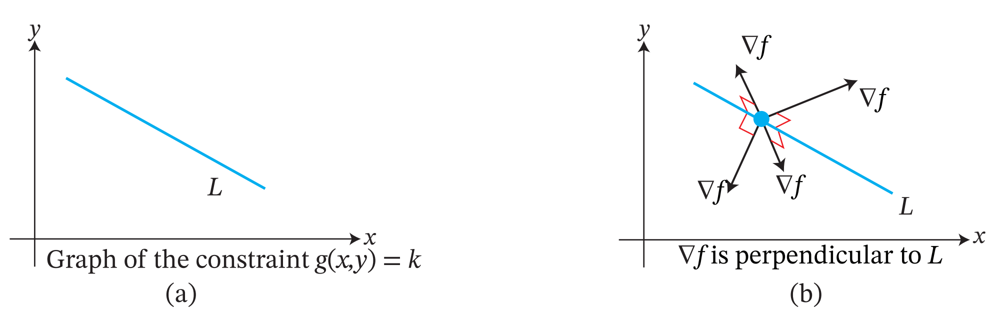


Figure 16.7.1

Assume the extremum occurs at P. Let $\nabla f(P)$ be the gradient of f evaluated at a point P. Assuming $\nabla f(P)$ is not the zero vector, what can we say about its direction? (See Figure 16.7.1(b).)

Suppose that $\nabla f(P)$ is not perpendicular to L. Let $\mathbf{u}$ be a unit vector parallel to L. Then $D_{\mathbf{u}} f(P) = \nabla f(P) \cdot \mathbf{u}$ is not zero. If $D_{\mathbf{u}} f(P)$ is positive then $f(x, y)$ is increasing in the direction $\mathbf{u}$, which is along L. In the direction $-\mathbf{u}$, $f(x, y)$ is decreasing. Therefore the point (a, b) could not provide either a maximum or a minimum. Since this contradicts the hypothesis that f has an extremum at P, the assumption that $\nabla f(P)$ is not perpendicular to L must be false. That means $\nabla f(P)$ must be perpendicular to L. Since $g(x, y) = k$ is a level curve of g, $\nabla g(P)$ is also perpendicular to L. This means $\nabla f(P)$ and $\nabla g(P)$ must be parallel vectors: there is a scalar λ, called a **Lagrange multiplier**, such that

$$\nabla f(P) = \lambda \nabla g(P). \tag{16.7.1}$$

The Greek letter λ, pronounced *lam-da*, corresponds to the lowercase letter "l" (ell).

EXAMPLE 1. Find the minimum of $x^2 + 2y^2$ on the line $x + y = 2$.

SOLUTION Since $x^2 + 2y^2$ increases without bound in both directions along the line it must have a minimum somewhere on the line.

Here $f(x, y) = x^2 + 2y^2$ and $g(x, y) = x + y$. Therefore $\nabla f = 2x\mathbf{i} + 4y\mathbf{j}$ and $\nabla g = \mathbf{i} + \mathbf{j}$. At the minimum, the gradients of f and g are parallel. That is, there is a scalar λ such that $\nabla f = \lambda \nabla g$. This implies that

$$2x\mathbf{i} + 4y\mathbf{j} = \lambda(\mathbf{i} + \mathbf{j}). \tag{16.7.2}$$

This single vector equation (16.7.2) leads to the two scalar equations

$$2x = \lambda \qquad \text{(equating } \mathbf{i} \text{ components)} \tag{16.7.3}$$

$$4y = \lambda \qquad \text{(equating } \mathbf{j} \text{ components).} \tag{16.7.4}$$

But we also require that the constraint be satisfied:

$$x + y = 2. \tag{16.7.5}$$

From (16.7.3) and (16.7.4), $2x = 4y$ or $x = 2y$, which can be substituted into (16.7.5) to obtain $2y + y = 2$ or $y = 2/3$, and, then, $x = 2y = 4/3$. The minimum occurs at $(4/3, 2/3)$ and the minimum value of $x^2 + 2y^2$ on the line $x + y = 2$ is

$$f\left(\frac{4}{3}, \frac{2}{3}\right) = \left(\frac{4}{3}\right)^2 + 2\left(\frac{2}{3}\right)^2 = \frac{24}{9} = \frac{8}{3}.$$

◆

Observation 16.7.1:

We solved the constrained optimization problem in Example 1 without ever finding the value of the Lagrange multiplier, λ. This is not unusual, but sometimes it is necessary to find λ.

The General Method

Let us see why Lagrange's method works when the constraint is not a line, but a curve.

The problem is to maximize or minimize a function $u = f(x, y)$ subject to a constraint $g(x, y) = k$.

The graph of the constraint $g(x, y) = k$ is, in general, a curve C, as shown in Figure 16.7.2. Let C be parameterized by $\mathbf{G}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}$. Assume that f, considered only on points of C, takes a maximum (or minimum) value at P_0. Let $\mathbf{G}(t_0) = \overrightarrow{OP_0}$. Then u is a function of t:

$$u = f(x(t), y(t)),$$

and, as shown in the proof of Theorem 16.5.1 of Section 16.5,

$$\frac{du}{dt} = \nabla f(x(t), y(t)) \cdot \mathbf{G}'(t). \tag{16.7.6}$$

Since f, considered only on C, has an extremum at $\mathbf{G}(t_0)$,

$$\frac{du}{dt} = 0 \qquad \text{at } t = t_0.$$

Thus, by (16.7.6),

$$\nabla f(P_0) \cdot \mathbf{G}'(t_0) = 0.$$

This means that $\nabla f(P_0)$ is perpendicular to $\mathbf{G}'(t_0)$. But, since the gradient ∇g is perpendicular to the level curve $g(x, y) = k$, $\nabla g(P_0)$ is also perpendicular to $\mathbf{G}'(t_0)$. (We implicitly assume that $\nabla g(P_0)$ is not $\mathbf{0}$.) See Figure 16.7.3. Then, because $\nabla f(P_0)$ is parallel to $\nabla g(P_0)$, there is a scalar λ such that $\nabla f(P_0) = \lambda \nabla g(P_0)$.

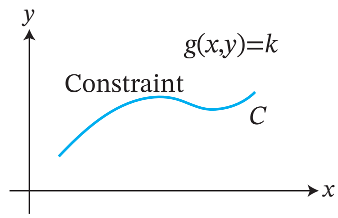


Figure 16.7.2

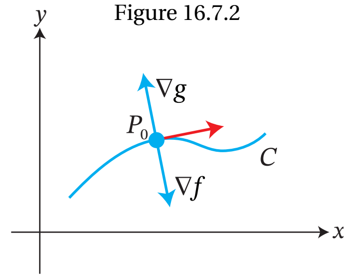


Figure 16.7.3

EXAMPLE 2. Maximize the function $x^2 y$ for points (x, y) on the unit circle $x^2 + y^2 = 1$.

SOLUTION Let $g(x, y) = x^2 + y^2$. We wish to maximize $f(x, y) = x^2 y$ for points on the circle $g(x, y) = 1$. The gradients of these two functions are

$$\nabla f = \nabla(x^2 y) = 2xy\mathbf{i} + x^2\mathbf{j} \qquad \text{and} \qquad \nabla g = \nabla(x^2 + y^2) = 2x\mathbf{i} + 2y\mathbf{j}.$$

At an extreme point for this problem, $\nabla f = \lambda \nabla g$ for some scalar λ. This gives us two scalar equations:

$$2xy = \lambda(2x) \qquad \text{(equating } \mathbf{i} \text{ components)} \tag{16.7.7}$$

$$x^2 = \lambda(2y) \qquad \text{(equating } \mathbf{j} \text{ components)} \tag{16.7.8}$$

The third equation is the constraint,

$$x^2 + y^2 = 1 \qquad \text{(the constraint)}. \tag{16.7.9}$$

Since the maximum value will be positive, it does not occur when $x = 0$, and we may safely assume x is not 0. Then, dividing both sides of (16.7.7) by $2x$, we get $y = \lambda$. Thus (16.7.8) becomes

$$x^2 = 2y^2. \tag{16.7.10}$$

Combining this with (16.7.9), we have $2y^2 + y^2 = 1$ or $y^2 = 1/3$. Thus $y = \sqrt{3}/3$ or $y = -\sqrt{3}/3$. By (16.7.10),

$$x = \sqrt{2}y \qquad \text{or} \qquad x = -\sqrt{2}y.$$

The maximum must occur at one (or more) of the following four points on the circle:

$$\left(\frac{\sqrt{6}}{3}, \frac{\sqrt{3}}{3}\right), \left(\frac{-\sqrt{6}}{3}, \frac{\sqrt{3}}{3}\right), \left(\frac{-\sqrt{6}}{3}, \frac{-\sqrt{3}}{3}\right), \left(\frac{\sqrt{6}}{3}, \frac{-\sqrt{3}}{3}\right).$$

At the first and second points, with $y > 0$, $x^2 y$ is positive (and equal), while at the third and fourth points, with $y < 0$, $x^2 y$ is negative (and, again, equal). The first two points provide the maximum value of $x^2 y$ on the circle $x^2 + y^2 = 1$:

$$f\left(\frac{\pm\sqrt{6}}{3}, \frac{\sqrt{3}}{3}\right) = \left(\frac{\pm\sqrt{6}}{3}\right)^2 \frac{\sqrt{3}}{3} = \frac{2\sqrt{3}}{9}.$$

The third and fourth points provide the minimum value of $x^2 y$ on the unit circle:

$$f\left(\frac{\pm\sqrt{6}}{3}, \frac{-\sqrt{3}}{3}\right) = \frac{-2\sqrt{3}}{9}.$$

♦

The attentive reader will have noticed that the development of Lagrange multipliers starts with the assumption that the constrained optimization problem has a solution. From this assumption we found that equation (16.7.1) must be satisfied. To show that the solution found by this method is, in fact, a solution of the original constrained optimization problem involves ideas and methods that are more commonly addressed in future mathematics courses.

More Variables

In the preceding examples we examined the maximum and minimum of $f(x,y)$ on a curve $g(x,y)=k$. The same method works for finding extreme values of $f(x,y,z)$ on a surface $g(x,y,z)=k$. If f has, say, a minimum at $P=(a,b,c)$, then it does on any curve through P on the surface $g(x,y,z)=k$. Thus $\nabla f(P)$ is perpendicular to any curve on the surface through P. But so is $\nabla g(P)$. Thus $\nabla f(P)$ and $\nabla g(P)$ are parallel, and, assuming $\nabla g(P)\neq \mathbf{0}$, there is a scalar λ such that $\nabla f(P)=\lambda\nabla g(P)$. There will be four scalar equations: three from the vector equation $\nabla f(P)=\lambda\nabla g(P)$ and one from the constraint $g(x,y,z)=k$. That gives four equations in four unknowns, x, y, z, and λ, even though it is not necessary to find λ.

EXAMPLE 3. Find the rectangular box with the largest volume, if its surface area is 96 square feet.

SOLUTION

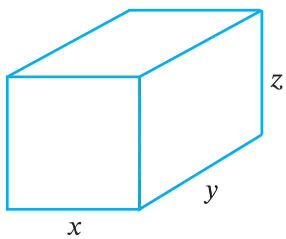


Figure 16.7.4

Let the dimensions be x, y, and z and the volume be V, which equals xyz. The surface area is $2xy+2xz+2yz$. See Figure 16.7.4.

We wish to maximize $V(x,y,z)=xyz$ subject to the constraint

$$g(x,y,z)=2xy+2xz+2yz=96. \qquad (16.7.11)$$

In this case, $\nabla V = yz\mathbf{i}+xz\mathbf{j}+xy\mathbf{k}$ and $\nabla g=(2y+2z)\mathbf{i}+(2x+2z)\mathbf{j}+(2x+2y)\mathbf{k}$. The vector equation $\nabla V=\lambda\nabla g$ provides three scalar equations:

$$yz=\lambda(2y+2z) \qquad (16.7.12)$$
$$xz=\lambda(2x+2z) \qquad (16.7.13)$$
$$xy=\lambda(2x+2y).$$

The fourth equation is the constraint,

$$2xy+2xz+2yz=96.$$

Solving for λ in (16.7.12) and in (16.7.13), and equating the results gives

$$\frac{yz}{2y+2z}=\frac{xz}{2x+2z}.$$

Since $z\neq 0$, we have

$$\frac{y}{2y+2z}=\frac{x}{2x+2z}.$$

Clearing denominators gives

$$2xy+2yz=2xy+2xz,$$

hence,

$$2yz=2xz.$$

Since $z\neq 0$, we conclude that $x=y$.

Because x, y, and z play the same roles in both the volume xyz and in the surface area, $2(xy+xz+yz)$, we conclude also that $x=z$. Then $x=y=z$. The box of maximum volume is a cube.

To find its dimensions use the constraint, which tells us that $6x^2=96$ or $x=4$. Hence y and z are 4 also. ◆

Multiple Constraints

Lagrange multipliers can also be used to maximize a function $f(x,y,z)$ subject to more than one constraint. For example, in the case of two constraints they would have the form

$$g(x,y,z)=k_1 \qquad \text{and} \qquad h(x,y,z)=k_2. \qquad (16.7.14)$$

The two surfaces (16.7.14) in general meet in a curve C, as shown by the intersection of the pink and purple surfaces in Figure 16.7.5. Assume that C is parameterized by the function $\mathbf{G}$. Then at a maximum (or minimum) of f at a point $P_0 = (x_0, y_0, z_0)$ on C, and $\mathbf{G}(t_0) = \overrightarrow{OP_0}$,

$$\nabla f(P_0) \cdot \mathbf{G}'(t_0) = 0.$$

Thus ∇f, evaluated at P_0, is perpendicular to $\mathbf{G}'(t_0)$. But $\nabla g(P_0)$ and $\nabla h(P_0)$, being normal vectors at P_0 to the level surfaces $g(x,y,z) = k_1$ and $h(x,y,z) = k_2$, respectively, are both perpendicular to $\mathbf{G}'(t_0)$. Thus, as Figure 16.7.5 shows,

∇f, $\nabla g(P_0)$, and $\nabla h(P_0)$ are all perpendicular to $\mathbf{G}'(t_0)$ at $P_0 = (x_0, y_0, z_0)$.

Consequently, $\nabla f(P_0)$ lies in the plane determined by the vectors $\nabla g(P_0)$ and $\nabla h(P_0)$ that we assume are not parallel. Hence there are scalars λ and μ such that

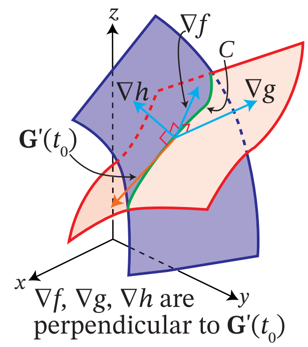


Figure 16.7.5

$$\nabla f(P_0) = \lambda \nabla g(P_0) + \mu \nabla h(P_0).$$

This vector equation provides three scalar equations in x, y, z, λ, and μ. The two constraints, (16.7.14), give two more equations. So we have five equations in five unknowns. As before, we find λ and μ only if they assist the algebra.

The Greek letter μ, pronounced *mew*, corresponds to the lowercase letter "m".

If a maximum occurs at an endpoint of the curves or if the two surfaces do not meet in a curve or if ∇g and ∇h are parallel, this method does not apply. We will content ourselves by illustrating the method with an example in which there are two constraints. A rigorous development of the material in this section belongs in an advanced calculus course.

EXAMPLE 4. Minimize the quantity $x^2 + y^2 + z^2$ subject to the constraints $x + 2y + 3z = 6$ and $x + 3y + 9z = 9$.

SOLUTION There are three variables and two constraints. Each constraint is a plane. Together they give a line, which we will not find explicitly. The function $x^2 + y^2 + z^2$ is the square of the distance from (x, y, z) to the origin. So the problem can be rephrased as how far is the origin from a certain line? (It could also be solved by vector algebra. See Exercise 23.) When viewed this way, the problem certainly has a solution.

To find the solution by the method of Lagrange multipliers, define

$$f(x,y,z) = x^2 + y^2 + z^2$$
$$g(x,y,z) = x + 2y + 3z$$
$$h(x,y,z) = x + 3y + 9z.$$

To find a minimizer of $f(x,y,z)$ on the planes $g(x,y,z) = 6$ and $h(x,y,z) = 9$ we need the gradients of f, g, and h:

$$\nabla f = 2x\mathbf{i} + 2y\mathbf{j} + 2z\mathbf{k}$$
$$\nabla g = \mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$$
$$\nabla h = \mathbf{i} + 3\mathbf{j} + 9\mathbf{k}.$$

Then, setup the equations with Lagrange multipliers (constants) λ and μ such that

$$\nabla f = \lambda \nabla g + \mu \nabla h.$$

Therefore, the five equations for x, y, z, λ, and μ are

$$2x = \lambda + \mu \quad (\text{equating } \mathbf{i} \text{ components}) \tag{16.7.15}$$
$$2y = 2\lambda + 3\mu \quad (\text{equating } \mathbf{j} \text{ components}) \tag{16.7.16}$$
$$2z = 3\lambda + 9\mu \quad (\text{equating } \mathbf{k} \text{ components}) \tag{16.7.17}$$
$$x + 2y + 3z = 6 \quad (\text{first constraint}) \tag{16.7.18}$$
$$x + 3y + 9z = 9 \quad (\text{second constraint}). \tag{16.7.19}$$

By (16.7.15), (16.7.16), and (16.7.17),

$$x = \frac{\lambda+\mu}{2}, \qquad y = \frac{2\lambda+3\mu}{2}, \qquad z = \frac{3\lambda+9\mu}{2}.$$

Equations (16.7.18) and (16.7.19) then become

$$\frac{\lambda+\mu}{2} + \frac{2(2\lambda+3\mu)}{2} + \frac{3(3\lambda+9\mu)}{2} = 6$$

and

$$\frac{\lambda+\mu}{2} + \frac{3(2\lambda+3\mu)}{2} + \frac{9(3\lambda+9\mu)}{2} = 9,$$

which simplify to

$$14\lambda + 34\mu = 12 \tag{16.7.20}$$

and

$$34\lambda + 91\mu = 18. \tag{16.7.21}$$

Solving (16.7.20) and (16.7.21) gives

$$\lambda = \frac{240}{59} \qquad \text{and} \qquad \mu = -\frac{78}{59}.$$

Thus

$$\begin{aligned} x &= \frac{\lambda+\mu}{2} = \frac{81}{59} \approx 1.37 \\ y &= \frac{2\lambda+3\mu}{2} = \frac{123}{59} \approx 2.08 \\ z &= \frac{3\lambda+9\mu}{2} = \frac{9}{59} \approx 0.15. \end{aligned}$$

The minimum of $x^2+y^2+z^2$ is thus How do we know this is a minimum, and not a maximum?

$$\left(\frac{81}{59}\right)^2 + \left(\frac{123}{59}\right)^2 + \left(\frac{9}{59}\right)^2 = \frac{21,771}{3,481} = \frac{369}{59} \approx 6.2542.$$

◆

In Example 4 there were three variables, x, y, and z, and two constraints. There may be many variables, x_1, x_2, $\dots x_n$, and many constraints. If there are m constraints, g_1, g_2, ..., g_m, introduce m Lagrange multipliers λ_1, λ_2, ..., λ_m, one for each constraint. So there would be $m+n$ equations, n from the n components in

$$\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 + \cdots + \lambda_m \nabla g_m$$

and m more from the m constraints. There would be $m+n$ unknowns, $x_1, x_2, \dots, x_n, \lambda_1, \lambda_2, \dots, \lambda_m$.

Summary

The basic idea of the Lagrange multiplier method is that if $f(x,y)$ has an extremum at P_0 on the curve $g(x,y)=k$, then ∇f and ∇g are parallel: there is a scalar λ such that $\nabla f = \lambda \nabla g$.

Similarly, if $f(x,y,z)$ has an extreme value on the surface $g(x,y,z)=k$, then ∇f and ∇g are parallel at a point where the extreme value occurs. That is, there is a scalar λ such that $\nabla f = \lambda \nabla g$. If there are two constraints $g(x,y,z)=k_1$ and $h(x,y,z)=k_2$, then ∇f lies in the plane of ∇g and ∇h: there are scalars λ and μ such that $\nabla f = \lambda \nabla g + \mu \nabla h$.

These ideas extend in a natural way to find the extrema of a function of n variables subject to m constraints. In this case the method of Lagrange multipliers produces a collection of $m+n$ simultaneous equations in $m+n$ variables that must be solved.

CAUTION When solving a system of equations, by hand or with technology, be sure to find all solutions.

EXERCISES for Section 16.7

In Exercises 1 to 14, use the method of Lagrange multipliers unless advised otherwise.

NOTE: When minimizing a distance, square roots can be avoided by minimizing the square of the distance.

1. Maximize xy for points on the circle $x^2 + y^2 = 4$.

2. Minimize $x^2 + y^2$ for points on the line $2x+3y=6$.

3. Minimize $2x+3y$ on the portion of the hyperbola $xy=1$ in the first quadrant.

4. Maximize $x+2y$ on the ellipse $x^2+2y^2=8$.

5. Find the largest area of rectangles whose perimeters are 12 centimeters.

6. A rectangular box is to have a volume of 1 cubic meter. Find its dimensions if its surface area is minimal.

7. Find the point on the plane $x+2y+3z=6$ that is closest to the origin
(a) using Lagrange multipliers and (b) using vector algebra.

8. Maximize $x+y+2z$ on the sphere $x^2+y^2+z^2=9$.

9. Minimize the distance from (x, y, z) to $(1,3,2)$ for points on the plane $2x+y+z=5$.
(a) using Lagrange multipliers and (b) using vector algebra.

10. Find the dimensions of the box of largest volume whose surface area is 6 square inches.

11. Maximize $x^2y^2z^2$ subject to $x^2+y^2+z^2=1$.

12. Find the points on the surface $xyz=1$ closest to the origin.

13. Minimize $x^2+y^2+z^2$ on the line common to the two planes $x+2y+3z=0$ and $2x+3y+z=4$.

14. The plane $2y+4z-5=0$ meets the cone $z^2=4(x^2+y^2)$ in a curve. Find the point on it nearest the origin.

---

15. (a) Sketch the elliptical paraboloid $z=x^2+2y^2$.
(b) Sketch the plane $x+y+z=1$.
(c) Sketch the intersection of the surfaces in (a) and (b).
(d) Find the highest point on the intersection in (c).

16. (a) Sketch the ellipsoid $x^2+y^2/4+z^2/9=1$ and the point $P(2,1,3)$.
(b) Find the point Q on the ellipsoid that is nearest P.
(c) What is the angle between PQ and the tangent plane at Q?

17. (a) Sketch the hyperboloid $x^2-y^2/4-z^2/9=1$. How many sheets does it have?
(b) Plot the point $(1,1,1)$. Is it above or below the hyperboloid?
(c) Find the point on the hyperboloid nearest P.

18. Maximize $x^3+y^3+2z^3$ on the intersection of the spheres $x^2+y^2+z^2=4$ and $(x-3)^2+y^2+z^2=4$.

In Exercises 19 to 22 solve the indicated exercise in Section 16.6 by Lagrange multipliers.

19. Exercise 25 **20.** Exercise 26 **21.** Exercise 29 **22.** Exercise 30

---

23. Solve Example 4 by vector algebra.

24. Solve Exercise 13 by vector algebra.

---

25. Show that a triangle in which the product of the sines of the three angles is maximized is equilateral.

26. Solve Exercise 25 by labeling the angles x, y, and $\pi - x - y$ and minimizing a function of x and y by the method of Section 16.6.

27. Maximize $x + 2y + 3z$ subject to the constraints $x^2 + y^2 + z^2 = 2$ and $x + y + z = 0$.

28. (a) Maximize $x_1 x_2 \cdots x_n$ subject to the constraints that $\sum_{i=1}^{n} x_i = 1$ and all $x_i \geq 0$.

(b) Deduce that for nonnegative numbers $a_1, a_2, \ldots, a_n$, $\sqrt[n]{a_1 a_2 \cdots a_n} \leq (a_1 + a_2 + \cdots + a_n)/n$.

Observation 16.7.2: ***Arithmetic-Geometric Mean Inequality***

Exercise 28 proves that the **geometric mean** of n nonnegative numbers is less than or equal to the **arithmetic mean**. (See also Exercise 50 in Section 16.6.)

29. (a) Maximize $\sum_{i=1}^{n} x_i y_i$ subject to the constraints $\sum_{i=1}^{n} x_i^2 = 1$ and $\sum_{i=1}^{n} y_i^2 = 1$.

(b) Deduce the **Cauchy-Schwarz inequality**: for numbers $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots b_n$,

$$\sum_{i=1}^{n} a_i b_i \leq \left(\sum_{i=1}^{n} a_i^2\right)^{1/2} \left(\sum_{i=1}^{n} b_i^2\right)^{1/2}. \tag{16.7.22}$$

(c) How would you justify the inequality in (b), for $n = 3$, using vectors?

30. Let $a_1, a_2, \ldots, a_n$ be fixed nonzero numbers. Maximize $\sum_{i=1}^{n} a_i x_i$ subject to $\sum_{i=1}^{n} x_i^2 = 1$.

31. Let p and q be positive numbers that satisfy the equation $\frac{1}{p} + \frac{1}{q} = 1$. Obtain **Hölder's inequality** for nonnegative numbers a_i and b_i,

$$\sum_{i=1}^{n} a_i b_i \leq \left(\sum_{i=1}^{n} a_i^p\right)^{1/p} \left(\sum_{i=1}^{n} b_i^q\right)^{1/q}, \tag{16.7.23}$$

as follows.

(a) Maximize $\sum_{i=1}^{n} x_i y_i$ subject to $\sum_{i=1}^{n} x_i^p = 1$ and $\sum_{i=1}^{n} y_i^q = 1$.

(b) By letting $x_i = \dfrac{a_i}{\left(\sum_{i=1}^{n} a_i^p\right)^{1/p}}$ and $y_i = \dfrac{b_i}{\left(\sum_{i=1}^{n} b_i^q\right)^{1/q}}$, obtain Hölder's inequality.

Observation 16.7.3: ***A Connection Between Cauchy-Schwarz and Hölder's Inequalities***

Hölder's inequality, (16.7.23), with $p = 2$ and $q = 2$, reduces to the Cauchy-Schwarz inequality, (16.7.22).

16.8 Thermodynamics and Partial Derivatives

Some basic equations of thermodynamics follow from the chain rule and the equality of mixed partial derivatives. We will describe the mathematics within the thermodynamics context. This section may serve as a review or as a reference.

Implications of the Chain Rule

We start with a function of three variables, $f(x, y, z)$, which we assume has first partial derivatives

$$\left.\frac{\partial f}{\partial x}\right|_{y,z} \qquad \left.\frac{\partial f}{\partial y}\right|_{x,z} \qquad \left.\frac{\partial f}{\partial z}\right|_{x,y}.$$

The subscripts denote the variables held fixed. Until now, these subscripts have been omitted because there was no possible confusion about the variables that were being held constant. However, in this section we will encounter cases where the subscripts must be used for the sake of clarity.

Assume that z is a function of x and y, $z = g(x, y)$. Then $f(x, y, z) = f(x, y, g(x, y))$ is a function of two variables. We call it $h(x, y)$: $h(x, y) = f(x, y, g(x, y))$. There are only two first-order partial derivatives of h:

$$\left.\frac{\partial h}{\partial x}\right|_{y} \quad \text{and} \quad \left.\frac{\partial h}{\partial y}\right|_{x}.$$

Let the value of $f(x, y, z)$ be called u, so $u = f(x, y, z)$. But x, y, and z are functions of x and y: $x = x$, $y = y$, and $z = g(x, y)$. So, actually, $u = h(x, y)$.

Figure 16.8.1 provides a pictorial view of the relationship between the variables. Both x and y appear as middle and independent variables. We have $u = f(x, y, z)$ and also $u = h(x, y)$. By the chain rule

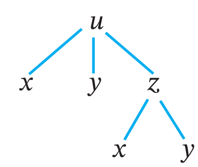


Figure 16.8.1

$$\left.\frac{\partial h}{\partial x}\right|_{y} = \left.\frac{\partial f}{\partial x}\right|_{y,z} \left.\frac{\partial x}{\partial x}\right|_{y} + \left.\frac{\partial f}{\partial y}\right|_{x,z} \left.\frac{\partial y}{\partial x}\right|_{y} + \left.\frac{\partial f}{\partial z}\right|_{x,y} \left.\frac{\partial g}{\partial x}\right|_{y}.$$

We know $\partial x/\partial x = 1$. Because x and y are independent variables, $\partial y/\partial x = 0$ and we have

$$\left.\frac{\partial h}{\partial x}\right|_{y} = \left.\frac{\partial f}{\partial x}\right|_{y,z} + \left.\frac{\partial f}{\partial z}\right|_{x,y} \left.\frac{\partial g}{\partial x}\right|_{y}, \tag{16.8.1}$$

or

$$\frac{\partial h}{\partial x} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial g}{\partial x}. \tag{16.8.2}$$

When the subscripts are omitted, we look back at the definitions of f, g, and h to see which variables are held fixed.

Observation 16.8.1: *Changing One Variable can Affect Multiple Variables*

A change in x affects f directly, but a change in x also causes a change in z, thus indirectly affecting f.

EXAMPLE 1. We check (16.8.2) for $h(x, y) = f(x, y, z)$ when $f(x, y, z) = x^2y^3z^5$ and $z = g(x, y) = 2x + 3y$.

SOLUTION In this case $h(x, y) = f(x, y, g(x, y)) = x^2y^3(2x+3y)^5$. Computing $\partial h/\partial x$ directly gives

$$\begin{aligned}\frac{\partial h}{\partial x} &= \frac{\partial}{\partial x}\left(x^2y^3(2x+3y)^5\right)\\ &= y^3\frac{\partial}{\partial x}\left(x^2(2x+3y)^5\right)\\ &= y^3\left(2x(2x+3y)^5 + x^2\left(5(2x+3y)^4(2)\right)\right)\\ &= 2xy^3(2x+3y)^5 + 10x^2y^3(2x+3y)^4. \end{aligned} \tag{16.8.3}$$

Now, let us find $\partial h/\partial x$ with the aid of (16.8.2). We have $\partial f/\partial x = 2xy^3z^5$, $\partial f/\partial z = 5x^2y^3z^4$, and $\partial g/\partial x = 2$. Thus

$$\begin{aligned}\frac{\partial h}{\partial x} &= \frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial g}{\partial x}\\ &= 2xy^3z^5 + (5x^2y^3z^4)(2)\\ &= 2xy^3(2x+3y)^5 + 10x^2y^3(2x+3y)^4,\end{aligned}$$

which agrees with (16.8.3). ♦

What If $z = g(x, y)$ Makes $f(x, y, z)$ Constant?

Assume that when z is replaced by $g(x, y)$, $h(x, y) = f(x, y, g(x, y))$ is constant: $h(x, y) = f(x, y, g(x, y)) = C$. This happens when we use the equation $f(x, y, z) = C$ to determine z implicitly as a function of x and y. Then

$$\left.\frac{\partial h}{\partial x}\right|_y = 0 \quad \text{and} \quad \left.\frac{\partial h}{\partial y}\right|_x = 0.$$

In this case, which occurs frequently in thermodynamics, (16.8.1) becomes

$$0 = \left.\frac{\partial f}{\partial x}\right|_{y,z} + \left.\frac{\partial f}{\partial z}\right|_{x,y}\left.\frac{\partial z}{\partial x}\right|_y. \tag{16.8.4}$$

PREVIEW: Equation (16.8.4) is the foundation for (16.8.8) and (16.8.9), two key equations in thermodynamics.

Solving (16.8.4) for $(\partial z/\partial x)|_y$ we obtain

$$\left.\frac{\partial z}{\partial x}\right|_y = \frac{-\left.\frac{\partial f}{\partial x}\right|_{y,z}}{\left.\frac{\partial f}{\partial z}\right|_{x,y}}. \tag{16.8.5}$$

Equation (16.8.5) expresses the partial derivative of $g(x, y)$ with respect to x in terms of the partial derivatives of the original function $f(x, y, z)$.

NOTATION: Writing the left-hand side of (16.8.5) as $\left.\frac{\partial z}{\partial x}\right|_y$, not $\frac{\partial g(x,y)}{\partial x}$, helps to remember this result.

EXAMPLE 2. Let $f(x, y, z) = x^3y^5z^7$. Define $z = g(x, y)$ implicitly by $x^3y^5(g(x, y))^7 = 1$. That is, $z = g(x, y) = x^{-3/7}y^{-5/7}$. Verify (16.8.5).

SOLUTION We begin by computing some partial derivatives of $f(x, y, z)$ and $z = g(x, y)$:

$$\left.\frac{\partial z}{\partial x}\right|_y = \frac{-3}{7}x^{-10/7}y^{-5/7}, \qquad \left.\frac{\partial f}{\partial x}\right|_{y,z} = 3x^2y^5z^7, \quad \text{and} \quad \left.\frac{\partial f}{\partial z}\right|_{x,y} = 7x^3y^5z^6.$$

Substituting in the right-hand side of (16.8.5), yields

$$\begin{aligned}
\frac{-\left.\frac{\partial f}{\partial x}\right|_{y,z}}{\left.\frac{\partial f}{\partial z}\right|_{x,y}} &= \frac{-\left(3x^2y^5z^7\right)}{7x^3y^5z^6} \\
&= -\frac{3}{7}x^{-1}z && (\text{ simplifying powers }) \\
&= -\frac{3}{7}x^{-1}\left(x^{-3/7}y^{-5/7}\right) && (\text{ because } x^3y^5z^7 = 1 \text{ we find } z = x^{-3/7}y^{-5/7}\text{ }) \\
&= -\frac{3}{7}x^{-10/7}y^{-5/7} && (\text{ combining powers of } x\text{ }) \\
&= \left.\frac{\partial z}{\partial x}\right|_{y} && (\text{ because, again, } z = x^{-3/7}y^{-5/7}\text{ }).
\end{aligned}$$

Thus, (16.8.5) is satisfied. ♦

The Reciprocity Relations

A thermodynamics text has equations of the form $\left.\frac{\partial x}{\partial z}\right|_{y} = \frac{1}{\left.\frac{\partial z}{\partial x}\right|_{y}}$.

In this section, we will explain where this comes from, presenting the mathematical details often glossed over in the applied setting. The fact that a certain quantity is conserved is frequently expressed by saying a function $f(x,y,z)$ has a constant value,C: $f(x,y,z) = C$.

It is typically assumed that this equation determines z as a function of x and y, or, similarly, determines x as a function of y and z, or y as a function of x and z. There are six first partial derivatives:

$$\left.\frac{\partial z}{\partial x}\right|_{y}, \quad \left.\frac{\partial z}{\partial y}\right|_{x}, \quad \left.\frac{\partial x}{\partial y}\right|_{z}, \quad \left.\frac{\partial x}{\partial z}\right|_{y}, \quad \left.\frac{\partial y}{\partial x}\right|_{z}, \text{ and } \quad \left.\frac{\partial y}{\partial z}\right|_{x}. \tag{16.8.6}$$

See Exercise 5 for an interesting relationship between these six derivatives.

An equation analogous to (16.8.5) holds for each partial derivative in (16.8.6). For instance,

$$\left.\frac{\partial x}{\partial z}\right|_{y} = \frac{-\left.\frac{\partial f}{\partial z}\right|_{x,y}}{\left.\frac{\partial f}{\partial x}\right|_{y,z}}. \tag{16.8.7}$$

Combining equations (16.8.5) and (16.8.7) yields another reciprocity relation, which we state as a theorem.

Theorem 16.8.2: A Reciprocity Relation

Suppose x, y, and z satisfy a relationship $f(x,y,z) = C$ for a differentiable function f and for some constant C. Then, for example,

$$\left.\frac{\partial x}{\partial z}\right|_{y} = \frac{1}{\left.\frac{\partial z}{\partial x}\right|_{y}}. \tag{16.8.8}$$

Similar relationships hold for other pairs of variables.

Observation 16.8.3: *Motivation for Theorem 16.8.2*

Theorem 16.8.2 is to be expected, for $\Delta z/\Delta x$ is the reciprocal of $\Delta x/\Delta z$.

Equation (16.8.8) is an example of a **reciprocity relation**: The partial derivative of one variable with respect to a second variable is the reciprocal of the partial derivative of the second variable with respect to the first variable.

EXAMPLE 3. Let $2x+3y+5z=12$. Verify that $\partial z/\partial x$ is the reciprocal of $\partial x/\partial z$.

SOLUTION In this case it is easy to solve for z in terms of x and y: $z=(12-2x-3y)/5$. Then $\partial z/\partial x=-2/5$.

Also, solving $2x+3y+5z=12$ for x in terms of y and z: $x=(12-3y-5z)/2$, so $\partial x/\partial z=-5/2$, which is, as it should be, the reciprocal of $\partial z/\partial x$. ♦

When $f(x,y,z)=C$ for a nonlinear function f it may not be so easy to find an explicit formula for each variable in terms of the other two variables. In these cases, use an formula like (16.8.7) to compute the needed partial derivative.

The Cyclic Relations

With the aid of equations like (16.8.7) we can establish the surprising relation

Theorem 16.8.4: The Cyclic Relations

Suppose x, y, and z satisfy a relationship $f(x,y,z)=C$ for a differentiable function f and for some constant C. Then

$$\left.\frac{\partial x}{\partial y}\right|_z \left.\frac{\partial y}{\partial z}\right|_x \left.\frac{\partial z}{\partial x}\right|_y = -1. \tag{16.8.9}$$

VOCABULARY: *The cyclic relation are sometimes referred to by other names, including the Triple Product Rule, the Cyclic Chain Rule, or Euler's Chain Rule.*

Proof of Theorem 16.8.4

Equation (16.8.9) results from the use of three versions of (16.8.7). The left-hand side of (16.8.9) can be expressed as

$$\left(\frac{-\left.\frac{\partial f}{\partial y}\right|_{x,z}}{\left.\frac{\partial f}{\partial x}\right|_{y,z}}\right)\left(\frac{-\left.\frac{\partial f}{\partial z}\right|_{x,y}}{\left.\frac{\partial f}{\partial y}\right|_{x,z}}\right)\left(\frac{-\left.\frac{\partial f}{\partial x}\right|_{y,z}}{\left.\frac{\partial f}{\partial z}\right|_{x,y}}\right). \tag{16.8.10}$$

Cancellation simplifies (16.8.10) to -1. ●

While the relationship between x, y, and z in the next example is very simple, do not let the fact that the manipulations are extremely simple detract from the significance of the result.

EXAMPLE 4. Let $f(x,y,z)=2x+3y+5z=12$. The equation determines implicitly each variable in terms of the others. Verify (16.8.9) in this case.

SOLUTION From $2x+3y+5z=12$,

$$x=\frac{12-3y-5z}{2}, \qquad y=\frac{12-2x-5z}{3}, \text{ and } \quad z=\frac{12-2x-3y}{5}.$$

Then $\partial x/\partial y = -3/2$, $\partial y/\partial z = -5/3$, and $\partial z/\partial x = -2/5$, and we have

$$\frac{\partial x}{\partial y}\frac{\partial y}{\partial z}\frac{\partial z}{\partial x} = \left(\frac{-3}{2}\right)\left(\frac{-5}{3}\right)\left(\frac{-2}{5}\right) = -1.$$

♦

Observation 16.8.5: ***Important Use of the Cyclic Relations***

If two of the three partial derivatives in (16.8.9) are easy to calculate, then we can use (16.8.9) to find the third, which may otherwise be hard to calculate.

The next example illustrates the use of the cyclic relation in thermodynamics.

EXAMPLE 5. In van der Waal's equation p, T, and v are related by

v is the reciprocal of density

$$p = \frac{RT}{v-b} - \frac{a}{v^2}, \tag{16.8.11}$$

where R, a and b are constants. Find $\left.\frac{\partial v}{\partial T}\right|_p$.

Exercises 13 and 14 present other ways to complete Example 5.

SOLUTION It is not reasonable to solve (16.8.11) for v. Instead, we consider using the cyclic relation

$$\left.\frac{\partial v}{\partial T}\right|_p \left.\frac{\partial T}{\partial p}\right|_v \left.\frac{\partial p}{\partial v}\right|_T = -1. \tag{16.8.12}$$

Looking at (16.8.11), we see that $(\partial p/\partial T)\big|_v$ is easier to calculate than $(\partial T/\partial p)\big|_v$. The reciprocity relation for $(\partial T/\partial p)\big|_v$ can be used to rewrite (16.8.12) as

$$\frac{\left.\frac{\partial v}{\partial T}\right|_p \left.\frac{\partial p}{\partial v}\right|_T}{\left.\frac{\partial p}{\partial T}\right|_v} = -1$$

and therefore

$$\left.\frac{\partial v}{\partial T}\right|_p = -\frac{\left.\frac{\partial p}{\partial T}\right|_v}{\left.\frac{\partial p}{\partial v}\right|_T}. \tag{16.8.13}$$

Since p is given as a function of v and T, the numerator and denominator in (16.8.13) are easy to find. They are

$$\left.\frac{\partial p}{\partial T}\right|_v = \frac{R}{v-b} \qquad \text{and} \qquad \left.\frac{\partial p}{\partial v}\right|_T = \frac{-RT}{(v-b)^2} + \frac{2a}{v^3}.$$

Thus, by (16.8.13),

$$\left.\frac{\partial v}{\partial T}\right|_p = \frac{-R/(v-b)}{-RT/(v-b)^2 + 2a/v^3}.$$

♦

Observation 16.8.6: ***There are Other Equations of State***

van der Waal's equation is only one example of an **equation of state**. Other equations of state for p, T, and v are considered in Exercises 11 and 12.

Using the Equality of the Mixed Partial Derivatives

Having shown how the chain rule provides some of the basic equations in thermodynamics, let us show how the equality of the mixed partials is used.

In a thermodynamic process there may be a pressure p, a temperature T, and a volume per unit mass v. Other common variables are listed in Table 16.8.1.

Variable	Interpretation
u	thermal energy per unit mass
s	entropy per unit mass
a	Helmholtz free energy per unit mass
g	Gibbs free energy per unit mass
h	enthalpy per unit mass

Table 16.8.1

That brings the total number of variables to eight. If they were independent, the states would be part of an eight-dimensional space, but they are not. In fact, any two determine all the others.

For instance, u may be viewed as a function of s and v, and we have $(\partial u/\partial s)|_v$, which is the definition of temperature, T. Thermodynamic texts either state or derive the **Gibbs relation**:

$$du = T\,ds - p\,dv.$$

It tells us that u is viewed as a function of s and v, and that

$$\left.\frac{\partial u}{\partial s}\right|_v = T \quad \text{and} \quad \left.\frac{\partial u}{\partial v}\right|_s = -p.$$

When you look at a thermometer, you are gazing at the value of a partial derivative.

We then have

$$\begin{aligned}
\frac{\partial^2 u}{\partial v \partial s} &= \frac{\partial^2 u}{\partial s \partial v} && (\text{ equality of mixed partials of } u(s,v)\text{)}\\
\frac{\partial}{\partial v}\left(\frac{\partial u}{\partial s}\right) &= \frac{\partial}{\partial s}\left(\frac{\partial u}{\partial v}\right) &&\\
\left.\frac{\partial T}{\partial v}\right|_s &= \left.\frac{\partial(-p)}{\partial s}\right|_v && (\text{ because } (\partial u/\partial s)|_v = T \text{ and } (\partial u/\partial v)|_s = -p\text{)}\\
\left.\frac{\partial T}{\partial v}\right|_s &= -\left.\frac{\partial p}{\partial s}\right|_v. &&
\end{aligned}$$

Several thermodynamic statements that equate two partial derivatives, such as the final equation above, are obtained this way. The starting point is

$$dz = M\,dx + N\,dy$$

where M is $(\partial z/\partial x)|_y$ and N is $(\partial z/\partial y)\big|_x$. Then, because

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x},$$

it follows that

$$\left.\frac{\partial M}{\partial y}\right|_x = \left.\frac{\partial N}{\partial x}\right|_y.$$

Summary

We showed how the chain rule in the special case where an intermediate variable is also a final variable justifies certain identities, namely, the reciprocal and cyclic relations. Then we showed how the equality of the mixed partial derivatives is used to derive other relationships involving partial derivatives.

EXERCISES for Section 16.8

1. Let $u = x^2 + y^2 + z^2$ and let $z = x + y$.

(a) The symbol $\frac{\partial u}{\partial x}$ has two interpretations. What are they?

(b) Evaluate $\frac{\partial u}{\partial x}$ in both cases.

(c) Using subscripts, distinguish the two partial derivatives.

2. Let $z = rst$ and let $r = st$.

(a) The symbol $\frac{\partial z}{\partial t}$ has two interpretations. What are they?

(b) Evaluate $\frac{\partial z}{\partial t}$ in both cases.

(c) Using subscripts, distinguish the two partial derivatives.

3. Let $u = f(x, y, z)$ and $z = g(x, y)$. Then u is indirectly a function of x and y. Express $\left.\frac{\partial u}{\partial y}\right|_x$ in terms of partial derivatives of f. Supply all the steps.

4. Assume that the equation $f(x, y, z) = C$, C a constant, determines x as a function of y and z: $x = h(y, z)$. Express $\left.\frac{\partial x}{\partial y}\right|_z$ in terms of partial derivatives of f. Supply all the steps.

5. What is the product of the six partial derivatives in (16.8.6)?

6. Using f from Example 2, verify the analog of (16.8.7) for $\left.\frac{\partial z}{\partial y}\right|_x$.

7. Let $f(x, y, z) = 2x + 4y + 3z$. The equation $f(x, y, z) = 7$ determines one variable as a function of the other two. Verify (16.8.7), where x is viewed as a function of y and z.

8. Obtain the cyclic relation $\left.\frac{\partial x}{\partial z}\right|_y \left.\frac{\partial z}{\partial y}\right|_x \left.\frac{\partial y}{\partial x}\right|_z = -1$.

9. Verify (16.8.9) for $f(x, y, z) = x^3y^5z^7 = 1$.

10. Verify (16.8.9) for $f(x, y, z) = 2x + 4y + 3z = 7$.

11. The equation of state for an ideal gas is $pv = RT$, where R is a constant. Find $\left.\frac{\partial v}{\partial T}\right|_p$.

12. The Redlich-Kwang equation $p = \frac{RT}{v - b} - \frac{a}{v(v + b)T^{1/2}}$ is an improvement upon van der Waal's equation of state (16.8.11) for gases and liquids. Find $\left.\frac{\partial v}{\partial T}\right|_p$.

13. Find $\left.\frac{\partial v}{\partial T}\right|_p$ in Example 5 by differentiating both sides of (16.8.11) with respect to T, holding p constant.

14. One way to find $\left.\frac{\partial v}{\partial T}\right|_p$ in Example 5 is by first finding an equation that expresses v in terms of T and p. What difficulty occurs in this approach?

15. In Example 5, find $\left.\frac{\partial v}{\partial p}\right|_T$, $\left.\frac{\partial T}{\partial v}\right|_p$, and $\left.\frac{\partial T}{\partial p}\right|_v$.

16. In thermodynamics there is the Gibbs relation $dh = T\,ds + v\,dp$. It is understood that $\left.\frac{\partial h}{\partial s}\right|_p = T$ and $\left.\frac{\partial h}{\partial p}\right|_s = v$. Deduce that $\left.\frac{\partial T}{\partial p}\right|_s = \left.\frac{\partial v}{\partial s}\right|_p$.

In Exercises 17 to 21 use these five steps as a guide to deriving the given equation.

(a) What are the dependent variables?
(b) What are the independent variables?
(c) What are the intermediate variables?
(d) Draw a diagram showing the paths from the dependent variables to the independent variables.
(e) Use the chain rule to complete the derivation.

17. $\left.\frac{\partial E}{\partial T}\right|_v = \left.\frac{\partial E}{\partial T}\right|_p + \left.\frac{\partial E}{\partial p}\right|_T \left.\frac{\partial p}{\partial T}\right|_v$

18. $\left.\frac{\partial E}{\partial v}\right|_p = \left.\frac{\partial E}{\partial T}\right|_p \left.\frac{\partial T}{\partial v}\right|_p$

19. $\left.\frac{\partial p}{\partial T}\right|_v = \frac{-\left.\frac{\partial v}{\partial T}\right|_p}{\left.\frac{\partial v}{\partial p}\right|_T}$

20. $\left.\frac{\partial p}{\partial T}\right|_v \left.\frac{\partial T}{\partial p}\right|_v = 1$

21. $\left.\frac{\partial E}{\partial p}\right|_v = \left.\frac{\partial E}{\partial T}\right|_p \left.\frac{\partial T}{\partial p}\right|_v + \left.\frac{\partial E}{\partial p}\right|_T$

---

22. Show that van der Waal's equation, (16.8.11) in Example 5, leads to $\left.\frac{\partial p}{\partial T}\right|_v \left.\frac{\partial T}{\partial v}\right|_p \left.\frac{\partial v}{\partial p}\right|_T = -1$.

23. Let $u = F(x,y,z)$ and $z = f(x,y)$. Thus u is a composite function of x and y: $u = G(x,y) = F(x,y,f(x,y))$. Assume that $G(x,y) = x^2 y$. Obtain a formula for $\frac{\partial f}{\partial x}$ in terms of $\frac{\partial F}{\partial x}$, $\frac{\partial F}{\partial y}$, and $\frac{\partial F}{\partial z}$. (All three need not appear in your answer.)

24. Let $u = F(x,y,z)$ and $x = f(y,z)$. Thus u is a composite function of y and z: $u = G(y,z) = F(f(y,z),y,z)$. Assume that $G(y,z) = 2y + z^2$. Obtain a formula for $\frac{\partial f}{\partial z}$ in terms of $\frac{\partial F}{\partial x}$, $\frac{\partial F}{\partial y}$, and $\frac{\partial F}{\partial z}$. (All three need not appear in your answer.)

25. Two functions u and v of x and y are defined implicitly by $F(u,v,x,y) = 0$ and $G(u,v,x,y) = 0$. Assuming all necessary differentiability, find a formula for $\frac{\partial u}{\partial x}$ in terms of partial derivatives of F and G.

16.9 Magnification of a Mapping

A function from a surface to a surface or from a solid region to a solid region that is one-to-one except perhaps on a boundary is called a mapping. This section defines the magnification of a mapping, which may vary from point to point, and develops ways to compute it for mappings between surfaces. How to compute it for mappings between intervals or between solid regions is dealt with in the exercises. Magnification will be used in the next chapter to evaluate integrals.

Mappings

A **mapping**, or **transformation**, is a one-to-one function F from a set $\mathscr{R}$ to a set $\mathscr{S}$ such that each point in $\mathscr{S}$ is an image of a point in $\mathscr{R}$. In the mappings of interest in this section, the sets will be planar or curved surfaces. On the boundaries of such regions the mappings may not be one-to-one. We will assume that the partial derivatives that appear are continuous. If $\mathscr{A}$ is a subset of $\mathscr{R}$, $F(\mathscr{A})$ denotes the subset of $\mathscr{S}$ consisting of the points that can be written as $F(P)$ for some point P in $\mathscr{A}$. In particular, $F(\mathscr{R}) = \mathscr{S}$.

The projection of a slide on a screen is a real-life example of a mapping. It projects a point on the slide to a point on the screen.

EXAMPLE 1. The mapping F connects the point (u, v) in the uv-plane and the point $(2u, 3v)$ in the xy-plane, that is $F(u, v) = (2u, 3v)$.

(a) Describe the mapping geometrically.
(b) Find the formula for $\text{inv}F = F^{-1}$.
(c) Show that the image of a line is a line.
(d) Find the image of the square whose vertices are $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$.
(e) Find the image of the unit disk, $u^2 + v^2 \le 1$.

SOLUTION

(a) If $F(u, v) = (x, y)$, we have $x = 2u$ and $y = 3v$. That implies that F magnifies horizontal distances by a factor of 2 and vertical distances by a factor of 3. Hence areas are magnified by a factor of 6.
(b) To find a formula for $\text{inv}F$ solve for u and v in terms of x and y. Because $x = 2u$ and $y = 3v$ we have $u = x/2$ and $v = y/3$. Thus $\text{inv}F$ maps (x, y) to $(x/2, y/3)$.
(c) A line L in the uv-plane has an equation of the form $au + bv + c = 0$ where not both a and b are zero. If (x, y) is in the image of the line, then $(\text{inv}F)(x, y)$ lies on L. Thus $(x/2, y/3)$ lies on L in the uv-plane. This implies that

$$a\frac{x}{2} + b\frac{y}{3} + c = 0.$$

Clearing denominators, we conclude that the image of L, $F(L)$, is described by the equation $3ax + 2by + 6c = 0$; the image of L is another line.
(d) The square $\mathscr{R}$ in Figure 16.9.1(a) is bounded by four lines. So its image is also bounded by four lines (but may not be a square). F takes the four corners of the square to $F(0,0) = (0,0)$, $F(1,0) = (2,0)$, $F(1,1) = (2,3)$, and $F(0,1) = (0,3)$. The image is the rectangle $\mathscr{S}$ in Figure 16.9.1(a). As expected, the area of $\mathscr{S}$ is six times the area of square $\mathscr{R}$.
(e) If $u^2 + v^2 \le 1$ and $(x, y) = F(u, v)$, then $(x/2)^2 + \left(y/3\right)^2 \le 1$. Figure 16.9.1(b) shows that F maps the unit circle in the uv-plane to an ellipse in the xy-plane.

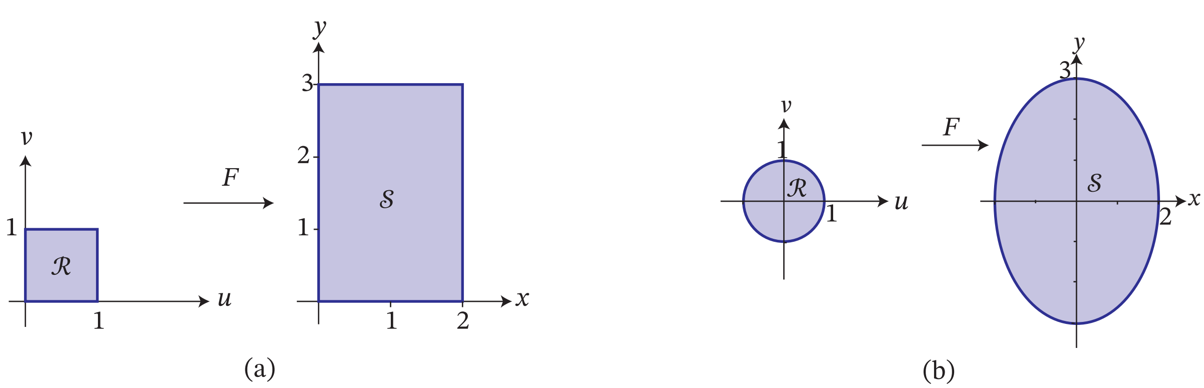


Figure 16.9.1

♦

The next example shows how coordinate systems, other than rectangular coordinates, can be introduced. We begin with polar coordinates. Other coordinate systems are studied and developed in the remaining sections.

EXAMPLE 2.

(a) Analyze the mapping F given by $F(r,\theta) = (r\cos(\theta), r\sin(\theta))$ from the $r\theta$-plane to the xy-plane.

(b) Sketch the set $\mathscr{R} = \left\{1 \le r \le 2, \frac{\pi}{6} \le \theta \le \frac{\pi}{4}\right\}$ on the $r\theta$-plane and its image $\mathscr{S}$ in the xy-plane.

SOLUTION

(a) First of all, r and θ, in spite of their appearance, are rectangular coordinates in the $r\theta$-plane. They play the role usually played by u and v. In the domain of F, r and θ are rectangular coordinates. In the range of F we have $x = r\cos(\theta)$ and $y = r\sin(\theta)$. So, in the xy-plane, r and θ are polar coordinates of (x, y). We may write $F(r,\theta) = (r,\theta)$ as long as we keep in mind that on the left-hand side of that equation (r,θ) are rectangular coordinates in the $r\theta$-plane, but on the right-hand side (r,θ) are polar coordinates in the xy-plane.

(b) In the $r\theta$-plane the set described by the inequalities $1 \le r \le 2$, $\pi/6 \le \theta \le \pi/4$ is the rectangle $\mathscr{S}$ shown in Figure 16.9.2. In the xy-plane it is not a rectangle but a sector with two straight lines and two curved sides. The curved sides lie on circles whose radii are 1 and 2.

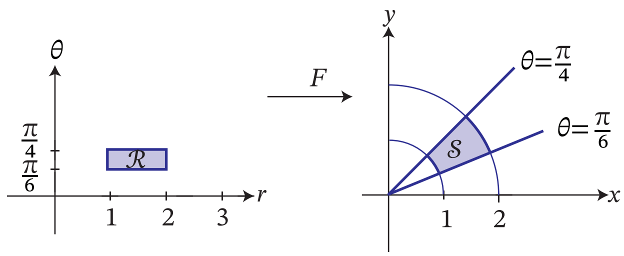


Figure 16.9.2

The mapping lifts the coordinates from the $r\theta$-plane, where they are ordinary rectangular coordinates, and places them like tags on another plane, where they become polar coordinates in the xy-plane.

Magnification of a Mapping

The mapping F given by $F(u, v) = (2u, 3v)$ in Example 1 magnifies all areas by a factor of 6. In the next example the magnification varies from point to point.

EXAMPLE 3. Determine the magnification of the mapping in Example 2.

SOLUTION

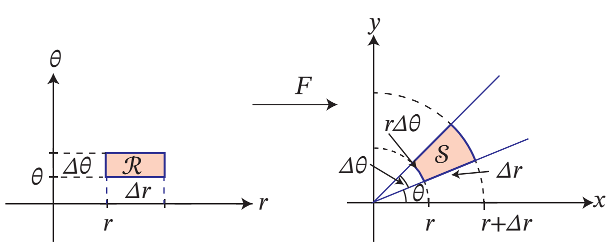


Figure 16.9.3

To find the magnification at (r,θ) in the $r\theta$-plane we compare the area of a small rectangle which has one corner at (r,θ) with the area of its image $\mathscr{S} = F(\mathscr{R})$. The dimensions of $\mathscr{R}$ are Δr and $\Delta\theta$. Thus the area of $\mathscr{R}$ is their product, $\Delta r\,\Delta\theta$. See Figure 16.9.3.

The area of $\mathscr{S}$ is a bit more complicated, until it is recognized as a sector of the washer with outer radius $r + \Delta r$ and inner radius r. As such, its area is $\left((r+\Delta r)^2/2 - r^2/2\right)\Delta\theta = \left(r\,\Delta r + (\Delta r)^2/2\right)\Delta\theta$.

The quotient of the two areas is

$$\frac{\left(r\,\Delta r + \frac{1}{2}(\Delta r)^2\right)\Delta\theta}{\Delta r\,\Delta\theta} = r + \frac{1}{2}\Delta r.$$

To find the magnification of F at the point (r,θ), we take the limit as the rectangle $\mathscr{R}$ shrinks to the point (r,θ), that is, both Δr and $\Delta\theta$ decrease to zero. For this reason we conclude that the magnification of $F(r,\theta) = (r\cos(\theta), r\sin(\theta))$ at (r,θ) is r.

We found the magnification in Example 3 with the aid of simple geometric figures. We now obtain a general formula for the magnification of a mapping by answering the question: "If $F(u,v) = \big(f(u,v), g(u,v)\big)$, by what factor does it magnify or shrink the area of a small patch near a point (u_0, v_0)?"

To answer this we will determine how much the mapping F magnifies the area of a small rectangle near (u_0, v_0). For positive changes Δu in u and Δv in v define $\mathscr{B}$ to be the rectangle in the uv-plane whose vertices are (u_0, v_0), $(u_0+\Delta u, v_0)$, $(u_0+\Delta u, v_0+\Delta v)$, $(u_0, v_0+\Delta v)$ shown in Figure 16.9.4. Its area is $\Delta u \Delta v$. The image $F(\mathscr{B})$, which we call $\mathscr{C}$, is bounded by curves that are the images of the edges of $\mathscr{B}$. They need not be straight lines. However, when $F(\mathscr{B})$ is small they resemble straight lines.

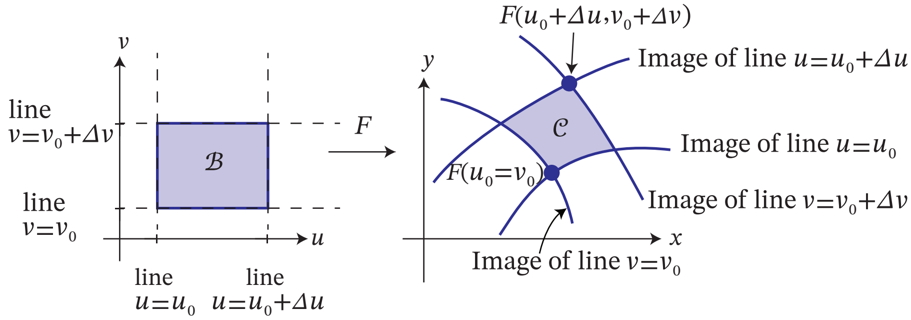


Figure 16.9.4

The magnification of F at (u_0, v_0) is defined as a limit:

$$\lim_{\Delta u, \Delta v \to 0} \frac{\text{Area of } \mathscr{C}}{\text{Area of } \mathscr{B}} = \lim_{\Delta u, \Delta v \to 0} \frac{\text{Area of } \mathscr{C}}{\Delta u \Delta v}$$

To express the limit in terms of $f(u,v)$ and $g(u,v)$, we will estimate the area of $\mathscr{C}$. using vectors and the fact that $|\mathbf{A} \times \mathbf{B}|$ is the area of the parallelogram spanned by the vectors $\mathbf{A}$ and $\mathbf{B}$.

If O is the origin of the xy-coordinates, define $\mathbf{r}$ to be the vector $\overrightarrow{OF}(u,v)$. The change in $\mathbf{r}$ due to a change Δu in u is $\Delta\mathbf{r} = \mathbf{r}(u_0+\Delta u, v_0) - \mathbf{r}(u_0, v_0)$. Now, $\Delta\mathbf{r}/\Delta u$ approximates $\partial\mathbf{r}/\partial u$ at (u_0, v_0). If follows that $\partial\mathbf{r}/\partial u$ approximates $\Delta\mathbf{r}/\Delta u$. Thus

$$\Delta\mathbf{r} = \frac{\Delta\mathbf{r}}{\Delta u}\Delta u \approx \left(\frac{\partial\mathbf{r}}{\partial u}\right)\Delta u.$$

Similarly, $(\partial\mathbf{r}/\partial v)\,\Delta v$ approximates the change in $\mathbf{r}$ due to the change Δv in v. With these observations, we are ready to estimate the area of $\mathscr{C}$.

The vector $(\partial\mathbf{r}/\partial u)\,\Delta u$ is tangent to the curve between $F(u_0, v_0)$ and $F(u_0+\Delta u, v_0)$ and approximates the vector $\mathbf{r}(u_0+\Delta u, v_0) - \mathbf{r}(u_0, v_0)$. (See Figure 16.9.5.) Similarly, $(\partial\mathbf{r}/\partial v)\,\Delta v$ approximates the vector from $F(u_0, v_0)$ to $F(u_0, v_0+\Delta v)$. Then the area of the parallelogram spanned by the vectors $(\partial\mathbf{r}/\partial u)\,\Delta u$ and $(\partial\mathbf{r}/\partial v)\,\Delta v$ approximates the area of $\mathscr{C}$. For this reason we expect that

$$\begin{aligned}
\text{Magnification of } F \text{ at } (u_0, v_0) &= \lim_{\Delta u, \Delta v \to 0} \frac{\text{Area of parallelogram}}{\Delta u \Delta v} \\
&= \lim_{\Delta u, \Delta v \to 0} \frac{\left|\dfrac{\partial\mathbf{r}}{\partial u}\Delta u \times \dfrac{\partial\mathbf{r}}{\partial v}\Delta v\right|}{\Delta u \Delta v} \\
&= \lim_{\Delta u, \Delta v \to 0} \frac{\left|\dfrac{\partial\mathbf{r}}{\partial u} \times \dfrac{\partial\mathbf{r}}{\partial v}\right|\Delta u \Delta v}{\Delta u \Delta v} \\
&= \left|\frac{\partial\mathbf{r}}{\partial u} \times \frac{\partial\mathbf{r}}{\partial v}\right|.
\end{aligned}$$

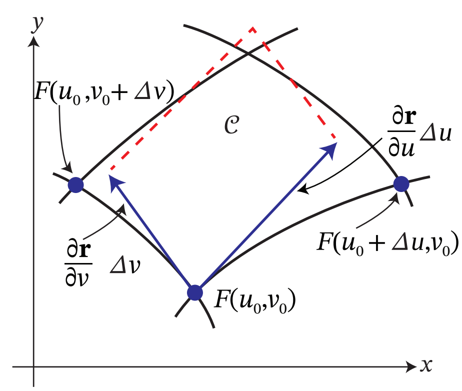


Figure 16.9.5

Because $\partial\mathbf{r}/\partial u$ and $\partial\mathbf{r}/\partial v$ are evaluated at (u_0, v_0) we have

Definition: *Magnification and a Cross Product*

The **magnification** of the mapping $(x, y) = F(u, v)$, with $\mathbf{r}(u, v) = x\mathbf{i} + y\mathbf{j}$, at a point (u_0, v_0) is

$$\text{Magnification of } F \text{ at } (u_0, v_0) = \left|\frac{\partial \mathbf{r}}{\partial u}(u_0, v_0) \times \frac{\partial \mathbf{r}}{\partial v}(u_0, v_0)\right|. \tag{16.9.1}$$

For example, the magnification of $F(u, v) = \big(f(u, v), g(u, v)\big)$ at (u_0, v_0) is

$$\left|\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial f}{\partial u} & \frac{\partial g}{\partial u} & 0 \\ \frac{\partial f}{\partial v} & \frac{\partial g}{\partial v} & 0 \end{pmatrix}\right| = \left|0\mathbf{i} - 0\mathbf{j} + \det\begin{pmatrix} \frac{\partial f}{\partial u} & \frac{\partial g}{\partial u} \\ \frac{\partial f}{\partial v} & \frac{\partial g}{\partial v} \end{pmatrix}\mathbf{k}\right| = \left|\left(\frac{\partial f}{\partial u}\frac{\partial g}{\partial v} - \frac{\partial f}{\partial v}\frac{\partial g}{\partial u}\right)\mathbf{k}\right| = \left|\frac{\partial f}{\partial u}\frac{\partial g}{\partial v} - \frac{\partial f}{\partial v}\frac{\partial g}{\partial u}\right|.$$

The magnification is the absolute value of the determinant of the 2×2 matrix

$$\begin{pmatrix} \frac{\partial f}{\partial u} & \frac{\partial g}{\partial u} \\ \frac{\partial f}{\partial v} & \frac{\partial g}{\partial v} \end{pmatrix}. \tag{16.9.2}$$

The magnification at a point P will be denoted $M(P)$. Where the magnification is greater than 1, F increases area; where it is less than 1 it shrinks area. But, a magnification can never be negative. (Why? What has to happen for the magnification to be zero?)

TERMINOLOGY: It may sound odd to treat shrinking area as a form of magnification, but we have no trouble understanding when someone talks about going down in an elevator or when a football team gains −3 yards on a play.

EXAMPLE 4. Find the magnification of the mapping $F(u, v) = (u\cos(v), u\sin(v))$ Assume u is positive.

SOLUTION First, observe that this is the mapping in Example 3 with u and v playing the roles of r and θ, respectively. We have $x = u\cos(v)$ and $y = u\sin(v)$, hence the mapping is $F(u.v) = (f(u, v), g(u, v))$ where $f(u, v) = u\cos(v)$ and $g(u, v) = u\sin(v)$. The matrix (16.9.2) becomes

$$\begin{pmatrix} \cos(v) & \sin(v) \\ -u\sin(v) & u\cos(v) \end{pmatrix}$$

whose determinant is $u\cos^2(v) + u\sin^2(v) = u$, in agreement with Example 3. ♦

In the next example only one of the regions is planar. It is a preview of what will encounter when discussing integrals over curved surfaces in Section 17.7. See also Exercise 16.

EXAMPLE 5. The function $f(x, y)$ is defined in a region $\mathscr{R}$ in the xy-plane and $\mathscr{S}$ is the surface $z = f(x, y)$ above or below $\mathscr{S}$. $F(x, y) = (x, y, f(x, y))$ is a mapping from $\mathscr{R}$ to $\mathscr{S}$. Find the magnification $M_F(x, y)$.

SOLUTION The mapping F, shown in Figure 16.9.6, is the inverse of the projection from the surface to the xy-plane which sends $(x, y, f(x, y))$ to (x, y). We have

$$\mathbf{r}(x, y) = x\mathbf{i} + y\mathbf{j} + f(x, y)\mathbf{k}.$$

The magnification is the length of $\partial\mathbf{r}/\partial \times \times\partial\mathbf{r}/\partial y$. The cross product is

$$\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial x}{\partial x} & \frac{\partial y}{\partial x} & \frac{\partial f}{\partial x} \\ \frac{\partial x}{\partial y} & \frac{\partial y}{\partial y} & \frac{\partial f}{\partial y} \end{pmatrix} = \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & \frac{\partial f}{\partial x} \\ 0 & 1 & \frac{\partial f}{\partial y} \end{pmatrix} = -\frac{\partial f}{\partial x}\mathbf{i} - \frac{\partial f}{\partial y}\mathbf{j} + \mathbf{k}.$$

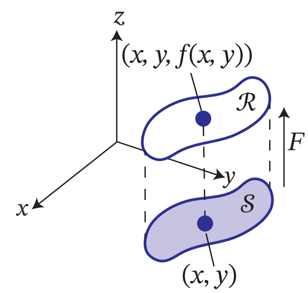


Figure 16.9.6

Therefore the magnification is

$$\left|-\frac{\partial f}{\partial x}\mathbf{i} - \frac{\partial f}{\partial y}\mathbf{j} + \mathbf{k}\right| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial x}\right)^2 + 1}.$$

♦

Summary

This section defined the magnification of a mapping between surfaces and discussed how to compute it. The most general formula for the magnification of a mapping F is the magnitude of $\partial\mathbf{r}/\partial u \times \partial\mathbf{r}/\partial v$, where $\mathbf{r}$ is the position vector of the mapping F, $\mathbf{r} = \overrightarrow{OF}$, and O is the origin.

If the image of F lies in the xy-plane, it can be written in vector form as $F(u,v) = f(u,v)\mathbf{i} + g(u,v)\mathbf{j} + 0\mathbf{k}$. The magnification, M_F, in this case is

$$\left|\det\begin{pmatrix} \frac{\partial f}{\partial u} & \frac{\partial g}{\partial u} \\ \frac{\partial f}{\partial v} & \frac{\partial g}{\partial v} \end{pmatrix}\right|. \tag{16.9.3}$$

The 2×2 matrix in (16.9.3) is commonly referred to as the **Jacobian matrix** for the mapping $(x,y) = F(u,v)$.

More generally, if the image of F is in xyz-space, the magnification is the determinant of a 3×3 Jacobian matrix:

$$M_F = \left|\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial f}{\partial u} & \frac{\partial g}{\partial u} & \frac{\partial h}{\partial u} \\ \frac{\partial f}{\partial v} & \frac{\partial g}{\partial v} & \frac{\partial h}{\partial v} \end{pmatrix}\right|. \tag{16.9.4}$$

An important concept to remember is that (16.9.3) and (16.9.4) rest on the fact that the area of the parallelogram spanned by $\mathbf{A}$ and $\mathbf{B}$ is $|\mathbf{A}\times\mathbf{B}|$.

EXERCISES for Section 16.9

NOTE: Unless specified otherwise, assume that all functions have continuous derivatives.

1. Assume $F(u,v) = (2u+3v, u-v) = (x,y)$. Find
   (a) $M(u,v)$, (b) $\mathrm{inv}F$ by solving for u and v in terms of x and y, and (c) $M_{\mathrm{inv}F}$.
2. Assume $F(x,y) = (x, y, xy)$ for $x,\ y > 0$.
   (a) Describe F geometrically, with a diagram. (b) Explain why F is one-to-one. (c) Find $M(x,y)$.
3. Assume $F(u,v) = (e^u\cos(v), e^u\sin(v))$.
   (a) Draw $F(\mathscr{R})$ for the rectangle $\mathscr{R}$ with $0 \le u \le 1$ and $0 \le \theta \le \frac{\pi}{6}$. (b) Find $M(u,v)$.
4. For a mapping F, $F(P_0) = Q_0$ and $M(P_0) = 3$. What is the magnification of $\mathrm{inv}F$ at P_0?
5. Use (16.9.3) to find the magnification of the mapping $F(u,v) = (2u, 3v)$. As noticed in Example 1, it should be 6.

6. In Example 1 the mapping $F(u,v) = (2u,3v)$ magnifies all areas by 6.

(a) Assuming the area of a circle with radius 1 is π, show that the area within the ellipse $\frac{x^2}{2^2} + \frac{y^2}{3^2} = 1$ is 6π.

(b) Show, by the same reasoning that the area within the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ is πab.

Exercises 7 to 10 concern mappings between intervals rather than between surfaces. Assume that the function f is strictly increasing on $[a,b]$ and that the range of f is $[c,d]$. (Assume $b > a$ and $d > c$.)

7. (a) Define the magnification of f at x_0. (b) Show that the magnification of f at x_0 is $f'(x_0)$.

8. What is the magnification, $M(x)$, of $f(x) = x^3$ at $x = \frac{1}{2}$?

9. (a) Compute $\int_a^b M(x)\,dx$. (b) Show that the average value of $M(x)$ in $[a,b]$ is $\frac{d-c}{b-a}$.

10. Show that there is a number x_0 in $[a,b]$ where $M(x_0)$ equals $\frac{d-c}{b-a}$, which can be viewed as the "global" magnification of f.

---

11. The inverse of the mapping F in Example 2 assigns to (x,y) $(x>0,\ y>0)$ the point (r,θ) $(r>0,\ 0<\theta<\pi/2)$ such that $F(r,\theta) = (x,y)$. Denote by $\mathrm{inv}F$ the inverse of F.

(a) Show that $(\mathrm{inv}F)(x,y) = \left(\sqrt{x^2+y^2}, \arctan\left(\frac{y}{x}\right)\right)$.

(b) Show that $M_{\mathrm{inv}F}(x,y)$ equals $\frac{1}{\sqrt{x^2+y^2}}$.

(c) Why is the result in (b) to be expected?

12. (a) Sketch the parabola $\mathscr{R}$ whose equation is $v = \frac{1}{4}u^2$ and the parabola $\mathscr{S}$ whose equation is $y = 4x^2$.

(b) Do the two parabolas seem to have the same shape, with one being a larger version of the other?

(c) Find a mapping F from uv-space to xy-space such that $F(\mathscr{R}) = \mathscr{S}$ with $F(u,v) = (ku,kv)$ for some constant k.

(d) What does (c) imply about (b)?

13. If F is a mapping from uv-space to xy-space and G is a mapping from xy-space to st-space, then the composition $H = G\circ F$ is a mapping from uv-space to st-space.

(a) How do you think the magnification of H is related to the magnifications of F and of G?

(b) Use the definition of magnification as a limit to answer (a).

(c) Use (16.9.2) to answer (a).

14. Deduce, as a special case of Exercise 13, that if $F(P) = Q$, then $M_{\mathrm{inv}F}(Q)$ is the reciprocal of $M_F(P)$.

15. Assume F is a mapping from surface $\mathscr{R}$ to surface $\mathscr{S}$ and that you know $M_F(P)$ at each point P of $\mathscr{R}$. How could you use this information to estimate the surface area of $\mathscr{S}$ as accurately as you please?

16. Example 5 determined the magnification of the mapping F defined by $F(x,y) = (x,y,f(x,y))$. The cross product was used. This exercise outlines another way to determine the magnification.

(a) At a point P on the surface $z = f(x,y)$ choose the normal to the surface that makes an acute angle γ with the unit vector $\mathbf{k}$ on the z-axis. Define the mapping G from the surface $z = f(x,y)$ to the xy-plane by $G(x,y,f(x,y)) = (x,y)$. Show that M_G equals $\cos(\gamma)$.

(b) Using (a), show that the magnification at (x,y) equals $\frac{1}{\cos(\gamma)}$, where γ is the normal at $G(x,y)$ such that the angle between γ and $\mathbf{k}$ is acute.

17. Assume F is a mapping from uvw-space to xyz-space: $F(u,v,w) = \big(f(u,v,w), g(u,v,w), h(u,v,w)\big)$.

(a) Define the magnification of F at the point (u_0, v_0, w_0) in terms of a limit.

(b) Show that this limit is the absolute value of the determinant $\det\begin{pmatrix} \frac{\partial f}{\partial u} & \frac{\partial g}{\partial u} & \frac{\partial h}{\partial u} \\ \frac{\partial f}{\partial v} & \frac{\partial g}{\partial v} & \frac{\partial h}{\partial v} \\ \frac{\partial f}{\partial w} & \frac{\partial g}{\partial w} & \frac{\partial h}{\partial w} \end{pmatrix}$.

18. (a) Find the magnification of the mapping F from uv-space to xyz-space defined by

$$F(u,v) = (3\sin(u)\cos(v), 3\sin(u)\sin(v), 3\cos(u)).$$

(b) Show that $F(u,v)$ lies on the sphere of radius 3 and center $(0,0,0)$.

19. A mapping F from the uv-plane to a parameterized surface in xyz-space has the form

$$F(u,v) = \big(f(u,v), g(u,v), h(u,v)\big).$$

Find the magnification of this mapping. Express this answer both in terms of (derivatives of) F and in terms of (derivatives of) f, g, and h.

20. The mapping involving polar coordinates examined in Examples 2, 3, and 4 can also be expressed in the language of vectors as $F(r,\theta) = r\cos(\theta)\mathbf{i} + r\sin(\theta)\mathbf{j}$.

(a) Compute $\dfrac{\partial F}{\partial r}$ and $\dfrac{\partial F}{\partial \theta}$.

(b) Using the algebra of vectors, show that the vectors found in (a) are perpendicular.

(c) Instead of the method in (b), show with the aid of a diagram that the vectors in (a) are perpendicular.

(d) Use (b) to estimate the area of the image of a small rectangle in the $r\theta$-plane of dimensions Δr and $\Delta\theta$.

Exercises 21 and 22 involve complex numbers, introduced in Section 12.5.

21. Inspired by Exercise 7(b), and naïve to the nuances of differentiation with respect to a complex-valued variable, Sam has an idea for finding the magnification of mappings given in terms of complex variables. Recall that if $y = f(x)$ then $\Delta y/\Delta x$ estimates the derivative $f'(x)$ when Δx is small. Hence $f'(x)$ also estimates $\Delta y/\Delta x$ when Δx is small. Thus $\Delta y = \frac{\Delta y}{\Delta x}\Delta x \approx f'(x)\Delta x$, and so $f'(x)$ measures the local magnification of intervals at x.

SAM: $F(z) = z^2$ is a mapping for z in the top half-plane. So its magnification at z is $|F'(z)| = |2z| = 2|z|$.

JANE: For z and $F(z)$ real I agree. I don't see why it should be $|F'(z)|$ for complex z and $F(z)$.

SAM: Why not? It's obvious to me.

(a) Determine the magnification of areas of the mapping $F(z) = z^2$. (This is the mapping $F(x,y) = \big(x^2 - y^2, 2xy\big)$ because $z^2 = (x+iy)^2 = \big(x^2 - y^2\big) + 2xyi$.

(b) Determine the magnification of areas under the mapping $F(z) = e^z$.

(c) Do you think Sam is right? Is Jane's skepticism justified?

(d) Conclude that Sam is "almost right" by conjecturing the simple formula involving $F'(z)$.

22. (a) Assume that $F(z)$ is a mapping and $F'(z_0)$ is not 0. Show that $\Delta F = F(z_0 + \Delta z) - F(z_0)$ is approximately $F'(z_0)\Delta z$.

(b) Show that if $\mathscr{S}$ is a small disk with center z_0 and radius r, the $F(\mathscr{S})$ is approximately a small disk with center $F(z_0)$ and radius $\left|F'(z_0)\right| r$.

(c) Using (b), express $M_F(z_0)$ in terms of $F'(z_0)$.

16.S Chapter Summary

This chapter extended the notion of the derivative to functions of two or more variables. For a function of several variables a partial derivative is the derivative with respect to one of the variables when the other variables are held constant.

The definition of any derivative rests on a limit. The partial derivative with respect to x of $f(x,y)$ at (a,b) is

$$\frac{\partial f}{\partial x}(a,b) = f_x(a,b) = \lim_{\Delta x\to 0}\frac{f(a+\Delta x,b)-f(a,b)}{\Delta x}.$$

As there are higher-order derivatives of a function of one variable, there are higher-order partial derivatives of functions with two or more variables:

$$\frac{\partial^2 f}{\partial x\partial x} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = f_{xx},\quad \frac{\partial^2 f}{\partial x\partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = f_{yx},\quad \frac{\partial^2 f}{\partial y\partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = f_{xy},\quad \text{and}\quad \frac{\partial^2 f}{\partial y\partial y} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = f_{yy}.$$

For functions usually encountered in applications the two mixed partials, f_{yx} and f_{xy}, are equal; the order of differentiation does not matter. (This is never a problem when the second-order partial derivatives are continuous functions.)

Also, for common functions we can differentiate under the integral sign:

$$\text{if } g(y) = \int_a^b f(x,y)\,dx, \text{ then } \frac{dg}{dy} = \int_a^b \frac{\partial f}{\partial y}(x,y)\,dx.$$

For a function of one variable, $f(x)$, with a continuous derivative, the differential change in f due to a change Δx in x is:

$$\Delta f = f(a+\Delta x)-f(a) = f'(c)\Delta x = (f'(a)+\epsilon)\Delta x = f'(a)\Delta x+\epsilon\Delta x. \tag{16.S.1}$$

Here c is in $[a,a+\Delta x]$ and $\epsilon\to 0$ as $\Delta x\to 0$. The analog of (16.S.1) for a function of two or more variables is the basis for the chain rule for functions of several variables:

$$\begin{aligned}\Delta f &= f(a+\Delta x,b+\Delta y)-f(a,b)\\ &= f(a+\Delta x,b+\Delta y)-f(a,b+\Delta y)+f(a,b+\Delta y)-f(a,b)\\ &= \frac{\partial f}{\partial x}(a,b)\,\Delta x+\frac{\partial f}{\partial y}(a,b)\,\Delta y+\epsilon_1\Delta x+\epsilon_2\Delta y,\end{aligned}$$

where ϵ_1 and $\epsilon_2\to 0$ as $\Delta x\to 0$ and $\Delta y\to 0$.

The gradient is a vector function. The gradient of a function $f(x,y)$ is $\nabla f = \langle f_x, f_y\rangle$ or, for a function of three variables, $\nabla f = \langle f_x, f_y, f_z\rangle$. The all cases, the gradient points in the direction a function increases most rapidly. The rate at which $f(x,y)$ changes in the direction of a unit vector $\mathbf{u}$ is $\nabla f\cdot\mathbf{u}$. The gradient evaluated at a point on the level curve $f(x,y)=k$ or level surface $f(x,y,z)=k$ is perpendicular to the level curve or level surface that passes through that point. At a critical point the gradient is $\mathbf{0}$, the zero vector.

For a function of one variable the sign of the second derivative helps tell whether a critical point is a maximum or a minimum. For a function of two variables, the test involves three second-order partial derivatives. The signs of f_{xx} and $D = f_{xx}f_{yy}-(f_{xy})^2$ are important.

The Lagrange multiplier method (Section 16.7)of finding an extremum of f subject to constraints $g_1=0$, $g_2=0$, $\dots$, $g_n=0$ depends on the observation that at an extremum ∇f can be written as $\lambda_1\nabla g_1+\lambda_2\nabla g_2+\cdots+\lambda_n\nabla g_n$.

Section 16.8 applied the chain rule and the equality of mixed partial derivatives to develop some fundamental equations in thermodynamics.

And, in the final section, partial derivatives were used to calculate the magnification of a mapping.

EXERCISES for Section 16.S

1. Let $f(x,y) = x^2 - y^2$ and $g(x,y) = 2xy$. Show that
(a) $\frac{\partial f}{\partial x} = \frac{\partial g}{\partial y}$ (b) $\frac{\partial f}{\partial y} = -\frac{\partial g}{\partial x}$ (c) $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$ (d) $\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} = 0$

2. Repeat Exercise 1 for $f(x,y) = \ln\left(\sqrt{x^2+y^2}\right)$ and $g(x,y) = \arctan\left(y/x\right)$.

3. In estimating the volume of a right circular cylinder tree trunk, there may be a 5 percent error in estimating the diameter and a 3 percent error in measuring the height. How large an error may occur in the estimate?

4. Let T denote the time it takes for a pendulum to complete a back-and-forth swing. If its length is L and g is the acceleration due to gravity, then $T = 2\pi\sqrt{\frac{L}{g}}$. A 3 percent error may be made in measuring L and a 2 percent error in measuring g. How large an error may we make in estimating T?

5. Let $u = f(x,y,z)$ and $\mathbf{r} = \mathbf{G}(t)$. Then u is a composite function of t. Show that $\frac{du}{dt} = \nabla f \cdot \mathbf{G}'(t)$, where ∇f is evaluated at $\mathbf{G}(t)$. For instance, let $u = f(x,y,z)$ and let $\mathbf{G}$ describe the path of a particle. Then the rate of change in the temperature on the path is the dot product of the temperature gradient ∇f and the velocity vector $\mathbf{v} = \mathbf{G}'$.

In Exercises 6 to 12 assume the functions are defined throughout the xy-plane and have continuous partial derivatives.

6. The function $3x + g(y)$, for any differentiable function $g(y)$, satisfies the partial differential equation $\partial f/\partial x = 3$. Are there any other solutions to that equation? Explain your answer.

7. Show that there is no function $f(x,y)$, with continuous second-order partial derivatives, such that $\frac{\partial f}{\partial x} = 3y$ and $\frac{\partial f}{\partial y} = 4x$.

In Exercises 8 to 12 find all functions such that

8. $\frac{\partial f}{\partial x} = 3$ and also $\frac{\partial f}{\partial y} = 4$.

9. $f_{xx}(x,y) = 0$.

10. $f_{xx}(x,y) = 0$ and $f_{yy}(x,y) = 0$.

11. $f_{xy}(x,y) = 0$.

12. $\frac{\partial^2 f}{\partial x \partial y}(x,y) = 1$.

---

13. Show that for a polynomial $P(x,y)$, P_{yx} equals P_{xy}.

NOTE: It is enough to consider monomials $ax^m y^n$, where a is constant and m and n are nonnegative integers.

14. A hiker is at the origin on a hill whose surface has the equation $z = x$. If he walks south, above the positive x-axis the slope of his path would be 1. If he walked along the y-axis, the slope would be 0.
(a) If he walked northeast what would the slope of his path be?
(b) In what direction should he walk so his path would have a slope of 0.2?

15. Let f and g be functions of x and y that have continuous second derivatives. Assume the first partial derivatives of f and g satisfy

$$\frac{\partial f}{\partial x} = \frac{\partial g}{\partial y} \quad \text{and} \quad \frac{\partial f}{\partial y} = -\frac{\partial g}{\partial x}. \tag{16.S.2}$$

Show that

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0 \qquad \text{and} \qquad \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} = 0. \tag{16.S.3}$$

16. Let $V(x, y, z) = xyz$ be the volume of a box of sides x, y, and z. Compute ΔV and dV and show them in Figure 16.S.1.

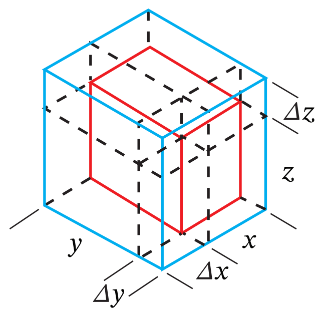


Figure 16.S.1

Exercises 17 to 20 concern the definition of $\lim_{(x,y)\to P_0} f(x, y)$.

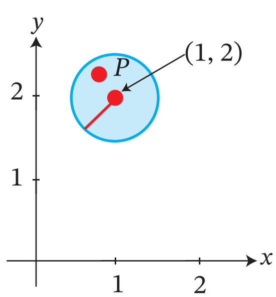


Figure 16.S.2

17. Let $f(x, y) = x + y$.

(a) Show that if $P = (x, y)$ lies within a distance 0.01 of $(1, 2)$, then $|x - 1| < 0.01$ and $|y - 2| < 0.01$. (See Figure 16.S.2).

(b) Show that if $|x - 1| < 0.01$ and $|y - 2| < 0.01$, then $|f(x, y) - 3| < 0.02$.

(c) Find a number $\delta > 0$ such that if $P = (x, y)$ is in the disk with center $(1, 2)$ and radius δ, then $|f(x, y) - 3| < 0.001$.

(d) Show that for any positive number ϵ, no matter how small, there is a positive number δ such that when $P = (x, y)$ is in the disk with radius δ and center $(1, 2)$, then $|f(x, y) - 3| < \epsilon$. Give δ as a function of ϵ.

(e) What does (d) imply about the function $f(x, y) = x + y$?

18. Let $f(x, y) = 2x + 3y$.

(a) Find a disk with center $(1, 1)$ such that whenever P is in it, $|f(P) - 5| < 0.01$

(b) Let ϵ be a positive number. Show that there is a disk with center $(1, 1)$ such that whenever P is in it, $|f(P) - 5| < \epsilon$. Give the radius δ as a function of ϵ.

(c) What may we conclude from (b)?

19. Let $f(x, y) = \dfrac{x^2 y}{x^4 + 2y^2}$.

(a) What is the domain of f?

(b) Fill in the three missing values in the table:

(x, y)	$(0.01, 0.01)$	$(0.01, 0.02)$	$(0.001, 0.003)$
$f(x, y)$			

(c) From (b), do you think $\lim_{P\to(0,0)} f(P)$ exists? If so, what is it?

(d) Fill in the three missing values in the table:

(x, y)	$(0.5, 0.25)$	$(0.1, 0.01)$	$(0.001, 0.000001)$
$f(x, y)$			

(e) From (d), do you think $\lim_{P\to(0,0)} f(P)$ exists? If so, what is it?

(f) Does $\lim_{P\to(0,0)} f(P)$ exist? If so, what is it? Explain.

20. Let $f(x,y) = \dfrac{5x^2y}{2x^4+3y^2}$.

(a) What is the domain of f?
(b) As P approaches $(0,0)$ on the line $y = 2x$, what happens to $f(P)$?
(c) As P approaches $(0,0)$ on the line $y = 3x$, what happens to $f(P)$?
(d) As P approaches $(0,0)$ on the parabola $y = x^2$, what happens to $f(P)$?
(e) Does $\lim_{P\to(0,0)} f(P)$ exist? If so, what is it? Explain.

---

21. This exercise outlines a proof of the fact that the mixed partials of $f(x,y)$ are generally equal. It suffices to show that $f_{xy}(0,0) = f_{yx}(0,0)$. We assume that the first and second partial derivatives are continuous in some disk with center $(0,0)$.

(a) Why is $f_{xy}(0,0)$ equal to $\lim_{k\to 0} \dfrac{f_x(0,k)-f_x(0,0)}{k}$?
(b) Why is the limit in (a) equal to $\lim_{k\to 0}\left(\lim_{h\to 0} \dfrac{(f(h,k)-f(0,k))-(f(h,0)-f(0,0))}{hk}\right)$?
(c) Let $u(y) = f(h,y) - f(0,y)$. Show that the fraction in (b) equals $\dfrac{u(k)-u(0)}{hk}$, and it equals $\dfrac{u'(K)}{h}$ for some K between 0 and k.
(d) Why is $u'(K) = f_y(h,K) - f_y(0,K)$?
(e) Why is $\dfrac{u'(K)}{h}$ equal to $(f_y)_x(H,K)$ for some H between 0 and h?
(f) Deduce that $f_{xy}(0,0) = f_{yx}(0,0)$.
(g) Did this derivation use the continuity of f_{yx}? If so, how?
(h) Did this derivation use the continuity of f_{xy}? If so, how?
(i) Did we need to assume f_{xy} exists? If so, where was the assumption used?
(j) Did we need to assume f_{yx} exists? If so, where was the assumption used?

22. The assertion that we can differentiate across the integral sign says that

$$\frac{d}{dt}\int_a^b f(x,t)\,dx = \int_a^b \frac{\partial}{\partial t} f(x,t)\,dx. \tag{16.S.4}$$

(a) Why is the derivative on the left an ordinary derivative but the derivative on the right is a partial derivative?
(b) Rewrite (16.S.4) by replacing both derivatives with appropriate limits.
(c) Verify (16.S.4) for $f(x,t) = x^7t^4$.
(d) Verify (16.S.4) for $f(x,t) = \cos(xt)$.

23. Assume that $f(x,y,z)$ has an extreme value at $\mathbf{p}_0$ on the level surface $g(x,y,z) = k$. Why are ∇g and ∇f evaluated at $\mathbf{p}_0$ both perpendicular to the surface at $\mathbf{p}_0$?

---

24. (*Computer graphics*) Justify these two potential recipes for finding the minimum distance from a point Q to the surface $z = F(u,v)$:

(a) For a given point Q, let $\mathbf{q} = \overrightarrow{OQ}$. Find all points P on the surface $z = F(u,v)$ such that $\mathbf{p} = \overrightarrow{OP}$ satisfies the equation $(\mathbf{p}-\mathbf{q}) \times \left(\dfrac{\partial \mathbf{p}}{\partial u} \times \dfrac{\partial \mathbf{p}}{\partial v}\right) = \mathbf{0}$.
(b) For a given point Q, let $\mathbf{q} = \overrightarrow{OQ}$. Find all points P on the surface $z = F(u,v)$ such that $\mathbf{p} = \overrightarrow{OP}$ satisfies the equation $\mathbf{p}-\mathbf{q} = c\left(\dfrac{\partial \mathbf{p}}{\partial u} \times \dfrac{\partial \mathbf{p}}{\partial v}\right)$ for some scalar c.

25. (*Computer graphics*) The **silhouette edge** of a surface projected onto a display plane is the collection of points whose outward normal vector is in the display plane.

Justify the following approach to finding the border of the silhouette edge of the surface $z = \mathbf{F}(u, v)$ on the xy-plane: Find all points P on the surface $z = F(u, v)$ such that $\mathbf{p} = \overrightarrow{OP}$ satisfies $\mathbf{k} \cdot \left(\frac{\partial \mathbf{p}}{\partial u} \times \frac{\partial \mathbf{p}}{\partial v}\right) = 0$.

Exercises 26 to 28 involve functions of more than three variables. The gradient, basic geometry of normal vectors, and Lagrange multipliers all extend naturally to higher dimensions. (Note that the cross product is defined only in three dimensions, and the second-partial-derivative test to classify a critical point is more easily stated in terms of properties of the matrix of second-order partial derivatives.)

26. (*Computer science*) This problem arises in the design of efficient bucket sorts. (A **bucket sort** is a way of rearranging (sorting) information in a database.) Let $p_1, p_2, \ldots, p_k$ and B be positive constants. Let $b_1, b_2, \ldots, b_k$ be k nonnegative variables satisfying $\sum_{j=1}^{k} b_j = B$. The quantity $\sum_{j=1}^{k} p_j \cdot 2^{B-b_j}$ represents the expected search time. What values of $b_1, b_2, \ldots, b_k$ does the method of Lagrange multipliers give for the minimum expected search time?

REFERENCE: J. D. Ullman, *Principles of Database Systems*, pp. 82–83, Computer Sci Press, Potomac, Md., 1980.

27. A consumer has a budget of B dollars and may purchase n different items. The price of the i^{th} item is p_i dollars. When the consumer buys x_i units of the i^{th} item, the total cost is $\sum_{i=1}^{n} p_i x_i$. Assume that $\sum_{i=1}^{n} p_i x_i = B$ and that the consumer wishes to maximize her utility $u(x_1, x_2, \ldots, x_n)$. (a) Show that when $x_1, \ldots, x_n$, are chosen to maximize utility, then $\frac{1}{p_i}\frac{\partial u}{\partial x_i} = \frac{1}{p_j}\frac{\partial u}{\partial x_j}$, for all i, $j = 1, \ldots, n$. (b) Explain the result in (a) using economic intuition.

28. The following is quoted from a bioeconomics text

> [S]uppose there are N fishing grounds. Let $H^i = H^i(R^i, E^i)$ denote the production function for the total harvest H^i on the i^{th} ground as a function of the recruited stock level R^i and effort E^i on the i^{th} ground. The problem is to determine the least total cost $\sum_{i=1}^{N} c_i E^i$ at which a given total harvest $H = \sum_{i=1}^{n} H^i$ can be achieved. This problem can be easily solved by Lagrange multipliers. The result is simply $\frac{1}{c_i}\frac{\partial H^i}{\partial E^i}$ = constant [independent of i].
>
> REFERENCE: Colin W. Clark, *Mathematical Bioeconomics*, Wiley, New York, 1976.

Verify the assertion.

Each c_i is a constant. The superscripts are used to name the functions, not exponents.

Calculus is Everywhere # 23

The Wave in a Rope

We will develop what may be the most famous partial differential equation: the **wave equation**. We will solve the wave equation in CIE 24 at the end of Chapter 17 and then, in CIE 26 at the end of Chapter 18, see that the wave equation arises in the study of electric and magnetic fields.

As Morris Kline writes in *Mathematical Thought from Ancient to Modern Times*, "The first real success with partial differential equations came in renewed attacks on the vibrating string problem, typified by the violin string. The approximation that the vibrations are small was imposed by d'Alembert (1717-1783) in his papers of 1746."

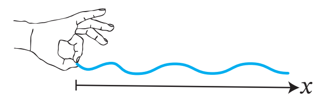


Figure C.23.1

Imagine gently shaking the end of a rope, as in Figure C.23.1. Even though each (short) section of the string moves up and down, the wave moves to the right. (It may, eventually, come back to the left, but that depends on how the ends of the rope are constrained, and is not our concern here.) The important point is that a vertical displacement starts a wave moving horizontally throughout the length of the rope.

To develop the mathematics of the wave, we make some simplifying assumptions. We suppose that each segment moves only up and down, the distance it moves is very small, and the slope of the curve assumed by the rope remains close to zero. (Note that this means the actual length of each segment changes, but only slightly.)

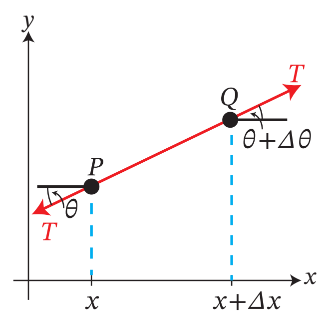


Figure C.23.2

At time t the vertical position of the segment whose x-coordinate is x is $y = y(x,t)$ for it depends on both location x and time t. Assume that the tension T throughout the rope is constant. For a short section of the rope at time t, shown as PQ in Figure C.23.2, we apply Newton's Second Law, which implies force equals mass times acceleration.

Denote the length of the segment as Δx. While, as noted above, the actual length of each segment changes as its endpoints move vertically, the assumption that displacements are small makes this assumption valid (at least here). If the linear density of the rope is λ, the mass of the segment is λ times its length: $\lambda\Delta x$ The upward force exerted by the rope on the segment is $T\sin(\theta+\Delta\theta)$ and the downward force is $T\sin(\theta)$. The net vertical force is $T\sin(\theta+\Delta\theta) - T\sin(\theta)$. Thus

$$\underbrace{T\sin(\theta+\Delta\theta) - T\sin(\theta)}_{\text{net vertical force}} = \underbrace{\lambda\Delta x}_{\text{mass}} \quad \underbrace{\frac{\partial^2 y}{\partial t^2}}_{\text{acceleration}} \quad . \tag{C.23.1}$$

Note that because y is a function of x and t, the acceleration (second derivative of position with respect to time) is a partial derivative, not an ordinary derivative.

It remains to express $\sin(\theta)$ and $\sin(\theta+\Delta\theta)$ in terms of the partial derivative $\partial y/\partial x$. Because θ is near 0, $\cos(\theta)$ is near 1. Thus $\sin(\theta)$ is approximately $\sin(\theta)/\cos(\theta) = \tan(\theta)$, the slope of the rope at time t above (or below) x, which is $\partial y/\partial x$ at x and t. Similarly, $\sin(\theta+\Delta\theta)$ is approximately $\partial y/\partial x$ at $x+\Delta x$ and t. So (C.23.1) is approximated by

$$T\frac{\partial y}{\partial x}(x+\Delta x, t) - T\frac{\partial y}{\partial x}(x,t) = \lambda\Delta x\frac{\partial^2 y}{\partial t^2}(x,t).$$

Dividing by Δx gives

$$\frac{T\left(\frac{\partial y}{\partial x}(x+\Delta x,t) - \frac{\partial y}{\partial x}(x,t)\right)}{\Delta x} = \lambda\frac{\partial^2 y}{\partial t^2}(x,t).$$

Letting Δx approach 0, we obtain

$$T\frac{\partial^2 y}{\partial x^2}(x,t) = \lambda\frac{\partial^2 y}{\partial t^2}(x,t).$$

Since both T and λ are positive, $Ts/\lambda = c^2$ for some constant c, and we can write (23) in the traditional form

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 y}{\partial t^2}.$$

This is the famous **wave equation**.

Observation C.23.1: ***A Geometric Interpretation of the Wave Equation***

The curvature of the rope at any point can be expressed as

$$\frac{\frac{\partial^2 y}{\partial x^2}}{\left(\sqrt{1+\left(\frac{\partial y}{\partial x}\right)^2}\right)^3}.$$

Since we assume that the slope of the rope remains near 0, the curvatuve is reasonably approximated by $\partial^2 y/\partial x^2$. And this, by the wave equation, is a constant multiple of the acceleration, $\partial^2 y/\partial t^2$.

At the curvier part of the rope, the acceleration is greater. The wave moves along the rope, but the molecules of the rope move up and down perpendicular to the direction of the waves.

As will be shown in CIE 24, "Solving the Wave Equation", at the end of the next chapter, the CIE in the next chapter shows, the constant c turns out to be the velocity of the wave.

EXERCISES for CIE C.23

1. In the discussion in the text what are the meanings of (a) $\frac{\partial y}{\partial x}$, (b) $\frac{\partial y}{\partial t}$, (c) $\frac{\partial^2 y}{\partial t^2}$, and (d) $\frac{\partial^2 y}{\partial x^2}$.

2. The argument used in this section depended on the approximation of $\sin(\theta)$ by $\tan(\theta)$ for small θ. Check this estimate by computing $\frac{\sin(x)}{\tan(x)}$ for (a) $x = 0.3$ radians (17 degrees), (b) $x = 0.1$ radians, and (c) $x = 0.01$ radians.

Chapter 17

Plane and Solid Integrals

In Chapter 3 we introduced the derivative, one of the two main concepts in calculus. In Chapter 16 we extended the derivative to higher dimensions. In the present chapter, we generalize the concept of the definite integral, introduced in Chapter 6, to higher dimensions.

Instead of using the notation of Chapter 6 to define the definite integral, we will restate the definition so that it easily generalizes to higher dimensions.

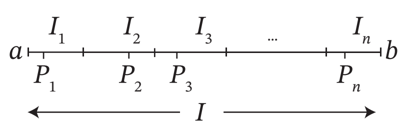


Figure 17.0.1

We started with an interval $[a, b]$, which we will call I, and a continuous function f defined at each point P of I. Then we cut I into n short intervals $I_1, I_2, \ldots, I_n$, and chose a point P_1 in I_1, P_2 in $I_2, \ldots, P_n$ in I_n. See Figure 17.0.1. Denoting the length of I_i by L_i, we formed

$$\sum_{i=1}^{n} f(P_i) L_i.$$

The limit of the sums as all the subintervals are chosen shorter and shorter is the definite integral of f over interval I. We denoted it $\int_a^b f(x)\ dx$. We now denote it $\int_I f(P)\ dL$. This notation tells us that we are integrating a function, f, over an interval I. The dL reminds us that the integral is the limit of approximations formed as the sum of products of a function value and the length of a short interval.

We will define integrals of functions over plane regions (Section 17.1), such as squares and disks, over solid regions (Section 17.4), such as cubes and balls, and over surfaces (Section 17.7) such as the surface of a ball, in the same way. Some applications involving multiple integrals are discussed in Section 17.8.

It is one thing to define these higher-dimensional integrals. It is another to calculate them. Most of our attention will be devoted to seeing how to compute them with the aid of integrals over intervals, the type defined in Chapter 6. This generally means working in an appropriate coordinate system.

In two dimensions the most common coordinate systems used to evaluate a double integral are rectangular coordinates (Section 17.2) and polar coordinates (Section 17.3).

In three dimensions, in addition to rectangular coordinates, we introduce cylindrical and spherical coordinates in Section 17.5, and then show how to use these coordinates to evaluate some triple integrals in Section 17.6. A more general treatment of multiple integration in other coordinate systems can be found in Section 16.9.

17.1 The Double Integral: Integrals Over Plane Areas

The goal of this section is to introduce the integral of a function defined in a region of a plane. We assume that the reader is familiar with the remarks in the introduction to this chapter.

Volume Approximated by Sums

Let $\mathscr{R}$ be a region in the xy-plane bounded by curves. For convenience, assume $\mathscr{R}$ is convex, for example an ellipse, a disk, a parallelogram, a rectangle, or a square. We draw $\mathscr{R}$ in perspective in Figure 17.1.1(a).

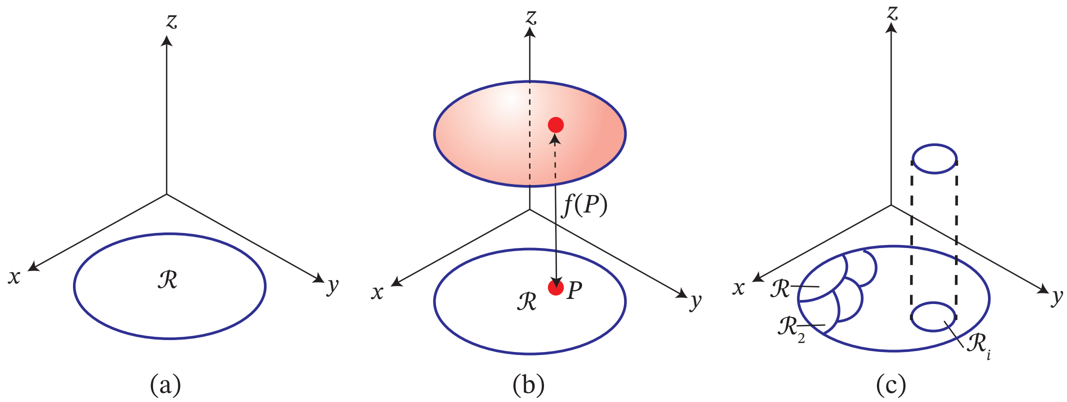


Figure 17.1.1

Suppose that there is a surface above $\mathscr{R}$ and that $f(P)$ is the height of the surface above any point P on $\mathscr{R}$, as shown in Figure 17.1.1(b)

If we know $f(P)$ for every point P could we estimate the volume, V, of the solid under the surface and above $\mathscr{R}$? As we used rectangles to estimate the area of a region in Section 6.1, we will use cylinders to estimate the volume of a solid. Just as the area of a rectangle is the product of the length of its base and its height, the volume of a cylinder is the product of the area of its base and its height.

Cut $\mathscr{R}$ into n regions $\mathscr{R}_1, \mathscr{R}_2, \dots, \mathscr{R}_n$. Each $\mathscr{R}_i$ has area A_i. Choose points P_1 in $\mathscr{R}_1$, P_2 in $\mathscr{R}_2$, …, P_n in $\mathscr{R}_n$. Over each region $\mathscr{R}_i$ put a cylinder with height $f(P_i)$ and base $\mathscr{R}_i$. The total volume of the n cylinders,

$$\sum_{i=1}^{n} f(P_i)A_i, \tag{17.1.1}$$

is an approximation to the volume V. If f is a continuous function and $\mathscr{R}$ is cut into more regions (in such a way that area of the largest subregion becomes smaller), the sum (17.1.1) approaches the volume V.

EXAMPLE 1. Estimate the volume of the solid under the paraboloid $z = (x^2 + y^2)/4$ and above the rectangle $\mathscr{R}$ whose vertices are $(0,0)$, $(3,0)$, $(3,2)$, and $(0,2)$.

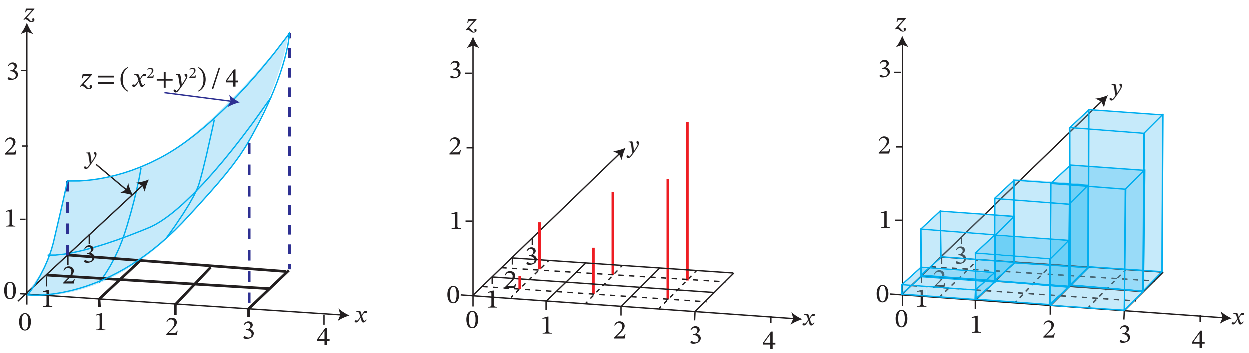
SOLUTION Figure 17.1.2(a) shows the graph of the paraboloid over the specified rectangle. The highest point is above $(3,2)$, where $z = 13/4$. So the solid fits in a box whose height is 3.25 and whose base has area 6. Therefore we know its volume is at most $3.25 \cdot 6 = 19.5$.

To estimate the volume we cut the rectangular base into six 1×1 squares and evaluate $z = (x^2 + y^2)/4$ at, say, the center of each square, as shown in Figure 17.1.2(b).

Then we form a cylinder for each square. The base is the square and the height is the value of $(x^2 + y^2)/4$ at the center of the square, as shown in Figure 17.1.2(c).

Then the total volume of these six rectangular cylinders is

$$\begin{aligned} &\underbrace{\frac{1}{8}}_{\text{height}} \cdot \underbrace{1}_{\text{area of base}} + \underbrace{\frac{5}{8}}_{\text{height}} \cdot \underbrace{1}_{\text{area of base}} + \underbrace{\frac{13}{8}}_{\text{height}} \cdot \underbrace{1}_{\text{area of base}} \\ &+ \underbrace{\frac{5}{8}}_{\text{height}} \cdot \underbrace{1}_{\text{area of base}} + \underbrace{\frac{9}{8}}_{\text{height}} \cdot \underbrace{1}_{\text{area of base}} + \underbrace{\frac{17}{8}}_{\text{height}} \cdot \underbrace{1}_{\text{area of base}} = \frac{25}{4}. \end{aligned} \tag{17.1.2}$$



(a) (b) (c)

Figure 17.1.2

This estimate is 25/4 cubic units. By cutting the base into more pieces (with smaller areas) and using more cylinders we could make a more accurate estimate. The exact volume will be found in Example 5 in Section 17.2. ♦

Density

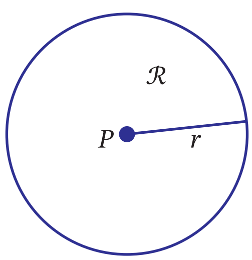


Figure 17.1.3

Before we consider a total mass problem we must define the concept of density. Imagine that we have a piece of sheet metal, which we view as part of a plane, that is homogeneous, the same everywhere. Let $\mathscr{R}$ be any region in it, of area A and mass m. The quotient m/A is the same for all regions $\mathscr{R}$, and is called the density. A material, unlike sheet metal, may not be uniform. As $\mathscr{R}$ varies, the quotient m/A, or average density in $\mathscr{R}$, also varies.

Physicists define **density at a point** as follows: Consider a small disk $\mathscr{R}$ of radius r and center at P, as in Figure 17.1.3. If $m(r)$ is the mass in the disk and $A(r)$ is its area (πr^2), then

$$\text{Density at } P = \lim_{r\to 0} \frac{m(r)}{A(r)}.$$

Density is denoted $\sigma(P)$, which is read as "sigma of P". Like the physicists, we assume the density, $\sigma(P)$, exists at each point, P, and is a continuous function. This implies that if $\mathscr{R}$ is a region of area A and P is a point in $\mathscr{R}$ then the product $\sigma(P)A$ is an approximation of the mass in $\mathscr{R}$.

The Greek letter σ corresponds to our letter "s", the initial letter of "surface"; $\sigma(P)$ is the density of a surface at point P.

Total Mass Approximated by Sums

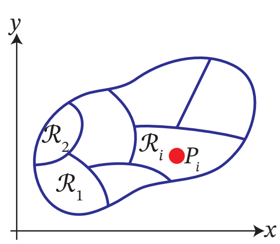


Figure 17.1.4

Assume that a flat region $\mathscr{R}$ is occupied by a material of varying density. The density at point P in $\mathscr{R}$ is $\sigma(P)$. Estimate M, the total mass in $\mathscr{R}$.

We cut $\mathscr{R}$ into n small regions $\mathscr{R}_1, \mathscr{R}_2, \ldots, \mathscr{R}_n$. Each $\mathscr{R}_i$ has area A_i. We choose points P_1 in $\mathscr{R}_1$, P_2 in $\mathscr{R}_2, \ldots, P_n$ in $\mathscr{R}_n$. Then we estimate the mass in each region $\mathscr{R}_i$, as shown in Figure 17.1.4. The mass in $\mathscr{R}_i$ is approximately

$$\underbrace{\sigma(P_i)}_{\text{density}} \cdot \underbrace{A_i}_{\text{area}}$$

and the corresponding estimate of the total mass of region $\mathscr{R}$ is

$$\sum_{i=1}^{n} \sigma(P_i) A_i. \tag{17.1.3}$$

If σ is a continuous function then, as we divide $\mathscr{R}$ into smaller and smaller regions, the sums (17.1.3) approach the total mass M.

A **lamina** is a thin plate, sheet, or layer.

EXAMPLE 2. A rectangular **lamina** of varying density occupies the rectangle with corners at $(0,0)$, $(3,0)$, $(3,2)$, and $(0,2)$ in the xy-plane with units of centimeters on each axis. Its density at (x,y) is $(x^2+y^2)/4$ grams per square centimeter.

Estimate its mass by cutting it into six 1×1 squares and evaluating the density at the center of each square.

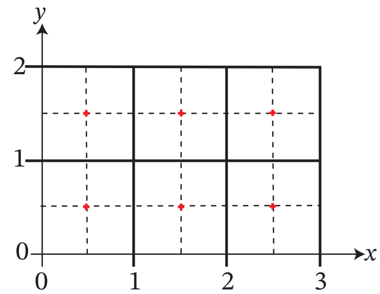


Figure 17.1.5

SOLUTION The six squares are shown in Figure 17.1.5. The density at the center of the square bounded by $x=1$, $x=2$, $y=0$, and $y=1$ is $((3/2)^2+(1/2)^2)/4 = 5/8$. Since its area is $1\times 1 = 1$, an estimate of its mass is

$$\underbrace{\frac{5}{8}}_{\text{density}} \cdot \underbrace{1}_{\text{area}} = \frac{5}{8} \text{ grams.}$$

Similar estimates for the remaining small squares give a total estimate of

$$\frac{1}{8}\cdot 1+\frac{5}{8}\cdot 1+\frac{13}{8}\cdot 1+\frac{5}{8}\cdot 1+\frac{9}{8}\cdot 1+\frac{17}{8}\cdot 1=\frac{25}{4} \text{ grams.}$$

This sum is identical to the sum (17.1.2), which estimates a volume. ◆

The arithmetic in Examples 1 and 2 shows that unrelated problems, one about volume, the other about mass, lead to the same estimates. Moreover, as the rectangle is cut into smaller pieces, the estimates would become closer and closer to the volume or the mass. These estimates, similar to the estimates $\sum_{i=1}^{n} f(P_i)L_i$ that appear in the definition of the definite integral $\int_a^b f(x)\,dx$, bring us to the definition of a double integral. It is called a double integral because the domain of the function is in the two-dimensional plane.

The Double Integral

The definition of the double integral is almost the same as that of $\int_a^b f(x)\,dx$, the integral over an interval. The differences are:

1. instead of dividing an interval into smaller intervals, we divide a plane region into smaller plane regions,
2. instead of a function defined on an interval, we have a function defined on a plane region, and
3. we need a quantitative way to say that a little region is small.

To give meaning to "small", we introduce the diameter of a plane region. The **diameter** of a region bounded by a curve is the maximum distance between two points in the region. For instance, the diameter of a square S of side s is $\sqrt{2}s$ and the diameter of a disk is the same as its diameter that we know from geometry.

Definition: *Double Integral*

Assume that $\mathscr{R}$ is a plane region bounded by curves and f is a continuous function defined on $\mathscr{R}$. Partition $\mathscr{R}$ into n smaller regions $\mathscr{R}_1, \mathscr{R}_2, \dots, \mathscr{R}_n$ with areas $A_1, A_2, \dots, A_n$ and sampling points $P_1, P_2, \dots, P_n$. The limit of sums of the form

$$\sum_{i=1}^{n} f(P_i)A_i, \tag{17.1.4}$$

as the diameter of each $\mathscr{R}_i$ approaches 0, exists (no matter how the sampling points P_i are chosen). The limiting value is called the **integral of f over $\mathscr{R}$** or the **double integral of f over $\mathscr{R}$** and is denoted

$$\int_{\mathscr{R}} f(P)\,dA.$$

It is proved in advanced courses that when f is continuous, the sums (17.1.4) approach a limit. That is, if f is continuous on $\mathscr{R}$ then the double integral $\int_{\mathscr{R}} f(P)\,dA$ exists.

Volume Expressed as a Double Integral

With this definition of double integral, Example 1 leads to the following definition of volume of a solid.

Definition: *The Double Integral of Cross Section is the Volume*

Let S be a solid and $\mathscr{R}$ its projection on a plane, as in Figure 17.1.6(a). Assume that for each point P in $\mathscr{R}$ the line through P perpendicular to $\mathscr{R}$ intersects S in a line segment of length $c(P)$. Then

$$\text{Volume of } S = \int_{\mathscr{R}} c(P)\, dA.$$

Mass Expressed as a Double Integral

Likewise, Example 2 leads to a definition of the mass of plane region.

Definition: *The Double Integral of Density is the Total Mass*

For a plane distribution of mass through a region $\mathscr{R}$ the density may vary. As shown in Figure 17.1.6(b), denote the density at P by $\sigma(P)$. Then

$$\text{Mass in } \mathscr{R} = \int_{\mathscr{R}} \sigma(P)\, dA.$$

Average Value as a Double Integral

The average value of $f(x)$ for x in the interval $[a,b]$ was defined in Section 6.3 as

$$\frac{1}{\text{Length of interval}} \int_a^b f(x)\, dx.$$

We make a similar definition for a function defined on a two-dimensional region.

Definition: *Average Value of a Function in a Plane Region*

The **average value** of f over the region $\mathscr{R}$ in the xy-plane is

$$\frac{1}{\text{Area of } \mathscr{R}} \int_{\mathscr{R}} f(P)\, dA.$$

If $f(P)$ is positive for all P in $\mathscr{R}$, there is a geometric interpretation of the average of f over $\mathscr{R}$. Let S be the solid below the graph of f (a surface) and above $\mathscr{R}$. The average value of f over $\mathscr{R}$ is the height of the cylinder whose base is $\mathscr{R}$ and whose volume is the same as the volume of S. (See Figure 17.1.6(c).)

The integral $\int_{\mathscr{R}} f(P)\, dA$ is called an integral over a plane region to distinguish it from $\int_a^b f(x)\, dx$, which is called an integral over an interval. Another notation is $\int\int_{\mathscr{R}} f(x,y)\, dA$. We prefer the notation $\int_{\mathscr{R}} f(P)\, dA$ because it does not favor any particular coordinate system and the repeated integral sign is reserved for a different concept.

In the case of the constant function $f(P) = 1$ we compute $\int_{\mathscr{R}} f(P)\, dA$. The approximating sum $\sum_{i=1}^n f(P_i) A_i$ equals $\sum_{i=1}^n 1 \cdot A_i = A_1 + A_2 + \cdots + A_n$, which is the area of the region $\mathscr{R}$ that is being partitioned. Since every approximating sum has this same value, it follows that

$$\lim_{n\to\infty} \sum_{i=1}^n f(P_i) A_i = \text{Area of } \mathscr{R}.$$

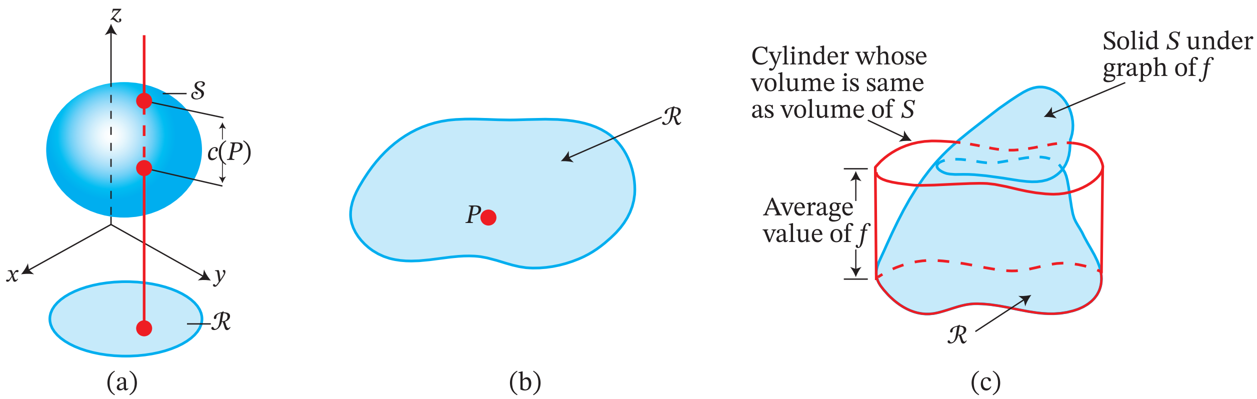


Figure 17.1.6

Definition: *The Double Integral of the Function* 1 *is the Area*

$$\int_{\mathcal{R}} 1\,dA = \text{Area of } \mathcal{R}.$$

This formula will come in handy on several occasions. The 1 is often omitted, in which case we write $\int_{\mathcal{R}} dA =$ Area of $\mathcal{R}$. A table summarizing some of the main applications of the double integral $\int_{\mathcal{R}} f(P)\,dA$ can be found in Section 17.S (see Table 17.S.1 on page 1005).

Properties of Double Integrals

Integrals over plane regions have properties similar to those of integrals over intervals:

1. $\int_{\mathcal{R}} cf(P)\,dA = c\int_{\mathcal{R}} f(P)\,dA$ for any constant c.
2. $\int_{\mathcal{R}} [f(P)+g(P)]\,dA = \int_{\mathcal{R}} f(P)\,dA + \int_{\mathcal{R}} g(P)\,dA$.
3. If $f(P) \le g(P)$ for all points P in $\mathcal{R}$, then $\int_{\mathcal{R}} f(P)\,dA \le \int_{\mathcal{R}} g(P)\,dA$.
4. If $\mathcal{R}$ is broken into two regions, $\mathcal{R}_1$ and $\mathcal{R}_2$, overlapping at most on their boundaries, then

$$\int_{\mathcal{R}} f(P)\,dA = \int_{\mathcal{R}_1} f(P)\,dA + \int_{\mathcal{R}_2} f(P)\,dA.$$

For instance, consider Property 3 when $f(P)$ and $g(P)$ are both positive. Then $\int_{\mathcal{R}} f(P)\,dA$ is the volume under the surface $z = f(P)$ and above $\mathcal{R}$ in the xy-plane. Similarly $\int_{\mathcal{R}} g(P)\,dA$ is the volume under $z = g(P)$ and above $\mathcal{R}$. Then Property 3 asserts that the volume of a solid is not larger than the volume of a solid that contains it. (See Figure 17.1.7.)

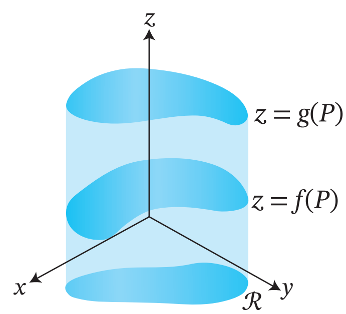


Figure 17.1.7

Summary

Just as $\int_a^b f(x)\,dx$, an integral over an interval, is defined as the limit of sums of the form $\sum_{i=1}^n f(P_i)L_i$, $\int_{\mathcal{R}} f(P)\,dA$, a double integral over a plane region, is defined as the limit of sums of the form $\sum_{i=1}^n f(P_i)A_i$. They arise in computing volumes, total mass, or average value. The sum $\sum_{i=1}^n f(P_i)A_i$ is an estimate of the double integral $\int_{\mathcal{R}} f(P)\,dA$.

EXERCISES for Section 17.1

1. In the estimates for the volume in Example 1, the centers of the squares were used as the P_i's. Make an estimate for the volume in Example 1 by using the same partition but taking as P_i (a) the lower left corner of $\mathscr{R}_i$ and (b) the upper right corner of $\mathscr{R}_i$. (c) What do (a) and (b) say about the volume of the solid?

2. Estimate the mass in Example 2 using the partition of $\mathscr{R}$ into six 1×1 squares and (a) using the upper left corner of $\mathscr{R}_i$ as the P_i's and (b) using the lower right corner of $\mathscr{R}_i$ as the P_i's.

3. Let $\mathscr{R}$ be a set in the plane whose area is A. Let f be the function such that $f(P)=5$ for every point P in $\mathscr{R}$.

(a) What can be said about any approximating sum $\sum_{i=1}^{n} f(P_i)A_i$ formed for this $\mathscr{R}$ and this f?

(b) What is the value of $\int_{\mathscr{R}} f(P)\,dA$?

4. Let $\mathscr{R}$ be the square with vertices $(1,1)$, $(5,1)$, $(5,5)$, and $(1,5)$. Let $f(P)$ be the distance from P to the y-axis.

(a) Estimate $\int_{\mathscr{R}} f(P)\,dA$ by partitioning $\mathscr{R}$ into four squares and using their centers as sampling points.

(b) Show that $16 \le \int_{\mathscr{R}} f(P)\,dA \le 80$.

5. Use the estimate obtained in Example 1 of the volume under the paraboloid $z=\frac{1}{4}(x^2+y^2)$ and above the rectangle $\mathscr{R}$ with vertices $(0,0)$, $(3,0)$, $(3,2)$, and $(0,2)$ to estimate the average height of the paraboloid over $\mathscr{R}$.

6. Assume that for P in $\mathscr{R}$, $m \le f(P) \le M$, where m and M are constants. Let A be the area of $\mathscr{R}$. By examining approximating sums, show that $mA \le \int_{\mathscr{R}} f(P)\,dA \le MA$.

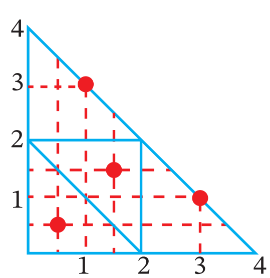


Figure 17.1.8

7. (a) Let $\mathscr{R}$ be the rectangle with vertices $(0,0)$, $(2,0)$, $(2,3)$, and $(0,3)$. Let $f(P)=f(x,y)=\sqrt{x+y}$. Estimate $\int_{\mathscr{R}} \sqrt{x+y}\,dA$ by partitioning $\mathscr{R}$ into six 1×1 squares and choosing the sampling points to be their centers.

(b) Use (a) to estimate the average value of f over $\mathscr{R}$.

8. (a) Let $\mathscr{R}$ be the square with vertices $(0,0)$, $(0.8,0)$, $(0.8,0.8)$, and $(0,0.8)$. Let $f(P)=f(x,y)=e^{xy}$. Estimate $\int_{\mathscr{R}} e^{xy}\,dA$ by partitioning $\mathscr{R}$ into sixteen squares and choosing the sampling points to be their centers.

(b) Use (a) to estimate the average value of $f(P)$ over $\mathscr{R}$.

(c) Show that $0.64 \le \int_{\mathscr{R}} f(P)\,dA \le 0.64e^{0.64}$.

9. (a) Let $\mathscr{R}$ be the triangle with vertices $(0,0)$, $(4,0)$, and $(0,4)$ shown in Figure 17.1.8. Let $f(x,y)=x^2y$. Use the partition into four triangles and sampling points shown in the diagram to estimate $\int_{\mathscr{R}} f(P)\,dA$.

(b) What is the maximum value of $f(x,y)$ in $\mathscr{R}$?

(c) From (b) obtain an upper bound on $\int_{\mathscr{R}} f(P)\,dA$.

10. (a) Sketch the surface $z = \sqrt{x^2 + y^2}$.
(b) Let V be the region in space below the surface in (a) and above the square $\mathscr{R}$ with vertices $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$. Let v be the volume of V. Show that $v \le \sqrt{2}$.
(c) Using a partition of $\mathscr{R}$ with sixteen squares, find an estimate for v that is too large.
(d) Using the partition in (c), find an estimate for v that is too small.

11. The amount of rain that falls at point P during one year is $f(P)$ inches. Let $\mathscr{R}$ be a region, and assume areas are measured in square inches. (a) What is the meaning of $\int_{\mathscr{R}} f(P)\, dA$? (b) What is the meaning of $\frac{\int_{\mathscr{R}} f(P)\, dA}{\text{Area of } \mathscr{R}}$?

12. A region $\mathscr{R}$ in the plane is divided into two regions $\mathscr{R}_1$ and $\mathscr{R}_2$. The function $f(P)$ is defined throughout $\mathscr{R}$. Assume that you know the areas, A_1 and A_2, of $\mathscr{R}_1$ and $\mathscr{R}_2$ and the averages, f_1 and f_2, of f over $\mathscr{R}_1$ and $\mathscr{R}_2$. Find the average of f over $\mathscr{R}$. (See Figure 17.1.9(a).)

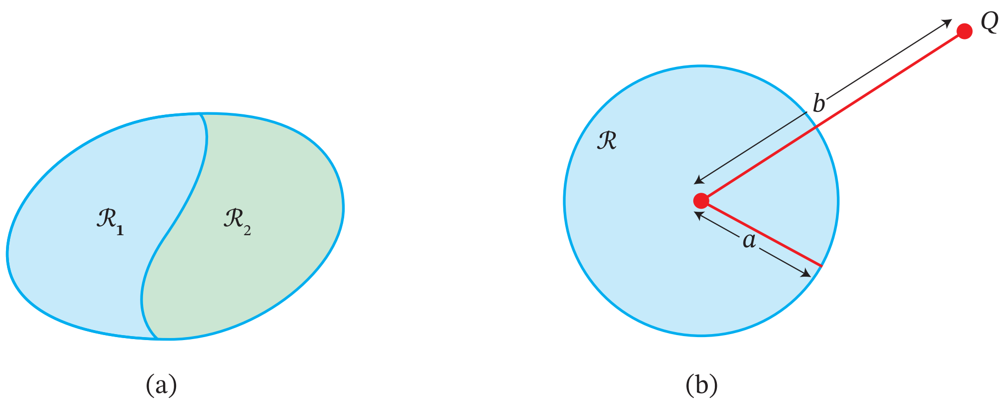


Figure 17.1.9

13. A point Q on the xy-plane is at a distance b from the center of a disk $\mathscr{R}$ of radius $a (a < b)$ in the xy-plane. For P in $\mathscr{R}$ let $f(P) = \frac{1}{|PQ|}$. Find positive numbers c and d such that: $c < \int_{\mathscr{R}} f(P)\, dA < d$. The numbers c and d depend on a and b. See Figure 17.1.9(b).

14. Figure 17.1.10(a) shows parts of some level curves of a function $z = f(x, y)$ and a square $\mathscr{R}$. Estimate $\int_{\mathscr{R}} f(P)\, dA$, and describe your reasoning.

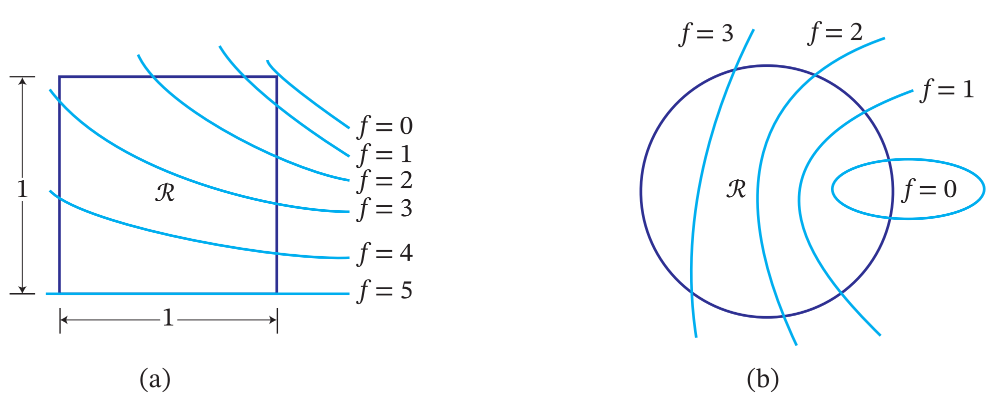


Figure 17.1.10

15. Figure 17.1.10(b) shows parts of some level curves of a function $z = f(x, y)$ and a unit disk $\mathscr{R}$. Estimate $\int_{\mathscr{R}} f(P)\, dA$, and describe your reasoning.

16. (a) Let $\mathscr{R}$ be a disk of radius 1. Let $f(P)$, for P in $\mathscr{R}$, be the distance from P to the center of the disk. By cutting $\mathscr{R}$ into narrow circular rings with centers at the center of the disk, evaluate $\int_{\mathscr{R}} f(P)\, dA$.

(b) Find the average of $f(P)$ over $\mathscr{R}$.

Exercises 17 and 18 introduce **Monte Carlo methods** for estimating a double integral using randomly chosen points. They tend to be inefficient because the error decreases on the order of $\frac{1}{\sqrt{n}}$, where n is the number of random points. That is a slow rate. They are used only when it's possible to choose n large.

17. This exercise involves estimating an integral by choosing points randomly. A computer can generate random numbers and thus random points in the plane that can be used to estimate definite integrals, as we now show. Say that a region $\mathscr{R}$ lies in the square whose vertices are $(0,0)$, $(2,0)$, $(2,2)$, and $(0,2)$, and a complicated function f is defined in $\mathscr{R}$. The machine generates 100 random points (x, y) in the square. Of these, 73 lie in $\mathscr{R}$. The average value of f for these 73 points is 2.31.

(a) What is a reasonable estimate of the area of $\mathscr{R}$? (b) What is a reasonable estimate of $\int_{R} f(P)\, dA$?

18. Let $\mathscr{R}$ be the disk in the xy-plane bounded by the unit circle $x^2 + y^2 = 1$. Let $f(x, y) = e^{x^2 y}$ be the temperature at the point (x, y).

(a) Estimate the average value of f over $\mathscr{R}$ by evaluating $f(x, y)$ at twenty random points in $\mathscr{R}$. (Adjust your program to select each of x and y randomly in the interval $[-1, 1]$. In this way you construct a random point (x, y) in the square whose vertices are $(1,1)$, $(-1,1)$, $(-1,-1)$, $(1,-1)$. Consider only those points that lie in $\mathscr{R}$.) (b) Use (a) to estimate $\int_{\mathscr{R}} f(P)\, dA$. (c) Show why $\frac{\pi}{e} \le \int_{\mathscr{R}} f(P)\, dA \le \pi e$.

---

19. Sam is heckling again.

SAM: As usual, the authors made this harder than necessary. They didn't need to introduce diameters. Instead they could have used good old area. They could have taken the limit as all the areas of the little pieces approached 0. I'll send them a note.

Is Sam right? Explain your response.

20. (a) Does a square of diameter d fit in a disk of diameter d? Explain.
(b) Does an equilateral triangle of diameter d fit in a disk of diameter d? Explain.
(c) What other regular n-gons of diameter d fit in a disk of diameter d? Explain.

In Exercises 21 and 24 a and b are constants. In each case verify that the derivative of the first expression is the second expression.

21. $\dfrac{ax^2 + b}{x^3}$, $\dfrac{-(ax^2 + 3b)}{x^4}$.

22. $\dfrac{x}{2} + \dfrac{\sin(2ax)}{4a}$, $\cos^2(ax)$.

23. $\dfrac{1}{a}\sin(ax) - \dfrac{1}{3a}\sin^3(ax)$, $\cos^3(ax)$.

24. $\dfrac{b}{a^2(ax+b)} + \dfrac{1}{a^2}\ln|ax+b|$, $\dfrac{x}{(ax+b)^2}$.

17.2 Computing Double Integrals Using Rectangular Coordinates

In this section we will show how to use rectangular coordinates to evaluate $\int_{\mathscr{R}} f(P)\, dA$, the integral of a function f over a plane region $\mathscr{R}$. The method requires that both $\mathscr{R}$ and f be described in rectangular coordinates. We first show how to describe plane regions $\mathscr{R}$ in rectangular coordinates.

Describing $\mathscr{R}$ in Rectangular Coordinates

Some examples illustrate how to describe plane regions by their cross sections in rectangular coordinates.

EXAMPLE 1. Describe a disk $\mathscr{R}$ of radius a in rectangular coordinates.

SOLUTION Introduce an xy-coordinate system with its origin at the center of the disk, as in Figure 17.2.1(a). The figure shows that x ranges from $-a$ to a. We now tell how y varies for each x in $[-a, a]$.

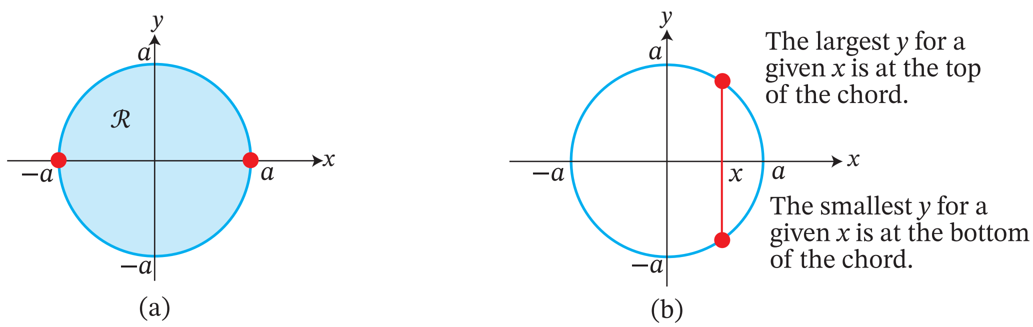


Figure 17.2.1

Figure 17.2.1(b) shows a cross section at x. The circle has the equation $x^2+y^2=a^2$. The top half has the equation $y=\sqrt{a^2-x^2}$ and the bottom half, $y=-\sqrt{a^2-x^2}$. So, for x in $[-a,a]$, y ranges from $-\sqrt{a^2-x^2}$ to $\sqrt{a^2-x^2}$. (As a check, test $x=0$. Are the bottom and top of the circle $y=-\sqrt{a^2-0^2}=-a$ and $y=\sqrt{a^2-0^2}=a$, respectively? Yes, as Figure 17.2.1(b) shows.)

Region $\mathscr{R}$ is described by vertical cross sections as

$$-a \le x \le a, \qquad -\sqrt{a^2-x^2} \le y \le \sqrt{a^2-x^2}.$$

◆

EXAMPLE 2. Let $\mathscr{R}$ be the region bounded by $y=x^2$, the x-axis, and the line $x=2$. Describe $\mathscr{R}$ by cross sections parallel to the y-axis.

SOLUTION Figure 17.2.2(a) shows that for points (x,y) in $\mathscr{R}$, x ranges from 0 to 2. To describe $\mathscr{R}$ completely, we describe the behavior of y for x in the interval $[0,2]$.

Because the cross section above $(x,0)$ extends from the x-axis to the curve $y=x^2$ the y-coordinate varies from 0 to x^2. The description of $\mathscr{R}$ by vertical cross sections is

$$0 \le x \le 2, \qquad 0 \le y \le x^2.$$

◆

EXAMPLE 3. Describe the region $\mathscr{R}$ of Example 2 by cross sections parallel to the x-axis, that is, by horizontal cross sections.

SOLUTION Figure 17.2.2(b) shows that y varies from 0 to 4. For y in $[0,4]$, x varies from a smallest value $x_1(y)$ to

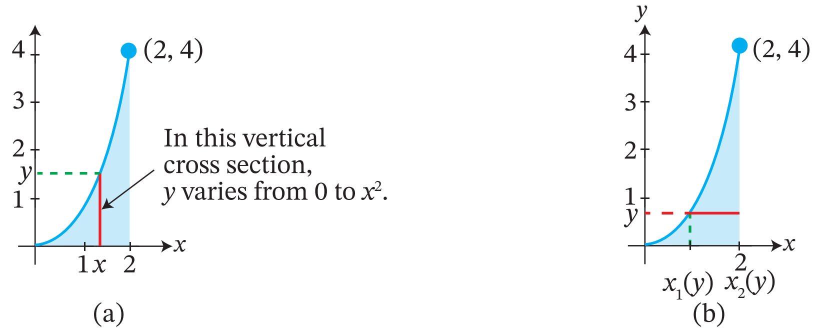


Figure 17.2.2

a largest value $x_2(y)$. For each value of y in $[0,4]$, $x_2(y) = 2$. To find $x_1(y)$, use the fact that the point $(x_1(y), y)$ is on the curve $y = x^2$, that is,

$$x_1(y) = \sqrt{y}.$$

The description of $\mathscr{R}$ in terms of horizontal cross sections is

$$0 \le y \le 4, \qquad \sqrt{y} \le x \le 2.$$

♦

EXAMPLE 4. Describe the region $\mathscr{R}$ whose vertices are $(0,0)$, $(6,0)$, $(4,2)$, and $(0,2)$ by vertical cross sections and then by horizontal cross sections.

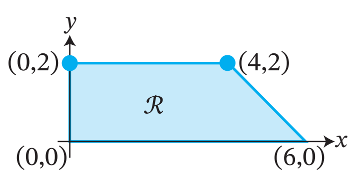


Figure 17.2.3

SOLUTION

In Figure 17.2.3 x varies between 0 and 6. For x in $[0,4]$, y ranges from 0 to 2. For x in $[4,6]$, y ranges from 0 to the value of y on the line through $(4,2)$ and $(6,0)$, which has the equation $y = 6 - x$. The description of $\mathscr{R}$ by vertical cross sections breaks into two parts:

$$0 \le x \le 4, \quad 0 \le y \le 2 \qquad \text{and} \qquad 4 \le x \le 6, \quad 0 \le y \le 6 - x.$$

Using horizontal cross sections provides a simpler description. First, y goes from 0 to 2. For y in $[0,2]$, x goes from 0 to the value of x on the line $y = 6 - x$. Solving this equation for x yields $x = 6 - y$.

The description in terms of horizontal cross sections is :

$$0 \le y \le 2, \qquad 0 \le x \le 6 - y,$$

which is, in fact, shorter than the corresponding description in terms of vertical cross sections. ♦

These examples are typical. First, determine the range of one coordinate, and then see how the other coordinate varies for any fixed value of the first coordinate. And, always pay attention to whether one representation is easier or more natural than the other.

Evaluating a Double Integral as an Iterated Integrals

We will offer an intuitive development of a formula for computing double integrals over plane regions.

We first develop a way for computing a double integral over a rectangle. After applying this in Example 5, we make the modification needed to evaluate double integrals over more general regions. The cross sections of a rectangular region $\mathscr{R}$ are described by

$$a \le x \le b, \qquad c \le y \le d,$$

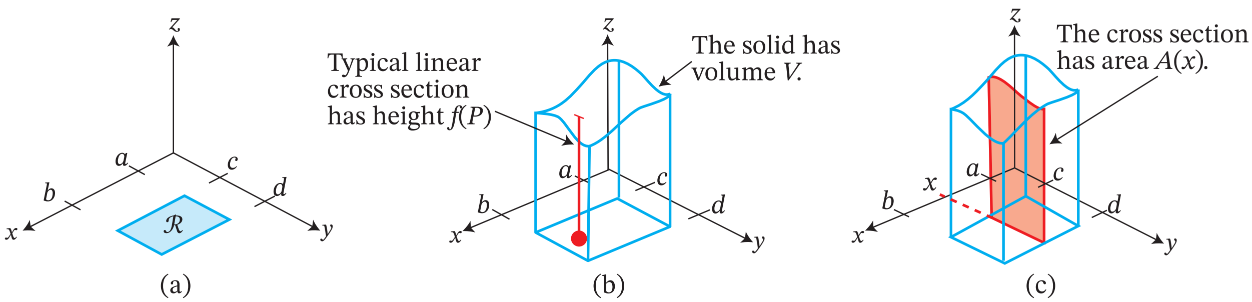


Figure 17.2.4

as shown in Figure 17.2.4(a). If $f(P) \geq 0$ for all P in $\mathcal{R}$, then $\int_{\mathcal{R}} f(P)\, dA$ is the volume V of the solid whose base is $\mathcal{R}$ and which has height $f(P)$ above P. (See Figure 17.2.4(b).) To find this volume, let $A(x)$ be the area of the cross section made by a plane perpendicular to the x-axis and having abscissa x, as in Figure 17.2.4(c). As was shown in Section 7.4,

$$V = \int_a^b A(x)\, dx.$$

The area $A(x)$ is expressible as a definite integral

$$A(x) = \int_c^d f(x,y)\, dy.$$

Here x is held fixed throughout this integration to find $A(x)$. This provides a way to evaluate $V = \int_{\mathcal{R}} f(P)\, dA$, namely,

$$\int_{\mathcal{R}} f(P)\, dA = V = \int_a^b A(x)\, dx = \int_a^b \left(\int_c^d f(x,y)\, dy \right) dx.$$

In short,

$$\int_{\mathcal{R}} f(P)\, dA = \int_a^b \left(\int_c^d f(x,y)\, dy \right) dx. \qquad (17.2.1)$$

Cross sections by planes perpendicular to the y-axis could be used. Then similar reasoning shows that

$$\int_{\mathcal{R}} f(P)\, dA = \int_c^d \left(\int_a^b f(x,y)\, dx \right) dy. \qquad (17.2.2)$$

The expressions 17.2.1 and 17.2.2 are called **iterated integrals**. Usually the parentheses are omitted and they are written $\int_a^b \int_c^d f(x,y)\, dy\, dx$ and $\int_c^d \int_a^b f(x,y)\, dx\, dy$, respectively. The order of dx and dy matters.Notice that the integral sign is like a left parenthesis and the corresponding differential is a right parenthesis, with the limits of integration corresponding with the variable in the differential. The differential on the left is the inner nesting; it tells which integration is performed first.

EXAMPLE 5. Compute the double integral $\displaystyle\int_{\mathcal{R}} f(P)\, dA$, where $\mathcal{R}$ is the rectangle shown in Figure 17.2.5(a) and f is defined by $f(P) = \dfrac{1}{4}|OP|^2$.

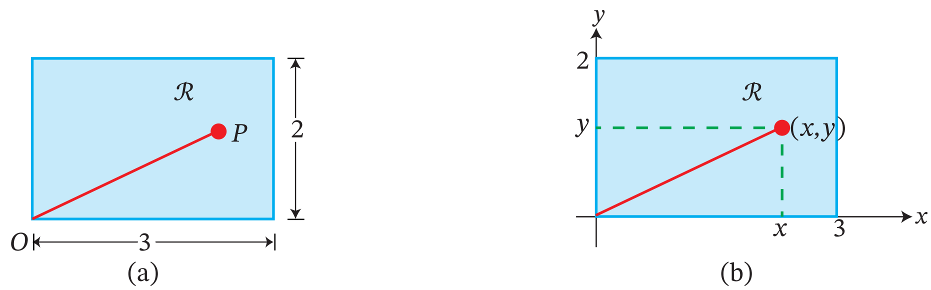


Figure 17.2.5

SOLUTION Introduce xy-coordinates as in Figure 17.2.5(b). In rectangular coordinates

$$f(x,y) = \frac{1}{4}|OP|^2 = \frac{1}{4}(x^2+y^2).$$

Because x takes all values from 0 to 3 and for each x the number y takes all values from 0 to 2,

$$\int_{\mathcal{R}} f(P)\,dA = \int_0^3 \left(\int_0^2 \frac{1}{4}(x^2+y^2)\,dy \right) dx.$$

NOTE: This double integral appeared in Examples 1 and 2, in Section 17.1, for a volume and for the mass of a solid, respectively.

We must first compute the inner integral $\int_0^2 (x^2+y^2)/4\,dy$, where x is fixed in $[0,3]$. To apply the fundamental theorem of calculus, find a function $F(x,y)$ such that $\partial F/\partial y = (x^2+y^2)/4$. Because x is constant during this first integration, $F(x,y) = x^2y/4 + y^3/12$ is such a function. By the fundamental theorem of calculus,

$$\begin{aligned} \int_0^2 \frac{1}{4}(x^2+y^2)\,dy &= \left(\frac{x^2y}{4} + \frac{y^3}{12}\right)\Bigg|_{y=0}^{y=2} && \text{(FTC I)} \\ &= \left(\frac{x^2\cdot 2}{4} + \frac{2^3}{12}\right) - \left(\frac{x^2\cdot 0}{4} + \frac{0^3}{12}\right) \\ &= \frac{x^2}{2} + \frac{2}{3}. \end{aligned}$$

The expression $x^2/2 + 2/3$ is the area $A(x)$ discussed earlier in this section.

Next we compute

$$\int_0^3 A(x)\,dx = \int_0^3 \left(\frac{x^2}{2} + \frac{2}{3}\right) dx.$$

By the fundamental theorem of calculus,

$$\begin{aligned} \int_0^3 \left(\frac{x^2}{2} + \frac{2}{3}\right) dx &= \left(\frac{x^3}{6} + \frac{2x}{3}\right)\Bigg|_0^3 && \text{(FTC I)} \\ &= \frac{27}{6} + \frac{6}{3} = \frac{39}{5} = \frac{13}{2}. \end{aligned}$$

Hence the double integral has the value $13/2 = 6.5$. The volume of the region in Example 1 of Section 17.1 is 6.5 cubic units. The mass in Example 2 is 6.5 grams. ♦

Observation 17.2.1: ***Double Integrals and Iterated Integrals***

Double integrals and iterated integrals are quite different objects, even when they represent the same quantity. To help distinguish them, double integrals are denoted by one integral sign (and a two-dimensional differential, such as dA) and an iterated integrals by two integral signs (each with a one-dimensional differential, such as dx and dy).

If $\mathscr{R}$ is not a rectangle, the iterated integral that equals $\int_{\mathscr{R}} f(P)\,dA$ differs from that for the case where $\mathscr{R}$ is a rectangle only in the intervals of integration. There are separate, but similar, formulas for iterated integrals when the region is described with vertical and horizontal cross sections.

Formula 17.2.1: ***Iterated Integral with Vertical Cross Sections***

If $\mathscr{R}$ has the description

$$a \le x \le b \qquad y_1(x) \le y \le y_2(x),$$

by cross sections parallel to the y-axis, as in Figure 17.2.6, then

$$\int_{\mathscr{R}} f(P)\,dA = \int_a^b \left(\int_{y_1(x)}^{y_2(x)} f(x,y)\,dy \right) dx.$$

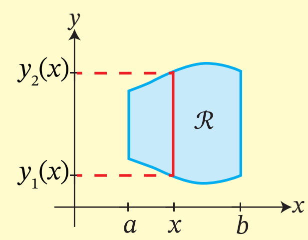


Figure 17.2.6

Formula 17.2.2: ***Iterated Integral with Horizontal Cross Sections***

If $\mathscr{R}$ has the description

$$c \le y \le d \qquad x_1(y) \le x \le x_2(y),$$

by cross sections parallel to the x-axis, then

$$\int_{\mathscr{R}} f(P)\,dA = \int_c^d \left(\int_{x_1(y)}^{x_2(y)} f(x,y)\,dx \right) dy.$$

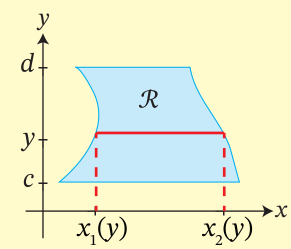


Figure 17.2.7

The intervals of integration are determined by the description of the region $\mathscr{R}$; the function f influences only the integrand. See Figure 17.2.7.

In the next example $\mathscr{R}$ is the region described in Examples 2 and 3.

EXAMPLE 6. Evaluate $\int_{\mathscr{R}} 3xy\,dA$ over the region $\mathscr{R}$ shown in Figure 17.2.8(a).

SOLUTION We can use cross sections parallel either to the y-axis or to the x-axis.

If, as shown in Figure 17.2.8(b), cross sections parallel to the y-axis are used, then $\mathscr{R}$ is described by

$$0 \le x \le 2 \qquad 0 \le y \le x^2.$$

Thus

$$\int_{\mathscr{R}} 3xy\,dA = \int_0^2 \left(\int_0^{x^2} 3xy\,dy \right) dx.$$

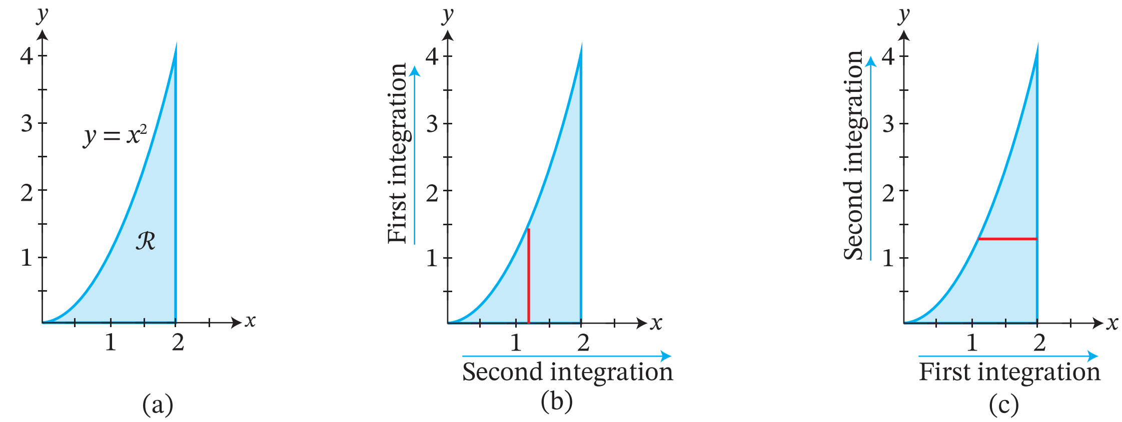


Figure 17.2.8

To compute the iterated integral, we start with the integral in which x is fixed and y goes from 0 to x^2. With x fixed,

$$\begin{aligned}\int_0^{x^2} 3xy\,dy &= \left(3x\frac{y^2}{2}\right)\Bigg|_{y=0}^{y=x^2} \qquad \text{(FTC I)}\\ &= 3x\frac{(x^2)^2}{2} - 3x\frac{0^2}{2}\\ &= \frac{3x^5}{2}.\end{aligned}$$

Then

$$\int_0^2 \frac{3x^5}{2}\,dx = \frac{3x^6}{12}\Bigg|_0^2 = 16.$$

Alternatively, the region $\mathscr{R}$ can, as shown in Figure 17.2.8(c), be described in terms of cross sections parallel to the x-axis:

$$0 \le y \le 4 \qquad \sqrt{y} \le x \le 2.$$

Then the double integral is evaluated by a different iterated integral,

$$\int_{\mathscr{R}} 3xy\,dA = \int_0^4 \left(\int_{\sqrt{y}}^2 3xy\,dx\right) dy,$$

which, as may be verified, also equals 16. ♦

The fact that $\int_{\mathscr{R}} 3xy\,dA = 16$ has many interpretations; here are three:

1. If at each point $P = (x, y)$ in $\mathscr{R}$ we erect a line segment above P of length $3xy$ cm, then the volume of this solid is 16 cm^3 (See Figure 17.2.9.)
2. If the density of matter at (x, y) in $\mathscr{R}$ is $3xy$ kg/cm^2, then the total mass in $\mathscr{R}$ is 16 kg.
3. If the temperature at (x, y) is $3xy$ degrees, then the average temperature in $\mathscr{R}$ is 16 degrees cm^2 divided by the area of $\mathscr{R}$. The area of $\mathscr{R}$ is:

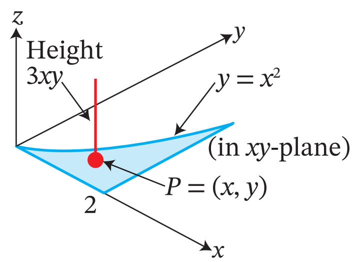


Figure 17.2.9

$$\begin{aligned}
\int_{\mathscr{R}} dA &= \int_0^2 \int_0^{x^2} dy\,dx && (\text{ express double integral as iterated integral })\\
&= \int_0^2 x^2\,dx && (\text{ evaluate inner definite integral (wrt } y) \text{ })\\
&= \frac{1}{3}x^3\Big|_0^2 && (\text{ FTC I })\\
&= \frac{8}{3}\text{cm}^2.
\end{aligned}$$

The average temperature of $\mathscr{R}$ is $16/(8/3) = 6$ degrees.

In Example 6 we could evaluate $\int_{\mathscr{R}} f(P)\,dA$ by cross sections in either direction. In the next example we do not have that option.

EXAMPLE 7. Find the mass of the triangular lamina with vertices at $(0,0)$, $(2,1)$, and $(0,1)$ when the density at (x,y) is e^{y^2}. That is, evaluate $\int_{\mathscr{R}} f(P)\,dA$, where $f(x,y) = e^{y^2}$ and $\mathscr{R}$ is the triangular region $\mathscr{R}$ is shown in Figure 17.2.10.

SOLUTION The description of $\mathscr{R}$ by vertical cross sections is

$$0 \le x \le 2, \qquad \frac{x}{2} \le y \le 1.$$

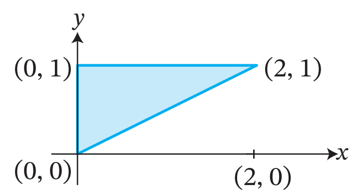


Figure 17.2.10

Hence

$$\int_{\mathscr{R}} f(P)\,dA = \int_0^2 \left(\int_{x/2}^1 e^{y^2}\,dy \right) dx.$$

Since e^{y^2} does not have an elementary antiderivative, the fundamental theorem of calculus is useless in computing

$$\int_{x/2}^1 e^{y^2}\,dy.$$

Do we have any other options to evaluate this double integral? Maybe we can try horizontal cross sections instead. With this approach the description of $\mathscr{R}$ is now

$$0 \le y \le 1, \qquad 0 \le x \le 2y.$$

This leads to a different iterated integral, namely

$$\int_{\mathscr{R}} f(P)\,dA = \int_0^1 \left(\int_0^{2y} e^{y^2}\,dx \right) dy.$$

For the first integration, $\int_0^{2y} e^{y^2}\,dx$, y is fixed; the integrand is constant. Thus

$$\begin{aligned}
\int_0^{2y} e^{y^2}\,dx = e^{y^2}\int_0^{2y} 1\,dx && (\text{ factor term in } y \text{ from integral wrt } x \text{ })\\
= e^{y^2}x\Big|_{x=0}^{x=2y} && (\text{ FTC I })\\
= e^{y^2}2y.
\end{aligned}$$

The second definite integral in the iterated integral is thus $\int_0^1 e^{y^2} 2y\,dy$, which can be evaluated by the fundamental theorem of calculus, since $\left(e^{y^2}\right)' = e^{y^2} 2y$:

$$\int_0^1 e^{y^2} 2y\,dy = e^{y^2}\Big|_0^1 = e^{1^2} - e^{0^2} = e - 1.$$

◆

Observation 17.2.2: *Choosing the Order of Integration in an Iterated Integral*

Computing a definite integral over a region $\mathcal{R}$ in the xy-plane involves a wise choice of an xy-coordinate system, a description of $\mathcal{R}$ and f relative to this coordinate system, and the computation of two definite integrals over intervals. The order of the integrations may affect the difficulty of the computation. That order is determined by the description of $\mathcal{R}$ by horizontal or vertical cross sections.

Summary

We showed that the integral of $f(P)$ over a plane region $\mathcal{R}$ can be evaluated by an iterated integral, where the limits of integration are determined by $\mathcal{R}$. If a line parallel to the y-axis meets the boundary of $\mathcal{R}$ in at most two points then $\mathcal{R}$ has the description

$$a \le x \le b, \qquad y_1(x) \le y \le y_2(x)$$

and double integrals can be evaluated as an iterated integral with the integral with respect to y evaluated before the integral with respect to x:

$$\int_{\mathcal{R}} f(P)\,dA = \int_a^b \left(\int_{y_1(x)}^{y_2(x)} f(x,y)\,dy \right) dx.$$

If a line parallel to the x-axis meets the boundary of $\mathcal{R}$ in at most two points, then, similarly, $\mathcal{R}$ can be described in the form

$$c \le y \le d \qquad x_1(y) \le x \le x_2(y)$$

and double integrals can be evaluated as an iterated integral with the integral with respect to x evaluated before the integral with respect to y:

$$\int_{\mathcal{R}} f(P)\,dA = \int_c^d \left(\int_{x_1(y)}^{x_2(y)} f(x,y)\,dx \right) dy.$$

Observation 17.2.3: *A Few More Words on Notation*

The notation $\int_{\mathcal{R}} f(P)\,dA$ is used for a (double) integral over a plane region. We will soon learn to use $\int_{\mathcal{S}} f(P)\,dS$ for an integral over a surface and $\int_{\mathcal{R}} f(P)\,dV$ for an integral over a region in space. The symbols dA, dS, and dV indicate the type of set over which the integral is defined.

Many people use repeated integral signs to distinguish dimensions. For instance they write $\int f(P)\,dA$ as $\iint f(P)\,dA$ or $\iint f(x,y)\,dx\,dy$. Similarly, they denote an integral over a region in space by $\iiint f(P)\,dx\,dy\,dz$.

We use the single integral sign and P for point for all integrals for (at least) three reasons:

1. it is free of any coordinate system
2. it is compact (uses the fewest symbols): $\int$ for "integral", $f(P)$ or f for the integrand, and dA, dS, or dV to indicate whether the domain of integration is a region, surface, or volume, and
3. it allows the symbols $\iint$ and $\iiint$ to be reserved for iterated integrals. An iterated integral is a different mathematical object; each integral in an iterated integral is an integral over an interval.

EXERCISES for Section 17.2

Exercises 1 to 12 provide practice in describing plane regions by cross sections in rectangular coordinates. Describe the region by (a) vertical cross sections and (b) horizontal cross sections.

1. The triangle with vertices $(0,0)$, $(2,1)$, $(0,1)$.

2. The triangle with vertices $(0,0)$, $(2,0)$, $(1,1)$.

3. The parallelogram with vertices $(0,0)$, $(1,0)$, $(2,1)$, $(1,1)$.

4. The parallelogram with vertices $(2,1)$, $(5,1)$, $(3,2)$, $(6,2)$.

5. The disk of radius 5 and center $(0,0)$.

6. The trapezoid with vertices $(1,0)$, $(3,2)$, $(3,3)$,$(1,6)$.

7. The triangle bounded by the lines $y=x$, $x+y=2$, and $x+3y=8$.

8. The region bounded by the ellipse $4x^2+y^2=4$.

9. The triangle bounded by the lines $x=0$, $y=0$, and $2x+3y=6$.

10. The region bounded by the curves $y=e^x$, $y=1-x$, and $x=1$.

11. The quadrilateral bounded by the lines $y=1$, $y=2$, $y=x$, and $y=\frac{x}{3}$.

12. The quadrilateral bounded by the lines $x=1$, $x=2$, $y=x$, and $y=5-x$.

In Exercises 13 to 16 draw the regions and describe them by horizontal cross sections.

13. $0\le x\le 2$, $2x\le y\le 3x$

14. $1\le x\le 2$, $x^3\le y\le 2x^2$

15. $0\le x\le \frac{\pi}{4}$, $0\le y\le \sin(x)$; $\frac{\pi}{4}\le x\le \frac{\pi}{2}$, $0\le y\le \cos(x)$

16. $1\le x\le e$, $\frac{x-1}{e-1}\le y\le \ln x$

In Exercises 17 to 22 evaluate each iterated integral.

17. $\displaystyle\int_0^1\int_0^x (x+2y)\,dy\,dx$

18. $\displaystyle\int_1^2\int_x^{2x} dy\,dx$

19. $\displaystyle\int_0^2\int_0^{x^2} xy^2\,dy\,dx$

20. $\displaystyle\int_1^2\int_0^y e^{x+y}\,dx\,dy$

21. $\displaystyle\int_1^2\int_0^{\sqrt{y}} yx^2\,dx\,dy$

22. $\displaystyle\int_0^1\int_0^x y\sin(\pi x)\,dy\,dx$

---

23. Complete the calculation of the second iterated integral in Example 6.

24. (a) Sketch the solid region $\mathscr{S}$ below the plane $z=1+x+y$ and above the triangle $\mathscr{R}$ in the xy-plane with vertices $(0,0)$, $(1,0)$, $(0,2)$. (b) Describe $\mathscr{R}$ in terms of coordinates. (c) Set up an iterated integral for the volume of $\mathscr{S}$. (d) Evaluate the expression in (c), and show in the manner of Figure 17.2.8(a) and 17.2.8(b) which integration you performed first. (e) Carry out (c) and (d) in the other order of integration.

25. Let $\mathscr{S}$ be the solid region below the paraboloid $z=x^2+2y^2$ and above the rectangle in the xy-plane with vertices $(0,0)$, $(1,0)$, $(1,2)$, $(0,2)$. Carry out the steps of Exercise 24.

26. Let $\mathscr{S}$ be the solid region below the saddle $z = xy$ and above the triangle in the xy-plane with vertices $(1,1)$, $(3,1)$, and $(1,4)$. Carry out the steps of Exercise 24 .

27. Let $\mathscr{S}$ be the solid region below the saddle $z = xy$ and above the region in the first quadrant of the xy-plane bounded by the parabolas $y = x^2$ and $y = 2x^2$ and the line $y = 2$. Carry out the steps of Exercise 24.

28. Find the mass of a lamina occupying the bounded region bounded by $y = 2x^2$ and $y = 5x-3$ and whose density at (x, y) is xy.

29. Find the mass of a thin lamina occupying the triangle whose vertices are $(0,0)$, $(1,0)$, $(1,1)$ and whose density at (x, y) is $\frac{1}{1+x^2}$.

30. The temperature at (x, y) is $T(x, y) = \cos(x+2y)$. Find the average temperature in the triangle with vertices $(0,0)$, $(1,0)$, $(0,2)$.

31. The temperature at (x, y) is $T(x, y) = e^{x-y}$. Find the average temperature in the triangle with vertices $(0,0)$, $(1,1)$, and $(3,1)$.

In Exercises 32 to 35 replace the iterated integral by an equivalent one with the order of integration reversed. First sketch the region $\mathscr{R}$ of integration.

32. $\displaystyle\int_0^2\int_0^{x^2} x^3 y\, dy\, dx$

33. $\displaystyle\int_0^{\pi/2}\int_0^{\cos x} x^2\, dy\, dx$

34. $\displaystyle\int_0^1\int_{x/2}^{x} xy\, dy\, dx + \int_1^2\int_{x/2}^{1} xy\, dy\, dx$

35. $\displaystyle\int_{-1/\sqrt{2}}^{0}\int_{-x}^{\sqrt{1-x^2}} x^3 y\, dy\, dx + \int_0^1\int_0^{\sqrt{1-x^2}} x^3 y\, dy\, dx$

In Exercises 36 to 39 evaluate the iterated integrals. First sketch the region of integration.

36. $\displaystyle\int_0^1\int_x^1 \sin(y^2)\, dy\, dx$

37. $\displaystyle\int_0^1\int_{\sqrt{x}}^1 \frac{dy}{\sqrt{1+y^3}}\, dx$

38. $\displaystyle\int_0^1\int_{\sqrt[3]{y}}^1 \sqrt{1+x^4}\, dx\, dy$

39. $\displaystyle\int_1^2\int_1^y \frac{\ln x}{x}\, dx\, dy + \int_2^4\int_{y/2}^2 \frac{\ln x}{x}\, dx\, dy$

---

40. Let $f(x, y) = y^2 e^{y^2}$ and let $\mathscr{R}$ be the triangle bounded by $y = a$, $y = x/2$, and $y = x$. Assume that a is positive. (a) Set up two iterated integrals for $\displaystyle\int_{\mathscr{R}} f(P)\, dA$. (b) Evaluate the easier one.

41. Let $\mathscr{R}$ be the finite region bounded by the curve $y = \sqrt{x}$ and the line $y = x$. Let $f(x, y) = \frac{\sin(y)}{y}$ if $y \neq 0$ and $f(x,0) = 1$. Compute $\displaystyle\int_{\mathscr{R}} f(P)\, dA$.

Exercises 42 to 44 are related.

42. Two points are picked at random in an interval of length a. What is the average value of the square of the distance between them?

(a) Let $\mathscr{R}$ be the square whose vertices are $(0,0)$, $(a,0)$, (a,a), and $(0,a)$. Define a function f on $\mathscr{R}$ by $f(x,y) = (y-x)^2$. Why is the answer to the question $\frac{1}{a^2}\int\limits_{\mathscr{R}} f(x,y)\,dA$?

(b) Why is the average less than a^2?

(c) Show that the average is $\frac{a^2}{6}$.

43. Two points are picked at random in an interval of length a. Show that the average value of the distance between them is $\frac{a}{3}$. In contrast to the previous problem, $f(x,y) = |y-x|$.

44. **Sam:** I can do Exercise 43 without calculus, in my head.

Jane: Impossible.

Sam: Take a typical point, say $\frac{a}{5}$. The average distance from it to points to the left of it is $\frac{a}{10}$. The average distance from it to points to the right of it is $\frac{2a}{5}$. The average of $\frac{a}{10}$ and $\frac{2a}{5}$ is $\frac{a}{4}$.

Jane: I believe you.

Sam: So the average distance between any two random points is $\frac{a}{4}$, not $\frac{a}{3}$, which Exercise 43 says.

Jane: There must be a mistake; $\frac{a}{3}$ and $\frac{a}{4}$ cannot both be correct.

Find the mistake.

17.3 Computing Double Integrals Using Polar Coordinates

This section shows how to evaluate a double integral, $\int_{\mathscr{R}} f(P)\,dA$, using polar coordinates. The method is appropriate when the region $\mathscr{R}$ has a simple description in polar coordinates, for instance if it is a disk or cardioid. As in Section 17.2, we first examine how to describe a region in polar coordinates. Then we develop the iterated integral in polar coordinates that equals $\int_{\mathscr{R}} f(P)\,dA$.

Describing $\mathscr{R}$ in Polar Coordinates

The description of a region $\mathscr{R}$ in polar coordinates begins by first determining the range of θ and then seeing how r varies for a fixed value of θ. (The reverse order is seldom useful.) Some examples show how to find how r varies.

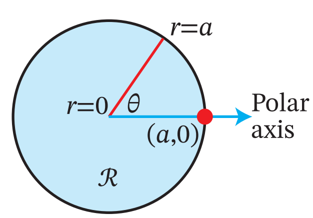


Figure 17.3.1

EXAMPLE 1. Let $\mathscr{R}$ be the disk of radius a and center at the pole of a polar coordinate system. (See Figure 17.3.1.) Describe $\mathscr{R}$ with cross sections by rays emanating from the pole. That is, in polar coordinates.

SOLUTION To sweep out $\mathscr{R}$, θ goes from 0 to 2π. On the ray for a fixed angle θ, r goes from 0 to a. (Again, see Figure 17.3.1.) The description is

$$0 \le \theta \le 2\pi, \qquad 0 \le r \le a.$$

◆

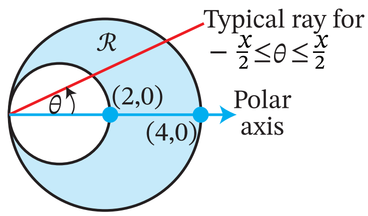


Figure 17.3.2

EXAMPLE 2. Let $\mathcal{R}$ be the region between the circles $r = 2\cos(\theta)$ and $r = 4\cos(\theta)$. Describe $\mathcal{R}$ in terms of cross sections by rays from the pole. (See Figure 17.3.2.)

SOLUTION To sweep out this region, use the rays from $\theta = -\pi/2$ to $\theta = \pi/2$. For each θ, r varies from $2\cos(\theta)$ to $4\cos(\theta)$. The description is

$$-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}, \qquad 2\cos(\theta) \le r \le 4\cos(\theta).$$

◆

As Examples 1 and 2 suggest, polar coordinates provide simple descriptions for regions bounded by circles centered at the origin or circles that pass through the origin. The next example shows that polar coordinates may also provide simple descriptions of regions bounded by straight lines, especially if some of them pass through the origin.

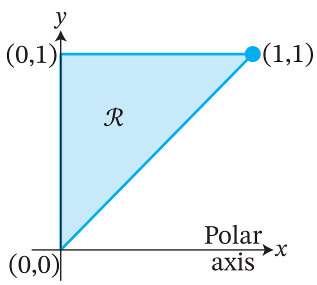


Figure 17.3.3

EXAMPLE 3. Let $\mathcal{R}$ be the triangular region whose vertices in rectangular coordinates are $(0,0)$, $(1,1)$, and $(0,1)$. Describe $\mathcal{R}$ in polar coordinates.

SOLUTION Inspection of $\mathcal{R}$ in Figure 17.3.3 shows that θ varies from $\pi/4$ to $\pi/2$. For each θ, r goes from 0 until the point (r,θ) is on the line $y = 1$, that is, on the line $r\sin(\theta) = 1$. Thus the upper limit of r for each θ is $1/\sin(\theta)$. The description of $\mathcal{R}$ is

$$\frac{\pi}{4} \le \theta \le \frac{\pi}{2}, \qquad 0 \le r \le \frac{1}{\sin(\theta)}.$$

◆

Observation 17.3.1: *Most Common Form for Descriptions of Regions by Rays*

In general, cross sections by rays lead to descriptions of plane regions of the form:

$$\alpha \le \theta \le \beta, \qquad r_1(\theta) \le r \le r_2(\theta).$$

A Basic Difference Between Rectangular and Polar Coordinates

Before we can set up an iterated integral in polar coordinates for $\int_{\mathcal{R}} f(P)\, dA$ we contrast certain properties of rectangular and polar coordinates.

The rectangle in Figure 17.3.4(a) is described by

$$x_0 \le x \le x_0 + \Delta x \qquad \text{and} \qquad y_0 \le y \le y_0 + \Delta y,$$

where x_0, Δx, y_0, and Δy are numbers with Δx and Δy positive. The area of the rectangle is the product of Δx and Δy:

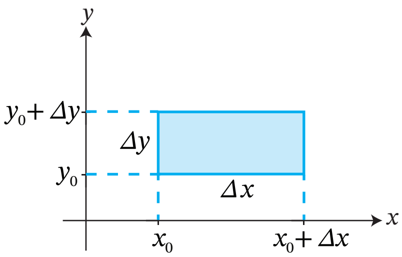
$$\text{Area of rectangle} = \Delta x\, \Delta y.$$

In polar coordinates the situation is much different. As we will show, the area of the little region corresponding to small changes Δr in r and $\Delta\theta$ in θ is not $\Delta r\, \Delta\theta$.

The shaded region in Figure 17.3.4(b) is the set in the plane consisting of the points (r,θ) such that

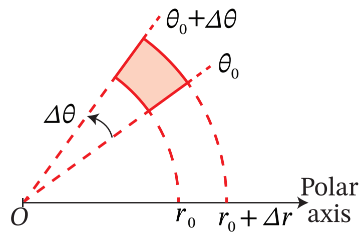
$$r_0 \le r \le r_0 + \Delta r \qquad \text{and} \qquad \theta_0 \le \theta \le \theta_0 + \Delta\theta,$$

where r_0, Δr, θ_0, and $\Delta\theta$ are positive numbers.



$x_0 \le x \le x_0+\Delta x$
$y_0 \le y \le y_0+\Delta y$

(a)



$r_0 \le r \le r_0+\Delta r$
$\theta_0 \le \theta \le \theta_0+\Delta\theta$

(b)

Figure 17.3.4

When Δr and $\Delta\theta$ are small, the set is approximately a rectangle, one side of which has length Δr and the other, $r_0\,\Delta\theta$. So its area is approximately $r_0\,\Delta r\,\Delta\theta$. In this case,

$$\text{Approximate area of polar rectangle} \approx r_0\,\Delta r\,\Delta\theta.$$

The area is not the product of Δr and $\Delta\theta$. (It could not be since $\Delta\theta$ is in radians, a dimensionless quantity. Thus $\Delta r\,\Delta\theta$ would have the dimension of length, not of area.) The exact area will be found in Exercise 30.

Observation 17.3.2: dA *in Polar Coordinates is* $r\,dr\,d\theta$

When setting up iterated integrals in polar coordinates remember to replace dA by $r\,dr\,d\theta$.

How to Evaluate $\int_{\mathscr{R}} f(P)\,dA$ by an Iterated Integral in Polar Coordinates

The method for computing $\int_{\mathscr{R}} f(P)\,dA$ for a region $\mathscr{R}$ described with polar coordinates involves an iterated integral where the dA is replaced by $r\,dr\,d\theta$ (or $r\,d\theta\,dr$). A more detailed explanation of why the r must be added is given at the end of this section.

Algorithm: Evaluating $\int_{\mathscr{R}} f(P)\,dA$ in Polar Coordinates

1. Express $f(P)$ as $f(r,\theta)$.
2. Describe the region r in polar coordinates: $\quad \alpha \le \theta \le \beta, \quad r_1(\theta) \le r \le r_2(\theta)$.
3. Evaluate the iterated integral $\displaystyle\int_\alpha^\beta \left(\int_{r_1(\theta)}^{r_2(\theta)} f(r,\theta)r\,dr \right) d\theta.$

REMINDER: Do not forget the factor r in the integrand.

EXAMPLE 4. Let $\mathscr{R}$ be the semicircle of radius a shown in Figure 17.3.5. Let $f(P)$ be the distance from a point P to the x-axis. Evaluate $\int_{\mathscr{R}} f(P)\,dA$ by an iterated integral in polar coordinates.

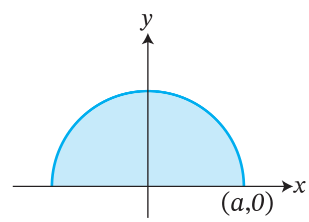


Figure 17.3.5

SOLUTION In polar coordinates $\mathscr{R}$ has the description

$$0 \le \theta \le \pi, \qquad 0 \le r \le a.$$

The distance from P to the x-axis is y in rectangular coordinates. Since $y = r\sin(\theta)$, $f(P) = r\sin(\theta)$. Thus,

$$\int_R f(P)\,dA = \int_0^{\pi}\left(\int_0^{a} \underbrace{(r\sin(\theta))}_{\text{Distance to } x\text{-axis.}} \underbrace{r}_{\text{Remember this factor.}}\, dr\right) d\theta.$$

From here on the calculation is like those in the preceding section. First, evaluate the inside integral (with respect to r):

$$\int_0^a r^2\sin(\theta)\,dr = \sin(\theta)\int_0^a r^2\,dr = \sin(\theta)\left(\frac{r^3}{3}\right)\Big|_0^a = \frac{a^3\sin(\theta)}{3}.$$

Next, evaluate the outer integral (with respect to θ):

$$\begin{aligned}\int_0^{\pi}\frac{a^3\sin(\theta)}{3}\,d\theta &= \frac{a^3}{3}\int_0^{\pi}\sin(\theta)\,d\theta = \frac{a^3}{3}(-\cos(\theta))\Big|_0^{\pi}\\ &= \frac{a^3}{3}[(-\cos(\pi))-(-\cos(0))] = \frac{a^3}{3}(1+1) = \frac{2a^3}{3}.\end{aligned}$$

Thus

$$\int_{\mathscr{R}} f(P)\,dA = \frac{2a^3}{3}.$$

◆

EXAMPLE 5. A ball of radius a has its center at the pole of a polar coordinate system. Find the volume of the part of the ball that lies above the plane region $\mathscr{R}$ bounded by the curve $r = a\cos(\theta)$. (See Figure 17.3.6(a).)

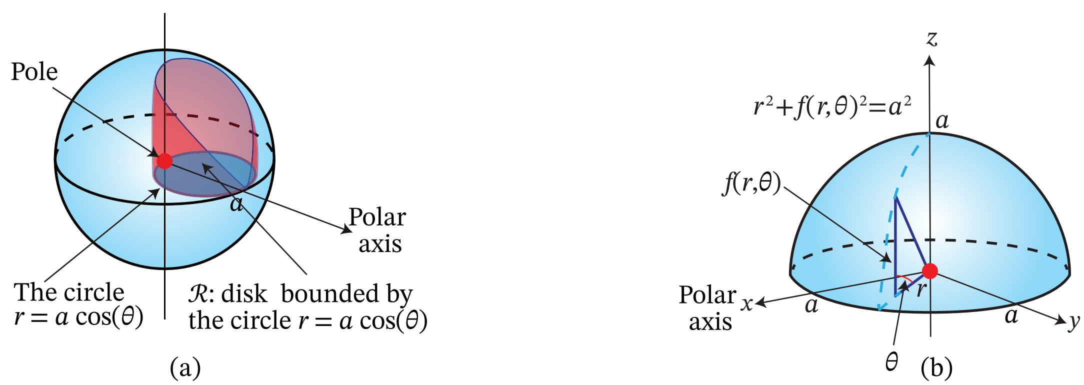


Figure 17.3.6

SOLUTION It is necessary to describe both the region $\mathscr{R}$ and the integrand f in polar coordinates, where $f(P)$ is the length of a cross section of the solid made by a vertical line through P. The region $\mathscr{R}$ is described as follows: r goes from 0 to $a\cos(\theta)$ for each θ in $[-\pi/2, \pi/2]$, that is,

$$-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}, \qquad 0 \le r \le a\cos(\theta).$$

We have to express $f(P)$ in polar coordinates. Figure 17.3.6(b) shows the top half of a ball of radius a. By the Pythagorean Theorem,

$$r^2 + (f(r,\theta))^2 = a^2.$$

Thus, because $f(r,\theta) \ge 0$,

$$f(r,\theta) = \sqrt{a^2 - r^2}.$$

Consequently,

$$\text{Volume} = \int_{\mathscr{R}} f(P)\, dA = \int_{-\pi/2}^{\pi/2} \left(\int_0^{a\cos(\theta)} \sqrt{a^2 - r^2}\; r\, dr \right) d\theta.$$

Exploiting symmetry, compute half the volume, keeping θ in $[0, \pi/2]$, and then double the result:

$$\begin{aligned}
\int_0^{a\cos(\theta)} \sqrt{a^2 - r^2}\; r\, dr &= \left.\frac{-(a^2 - r^2)^{3/2}}{3}\right|_0^{a\cos(\theta)} = -\left(\frac{(a^2 - a^2\cos^2(\theta))^{3/2}}{3} - \frac{(a^2)^{3/2}}{3} \right) \\
&= \frac{a^3}{3} - \frac{(a^2 - a^2\cos^2(\theta))^{3/2}}{3} = \frac{a^3}{3} - \frac{a^3(1 - \cos^2(\theta))^{3/2}}{3} \\
&= \frac{a^3}{3}(1 - \sin^3(\theta)).
\end{aligned}$$

NOTE: The trigonometric formula $\sin(\theta) = \sqrt{1 - \cos^2(\theta)}$ is true if $0 \le \theta \le \frac{\pi}{2}$, but not if $-\frac{\pi}{2} \le \theta < 0$.

Then comes the second integration:

$$\begin{aligned}
\int_0^{\pi/2} \frac{a^3}{3}\left(1 - \sin^3(\theta)\right) d\theta &= \frac{a^3}{3} \int_0^{\pi/2} \left(1 - \left(1 - \cos^2(\theta)\right)\sin(\theta)\right) d\theta \\
&= \frac{a^3}{3} \int_0^{\pi/2} \left(1 - \sin(\theta) + \cos^2(\theta)\sin(\theta)\right) d\theta \\
&= \frac{a^3}{3} \left.\left(\theta + \cos(\theta) - \frac{\cos^3(\theta)}{3}\right)\right|_0^{\pi/2} \\
&= \frac{a^3}{3}\left(\frac{\pi}{2} - \left(1 - \frac{1}{3}\right)\right) = a^3\left(\frac{3\pi - 4}{18}\right).
\end{aligned}$$

The total volume is twice as large,

$$V = \int_{-\pi/2}^{\pi/2} \int_0^{a\cos(\theta)} \sqrt{a^2 - r^2}\; r\, dr\, d\theta = 2\int_0^{\pi/2} \int_0^{a\cos(\theta)} \sqrt{a^2 - r^2}\; r\, dr\, d\theta = 2a^3\left(\frac{3\pi - 4}{18}\right) = a^3\left(\frac{3\pi - 4}{9}\right).$$

♦

EXAMPLE 6. A circular disk of radius a is formed of a material that has density $\sigma(P)$ at each point P equal to the distance from P to the disk's center.

(a) Set up an iterated integral in rectangular coordinates for the total mass of the disk.
(b) Set up an iterated integral in polar coordinates for the total mass of the disk.
(c) Compute the easier one.

SOLUTION The disk is shown in Figure 17.3.7.

(a) In rectangular coordinates, the density $\sigma(P)$ at the point $P = (x, y)$ is $\sqrt{x^2 + y^2}$.

The disk has the description

$$-a \le x \le a, \qquad -\sqrt{a^2 - x^2} \le y \le \sqrt{a^2 - x^2}.$$

Thus

$$\text{Mass} = \int_{\mathscr{R}} \sigma(P)\, dA = \int_{-a}^{a} \int_{-\sqrt{a^2-x^2}}^{\sqrt{a^2-x^2}} \sqrt{x^2 + y^2}\, dy\, dx.$$

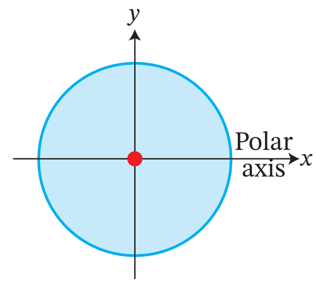


Figure 17.3.7

(b) In polar coordinates, the density $\sigma(P)$ at $P = (r, \theta)$ is r. The disk has the description

$$0 \le \theta \le 2\pi, \qquad 0 \le r \le a.$$

Thus

$$\text{Mass} = \int_{\mathscr{R}} \sigma(P)\, dA = \int_0^{2\pi} \int_0^a r \cdot r\, dr\, d\theta = \int_0^{2\pi} \int_0^a r^2\, dr\, d\theta.$$

(c) Even the first integration in the iterated integral in (a) would be difficult. However, the iterated integral in (b) is straightforward: The first integration (with respect to r) gives

$$\int_0^a r^2\, dr = \left.\frac{r^3}{3}\right|_0^a = \frac{a^3}{3}.$$

And, the second integration (with respect to θ) gives

$$\int_0^{2\pi} \frac{a^3}{3}\, d\theta = \left.\frac{a^3\theta}{3}\right|_0^{2\pi} = \frac{2\pi a^3}{3}.$$

The total mass is $\frac{2\pi}{3} a^3$. ◆

A More Complete Explanation of the Extra r in the Integrand

An estimate of the double integral $\int_{\mathscr{R}} f(P)\, dA$ for a region $\mathscr{R}$ in the xy-plane bounded by the circles $r = a$ and $r = b$ and the rays $\theta = \alpha$ and $\theta = \beta$, divide the region into n^2 pieces with the aid of the partition $r_0 = a, r_1, \dots, r_j, \dots, r_n = b$ and $\theta_0 = \alpha, \theta_1, \dots, \theta_i, \dots, \theta_n = \beta$. For convenience, assume that all $r_j - r_{j-1}$ are equal to Δr and all $\theta_i - \theta_{i-1}$ are equal to $\Delta\theta$. (See Figure 17.3.8(a).)

A typical piece, shown in Figure 17.3.8(b), has area exactly

$$A_{ij} = \frac{(r_j + r_{j-1})}{2}(r_j - r_{j-1})(\theta_i - \theta_{i-1}),$$

as shown in Exercise 30.

Let $P_{ij} = \big((r_j + r_{j-1})/2, (\theta_i + \theta_{i-1})/2\big)$. Then the sum of the n^2 terms of the form $f(P_{ij})A_{ij}$ is an estimate of $\int_{\mathscr{R}} f(P)\, dA$:

$$\int_{\mathscr{R}} f(P)\, dA \approx \sum_{i=1}^{n} \sum_{j=1}^{n} f(P_{ij}) A_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} f\left(\frac{r_j + r_{j-1}}{2}, \frac{\theta_i + \theta_{i-1}}{2}\right) \frac{r_j + r_{j-1}}{2} \Delta r\, \Delta\theta.$$

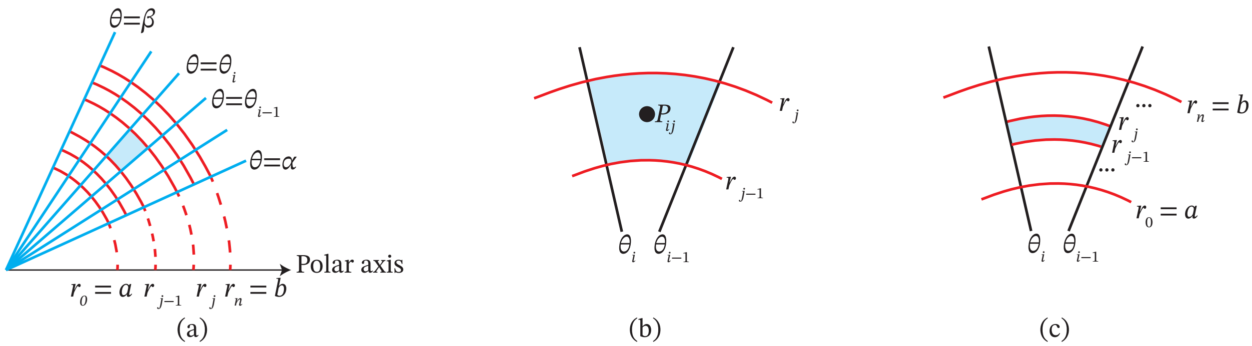


Figure 17.3.8

Let us look closely at the sum of the n pieces between the rays $\theta = \theta_{i-1}$ and $\theta = \theta_i$, as shown in Figure 17.3.8(c):

$$\sum_{j=1}^{n} f\left(\frac{r_j + r_{j-1}}{2}, \frac{\theta_i + \theta_{i-1}}{2}\right) \frac{r_j + r_{j-1}}{2} \Delta r\, \Delta\theta. \tag{17.3.1}$$

In (17.3.1), θ_i, θ_{i-1}, Δr, and $\Delta\theta$ are constants. If we define $g(r,\theta)$ to be $f(r,\theta)r$, then the sum in (17.3.1) is

$$\left(\sum_{j=1}^{n} g\left(\frac{r_j + r_{j+1}}{2}, \frac{\theta_i + \theta_{i-1}}{2}\right) \Delta r\right) \Delta\theta. \tag{17.3.2}$$

Observe that the sum in parentheses in (17.3.2) is an estimate of

$$\int_a^b g\left(r, \frac{\theta_i + \theta_{i-1}}{2}\right) dr.$$

Thus, the original double integral is now approximated by

$$\int_{\mathscr{R}} f(P)\, dA \approx \sum_{i=1}^{n} \sum_{j=1}^{n} f\left(\frac{r_j + r_{j-1}}{2}, \frac{\theta_i + \theta_{i-1}}{2}\right) \frac{r_j + r_{j-1}}{2} \Delta r\, \Delta\theta \approx \sum_{i=1}^{n} \left(\int_a^b g\left(r, \frac{\theta_i + \theta_{i-1}}{2}\right) dr\right) \Delta\theta. \tag{17.3.3}$$

For each $\alpha \le \theta \le \beta$, define $h(\theta) = \int_a^b g(r,\theta)dr$. Then the right-hand side of (17.3.3) simplifies to

$$\sum_{i=1}^{n} h\left(\frac{\theta_i + \theta_{i-1}}{2}\right) \Delta\theta.$$

This sum is an estimate of $\int_a^b h(\theta)d\theta$. Hence

$$\int_{\mathscr{R}} f(r,\theta)\, dA \approx \int_\alpha^\beta h(\theta)\, d\theta \approx \int_\alpha^\beta \int_a^b g(r,\theta)\, dr\, d\theta \approx \int_\alpha^\beta \int_a^b f(r,\theta)\, r\, dr\, d\theta. \tag{17.3.4}$$

The extra factor r appears when we obtained the first integral, $\int_a^b f(r,\theta) r\, dr$. The sum of the n^2 terms of the form $f(P_{ij})A_{ij}$, which we knew approximated the double integral $\int_{\mathscr{R}} f(P)\, dA$, we now see also approximates the iterated integral (17.3.4). Taking limits as $n \to \infty$ shows that approximations become equalities and the iterated integral equals the double integral, with an extra factor r in its integrand.

Summary

We saw how to calculate an integral $\int_{\mathscr{R}} f(P)\,dA$ by introducing polar coordinates. To do this, the plane region $\mathscr{R}$ is described in polar coordinates as

$$\alpha \le \theta \le \beta, \qquad r_1(\theta) \le r \le r_2(\theta).$$

Then

$$\int_{\mathscr{R}} f(P)\,dA = \int_{\alpha}^{\beta} \int_{r_1(\theta)}^{r_2(\theta)} f(r,\theta) r\,dr\,d\theta.$$

The extra factor of r in the integrand is due to the fact that a small region corresponding to changes dr and $d\theta$ has area approximately $r\,dr\,d\theta$ (not $dr\,d\theta$). Polar coordinates are convenient when either the function f or the region $\mathscr{R}$ has a simple description in terms of r and θ.

EXERCISES for Section 17.3

In Exercises 1 to 6 draw and describe the regions in the form $\alpha \le \theta \le \beta$, $r_1(\theta) \le r \le r_2(\theta)$.

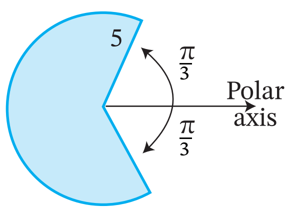


Figure 17.3.9

1. The region inside the curve $r = 3 + \cos(\theta)$.

2. The region between the curve $r = 3 + \cos(\theta)$ and the curve $r = 1 + \sin(\theta)$.

3. The triangle whose vertices have rectangular coordinates $(0,0),(1,1)$, and $(1,\sqrt{3})$.

4. The circle bounded by the curve $r = 3\sin(\theta)$.

5. The sector of the disk shown in Figure 17.3.9.

6. The region in the loop of the three-leaved rose, $r = \sin(3\theta)$, that lies in the first quadrant.

---

7. (a) Draw the region $\mathscr{R}$ bounded by the lines $y = 1$, $y = 2$, $y = x$, $y = \dfrac{x}{\sqrt{3}}$.
(b) Describe $\mathscr{R}$ in terms of horizontal cross sections.
(c) Describe $\mathscr{R}$ in terms of vertical cross sections.
(d) Describe $\mathscr{R}$ in terms of cross sections by polar rays.

8. (a) Draw the region $\mathscr{R}$ whose description is $-2 \le y \le 2$, $-\sqrt{4-y^2} \le x \le \sqrt{4-y^2}$.
(b) Describe $\mathscr{R}$ by vertical cross sections.
(c) Describe $\mathscr{R}$ by cross sections formed by using polar rays.

9. Describe in polar coordinates the square whose vertices have rectangular coordinates $(0,0)$, $(1,0)$, $(1,1)$, $(0,1)$.

10. Describe the trapezoid whose vertices have rectangular coordinates $(0,1)$, $(1,1)$, $(2,2)$, $(0,2)$ (a) by horizontal cross sections, (b) by vertical cross sections, and (c) in polar coordinates.

In Exercises 11 to 14 draw the region $\mathscr{R}$ and evaluate $\int_{\mathscr{R}} r^2\,dA$.

11. $-\dfrac{\pi}{2} \le \theta \le \dfrac{\pi}{2}$, $0 \le r \le \cos(\theta)$

12. $0 \le \theta \le \dfrac{\pi}{2}$, $0 \le r \le \sin^2(\theta)$

13. $0 \le \theta \le 2\pi$, $0 \le r \le 1 + \cos(\theta)$

14. $0 \le \theta \le \dfrac{\pi}{6}$, $0 \le r \le \sin(2\theta)$

In Exercises 15 to 18 draw the given region $\mathscr{R}$ and evaluate $\int_{\mathscr{R}} y^2\, dA$.

15. The disk of radius a, center at the pole.

16. The disk of radius a with center at $(a, 0)$ in polar coordinates.

17. The region within the cardioid $r = 1 + \sin(\theta)$.

18. The region within the leaf that is symmetric with respect to the ray $\theta = \pi/4$ of the four-leaved rose $r = \sin(2\theta)$.

The average of a function $f(P)$ over a region $\mathscr{R}$ in the plane is defined as $\int_{\mathscr{R}} f(P)\, dA$ divided by the area of $\mathscr{R}$. In Exercises 19 to 22, find this average.

19. $f(P)$ is the distance from P to the pole; $\mathscr{R}$ is one leaf of the three-leaved rose, $r = \sin(3\theta)$.

20. $f(P)$ is the distance from P to the x-axis; $\mathscr{R}$ is the region between the rays $\theta = \frac{\pi}{6}$ and $\theta = \frac{\pi}{4}$, and between the circles $r = 2$ and $r = 3$.

21. $f(P)$ is the distance from P to a fixed point on the border of a disk $\mathscr{R}$ of radius a.

22. $f(P)$ is the distance from P to the x-axis; $\mathscr{R}$ is the region within the cardioid $r = 1 + \cos(\theta)$.

In Exercises 23 to 26 evaluate the iterated integral using polar coordinates.

23. $\int_0^1 \int_0^x \sqrt{x^2 + y^2}\, dy\, dx$

24. $\int_0^1 \int_0^{\sqrt{1-x^2}} x^3\, dy\, dx$

25. $\int_0^{1/\sqrt{2}} \int_x^{\sqrt{1-x^2}} xy\, dy\, dx$

26. $\int_1^2 \int_{x/\sqrt{3}}^{\sqrt{3}x} (x^2 + y^2)^{3/2}\, dy\, dx$

---

27. Evaluate:

(a) $\int_{\mathscr{R}} \cos(x^2 + y^2)\, dA$ where $\mathscr{R}$ is the portion in the first quadrant of the disk of radius a centered at the origin.

(b) $\int_{\mathscr{R}} \sqrt{x^2 + y^2}\, dA$ where $\mathscr{R}$ is the triangle bounded by the line $y = x$, the line $x = 2$, and the x-axis.

28. Find the volume of the region above the paraboloid $z = x^2 + y^2$ and below the plane $z = x + y$.

29. The area of a region $\mathscr{R}$ is given by the double integral $\int_{\mathscr{R}} 1\, dA$. Use an appropriate description of $\mathscr{R}$ to find the area of a disk of radius a.

30. Find the area of the shaded region in Figure 17.3.4(b) as follows:

(a) Find the area of the ring between two circles, one of radius r_0, the other of radius $r_0 + \Delta r$.

(b) What fraction of the area in (a) is included between two rays whose angles differ by $\Delta\theta$?

(c) Show that the area of the shaded region in Figure 17.3.4(b) is $\left(r_0 + \frac{\Delta r}{2}\right) \Delta r\, \Delta\theta$.

31. In Example 5 we computed half of $\int_{-\pi/2}^{\pi/2} \int_0^{a\cos(\theta)} \sqrt{a^2 - r^2}\, r\, dr\, d\theta$ and doubled the result. Evaluate this iterated integral directly, without using symmetry. Is your result $\frac{3\pi - 4}{9} a^3$? CAUTION: Use trigonometric formulas with care.

Transportation problems can lead to integrals over plane sets, as Exercises 32 to 35 illustrate.

32. Show that the average distance from the center of a disk of area A to points in the disk is $\frac{2\sqrt{A}}{3\sqrt{\pi}} \approx 0.376\sqrt{A}$.

33. Show that the average distance from the center of a regular hexagon of area A to points in the hexagon is $\frac{\sqrt{2A}}{3^{3/4}}\left(\frac{1}{3} + \frac{\ln(3)}{4}\right) \approx 0.377\sqrt{A}$.

34. Show that the average distance from the center of a square of area A to points in the square is $\frac{\sqrt{A}}{6}\left(\sqrt{2} + \ln\left(\sqrt{2}+1\right)\right) \approx 0.383\sqrt{A}$.

35. Show that the average distance from the centroid of an equilateral triangle of area A to points in the triangle is $\frac{\sqrt{A}}{3^{9/4}}\left(2\sqrt{3} + \ln\left(\sqrt{3}+2\right)\right) \approx 0.404\sqrt{A}$. RECALL: A triangle's centroid is the intersection point of its medians.

In Exercises 32 to 35 distance is the ordinary straight-line distance. In cities the usual street pattern suggests that the **metropolitan distance** between the points (x_1, y_1) and (x_2, y_2) should be measured by $|x_1 - x_2| + |y_1 - y_2|$. Metropolitan distance is used in Exercises 36 and 37.

36. Show that if in Exercise 32 metropolitan distance is used, then the average is $\frac{8\sqrt{A}}{3\pi^{3/2}} \approx 0.479\sqrt{A}$.

37. Show that if in Exercise 34 metropolitan distance is used, then the average is $\frac{\sqrt{A}}{2}$.

NOTE: In most cities the metropolitan distance average tends to be about 25 percent larger than the direct-distance average.

Exercise 38 shows that $\int_0^{\infty} e^{-x^2}\, dx = \frac{\sqrt{\pi}}{2}$, an equation which connects the constants e and π and plays a key role in statistics. In spite of what is claimed in the following comment it is not as obvious to a mathematician as "twice two is four". Before beginning this exercise, read the following two historical quotes concerning this improper integral.

S. P. Thompson, in *Life of Lord Kelvin* (Macmillan, London, 1910), wrote:

> "Once when lecturing to a class he [the physicist Lord Kelvin] used the word 'mathematician' and then interrupting himself asked the class: 'Do you know what a mathematician is?' Stepping to his blackboard he wrote upon it: $\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$. Then putting his finger on what he had written, he turned to his class and said, 'A mathematician is one to whom this is as obvious as that twice two makes four is to you.' "

On the other hand, the mathematician Littlewood wrote in "Newton and the Attraction of the Sphere," *Mathematical Gazette*, vol. 63, 1948:

"Many things are not accessible to intuition at all, the value of $\int_0^\infty e^{-x^2}\,dx$ for instance."

38. Let $f(P) = e^{-r^2}$ where r is the distance from P to the origin. Hence, $f(r,\theta) = e^{-r^2}$ in polar coordinates and, in rectangular coordinates, $f(x, y) = e^{-x^2-y^2}$. Note that, as shown in Figure 17.3.10, $\mathscr{R}_1$ is inside $\mathscr{R}_2$ and $\mathscr{R}_2$ is inside $\mathscr{R}_3$.

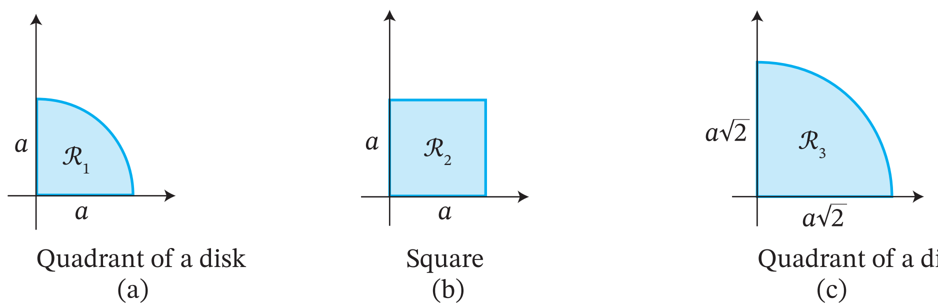


Figure 17.3.10

(a) Show that $\int_{\mathscr{R}_1} f(P)\,dA = \frac{\pi}{4}\left(1-e^{-a^2}\right)$.

(b) Show that $\int_{\mathscr{R}_3} f(P)\,dA = \frac{\pi}{4}\left(1-e^{-2a^2}\right)$.

(c) By considering $\int_{\mathscr{R}_2} f(P)\,dA$ and the results in (a) and (b), show that $\frac{\pi}{4}\left(1-e^{-a^2}\right) < \left(\int_0^a e^{-x^2}\,dx\right)^2 < \frac{\pi}{4}\left(1-e^{-2a^2}\right)$.

(d) Show that $\int_0^\infty e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}$.

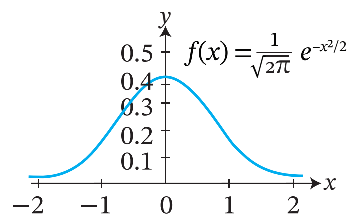


Figure 17.3.11

39. Figure 17.3.11 shows the **bell curve** or **normal curve**. Show that the area under it is 1. Use the information that $\int_0^\infty e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}$, established in Exercise 38.

---

40. **Sam:** The formula in this section for integrating in polar coordinates is wrong. I'll find the right formula. We don't need the factor r.

Jane: But the book's formula gives correct answers.

Sam: I don't care. Let $f(r,\theta)$ be positive and I'll show how to integrate over the set $\mathscr{R}$ bounded by $r = b$ and $r = a$, $b > a$, and $\theta = \beta$ and $\theta = \alpha$, $\beta > \alpha$. We have $\int_{\mathscr{R}} f(P)\,dA$ is the volume under the graph of f and above $\mathscr{R}$. Right?

Jane: Right.

Sam: The area of the cross section corresponding to a fixed angle θ is $\int_a^b f(r,\theta)\,dr$. Right?

Jane: Right.

Sam: So I just integrate cross-sectional areas as θ goes from α to β, and the volume is $\int_\alpha^\beta \int_a^b f(r,\theta)\,dr\,d\theta$. Perfectly straightforward. I hate to debunk a formula that's been known for three centuries.

What does Jane say next?

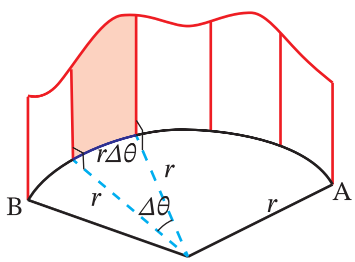
41. **JANE:** You looked at fixed θ. I'll use a fixed r. Look at the area under the graph of f and above the circle of radius r. I'll draw this fence for you (see Figure 17.3.12(a)).
To estimate its area I'll cut the arc $\widehat{AB}$ into n sections of equal length by angle $\theta_0 = \alpha, \dots, \theta_n = \beta$. Each short arc has length $r\ \Delta\theta$. (Remember, Sam, how radians are defined.)
The approximation to the shaded area looks like Figure 17.3.12(b) and resembles a rectangle of height $f(r,\theta)$ and width $r\ \Delta\theta$. So the local approximation to the area is $f(r,\theta)\ r\ \Delta\theta$ and the area of the fence is $\int_\alpha^\beta f(r,\theta)\ r\ \ d\theta$. Here r is fixed. Then I integrate this cross-sectional area as r goes from a to b. The total volume is then $\int_a^b \int_\alpha^\beta f(r,\theta)\ r\ d\theta\ dr$. That gives the volume, which equals $\int_{\mathscr{R}} f(r,\theta)\ dA$.

SAM: All right.

JANE: At least it gives the factor r.

SAM: Maybe we're both right.

What does Jane say?



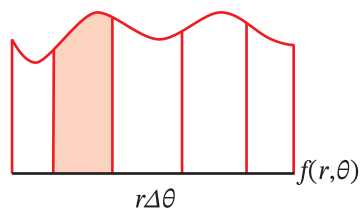


Figure 17.3.12

17.4 The Triple Integral: Integrals Over Solid Regions

In this section we define integrals over solid regions in space and show how to compute them by iterated integrals using rectangular coordinates. Throughout we assume the regions are bounded by smooth surfaces and the functions are continuous.

The Triple Integral

Let $\mathscr{R}$ be a region in space bounded by a surface and f a continuous function on $\mathscr{R}$. For instance, $\mathscr{R}$ could be a ball, a cube, or a tetrahedron. Partition $\mathscr{R}$ into n smaller regions $\mathscr{R}_1, \mathscr{R}_2, \dots, \mathscr{R}_n$ with volumes $V_1, V_2, \dots, V_n$ and sampling points $P_1, P_2, \dots, P_n$. The limit of sums of the form

$$\sum_{i=1}^{n} f(P_i)V_i,$$

as the diameter of each $\mathscr{R}_i$ approaches 0, exists (no matter how the sampling points P_i are chosen). The limiting value is called the integral of f over $\mathscr{R}$ or the **triple integral** of f over $\mathscr{R}$ and is denoted

$$\int_{\mathscr{R}} f(P)\ dV. \tag{17.4.1}$$

Double and triple integrals are both called **multiple integrals**. More generally, a multiple integral of a function f of n variables over an n-dimensional region $\mathcal{R}$ is the limit of Riemann sums of f over partitions of $\mathcal{R}$ as the diameter of the largest region in the partition approaches 0.

EXAMPLE 1. If $f(P) = 1$ for each point P in a solid region $\mathcal{R}$, compute $\int_{\mathcal{R}} f(P)\, dV$.

SOLUTION From the definition of a triple integral, with $f(P) = 1$ for all points P in $\mathcal{R}$,

$$\int_{\mathcal{R}} 1\, dV = \lim_{n\to\infty} \sum_{i=1}^{n} f(P_i)V_i.$$

where the V_i and P_i are as defined at the beginning of this section. But, for each positive integer n, the sum $\sum_{i=1}^{n} V_i$ has the same value

$$\sum_{i=1}^{n} 1\cdot V_i = V_1 + V_2 + \cdots + V_n = \text{Volume of } \mathcal{R}.$$

Combined, these two facts provide a formula for the volume of $\mathcal{R}$:

$$\int_{\mathcal{R}} 1\, dV = \lim_{n\to\infty} \sum_{i=1}^{n} f(P_i)V_i = \text{Volume of } \mathcal{R}.$$

◆

The **average value** of a function f defined on a region $\mathcal{R}$ in space is defined analogously to the definition in the plane:

$$\text{Average value of } f \text{ on } \mathcal{R} = \int_{\mathcal{R}} f(P)\, dV / (\text{Volume of } \mathcal{R}).$$

This is the analog of the definition of the average value of a function over an interval (Section 6.3) or the average value of a function over a plane region (Section 17.1).

If a mass is distributed in a region $\mathcal{R}$ its density at a point P is defined as a limit. For positive r let $V(r)$ be the volume of a ball of radius r centered at P, and $m(r)$ the mass in it. Then the density at P is

$$\lim_{r\to 0} \frac{m(r)}{V(r)}.$$

If f describes the density of matter in $\mathcal{R}$, then the average value of f is the density of a homogeneous solid occupying $\mathcal{R}$ and having the same total mass as the given solid. That is, if the average density, $\int_{\mathcal{R}} f(P)\, dV/(\text{Volume of } \mathcal{R})$, is multiplied by the volume of $\mathcal{R}$, the product is the total mass:

$$\text{Total mass} = \int_{\mathcal{R}} f(P)\, dV.$$

Describing a Solid Region in Space

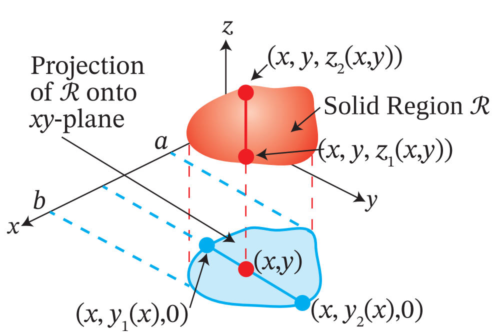


Figure 17.4.1

In order to evaluate triple integrals, it is necessary to describe solid regions in terms of an appropriate system of coordinates. A description of a solid region in rectangular coordinates has the form

$$a \le x \le b, \qquad y_1(x) \le y \le y_2(x), \qquad z_1(x,y) \le z \le z_2(x,y).$$

The inequalities on x and y describe the projection of the three-dimensional region onto the xy-plane. The inequalities for z then tell how z varies on a line parallel to the z-axis and passing through the point (x, y) in the projection. (See Figure 17.4.1.)

EXAMPLE 2. Describe in terms of x, y, and z the rectangular box shown in Figure 17.4.2(a).

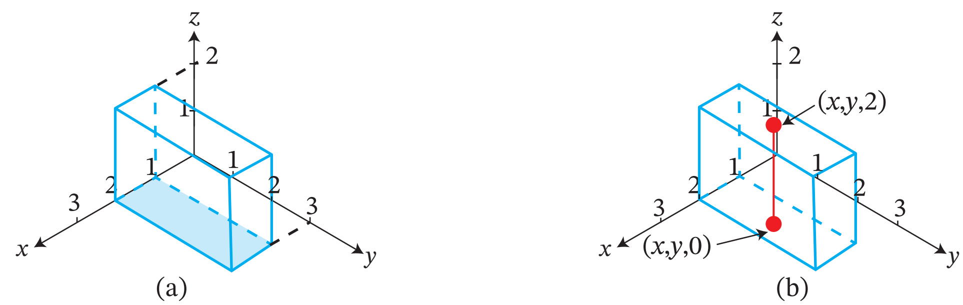


Figure 17.4.2

SOLUTION The projection of the box on the xy-plane has a description $1 \le x \le 2$, $0 \le y \le 3$. For each point in the (rectangular) projection on the xy-plane, z varies from 0 to 2, as shown in Figure 17.4.2(b). So the description of the box is

$$1 \le x \le 2, \qquad 0 \le y \le 3, \qquad 0 \le z \le 2,$$

which is read from left to right as "x goes from 1 to 2; for each x, y goes from 0 to 3; for each x and y, z goes from 0 to 2." ♦

Since both bounds on each of the three variables are constants, we could have changed the order of x and y in the description or projected the box on one of the other two coordinate planes. There are six possible descriptions.

When there are choices for the order of the variables in the description of a region, sometimes one order will be more advantageous for the calculations that need to be performed. In fact, we have seen this in Example 7 in Section 17.2 and in Example 6 in Section 17.3.

EXAMPLE 3. Describe in rectangular coordinates the tetrahedron bounded by the planes $x = 0$, $y = 0$, $z = 0$, and $x + y + z = 1$, as shown in Figure 17.4.3(a).

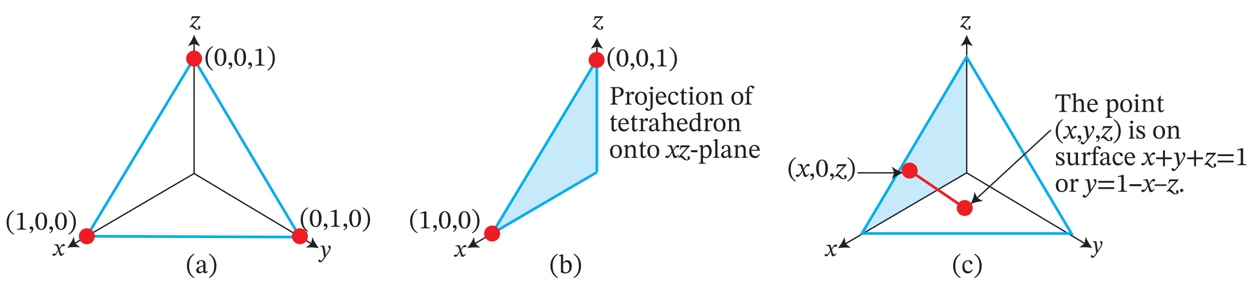


Figure 17.4.3

SOLUTION Project the tetrahedron onto the xz-plane. The projection is shown in Figure 17.4.3(b). A description of this triangular "shadow" in the xz-plane is

$$0 \le x \le 1, \qquad 0 \le z \le 1 - x,$$

since the slanted edge has the equation $x + z = 1$. For each point (x, z) in the projection, y ranges from 0 out to the value of y that satisfies the equation $x + y + z = 1$, that is, out to $y = 1 - x - z$. (See Figure 17.4.3(c).) A description of the tetrahedron is

$$0 \le x \le 1, \qquad 0 \le z \le 1 - x, \qquad 0 \le y \le 1 - x - z.$$

That is, x goes from 0 to 1; for each x, z goes from 0 to $1-x$; for each x and z, y goes from 0 to $1-x-z$. ♦

EXAMPLE 4. Describe in rectangular coordinates the ball of radius 4 whose center is at the origin.

SOLUTION The projection of the ball on the xy-plane is the disk of radius 4 and center $(0,0)$. Its description is

$$-4 \le x \le 4, \qquad -\sqrt{16-x^2} \le y \le \sqrt{16-x^2}.$$

Hold (x, y) fixed in the xy-plane and consider the way z varies on the line parallel to the z-axis that passes through the point $(x, y, 0)$. Since the surface of the ball has the equation $x^2+y^2+z^2=16$, for each (x, y), z varies from

$$-\sqrt{16-x^2-y^2} \qquad \text{to} \qquad \sqrt{16-x^2-y^2}.$$

This describes the line segment shown in Figure 17.4.4.

The ball, therefore, has a description

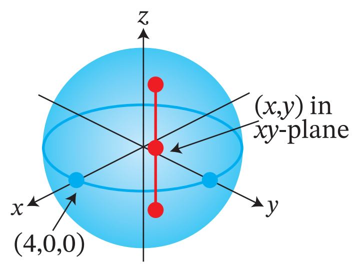


Figure 17.4.4

$$-4 \le x \le 4, \quad -\sqrt{16-x^2} \le y \le \sqrt{16-x^2}, \quad \sqrt{16-x^2-y^2} \le z \le \sqrt{16-x^2-y^2}.$$

♦

Evaluating a Triple Integral as an Iterated Integral

The iterated integral in rectangular coordinates for $\int_{\mathscr{R}} f(P)\, dV$ is similar to that for evaluating double integrals over plane sets. While triple integrals involve three integrations instead of two integrations for double integrals, the limits of integration are still determined by the description of the solid region $\mathscr{R}$. If $\mathscr{R}$ has the description

$$a \le x \le b, \qquad y_1(x) \le y \le y_2(x), \qquad z_1(x,y) \le z \le z_2(x,y),$$

then

$$\int_{\mathscr{R}} f(P)\, dV = \int_a^b \int_{y_1(x)}^{y_2(x)} \int_{z_1(x,y)}^{z_2(x,y)} f(x,y,z)\, dz\, dy\, dx.$$

An example illustrates how this formula is applied. In Exercise 31 an argument for its plausibility is presented.

EXAMPLE 5. Compute $\displaystyle\int_{\mathscr{R}} z\, dV$, where $\mathscr{R}$ is the tetrahedron in Example 3.

SOLUTION One description of the tetrahedron is $0 \le y \le 1$, $0 \le x \le 1-y$, $0 \le z \le 1-x-y$. Hence

$$\int_{\mathscr{R}} z\, dV = \int_0^1 \int_0^{1-y} \int_0^{1-x-y} z\, dz\, dx\, dy.$$

Compute the inner integral first, treating x and y as constants. By the fundamental theorem of calculus,

$$\begin{aligned}\int_0^{1-x-y} z\, dz &= \left.\frac{z^2}{2}\right|_{z=0}^{z=1-x-y} \\ &= \frac{(1-x-y)^2}{2}.\end{aligned}$$

The next integration, where y is fixed, is

$$\begin{aligned}\int_0^{1-y} \frac{(1-x-y)^2}{2}\,dx &= -\frac{(1-x-y)^3}{6}\bigg|_{x=0}^{x=1-y}\\ &= -\frac{0^3}{6}+\frac{(1-y)^3}{6}\\ &= \frac{(1-y)^3}{6}.\end{aligned}$$

The third integration is

$$\begin{aligned}\int_0^{1} \frac{(1-y)^3}{6}\,dy &= -\frac{(1-y)^4}{24}\bigg|_0^1\\ &= -\frac{0^4}{24}+\frac{1^4}{24}\\ &= \frac{1}{24}.\end{aligned}$$

♦

Interpreting Triple Integrals in the Physical World

Triple integrals appear in the study of gravitation, rotating bodies, centers of mass, and in electromagnetic theory, among others. The simplest way to think of them is to interpret $f(P)$ as the density at P of some distribution of matter. Then $\int_{\mathscr{R}} f(P)\,dV$ is the total mass in a region $\mathscr{R}$.

The single integral $\int_a^b 1\,dx$ gives the length of the interval $[a,b]$. Moving up to two dimensions, area can be found either as $\int_a^b f(x)\,dx$ where $f(x)$ is the height of a cross section or as a double integral $\int_{\mathscr{R}} 1\,dA$. And, in three dimensions, the volume of a three-dimensional solid region, $\mathscr{R}$, can be found as either as a double integral $\int_{\mathscr{S}} f(P)\,dA$ where $f(P)$ is the length of a cross section for each point in the region of integration, $\mathscr{S}$, or as a triple integral $\int_{\mathscr{R}} 1\,dV$. In four dimensions, if the integrand $f(P)$ is interpreted as the length of the "cross section" at each point of the region of integration, cR, then we are describing an object in four-dimensional space.

Note that $\int_a^b 1\,dx$ has the dimensions of length, L. The two integrals for area each have dimensions L^2. And, both of the integrals for volume have the dimension of L^3. The triple integral $\int_{\mathscr{R}} f(P)\,dV$ has dimensions L^4; in some contexts this will be interpreted as a **hypervolume**.

Observation 17.4.1: ***A Word about Four-Dimensional Space***

We can think of two-dimensional space (2-space) as the set of ordered pairs (x,y) of real numbers. The set of ordered triplets of real numbers (x,y,z) represents 3-space. The set of ordered quadruplets of real numbers (x,y,z,t) represents 4-space.

In 2-space the set of points of the form $(x,0)$, the x-axis, meets the set of points of the form $(0,y)$, the y-axis, in a point, namely the origin $(0,0)$. In 4-space the set of points of the form $(x,y,0,0)$ forms a plane congruent to the xy-plane. The set of points of the form $(0,0,z,t)$ forms another two-dimensional plane. Their intersection is a single point $(0,0,0,0)$. Can you picture two endless planes meeting in a single point? If so, please tell us how.

While visualizing objects in four (or more) dimensions is not natural, it is sometimes convenient to think in higher dimensional spaces. Two examples, in 4-space, are using the gradient vector to determine the normal vector to the tangent hyperplane at a point on a four-dimensional hypersurface and using a quadruple integral to find the hypervolume of a four-dimensional region.

Summary

We defined $\int_{\mathscr{R}} f(P)\,dV$, where $\mathscr{R}$ is a region in space. The volume of a solid region $\mathscr{R}$ is $\int_{\mathscr{R}} 1\,dV$ and, if $f(P)$ is the density of matter at P, then $\int_{\mathscr{R}} f(P)\,dV$ is the total mass of the object.

We also showed how to evaluate triple integrals by introducing rectangular coordinates and computing an iterated integral. There are six possible orders for the three variables x, y, and z. If $\mathscr{R}$ is described by

$$a \le x \le b, \qquad y_1(x) \le y \le y_2(x), \qquad z_1(x,y) \le z \le z_2(x,y).$$

Then

$$\int_{\mathscr{R}} f(P)\,dV = \int_a^b \int_{y_1(x)}^{y_2(x)} \int_{z_1(x,y)}^{z_2(x,y)} f(x,y,z)\,dz\,dy\,dx.$$

To use any of the other five orders requires a corresponding description of $\mathscr{R}$.

EXERCISES for Section 17.4

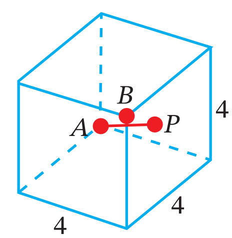
Exercises 1 to 4 concern the definition of $\int_{\mathscr{R}} f(P)\,dV$.

1. A cube of side 4 centimeters is made of a material of varying density. Near one corner, A, it is very light; at the opposite corner it is very dense. The density $f(P)$ (in grams per cubic centimeter) at a point P in the cube is the square of the distance from A to P (in centimeters). See Figure 17.4.5.

(a) Find upper and lower estimates for the mass of the cube by partitioning it into eight cubes.

(b) Using the same partition, estimate the mass of the cube, but select as the P_i's the centers of the cubes.

(c) What does (b) say about the average density in the cube?



Density at P is the square of the distance $|AP|$.
Figure 17.4.5

2. How would you define the average distance from points of a region $\mathscr{R}$ in space to a point P_0?

3. If $\mathscr{R}$ is a ball of radius r and $f(P) = 5$ for each point in $\mathscr{R}$, compute $\int_{\mathscr{R}} f(P)\,dV$ by examining approximating sums. The ball has volume $\frac{4}{3}\pi r^3$.

4. If $\mathscr{R}$ is a three-dimensional set and $f(P)$ is never more than 8 for all P in $\mathscr{R}$,

(a) what can we say about the maximum value of $\int_{\mathscr{R}} f(P)\,dV$?

(b) what can we say about the average of f over $\mathscr{R}$?

In Exercises 5 to 10 draw the solids.

5. $1 \le x \le 3$, $0 \le y \le 2$, $0 \le z \le x$

6. $0 \le x \le 1$, $0 \le y \le 1$, $1 \le z \le 1 + x + y$

7. $0 \le y \le 1$, $0 \le x \le y^2$, $y \le z \le 2y$

8. $0 \le y \le 1$, $y^2 \le x \le y$, $0 \le z \le x + y$

9. $0 \le z \le 3$, $0 \le y \le \sqrt{9 - z^2}$, $0 \le x \le \sqrt{9 - y^2 - z^2}$

10. $-1 \le z \le 1$, $-\sqrt{1 - z^2} \le x \le \sqrt{1 - z^2}$, $-\frac{1}{2} \le y \le \sqrt{1 - x^2 - z^2}$

In Exercises 11 to 14 evaluate the iterated integral.

11. $\int_0^1 \int_0^2 \int_0^x z\, dz\, dy\, dx$

12. $\int_0^1 \int_{x^3}^{x^2} \int_0^{x+y} z\, dz\, dy\, dx$

13. $\int_2^3 \int_x^{2x} \int_0^1 (x+z)\, dz\, dy\, dx$

14. $\int_0^1 \int_0^x \int_0^3 (x^2+y^2)\, dz\, dy\, dx$

---

15. Describe the solid cylinder of radius a and height h shown in Figure 17.4.6(a) in rectangular coordinates (a) in the order x, y, z and (b) in the order x, z, y.

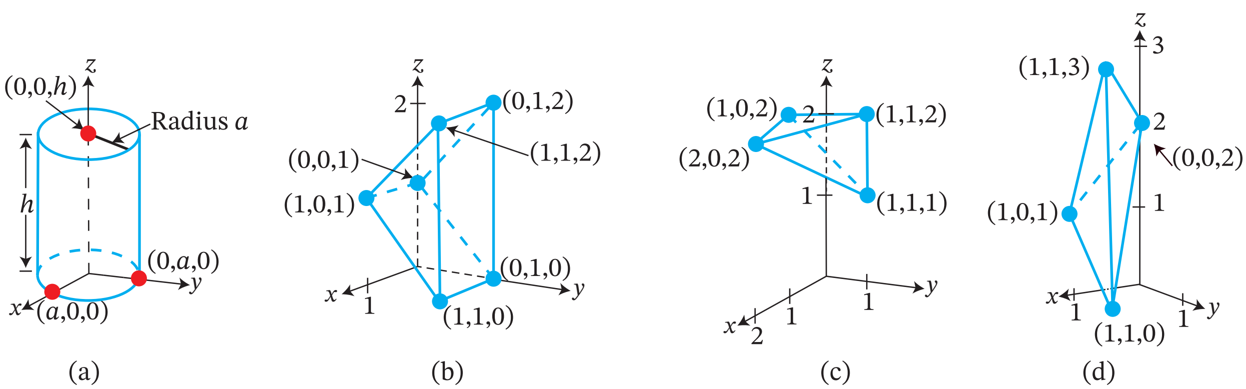


Figure 17.4.6

16. Describe the prism shown in Figure 17.4.6(b) in rectangular coordinates in two ways, (a) first projecting it onto the xy-plane and (b) first projecting it onto the xz-plane.

17. Describe the tetrahedron in Figure 17.4.6(c) in rectangular coordinates in two ways, (a) first projecting it onto the xy-plane and (b) first projecting it onto the xz-plane.

18. Describe the tetrahedron in Figure 17.4.6(d) with vertices at $(1,1,0)$, $(1,0,1)$, $(0,0,2)$, and $(1,1,3)$. (a) Draw its projection on the xy-plane. (b) Obtain equations for its top and bottom planes.

19. Let $\mathscr{R}$ be the tetrahedron whose vertices are $(0,0,0)$, $(a,0,0)$, $(0,b,0)$, and $(0,0,c)$, where a, b, and c are positive. (a) Sketch $\mathscr{R}$. (b) Find the equation of its top surface. (c) Compute $\int_{\mathscr{R}} z\, dV$.

20. Compute $\int_{\mathscr{R}} z\, dV$, where $\mathscr{R}$ is the region above the rectangle whose vertices are $(0,0,0)$, $(2,0,0)$, $(2,3,0)$, and $(0,3,0)$ and below the plane $z = x + 2y$.

21. Find the mass of the cube in Exercise 1. (See Figure 17.4.1.)

22. Find the average value of the square of the distance from a corner of a cube of side a to points in the cube.

23. Find the average of the square of the distance from a point in a cube of side a to the center of the cube.

24. A solid consists of points below the surface $z = xy$ that are above the triangle whose vertices are $(0,0,0)$, $(1,0,0)$, and $(0,2,0)$. If the density at (x,y,z) is $x+y$, find the total mass.

25. Compute $\int_{\mathscr{R}} xy\, dV$ for the tetrahedron of Example 3.

26. (a) Describe in rectangular coordinates the right circular cone $\mathscr{C}$ of radius r and height h if its axis is on the positive z-axis and its vertex is at the origin. Draw the cross sections for fixed x and for fixed x and y.

(b) Find $\int_{\mathscr{C}} z\,dV$.

(c) Find the average value of z in the cone in (a).

27. The temperature at (x,y,z) is e^{-x-y-z}. Find the average temperature in the tetrahedron whose vertices are $(0,0,0)$, $(1,1,0)$, $(0,0,2)$, and $(1,0,0)$.

28. The temperature at (x,y,z), $y>0$, is $\frac{e^{-x}}{\sqrt{y}}$. Find the average temperature in the region bounded by the cylinder $y=x^2$, the plane $y=1$, and the plane $z=2y$.

29. Without using an iterated integral, evaluate $\int_{\mathscr{R}} x\,dV$, where $\mathscr{R}$ is a ball of radius a with center at $(0,0,0)$.

30. The work done in lifting a weight of w pounds a vertical distance of x feet is wx foot-pounds. Imagine that through geological activity a mountain is formed consisting of material originally at sea level. Let the density of the material near point P in the mountain be $g(P)$ pounds per cubic foot and the height of P be $h(P)$ feet. What definite integral represents the total work expended in forming the mountain? This type of problem arises in the geological theory of mountain formation.

31. In Section 17.2 an intuitive argument was presented for the equality $\int_{\mathscr{R}} f(P)\,dA = \int_a^b \int_{y_1(x)}^{y_2(x)} f(x,y)\,dy\,dx$. Here is an intuitive argument for the equality $\int_{\mathscr{R}} f(P)\,dV = \int_{x_1}^{x_2}\int_{y_1(x)}^{y_2(x)}\int_{z_1(x,y)}^{z_2(x,y)} f(x,y,z)\,dz\,dy\,dx$. For this discussion, interpret $f(P)$ as the density at a point P.

(a) Let $\mathscr{R}(x)$ be the plane cross section consisting of all points in $\mathscr{R}$ with abscissa x. Show that the average density in $\mathscr{R}(x)$ is $\frac{1}{\text{Area of }\mathscr{R}(x)}\int_{y_1(x)}^{y_2(x)}\int_{z_1(x,y)}^{z_2(x,y)} f(x,y,z)\,dz\,dy$.

(b) Show that $\int_{y_1(x)}^{y_2(x)}\int_{z_1(x,y)}^{z_2(x,y)} f(x,y,z)\,dz\,dy\,\Delta x$ approximates the mass of $\mathscr{R}$ between the plane sections $\mathscr{R}(x)$ and $\mathscr{R}(x+\Delta x)$.

(c) From (b) obtain an iterated integral in rectangular coordinates for $\int_{\mathscr{R}} f(P)\,dV$.

17.5 Cylindrical and Spherical Coordinates

Rectangular coordinates provide convenient descriptions of solids bounded by planes. In this section we describe two other coordinate systems, cylindrical, ideal for describing circular cylinders, and spherical, ideal for describing spheres and cones. Both will be used in the next section to evaluate multiple integrals by iterated integrals.

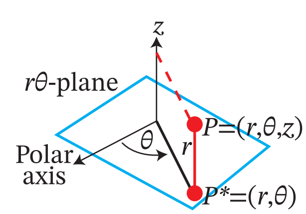


Figure 17.5.1

Cylindrical Coordinates

Cylindrical coordinates combine polar coordinates in the plane with the z of rectangular coordinates in space. A point P in space receives the name (r,θ,z) as in Figure 17.5.1. We are free to choose the direction of the polar axis; usually

it will coincide with the x-axis of an (x, y, z) system. The point (r,θ,z) is directly
above (or below) $P^* = (r,\theta)$ in the $r\theta$-plane. Since the set of points $P = (r,\theta,z)$ for which r is some constant is a circular cylinder, this coordinate system is convenient for describing such cylinders. As with polar coordinates, the cylindrical coordinates of a point are not unique.

Figure 17.5.2 shows the coordinate surfaces $\theta = k$, $r = k$, and $z = k$, respectively,where k is a positive number.

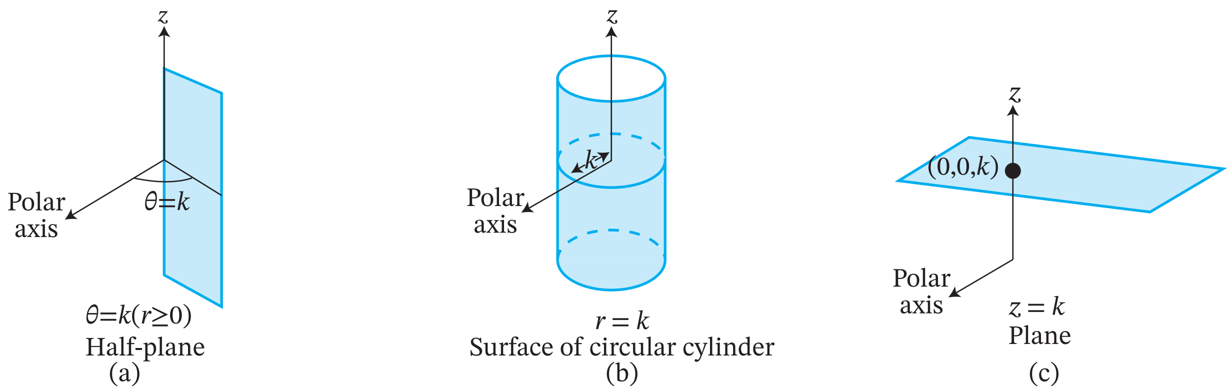


Figure 17.5.2

EXAMPLE 1. Describe a solid cylinder of radius a and height h in cylindrical coordinates. Assume that its axis is on the positive z-axis and its lower base has its center at the pole, as in Figure 17.5.3(a).

SOLUTION The projection of the cylinder on the $r\theta$-plane is the disk of radius a with center at the pole shown in Figure 17.5.3(b). Its description is

$$0 \le \theta \le 2\pi, \qquad 0 \le r \le a.$$

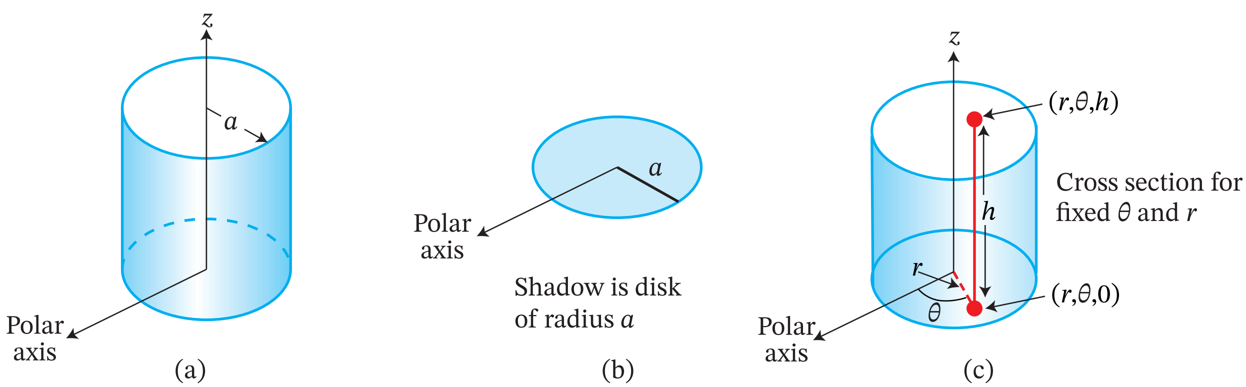


Figure 17.5.3

For each point (r,θ) in the projection, the line through it and parallel to the z-axis intersects the cylinder in a line segment. On each segment z varies from 0 to h. (See Figure 17.5.3(c).) Thus a description of the cylinder is

$$0 \le \theta \le 2\pi, \qquad 0 \le r \le a, \qquad 0 \le z \le h.$$

◆

EXAMPLE 2. Describe in cylindrical coordinates the region in space formed by the intersection of a solid (endless) cylinder of radius 3 with a ball of radius 5 whose center is on the axis of the cylinder. Place the cylindrical coordinate system as shown in Figure 17.5.4.

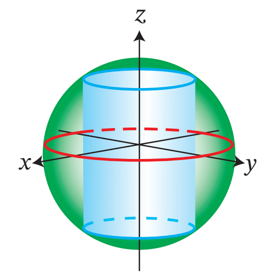


Figure 17.5.4

SOLUTION We start by observing that, by the Pythagorean theorem, $r^2 + z^2 = |OP|^2$. Thus the point $P = (r, \theta, z)$ is a distance $\sqrt{r^2 + z^2}$ from the origin O. We will use this fact in a moment.

Now we describe the solid. The projection of the cylinder on the $r\theta$-plane is a disk of radius 3, which is easily described in polar coordinates. For fixed θ and r, the cross section of the solid is a line segment determined by the sphere that bounds the ball. Because the sphere has radius 5, for any point (r, θ, z) on it

$$r^2 + z^2 = 25 \qquad \text{or} \qquad z = \pm\sqrt{25 - r^2}.$$

On the segment determined by r and θ, z varies from $-\sqrt{25 - r^2}$ to $\sqrt{25 - r^2}$.

Thus, the solid has the description

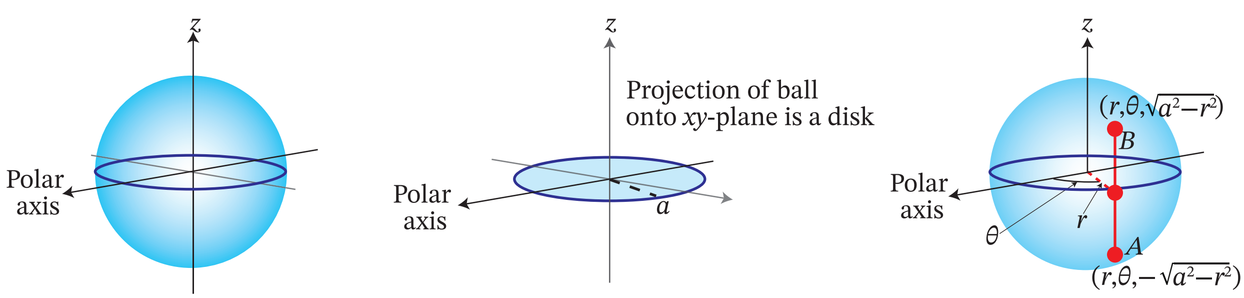
$$\underbrace{0 \le \theta \le 2\pi, \qquad 0 \le r \le 3}_{\text{the projection onto the } r\theta\text{-plane}}, \qquad \underbrace{-\sqrt{25 - r^2} \le z \le \sqrt{25 - r^2}}_{\text{range of } z \text{ for each } r \text{ and } \theta}.$$

♦

EXAMPLE 3. Describe a ball of radius a in cylindrical coordinates.

SOLUTION Place the origin at the center of the ball, as in Figure 17.5.5(a). The projection of the ball on the $r\theta$-plane is a disk of radius a, shown in Figure 17.5.5(b), which is described by

$$0 \le \theta \le 2\pi, \qquad 0 \le r \le a.$$



(a) (b) (c)
Figure 17.5.5

All that remains is to see how z varies for given r and θ. How does z vary on the line segment AB in Figure 17.5.5(c)? The bigger r is the shorter AB is.

The geometry is shown first in perspective, Figure 17.5.6(a), and then not in perspective, Figure 17.5.6(b).

From Figure 17.5.5(c) we see that z varies from $-\sqrt{a^2 - r^2}$ to $\sqrt{a^2 - r^2}$. As a quick check that this is correct, consider two easy cases, $r = 0$ and $r = a$. If r is a, then z varies from 0 to 0. If r is 0, then z varies from $-a$ to a.

When all of this information is combined, the description of the ball of radius a in cylindrical coordinates is

$$\underbrace{0 \le \theta \le 2\pi, \qquad 0 \le r \le a}_{\text{the projection onto the } r\theta\text{-plane}}, \qquad \underbrace{-\sqrt{a^2 - r^2} \le z \le \sqrt{a^2 - r^2}}_{\text{range of } z \text{ for each } r \text{ and } \theta}$$

♦

EXAMPLE 4. Draw the region $\mathcal{R}$ in the first octant bounded by the surfaces $r^2 + z^2 = a^2$, $\theta = \dfrac{\pi}{6}$, and $\theta = \dfrac{\pi}{3}$.

SOLUTION The surface $r^2 + z^2 = a^2$ is a sphere with radius a centered at the origin ($x^2 + y^2 + z^2 = a^2$ in rectangular coordinates). Figure 17.5.7(a) shows the part of it in the first octant.

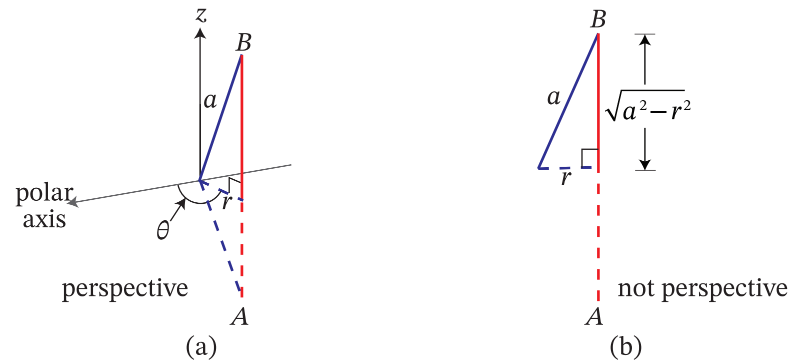


Figure 17.5.6

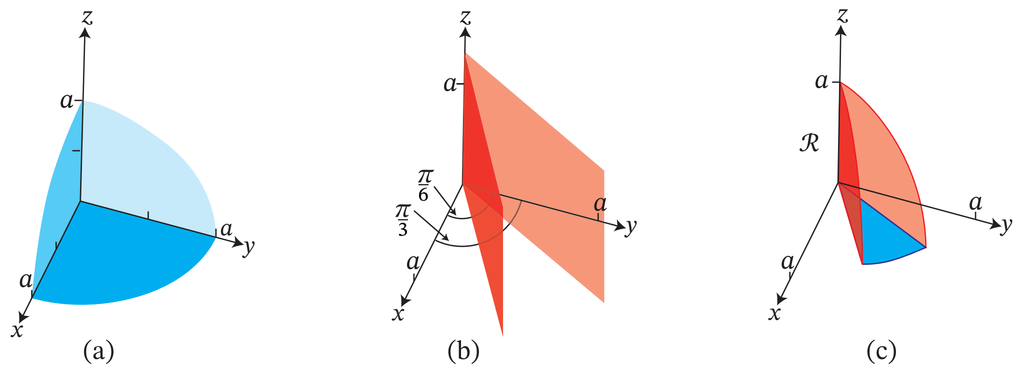


Figure 17.5.7

Next we draw the half planes $\theta = \pi/6$ and $\theta = \pi/3$, as in Figure 17.5.7(b), again showing only the parts in the first octant. Finally we put Figure 17.5.7(a) and (b) together in (c), to see that $\mathscr{R}$ is a wedge from a ball.

The boundary of $\mathscr{R}$ has three plane surfaces, $z = 0$, $\theta = \pi/6$, and $\theta = \pi/3$, and one curved surface, $r^2 + z^2 = a^2$.

The description of $\mathscr{R}$ in cylindrical coordinates is

$$\frac{\pi}{6} \le \theta \le \frac{\pi}{3}, \qquad 0 \le r \le a, \qquad 0 \le z \le \sqrt{a^2 - r^2}.$$

◆

The Volume Swept Out by Δr, $\Delta\theta$, and Δz

To use polar coordinates to evaluate an integral over a plane set we needed to know that the area of the region corresponding to small changes Δr and $\Delta\theta$ is roughly $r\Delta r\ \Delta\theta$. To evaluate integrals over solids using an iterated integral in cylindrical coordinates, we will need to estimate the volume of the region corresponding to small changes Δr, $\Delta\theta$, Δz in the three coordinates.

The set of points (r, θ, z) whose r-coordinates are between r_0 and $r_0 + \Delta r$, whose θ-coordinates are between θ_0 and $\theta_0 + \Delta\theta$, and whose z-coordinates are between z_0 and $z_0 + \Delta z$ is shown in Figure 17.5.8(a). It is a solid with four flat surfaces and two curved surfaces.

When Δr and $\Delta\theta$ are small, the area of the flat base of the solid is approximately $r\Delta r\ \Delta\theta$, as shown in Section 9.2 and as we saw when working with polar coordinates in the plane. Thus, when Δr, $\Delta\theta$, and Δz are small, the volume ΔV of the solid in Figure 17.5.8(b) is approximately

$$(\text{Area of base})(\text{height}) \approx r\ \Delta r\ \Delta\theta\ \Delta z.$$

This estimate leads us to the following formula for the volume of the region based at a point with cylidrical coordinates (r, θ, z) and extending a (short) distance Δr, $\Delta\theta$, and Δz in each cylindrical coordinate.

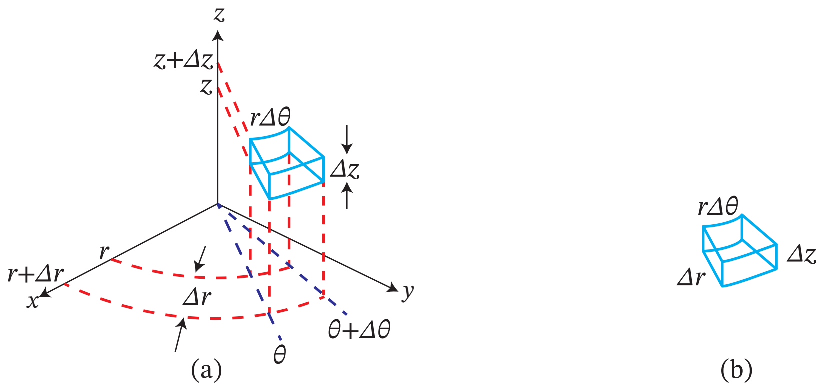


Figure 17.5.8

Formula 17.5.1: *Volume Swept Out by Δr, $\Delta\theta$, and Δz in Cylindrical Coordinates*

The approximate volume of the region consisting of all points (r, θ, z) whose r-coordinates are between r and $r + \Delta r$, whose θ-coordinates are between θ and $\theta + \Delta\theta$, and whose z-coordinates are between z and $z + \Delta z$ is

$$\Delta V \approx r\ \Delta r\ \Delta\theta\ \Delta z.$$

The same factor r that appears in iterated integrals in polar coordinates, also appears in iterated integrals in cylindrical coordinates.

Spherical Coordinates

In **spherical coordinates** a point P is described by three numbers: ρ, the distance from P to the origin O, θ, the same angle as in cylindrical coordinates, and ϕ, the angle between the positive z-axis and the ray from O to P.

The point P is denoted $P = (\rho, \theta, \phi)$. The angle ϕ is the same as the direction angle of OP with $\mathbf{k}$, $0 \le \phi \le \pi$. (See Figure 17.5.9.) For a positive constant k the coordinate surfaces $\rho = k$ (a sphere), $\theta = k$ (a half plane), and $\phi = k$ (a cone) are shown in Figure 17.5.10. Thus a point is described by the sphere, half plane, and cone on which it lies.

NOTE: ρ is pronounced "row"; it is the Greek letter for r. The Greek letter ϕ is pronounced "fee" or "fie."

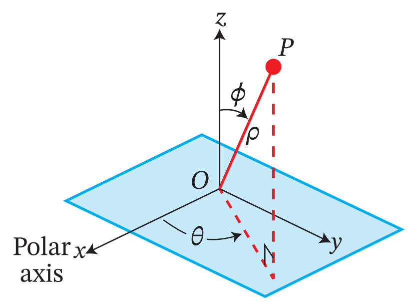


Figure 17.5.9

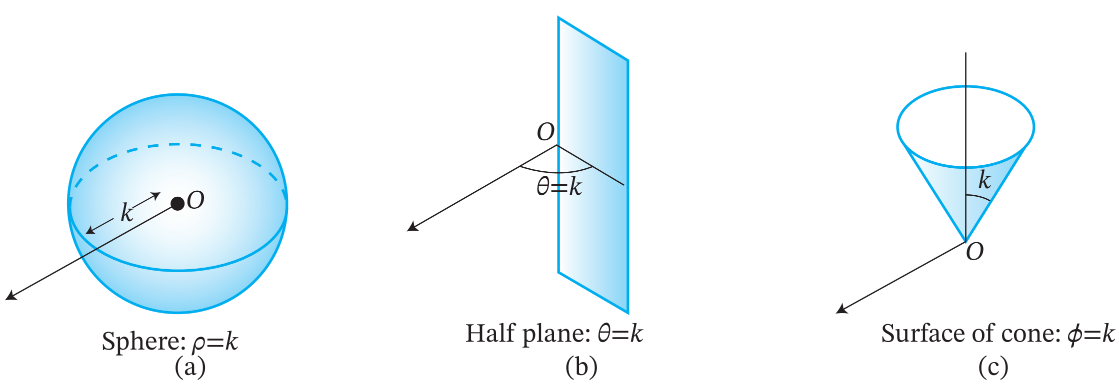


Figure 17.5.10

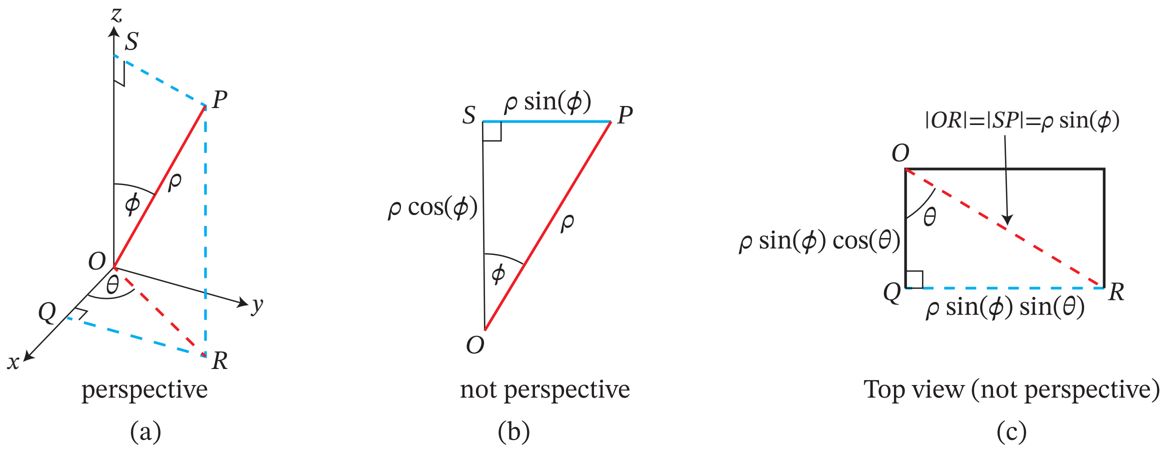


Figure 17.5.11

Relationship between Rectangular and Spherical Coordinates

Figure 17.5.11 displays the relation between spherical and rectangular coordinates of a point $P = (\rho, \theta, \phi) = (x, y, z)$.

The right triangle OSP has hypotenuse OP and a right angle at S, and the right triangle OQR has a right angle at Q. We have $z = |OS| = \rho\cos(\phi)$, $|OR| = |SP| = \rho\sin(\phi)$, $x = |OR|\cos(\theta) = \rho\sin(\phi)\cos(\theta)$, and $y = |OR|\sin(\theta) = \rho\sin(\phi)\sin(\theta)$. Hence, $x = \rho\sin(\phi)\cos(\theta)$, $y = \rho\sin(\phi)\sin(\theta)$, and $z = \rho\cos(\phi)$.

EXAMPLE 5. Figure 17.5.12 shows a point P given in spherical coordinates. Find its rectangular coordinates.

SOLUTION

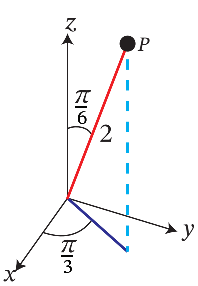


Figure 17.5.12

Because the spherical coordinates of P are $\rho = 2$, $\theta = \pi/3$, $\phi = \pi/6$,

$$x = 2\sin\left(\frac{\pi}{6}\right)\cos\left(\frac{\pi}{3}\right) = 2\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{2}$$
$$y = 2\sin\left(\frac{\pi}{6}\right)\sin\left(\frac{\pi}{3}\right) = 2\cdot\frac{1}{2}\cdot\frac{\sqrt{3}}{2} = \frac{\sqrt{3}}{2}$$
$$z = 2\cos\left(\frac{\pi}{6}\right) = 2\frac{\sqrt{3}}{2} = \sqrt{3}.$$

As a check, $x^2 + y^2 + z^2$ should equal ρ^2, and it does, for

$$\left(\frac{1}{2}\right)^2 + \left(\frac{\sqrt{3}}{2}\right)^2 + \left(\sqrt{3}\right)^2 = \frac{1}{4} + \frac{3}{4} + 3 = 4 = 2^2.$$

♦

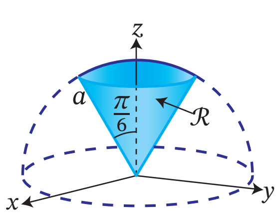


Figure 17.5.13

The next example uses spherical coordinates to describe a cone topped by part of a ball.

EXAMPLE 6. The region $\mathscr{R}$ consists of the portion of a ball of radius a that lies within a cone of half-angle $\frac{\pi}{6}$. The vertex of the cone is at the center of the ball. (See Figure 17.5.13.) Describe $\mathscr{R}$ in spherical coordinates.

SOLUTION The region $\mathscr{R}$ resembles an ice cream cone, a dry cone topped with a spherical scoop of ice cream.

Because $\mathscr{R}$ is a solid of revolution (around the z-axis), $0 \le \theta \le 2\pi$. The section of $\mathscr{R}$ corresponding to an angle θ is the intersection of $\mathscr{R}$ with a half plane, shown in Figure 17.5.14(a).

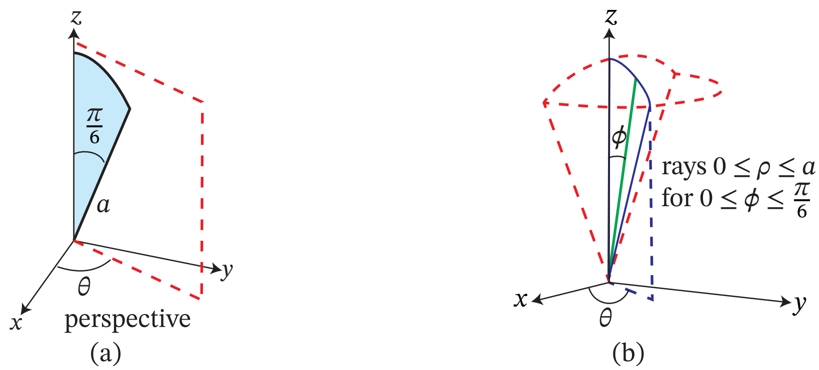


Figure 17.5.14

For each θ, the angle ϕ goes from 0 to $\pi/6$. Finally, θ and ϕ determine a ray on which ρ goes from 0 to a, as shown in Figure 17.5.14(b). The description in spherical coordinates is

$$0 \le \phi \le \frac{\pi}{6}, \qquad 0 \le \theta \le 2\pi, \qquad 0 \le \rho \le a.$$

♦

The next example describes a ball in rectangular and spherical coordinates.

EXAMPLE 7. Describe a ball of radius a in (a) rectangular coordinates and (b) spherical coordinates.

SOLUTION In each case we put the origin of the coordinate system at the center of the ball.

(a) Rectangular coordinates:

The projection of the ball on the xy-plane is a disk of radius a, described by

$$-a \le x \le a, \qquad -\sqrt{a^2 - x^2} \le y \le \sqrt{a^2 - x^2}.$$

For (x, y) in the projection of the ball onto the xy-plane, z varies along the segment AB in Figure 17.5.15.
The equation of the sphere is $x^2 + y^2 + z^2 = a^2$. Therefore, at A, z is $-\sqrt{z^2 - x^2 - y^2}$, and, at B, z is $\sqrt{a^2 - x^2 - y^2}$.

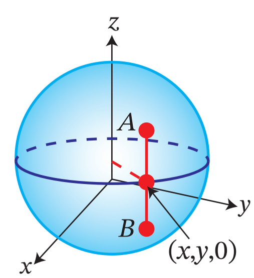


Figure 17.5.15

The entire description of the ball of radius a in rectangular coordinates is

$$-a \le x \le a, \quad -\sqrt{a^2 - x^2} \le y \le \sqrt{a^2 - x^2}, \quad -\sqrt{a^2 - x^2 - y^2} \le z \le \sqrt{z^2 - x^2 - y^2}.$$

(b) Spherical coordinates:
This time the projection on the xy-plane plays no role. Instead, we begin with

$$0 \le \theta \le 2\pi, \qquad 0 \le \phi \le \pi,$$

which sweeps out all the rays from the origin. On each ray ρ goes from 0 to a. The complete description involves only constants as bounds:

$$0 \le \phi \le \pi, \qquad 0 \le \theta \le 2\pi, \qquad 0 \le \rho \le a.$$

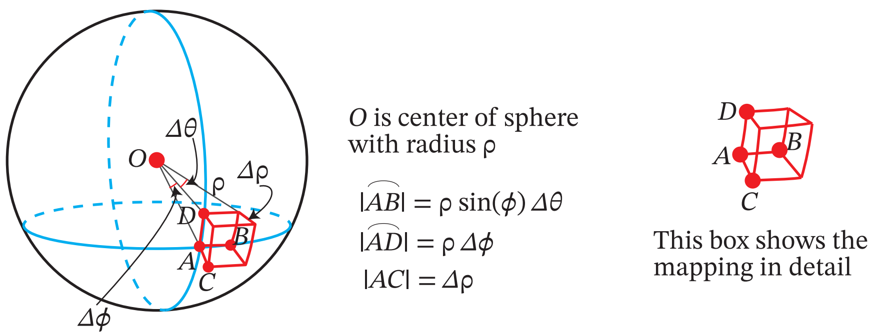
♦

NOTE: In Example 7, because the endpoints of each of the three variables are constants, the range of each variable is not influenced by the other variables. When this happens, the variables in the description can be given in any of the six possible orders.

The Volume Swept Out by $\Delta\rho$, $\Delta\theta$, and $\Delta\phi$

In the next section we will need an estimate of the volume of the little box-like region bounded by the spheres with radii ρ and $\rho+\Delta\rho$, the half-planes with angles θ and $\theta+\Delta\theta$, and the cones with half-angles ϕ and $\phi+\Delta\phi$. Two of its surfaces are flat, two are parts of spheres, and two are parts of cones. Arc $\widehat{AD}$ is part of a circle of radius ρ and arc $\widehat{AB}$ is part of a circle of radius $\rho\sin(\phi)$, as shown in Figure 17.5.16(a).

The product of the lengths of $\widehat{AB}$, AC, and $\widehat{AD}$ is an estimate of the volume of the little box. Figure 17.5.16 shows how to find each length. REMINDER: $\widehat{AB}$ and $\widehat{AD}$ are arcs of circles, while AC is a line segment.



(a) (b)

Figure 17.5.16

The result is

Formula 17.5.2: *Volume Swept Out by $\Delta\rho$, $\Delta\theta$, and $\Delta\phi$ in Spherical Coordinates*

The approximate volume of the region consisting of all points (ρ,θ,ϕ) whose ρ-coordinates are between ρ and $\rho+\Delta\rho$, whose θ-coordinates are between θ and $\theta+\Delta\theta$, and whose ϕ-coordinates are between ϕ and $\phi+\Delta\phi$ is

$$\Delta V \approx \rho^2\sin(\phi)\,\Delta\rho\,\Delta\phi\,\Delta\theta.$$

As we added an r to an integrand in polar or cylindrical coordinates we must add the factor $\rho^2\sin(\phi)$ to an integrand when using an iterated integral in spherical coordinates.

Summary

This section described cylindrical and spherical coordinates. The volume of the small box corresponding to small changes in the three cylindrical coordinates is approximately $r\,\Delta r\,\Delta\theta\,\Delta z$. Because of the presence of the factor r, we must adjoin an r to the integrand when using an iterated integral in cylindrical coordinates.

Similarly, $\rho^2\sin(\phi)$ must be added to an integrand when using an iterated integral in spherical coordinates.

Warning: Notational Differences Between Mathematics and Physics/Engineering

In many physics and engineering texts r is used instead of ρ and the roles of θ and ϕ are sometimes switched from the ones used in this section, and elsewhere in this book and in almost all other mathematical writing.

EXERCISES for Section 17.5

1. Fill in the blanks.
  (a) In rectangular coordinates in space a point is described as the intersection of ____________________.
  (b) In cylindrical coordinates a point is described as the intersection of a __________, a __________, and a __________.
  (c) In spherical coordinates a point is described as the intersection of a __________, a __________, and a __________.

2. On the region in Example 2 draw the set of points described by (a) $z = 2$, (b) $z = 3$, and (c) $z = 4.5$.

3. For the cylinder in Example 1 draw the set of points described by (a) $r = \frac{a}{2}$, (b) $\theta = \frac{\pi}{4}$, and (c) $z = \frac{h}{3}$.

4. (a) In the formula $\Delta V \approx r\,\Delta r\,\Delta\theta\,\Delta z$ which factors have the dimension of length?
  (b) Why would you expect three such factors?

5. (a) In the formula $\Delta V \approx \rho^2 \sin(\phi)\,\Delta\rho\,\Delta\theta\,\Delta\phi$, which factors have the dimension of length?
  (b) Why would you expect three such factors?

6. In one clear, large diagram, show how to express rectangular coordinates in terms of cylindrical coordinates.

7. In one clear, large diagram, show how to express rectangular coordinates in terms of spherical coordinates.

8. Find the cylindrical coordinates of $(x, y, z) = (3, 3, 1)$, including a clear diagram.

9. Find the spherical coordinates of $(x, y, z) = (3, 3, 1)$, including a clear diagram.

In Exercises 10 to 15 (a) draw the set of points described, and (b) describe the set of points in words.

See also Exercises 35 and 36.

10. r and z fixed, θ varies.
11. r and θ fixed, z varies.
12. θ and z fixed, r varies.
13. ρ and ϕ fixed, θ varies.
14. ρ and θ fixed, ϕ varies.
15. θ and ϕ fixed, ρ varies.

---

16. Find the equation of a sphere of radius a centered at the origin in
(a) spherical coordinates, (b) cylindrical coordinates, and (c) rectangular coordinates.

17. Explain why if $P = (x, y, z) = (\rho, \theta, \phi)$ in spherical coordinates, then $x^2 + y^2 + z^2 = \rho^2$.

18. Describe the region in Example 6 in cylindrical coordinates in the form $\alpha \le \theta \le \beta$, $r_1(\theta) \le r \le r_2(\theta)$, $z_1(r, \theta) \le z \le z_2(r, \theta)$.

19. Like Exercise 18, but in the order $a \le z \le b$, $\theta_1(z) \le \theta \le \theta_2(z)$, $r_1(\theta, z) \le r \le r_2(\theta, z)$.

20. Sketch the region in the first octant bounded by the planes $\theta = \frac{\pi}{6}$ and $\theta = \frac{\pi}{3}$ and the sphere $\rho = a$.

21. Estimate the area of the bottom face of the curvy box shown in Figure 17.5.16(b). Points A, B, and D lie on the sphere of radius ρ.

22. A cone of half-angle $\frac{\pi}{6}$ is cut by a plane perpendicular to its axis at a distance 4 from its vertex.
(a) Place it on a cylindrical coordinate system. (b) Describe it in cylindrical coordinates.

23. Repeat Exercise 22, but use spherical coordinates instead of cylindrical coordinates.

24. A solid, infinite cone has its vertex at the origin and its axis along the positive z-axis. It is made by revolving a line through the origin that has an angle A with the z-axis, about the z-axis. Describe it in
(a) spherical coordinates, (b) cylindrical coordinates, and (c) rectangular coordinates.

25. Use spherical coordinates to describe the surface in Figure 17.5.17. It is part of a cone of vertex half-angle α with the z-axis as its axis, situated within a sphere of radius a centered at the origin.

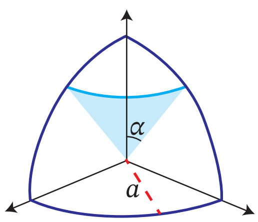


Figure 17.5.17

26. A ball of radius a has a diameter coinciding with the interval $[0,2a]$ on the x-axis. Describe the ball in spherical coordinates. (See Exercise 3(c) in Section 9.S.)

27. A ray is described in spherical coordinates by $\theta = \frac{\pi}{6}$ and $\phi = \frac{\pi}{4}$. (a) Draw a picture that shows the three direction angles of the ray. (b) Find $\cos(\alpha)$.

28. (a) If the region in Example 2 is described in the order $0 \le \theta \le 2\pi$, $z_1(\theta) \le z \le z_2(\theta)$, $r_1(\theta,z) \le r \le r_2(\theta,z)$, what complication arises?
(b) Describe the region using the order given in (a).

29. What is the distance between two points, $P_1 = (\rho_1,\theta_1,\phi_1)$ and $P_2 = (\rho_2,\theta_2,\phi_2)$, in spherical coordinates?

30. The points $P_1 = (\rho_1,\theta_1,\phi_1)$ and $P_2 = (\rho_1,\theta_2,\phi_2)$ both lie on a sphere of radius ρ_1. Assuming that both are in the first octant, find the great circle distance between them.

31. What is the speed of a particle moving along a curve that is at the point $(\rho(t),\theta(t),\phi(t))$ in spherical coordinates.

32. How far apart are the points (r_1,θ_1,z_1) and (r_2,θ_2,z_2) in the first octant? (a) Draw a large clear and well-labelled diagram of the two given points. (b) Find the distance between (r_1,θ_1,z_1) and (r_2,θ_2,z_2).

33. A circular cylinder has radius r and height h. One path for a bug crawling from $(r,0,h)$ to $(-r,0,0)$ (rectangular coordinates) on the surface goes from $(r,0,h)$ to $(-r,0,h)$ along a diameter, followed by the path straight down to $(-r,0,0)$. Another path stays on the curved surface. Which path is shorter?

34. Using a large clear diagram, estimate the volume of the region determined by small changes $\Delta\rho$, $\Delta\theta$, and $\Delta\phi$.

35. In cylindrical coordinates if two coordinates are held fixed and the third allowed to vary, one obtains a curve. Because there are three choices for the two fixed coordinates, there are three such curves at each point. Using a drawing, determine the angles between these curves. (See Exercises 10 to 12.)

36. In spherical coordinates if two coordinates are held fixed and the third allowed to vary, one obtains a curve. Because there are three choices for the two fixed coordinates, there are three such curves at each point. Using a drawing, determine the angles between these curves. (See Exercises 13 to 15.)

In Exercises 37 and 40 verify the equations by differentiating.

37. $\displaystyle\int \frac{dx}{x^3\sqrt{a^2+x^2}} = -\frac{\sqrt{a^2+x^2}}{2a^2x^2} + \frac{1}{2a^3}\ln\left|\frac{a+\sqrt{a^2+x^2}}{x}\right| + C.$

38. $\displaystyle\int \frac{x^2}{a^4-x^4}\,dx = \frac{1}{4a}\ln\left|\frac{a+x}{a-x}\right| - \frac{1}{2a}\arctan\left(\frac{x}{a}\right) + C.$

39. $\displaystyle\int \frac{\sqrt{ax+b}}{x}\,dx = 2\sqrt{ax+b} + b\int \frac{dx}{x\sqrt{ax+b}}.$

40. $\displaystyle\int \frac{dx}{x^2\sqrt{ax+b}} = -\frac{\sqrt{ax+b}}{bx} - \frac{a}{2b}\int \frac{dx}{x\sqrt{ax+b}}.$

In Exercises 41 to 42 verify the equations by integrating. (Show all steps.)

41. $\displaystyle\int \tan^2(ax)\,dx = \frac{1}{a}\tan(ax) - x + C.$

42. $\displaystyle\int \ln(ax)\,dx = x(\ln(ax)-1) + C.$

17.6 Computing Triple Integrals in Cylindrical or Spherical Coordinates

In Section 17.3 we learned how to evaluate some double integrals $\int_{\mathscr{R}} f(P)\, dA$ by two iterated integrals in polar coordinates. In this method it is necessary to multiply the integrand by an r because the area of the small piece determined by small increments in r and θ is not $\Delta r\ \Delta\theta$ but $r\ \Delta r\ \Delta\theta$. Similarly, when developing iterated integrals using cylindrical coordinates, an extra r must be adjoined to the integrand. For spherical coordinates we adjoin $\rho^2 \sin(\phi)$. These adjustments are based on the estimates of the volumes of the small curvy boxes made in the previous section.

A few examples will illustrate the general method, which is: Describe the solid $\mathscr{R}$ and the integrand in the most convenient coordinate system. Then use the description to set up an iterated integral, being sure to include the appropriate extra factor in the integrand. And, finally, evaluate the three iterated integrals.

Iterated Integrals in Cylindrical Coordinates

To evaluate $\int_{\mathscr{R}} f(P)\, dV$ in cylindrical coordinates express the integrand in cylindrical coordinates and describe the region $\mathscr{R}$ in cylindrical coordinates with dV replaced by $r\ dz\ dr\ d\theta$. While there are six possible orders of integration, the most common one has z varying first, then r, and finally θ.

Formula 17.6.1: *Evaluating $\int_{\mathscr{R}} f(P)\, dV$ in Cylindrical Coordinates*

When a region $\mathscr{R}$ in space can be described in cylindrical coordinates as

$$\theta_1 \le \theta \le \theta_2, \qquad r_1(\theta) \le r \le r_2(\theta), \qquad z_1(r,\theta) \le z \le z_2(r,\theta),$$

then

$$\int_{\mathscr{R}} f(P)\, dV = \int_{\theta_1}^{\theta_2} \int_{r_1(\theta)}^{r_2(\theta)} \int_{z_1(r,\theta)}^{z_2(r,\theta)} f(r,\theta,z)\ r\ dz\ dr\ d\theta.$$

Similar formulas exist for each of the other five orderings of the variables r, θ, and z.

EXAMPLE 1. Find the volume of a ball $\mathscr{R}$ of radius a using cylindrical coordinates.

SOLUTION Place the origin of a cylindrical coordinate system at the center of the ball, as in Figure 17.6.1. The volume of the ball is $\int_{\mathscr{R}} 1\ dV$. The description of $\mathscr{R}$ in cylindrical coordinates is

$$0 \le \theta \le 2\pi, \qquad 0 \le r \le a, \qquad -\sqrt{a^2-r^2} \le z \le \sqrt{a^2-r^2}.$$

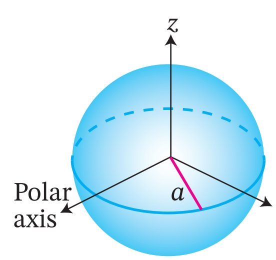


Figure 17.6.1

Remembering to insert the factor r as dV is converted to $r\ dz\ dr\ d\theta$, the iterated integral for the volume is

$$\text{Volume of ball} = \int_{\mathscr{R}} 1\ dV = \int_0^{2\pi} \int_0^a \int_{-\sqrt{a^2-r^2}}^{\sqrt{a^2-r^2}} 1 \cdot r\ dz\ dr\ d\theta.$$

The order of integration is determined by the order the variables appear in the description of the region $\mathscr{R}$.

This triple integral is evaluated from the inside out: z first, then r, and finally θ. Evaluation of the innermost (z) integral is straightforward:

$$\int_{-\sqrt{a^2-r^2}}^{\sqrt{a^2-r^2}} r\,dz = rz\Big|_{z=-\sqrt{a^2-r^2}}^{z=\sqrt{a^2-r^2}} = 2r\sqrt{a^2-r^2}.$$

Evaluation of the middle (r) integral uses a substitution $u = a^2 - r^2$:

$$\int_0^a 2r\sqrt{a^2-r^2}\,dr = \frac{-2(a^2-r^2)^{3/2}}{3}\Big|_{r=0}^{r=a} = \frac{2a^3}{3}.$$

Finally, evaluation of the outermost (θ) integral gives

$$\int_0^{2\pi} \frac{2a^3}{3}\,d\theta = \frac{2a^3}{3}\int_0^{2\pi} d\theta = \frac{2a^3}{3}\cdot 2\pi = \frac{4}{3}\pi a^3.$$

◆

EXAMPLE 2. Find the volume of the region $\mathscr{R}$ inside the cylinder $x^2+y^2=9$, above the xy-plane, and below the plane $z = x+2y+9$. Use cylindrical coordinates.

SOLUTION To see that the intersection of the plane $z = x+2y+9$ with the cylinder $x^2+y^2=9$ is always above the xy-plane, note that points (x, y) on the boundary of the cylinder can be written as $x = 3\cos(\theta)$ and $y = 3\sin(\theta)$ for $0 \le \theta \le 2\pi$. Then $z = x+2y+9 = 3\cos(\theta)+6\sin(\theta)+9$. For $0 \le \theta \le 2\pi$, $z \ge 3(-1)+6(-1)+9 = 0$, so z is always positive. (In fact, the minimum and maximum values of z are $9-3\sqrt{5} \approx 2.292$ and $9+3\sqrt{5} \approx 15.708$, respectively. See Exercise 33.)

We wish to evaluate $\int_{\mathscr{R}} 1\,dV$ over the region $\mathscr{R}$ described in cylindrical coordinates by

$$0 \le \theta \le 2\pi, \qquad 0 \le r \le 3, \qquad 0 \le z \le r\cos(\theta)+2r\sin(\theta)+9.$$

The corresponding iterated integrals take the form

Note the factor r.

$$\int_0^{2\pi}\int_0^3\int_0^{r\cos(\theta)+2r\sin(\theta)+9} 1\cdot r\,dz\,dr\,d\theta.$$

Evaluating the innermost integral, with respect to z, remembering that r and θ are both constant, gives

$$\begin{aligned}\int_0^{r\cos(\theta)+2r\sin(\theta)+9} r\,dz &= rz\Big|_{z=0}^{z=r\cos(\theta)+2r\sin(\theta)+9}\\ &= r^2\cos(\theta)+2r^2\sin(\theta)+9r.\end{aligned}$$

The middle integral, with respect to r, with θ constant, is next:

$$\begin{aligned}\int_0^3 \left(r^2\cos(\theta)+2r^2\sin(\theta)+9r\right)\,dr &= \left(\frac{r^3}{3}\cos(\theta)+\frac{2r^3}{3}\sin(\theta)+\frac{9r^2}{2}\right)\Big|_{r=0}^{r=3}\\ &= 9\cos(\theta)+18\sin(\theta)+\frac{81}{2}.\end{aligned}$$

Finally, the outermost integral, with respect to θ, is

$$\int_0^{2\pi} \left(9\cos(\theta) + 18\sin(\theta) + \frac{81}{2} \right) d\theta. \tag{17.6.1}$$

Explain, without referring to antiderivatives, why $\int_0^{2\pi} \cos(x)\, dx = 0$ and $\int_0^{2\pi} \sin(x)\, dx = 0$.

Because $\int_0^{2\pi} \cos(\theta)\, d\theta = 0$ and $\int_0^{2\pi} \sin(\theta)\, d\theta = 0$, (17.6.1) reduces to $\int_0^{2\pi} 81/2\, d\theta = 81\pi$. The volume is 81π. ♦

Iterated Integrals in Spherical Coordinates

To evaluate a triple integral $\int_{\mathscr{R}} f(P)\, dV$ in spherical coordinates, first describe the region $\mathscr{R}$ in spherical coordinates. In many problems the variables appear in the order

$$\phi_1 \le \phi \le \phi_2, \qquad \theta_1(\phi) \le \theta \le \theta_2(\phi), \qquad \rho_1(\theta,\phi) \le \rho \le \rho_2(\theta,\phi).$$

Each of the other five orderings of the variables can be encountered. For example, if θ and ϕ in the description of $\mathscr{R}$ are switched:

$$\theta_1 \le \theta \le \theta_2, \qquad \phi_1(\theta) \le \phi \le \phi_2(\theta), \qquad \rho_1(\theta,\phi) \le \rho \le \rho_2(\theta,\phi).$$

Then, to set up a triple integral as an equivalent iterated integral, express dV as $\rho^2 \sin(\phi)\, d\rho\, d\phi\, d\theta$. When θ and ϕ are interchanged, $dV = \rho^2 \sin(\phi)\, d\rho\, d\theta\, d\phi$. While these two expressions for dV are algebraically equivalent, the order of the three differentials is important when setting up the three iterated integrals.

Formula 17.6.2: *Evaluating $\int_{\mathscr{R}} f(P)\, dV$ in Spherical Coordinates*

When a region $\mathscr{R}$ in space can be described in spherical coordinates as

$$\theta_1 \le \theta \le \theta_2, \qquad \phi_1(\theta) \le \phi \le \phi_2(\theta), \qquad \rho_1(\theta,\phi) \le \rho \le \rho_2(\theta,\phi),$$

then,

$$\int_{\mathscr{R}} f(P)\, dV = \int_{\phi_1}^{\phi_2} \int_{\theta_1(\phi)}^{\theta_2(\phi)} \int_{\rho_1(\theta,\phi)}^{\rho_2(\theta,\phi)} f(\rho,\theta,\phi)\, \rho^2 \sin(\phi)\, d\rho\, d\theta\, d\phi.$$

Similar formulas exist for each of the other five orderings of the variables ρ, θ, and ϕ.

EXAMPLE 3. Find the volume of a ball of radius a, using spherical coordinates.

SOLUTION Place the origin of spherical coordinates at the center of the ball, as in Figure 17.6.2. The ball is described by

$$0 \le \phi \le \pi, \qquad 0 \le \theta \le 2\pi, \qquad 0 \le \rho \le a.$$

Then, replacing dV with $\rho^2 \sin(\phi)\, d\rho\, d\phi\, d\theta$, we arrive at the following expression for the volume of a sphere as a triple integral in spherical coordinates:

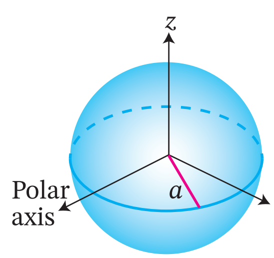


Figure 17.6.2

$$\text{Volume of ball} = \int_{\mathscr{R}} 1\, dV = \int_0^{\pi} \int_0^{2\pi} \int_0^{a} \rho^2 \sin(\phi)\, d\rho\, d\theta\, d\phi.$$

The inner integral, with respect to ρ with θ and ϕ held constant, is

$$\int_0^a \rho^2 \sin(\phi)\, d\rho = \sin(\phi) \int_0^a \rho^2\, d\rho = \sin(\phi)\, \frac{\rho^3}{3}\bigg|_0^a = \frac{a^3 \sin(\phi)}{3}.$$

The middle integral, with respect to θ with ϕ held constant, is

$$\int_0^{2\pi} \frac{a^3 \sin(\phi)}{3}\, d\theta = \frac{a^3 \sin(\phi)}{3}\theta\bigg|_0^{2\pi} = \frac{a^3 \sin(\phi)}{3}(2\pi) - \frac{a^3 \sin(\phi)}{3}(0) = \frac{2\pi a^3 \sin(\phi)}{3}.$$

The final integral, with respect to ϕ, is

$$\int_0^{\pi} \frac{2\pi a^3 \sin(\phi)}{3}\, d\phi = \frac{-2\pi a^3 \cos(\phi)}{3}\bigg|_0^{\pi} = \frac{-2\pi a^3(-1)}{3} - \frac{-2\pi a^3(1)}{3} = \frac{4}{3}\pi a^3.$$ ♦

Examples 1 and 3 confirm the well-known volume formula for a ball of radius a: $V = 4\pi a^3/3$.

An Integral in Gravity

The integral in the next example arises when computing a potential energy due to gravitational attraction. Its value implies that a homogeneous ball attracts a particle outside it as though all the mass in the ball were at its center.

Newton obtained this remarkable result in 1687.

EXAMPLE 4. A homogeneous ball of mass M and radius a occupies a region $\mathscr{R}$. Let A be a point at a distance H from the center of the ball, $H > a$. Compute $\int_{\mathscr{R}} \frac{\delta}{q(P)}\, dV$, where δ is the density of the ball and $q(P)$ is the distance from point P in $\mathscr{R}$ to A. (See Figure 17.6.3(a).)

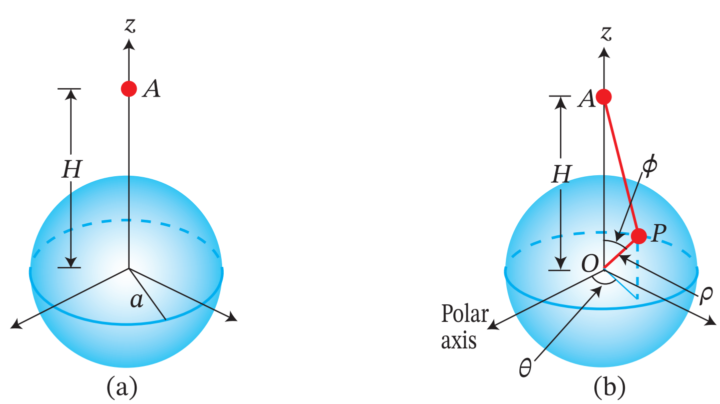


Figure 17.6.3

SOLUTION The presence of a triple integral over a sphere suggests using spherical coordinates to evaluate this triple integral. To express $q(P)$ in spherical coordinates, choose a spherical coordinate system whose origin is at the center of the sphere and such that the ϕ coordinate of the point A is 0. (See Figure 17.6.3(b).)

Let $P = (\rho, \theta, \phi)$ be a point in the ball. Applying the law of cosines to triangle AOP, we find that

$$q^2 = H^2 + \rho^2 - 2\rho H \cos(\phi).$$

Hence

$$q = \sqrt{H^2 + \rho^2 - 2\rho H \cos(\phi)}.$$

Since the ball is homogeneous, if its total mass is M, then

$$\delta = \frac{M}{\frac{4}{3}\pi a^3} = \frac{3M}{4\pi a^3}.$$

And so,

$$\int_{\mathscr{R}} \frac{\delta}{q(P)}\, dV = \int_{\mathscr{R}} \frac{3M}{4\pi a^3 q(P)}\, dV = \frac{3M}{4\pi a^3}\int_{\mathscr{R}} \frac{1}{q(P)}\, dV. \tag{17.6.2}$$

To evaluate

$$\int_{\mathscr{R}} \frac{1}{q(P)}\, dV$$

we choose to convert it to an iterated integral in spherical coordinates:

$$\int_{\mathscr{R}} \frac{1}{q(P)}\, dV = \int_0^{2\pi}\int_0^{a}\int_0^{\pi} \frac{\rho^2\sin(\phi)}{\sqrt{H^2+\rho^2-2\rho H\cos(\phi)}}\, d\phi\, d\rho\, d\theta.$$

While any of the six orderings of the variables is a valid expression for this triple integral, the ones starting with ρ would be quite challenging. We choose to integrate with respect to ϕ first because it is easy to evaluate. (The θ integral is trivial, but we save this step for the end.)

Evaluation of the innermost integral, where ρ and θ are constants, is accomplished with the aid of the substitution $u = H^2+\rho^2-2\rho H\cos(\phi)$ (so $du = 2\rho H\sin(\phi)\, d\phi$) and then the fundamental theorem of calculus:

$$\begin{aligned}\int_0^{\pi} \frac{\rho^2\sin\phi}{\sqrt{H^2+\rho^2-2\rho H\cos(\phi)}}\, d\phi &= \left.\frac{\rho\sqrt{H^2+\rho^2-2\rho H\cos(\phi)}}{H}\right|_{\phi=0}^{\phi=\pi}\\ &= \frac{\rho}{H}\left(\sqrt{H^2+\rho^2+2\rho H}-\sqrt{H^2+\rho^2-2\rho H}\right).\end{aligned}$$

Now, $\sqrt{H^2+\rho^2+2\rho H} = H+\rho$. Since $\rho \le a < H$, $H-\rho$ is positive and $\sqrt{H^2+\rho^2-2\rho H} = H-\rho$.

Thus the first integral equals

$$\frac{\rho}{H}((H+\rho)-(H-\rho)) = \frac{2\rho^2}{H}.$$

Evaluation of the middle integral, with respect to ρ, yields

$$\int_0^{a} \frac{2\rho^2}{H}\, d\rho = \frac{2a^3}{3H}.$$

Then, evaluation of the outermost integral gives

$$\int_0^{2\pi} \frac{2a^3}{3H}\, d\theta = \frac{4\pi a^3}{3H}.$$

From the evaluation of these three definite integrals we conclude that

$$\int_{\mathscr{R}} \frac{1}{q(P)}\, dV = \frac{4\pi a^3}{3H}.$$

By (17.6.2)

$$\int_{\mathscr{R}} \frac{\delta}{q(P)}\, dV = \frac{3M}{4\pi a^3}\frac{4\pi a^3}{3H} = \frac{M}{H}.$$

This result, M/H, is what we would get if all the mass were located at the center of the ball. ♦

The Moment of Inertia about a Line

In the study of rotation of an object about an axis, its **moment of inertia**, I, is used. It is defined as follows. The object occupies a region $\mathcal{R}$. Its density at a point P is $\delta(P)$, so its mass is $M = \int_{\mathcal{R}} \delta(P)\, dV$. Usually the density is constant, in which case it is M divided by the volume of $\mathcal{R}$ (or M divided by the area of $\mathcal{R}$ if $\mathcal{R}$ is planar). Let $r(P)$ be the distance from P to a line L. Then, by definition,

$$I = \text{Moment of Inertia about the line } L = \int_{\mathcal{R}} (r(P))^2 \delta(P)\, dV.$$

A similar definition holds for objects distributed on a plane region, with dV replaced by dA.

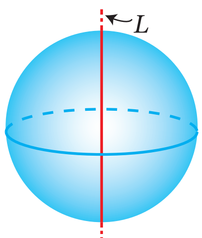


Figure 17.6.4

EXAMPLE 5. Compute the moment of inertia of a uniform object with mass M that has the form of a ball of radius a around a diameter L.

SOLUTION Since the object is uniform, the density $\delta(P)$ is constant, $\delta(P) = M/\left(4\pi a^3/3\right)$. We place the diameter L along the z-axis, as in Figure 17.6.4

We will compute the moment of inertia two ways. Because the distance $r(P)$ is r in cylindrical coordinates, we will first use those coordinates. Then we will calculate the moment of inertia in spherical coordinates.

Prior to evaluating any integrals, note that I must be less than Ma^2 since the maximum of $r(P)$ is a.

Cylindrical Coordinates

One description of the ball in cylindrical coordinates is

$$0 \le \theta \le 2\pi, \qquad -a \le z \le a, \qquad 0 \le r \le \sqrt{a^2 - z^2}.$$

Then, remembering the extra factor of r in the integrand,

$$I = \int_{\mathcal{R}} \frac{M}{\frac{4}{3}\pi a^3} r^2\, dV = \frac{3M}{4\pi a^3} \int_{\mathcal{R}} r^2\, dV = \frac{3M}{4\pi a^3} \int_0^{2\pi} \int_{-a}^{a} \int_0^{\sqrt{a^2-z^2}} r^3\, dr\, dz\, d\theta.$$

The innermost integration is

$$\int_0^{\sqrt{a^2-z^2}} r^3\, dr = \left.\frac{r^4}{4}\right|_0^{\sqrt{a^2-z^2}} = \frac{(a^2-z^2)^2}{4}.$$

The middle integral is then

$$\begin{aligned}
\int_{-a}^{a} \frac{(a^2-z^2)^2}{4}\, dz = \int_{-a}^{a} \frac{a^4 - 2a^2z^2 + z^4}{4}\, dz &= \frac{1}{4}\left.\left(a^4 z - \frac{2a^2z^3}{3} + \frac{z^5}{5}\right)\right|_{-a}^{a} \\
&= \frac{1}{4}\left(a^5 - \frac{2a^5}{3} + \frac{a^5}{5}\right) - \frac{1}{4}\left(-a^5 + \frac{2a^5}{3} - \frac{a^5}{5}\right) \\
&= \frac{4}{15}a^5.
\end{aligned}$$

And the outermost integral is

$$\int_0^{2\pi} \frac{4}{15}a^5\, d\theta = \frac{8\pi}{15}a^5.$$

Then, remembering to include the factor $3M/4\pi a^3$, we find the moment of inertia to be

$$I = \frac{3M}{4\pi a^3} \cdot \frac{8\pi}{15} a^5 = \frac{2}{5} M a^2.$$

Spherical Coordinates

Spherical coordinates provides an even simpler description of the ball. This does not always mean the resulting iterated integrals will be easier to evaluate.

The distance from a point P in the ball to the axis is, in spherical coordinates $r(P) = \sqrt{x^2 + y^2} = \rho \sin(\phi)$. The triple integral for the moment of inertia is

$$I = \frac{3M}{4\pi a^3} \int_{\mathscr{R}} (\rho \sin(\phi))^2 \, dV.$$

The corresponding iterated integral in spherical coordinates is

$$\int_0^{2\pi} \int_0^{\pi} \int_0^{a} (\rho \sin(\phi))^2 \rho^2 \sin(\phi) \, d\rho \, d\phi \, d\theta.$$

The innermost integral, with respect to ρ, is straightforward to evaluate:

$$\int_0^a \rho^4 \sin^3(\phi) \, d\rho = \frac{\rho^5}{5} \sin^3(\phi) \bigg|_{\rho=0}^{\rho=a} = \frac{a^5 \sin^3(\phi)}{5}.$$

The middle integral, with respect to ϕ, that awaits us is

$$\int_0^{\pi} \frac{a^5}{5} \sin^3(\phi) \, d\phi = \frac{a^5}{5} \int_0^{\pi} \sin^3(\phi) \, d\phi.$$

The odd power of $\sin(\phi)$ demands a little attention. Writing $\sin^3(\phi)$ as $(1 - \cos^2(\phi))\sin(\phi)$ allows this integral to be rewritten and evaluated as follows:

$$\begin{aligned}
\int_0^{\pi} \sin^3(\phi) \, d\phi &= \int_0^{\pi} (\sin(\phi) - \cos^2(\phi)\sin(\phi)) \, d\phi = \left(-\cos(\phi) + \frac{\cos^3(\phi)}{3} \right) \bigg|_0^{\pi} \\
&= \left(-(-1) + \frac{(-1)^3}{3} \right) - \left(-1 + \frac{1}{3} \right) = \frac{4}{3}.
\end{aligned}$$

For the final integration, remember the coefficient of $a^5/5$:

$$\frac{a^5}{5} \int_0^{2\pi} \frac{4}{3} \, d\theta = \frac{8\pi}{15}.$$

Bringing back the factor $3M/(4\pi a^3)$ gives the same moment of inertia as found with an iterated integral in cylindrical coordinates:

$$I = \frac{2}{5} M a^2.$$

◆

Observation 17.6.1: *The Result of Example 5 is Reasonable*

If the entire mass M was collected at a distance a from the axis line L, the moment of inertia would be Ma^2; so $I = 2Ma^2/5$ is plausible.

Looking back on the two approaches used to evaluate the triple integral in Example 5, note that the cylindrical coordinates provided the simpler description of the integrand and spherical coordinates provided the simpler description of the region. In this case, both iterated integrals were easy to evaluate.

PREVIEW: Exercise 27 shows that I plays the same role in a rotating body (such as a spinning skater) as mass does in an object moving along a line.

Summary

A triple integral $\int_{\mathscr{R}} f(P)\ dV$ may be evaluated by an iterated integral in cylindrical or spherical coordinates. In cylindrical coordinates the iterated integral takes the form

$$\int_{\theta_1}^{\theta_2}\int_{r_1(\theta)}^{r_2(\theta)}\int_{z_1(r,\theta)}^{z_2(r,\theta)} f(r,\theta,z)\ r\ dz\, dr\, d\theta.$$

The description of the region determines the limits of integration on the iterated integrals over intervals. (Changing the order of the description of $\mathscr{R}$ changes the order of the integrations.) The factor r must be inserted into the integrand.

In spherical coordinates the iterated integral usually appears as

$$\int_{\phi_1}^{\phi_2}\int_{\theta_1(\phi)}^{\theta_2(\phi)}\int_{\rho_1(\theta,\phi)}^{\rho_2(\theta,\phi)} f(\rho,\theta,\phi)\ \rho^2\sin(\phi)\ d\rho\ d\theta\ d\phi.$$

In this form, integration with respect to ρ is first, but as Example 4 illustrates, it may be convenient to integrate first with respect to ϕ or $theta$. The factor $\rho^2\sin(\phi)$ must be inserted in the integrand; sometimes this factor is precisely what is needed to make it possible to evaluate the triple integral.

EXERCISES for Section 17.6

In Exercises 1 to 4 (a) draw the region, (b) set up an iterated integral in cylindrical coordinates for the multiple integrals, and (c) evaluate the iterated integral.

1. $\int_{\mathscr{R}} r^2\ dV$, $\mathscr{R}$ is bounded by the cylinder $r = 3$ and the planes $z = 2x$ and $z = 3x$.

2. $\int_{\mathscr{R}} z\ dV$, $\mathscr{R}$ is bounded by the sphere $z^2 + r^2 = 25$, the plane $z = 0$, and the plane $z = 2$.

3. $\int_{\mathscr{R}} rz\ dV$, $\mathscr{R}$ is the part of the ball bounded by $r^2 + z^2 = 16$ in the first octant.

4. $\int_{\mathscr{R}} \cos(\theta)\ dV$, $\mathscr{R}$ is bounded by the cylinder $r = 2\cos(\theta)$, the paraboloid $z = r^2$, and the xy-plane.

5. Compute the volume of a right circular cone of height h and radius a using
(a) spherical coordinates, (b) cylindrical coordinates, and (c) rectangular coordinates.

6. Use cylindrical coordinates to find the volume of the region above $z = 0$ and below the paraboloid $z = 9 - r^2$.

7. A right circular cone of radius a and height h has a density at point P equal to the distance from P to the base of the cone. Use spherical coordinates to find its mass.

In Exercises 8 and 9 draw the region $\mathscr{R}$ and give a formula for the integrand $f(P)$ such that $\int_{\mathscr{R}} f(P)\, dV$ is described by the iterated integrals in spherical coordinates.

8. $\displaystyle\int_0^{\pi/2}\int_0^{\pi/4}\int_0^{\cos(\phi)} \rho^3 \sin^2(\theta)\sin(\phi)\, d\rho\, d\phi\, d\theta.$

9. $\displaystyle\int_0^{\pi/4} ; \int_{\pi/6}^{\pi/2}\int_0^{\sec(\theta)} \rho^3 \sin(\theta)\cos(\phi)\, d\rho\, d\phi\, d\theta.$

10. Let $\mathscr{R}$ be the solid region inside both the sphere $x^2 + y^2 + z^2 = 1$ and the cone $z = \sqrt{x^2 + y^2}$. Let the density at (x, y, z) be z. Set up iterated integrals for the mass in $\mathscr{R}$ using (a) rectangular coordinates, (b) cylindrical coordinates, (c) spherical coordinates. (d) Evaluate the iterated integral in (c).

11. Find the average temperature in a ball of radius a if the temperature is the square of the distance from a fixed equatorial plane.

In Exercises 12 and 13 evaluate the iterated integral in cylindrical coordinates.

12. $\displaystyle\int_0^{2\pi}\int_0^{1}\int_r^{1} zr^3\cos^2(\theta)\, dz\, dr\, d\theta$

13. $\displaystyle\int_0^{2\pi}\int_0^{a}\int_{-\sqrt{a^2-r^2}}^{\sqrt{a^2-r^2}} z^2 r\, dz\, dr\, d\theta$

14. Using cylindrical coordinates, find the volume of the region below the plane $z = y + 1$ and above the circle in the xy-plane whose center is $(0, 1, 0)$ and whose radius is 1. Include a drawing of the region.

15. Find the average distance from the center of a ball of radius a to other points of the ball by setting up iterated integrals in the three types of coordinate systems and evaluating the easiest.

16. A solid consists of that part of a ball of radius a that lies within a cone of vertex half-angle $\phi = \pi/6$, the vertex being at the center of the ball. Set up iterated integrals for $\int_{\mathscr{R}} z\, dV$ in three coordinate systems and evaluate the simplest.

In Exercises 17 to 22 evaluate the multiple integrals over a ball of radius a with center at the origin, without using an iterated integral.

17. $\displaystyle\int_{\mathscr{R}} \cos(\theta)\, dV$

18. $\displaystyle\int_{\mathscr{R}} \cos^2(\theta)\, dV$

19. $\displaystyle\int_{\mathscr{R}} z\, dV$

20. $\displaystyle\int_{\mathscr{R}} (3 + 2\sin(\theta))\, dV$

21. $\displaystyle\int_{\mathscr{R}} \sin^2(\phi)\, dV$

22. $\displaystyle\int_{\mathscr{R}} \cos^3(\phi)\, dV$

23. In polar, cylindrical, and spherical coordinates an extra factor appears in the integrand when writing an iterated integral. Why is that not necessary when using rectangular coordinates?

24. Is $\sqrt{a^2}$ always equal to a? Explain.

25. Using the method of Example 4 find the average value of q for all points P in the ball. The result is not the same as if the entire ball's mass were placed at its center.

26. (a) By integrating the function $f(P) = 1$, find the exact volume of the little curvy box corresponding to the changes $\Delta\rho$, $\Delta\theta$, $\Delta\phi$.

(b) Show that the ratio between that exact volume and our estimate, $\rho^2 \sin(\phi)\,\Delta\rho\,\Delta\theta\,\Delta\phi$ approaches 1 as $\Delta\rho$, $\Delta\theta$, and $\Delta\phi$ approach 0.

(c) Show that the exact volume in (a) can be written as $(\rho^*)^2 \sin(\phi^*)\ \Delta\rho\ \Delta\phi\ \Delta\theta$, where ρ^* is between ρ and $\rho + \Delta\rho$ and ϕ^* is between ϕ and $\phi + \Delta\phi$.

27. The kinetic energy of an object with mass m moving at the velocity v is $\frac{1}{2}mv^2$.

An object moving in a circle of radius r at the angular speed of ω radians per unit time has velocity $r\omega$. (Why?) Thus its kinetic energy is $\left(\frac{1}{2}mr^2\right)\omega^2$.

(a) The calculation of the kinetic energy of a mass M that occupies a region $\mathscr{R}$ in space involves an integral. The density of the mass is $\delta(P)$, which may vary from point to point. Let $r(P)$ be the distance from P to a fixed line L. If the mass is spinning around the axis L at the angular rate ω, show that its total kinetic energy is $\int\limits_{\mathscr{R}} \frac{1}{2}\delta(P)\ (r(P))^2\ \omega^2\ dV$.

(b) The formula found in (a) can be written as Kinetic Energy $= \frac{1}{2}I\omega^2$. What is the formula for I, the moment of inertia of the object about its axis L?

Observation 17.6.2:

Thus I plays the same role in rotational motion that mass m plays in linear motion in the formula $\frac{1}{2}mv^2$ for kinetic energy.

Every spinning ice skater knows this. When spinning with her arms extended she has a certain amount of kinetic energy. If she puts her arms to her sides she decreases her moment of inertia but has not destroyed her kinetic energy. That forces her angular speed to increase. The larger I is, the harder it is to stop it when it is spinning.

In Exercises 28 to 32 the objects have a homogeneous (constant density) mass M. Find the moment of inertia, I, of the given object relative to the given axis.

28. A rectangular box of dimensions, $a \times b \times c$, rotating around a line through the center of the box and perpendicular to the face of dimensions $a \times b$.

29. A solid cylinder of radius a and height h rotating around its axis.

30. A solid cylinder of radius a and height h rotating around a line (on its surface) parallel to the cylinder's axis.

31. A cylindrical tube of height h, inner radius a, and outer radius b, rotating around its axis.

32. A solid cylinder of radius a and height h rotating around a diameter in its base.

---

33. Verify the claims at the beginning of Example 2 about the minimum and maximum heights of points on the intersection of the plane $z = x + 2y + 9$ and the cylinder $x^2 + y^2 = 9$.

34. In Example 2 the region $\mathcal{R}$ was parameterized as $0 \le \theta \le 2\pi$, $0 \le r \le 3$, $0 \le z \le r\cos(\theta) + 2r\sin(\theta) + 9$. There are five other possible orderings of the variables for the parameterization of $\mathcal{R}$. Some of these orderings require splitting the region into two or more pieces to write the parameterization. Which, if any, of these orderings can parameterize the entire region in one piece? You do not have to find each parameterization to answer this question.

35. Solve Example 2 using rectangular coordinates.

36. Evaluate the moment of inertia in Example 5 when the region $\mathcal{R}$ has the description $0 \le \theta \le 2\pi$, $0 \le r \le a$, $-\sqrt{a^2 - r^2} \le z \le \sqrt{a^2 - r^2}$.

37. $\mathcal{R}$ is a solid ball of radius a with center at the origin of a coordinate system.

(a) Explain why $\int_{\mathcal{R}} x^2\, dV = \frac{1}{3} \int_{\mathcal{R}} (x^2 + y^2 + z^2)\, dV$.

(b) Evaluate the second integral by spherical coordinates.

(c) Use (b) to find $\int_{\mathcal{R}} x^2\, dV$.

38. If $\mathcal{R}$ is a ball centered at the origin of a rectangular coordinate system, show that $\int_{\mathcal{R}} (x^3 + y^3 + z^3)\, dV = 0$.

39. A homogeneous object with mass M occupies the region $\mathcal{R}$ between concentric spheres of radii a and b, $a < b$. Assume point A is at a distance H from their center, $H < a$. Evaluate $\int_{\mathcal{R}} \frac{\delta}{q}\, dV$, where δ is the density and $q = q(P)$ is the distance from H to any point P in $\mathcal{R}$.

OBSERVATION That the value of the integral does not involve H implies that a uniform hollow sphere exerts no gravitational force on objects in its interior.

Exercises 40 and 41 are a continuation of Example 4, and are related.

40. In Example 4, H is greater than a. Solve the same problem for H less than a.

41. Assume point A is in the plane of a disk but outside the disk. Is the average of the reciprocal of the distance from A to points in the disk equal to the reciprocal of the distance from A to the center of the disk?

---

42. Show that the result of Example 4 holds if the density $\delta(P)$ depends only on ρ, the distance to the center. Assume $\delta(P) = g(\rho)$.

NOTE: This is approximately the case with Earth, which is not homogeneous, but consists of concentric shells.

43. (a) A ball of radius a is not homogeneous. However, its density at P depends only on the distance from P to the center of the ball. That is, there is a function $g(\rho)$ such that the density at $P = (\rho, \theta, \phi)$ is $g(\rho)$. Using an iterated integral, show that the mass of the ball is

$$4\pi \int_0^a g(\rho)\rho^2\, d\rho.$$

(b) Obtain the same formula using the method of Chapter 6, that is, by an integral over an interval.

44. $\mathscr{R}$ is the part of a ball of radius a removed by a cylindrical drill of diameter a whose edge passes through the center of the sphere. (a) Sketch $\mathscr{R}$. (b) Find the volume of $\mathscr{R}$.

45. $\mathscr{R}$ is the ball of radius a. For any point P in the ball other than its center, define $f(P)$ to be the reciprocal of the distance from P to the origin. The average value of r over $\mathscr{R}$ involves an improper integral, since the function is unbounded near the origin. Does the improper integral converge or diverge? What is the average value of f over $\mathscr{R}$?

In Exercises 46 and 47 check the equations by differentiation.

46. $\int \frac{dx}{1+\cos(x)} = \tan\left(\frac{x}{2}\right) + C$

47. $\int \frac{x\,dx}{1+\cos(x)} = x\tan\left(\frac{x}{2}\right) + 2\ln\left|\cos\left(\frac{x}{2}\right)\right| + C$

17.7 Integrals Over Surfaces and Steradians

In this section we treat integrals over surfaces that may not be flat. We assume that the surface is smooth, or composed of a finite number of smooth pieces, and that the integrals we define exist.

Definition of a Surface Integral

Let $\mathscr{S}$ be a surface such as the surface of a ball (a sphere) or part of the saddle $z = xy$. If f is a scalar function defined on $\mathscr{S}$, we will define the integral $\int_{\mathscr{S}} f(P)\,dS$. The definition is practically identical with the definition of the double integral, which is the special case when the surface is a plane.

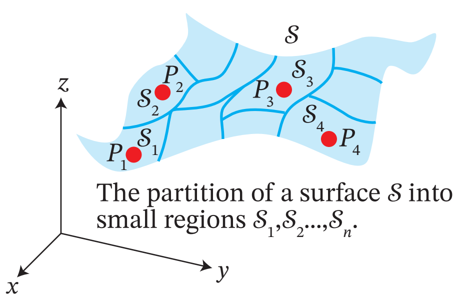


Figure 17.7.1

If $\mathscr{S}$ is a surface and $f(P)$ a scalar function defined on $\mathscr{S}$, the **surface integral** $\int_{\mathscr{S}} f(P)\,dS$ is defined just like a double integral over a flat surface. The only difference is that the partitions of $\mathscr{S}$ involve small curved surfaces, as shown in Figure 17.7.1.

If $f(P)$ is 1 for each point P in $\mathscr{S}$ then $\int_{\mathscr{S}} f(P)\,dS$ is the area of the surface $\mathscr{S}$. If $\mathscr{S}$ is occupied by material of surface density $\sigma(P)$ at P then $\int_{\mathscr{S}} \sigma(P)\,dS$ is the total mass of the surface $\mathscr{S}$.

If matter is distributed on the surface $\mathscr{S}$, its density at a point P in $\mathscr{S}$ is defined much the way density in a lamina was defined in Section 17.1. The only difference is that instead of considering a small disk around P one considers a small patch on $\mathscr{S}$ that contains P. "Small" means that the patch fits in a ball of radius r, and we let r approach 0.

First we show how to integrate over a sphere.

Integrating over a Sphere

If the surface $\mathscr{S}$ is a sphere or part of a sphere it is often convenient to evaluate an integral over it with the aid of spherical coordinates.

If the center of a spherical coordinate system (ρ, θ, ϕ) is at the center of a sphere of radius a, then ρ is constant on the sphere, $\rho = a$. As Figure 17.7.2 suggests, the area of the small region on the sphere corresponding to slight changes $d\theta$ and $d\phi$ is approximately

$$\left(a\,d\phi\right)\left(a\sin(\phi)\,d\theta\right) = a^2\sin(\phi)\,d\theta\,d\phi.$$

Thus we may write

$$dS = a^2 \sin(\phi)\, d\theta\, d\phi$$

and evaluate $\int_{\mathscr{S}} f(P)\, dS$ as an iterated integral in ϕ and θ. Example 1 illustrates how this works in a surface integral.

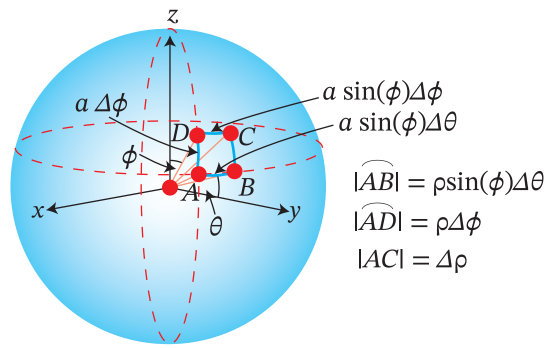


Figure 17.7.2

EXAMPLE 1. A sphere of radius a has its center at the origin of an xyz-coordinate system. The top half of the sphere, where z is positive, will be denoted as $\mathscr{S}$. Evaluate $\int_{\mathscr{S}} z\, dS$.

SOLUTION Since the sphere has radius a, $\rho = a$. The top half of the sphere is described by $0 \le \theta \le 2\pi$ and $0 \le \phi \le \pi/2$. In spherical coordinates $z = \rho\cos(\phi) = a\cos(\phi)$. Thus

$$\int_{\mathscr{S}} z\, dS = \int_{\mathscr{S}} (a\cos(\phi))\, dS = \int_0^{2\pi}\int_0^{\pi/2} (a\cos(\phi))\, a^2 \sin(\phi)\, d\phi\, d\theta.$$

First, the inner integral, with respect to ϕ, can be evaluated as follows:

$$\begin{aligned}\int_0^{\pi/2} (a\cos(\phi))a^2\sin(\phi)\, d\phi &= a^3\int_0^{\pi/2} \cos(\phi)\sin(\phi)\, d\phi \\ &= a^3 \frac{\sin^2(\phi)}{2}\Big|_0^{\pi/2} = \frac{a^3}{2}(1-0) = \frac{a^3}{2}.\end{aligned}$$

To complete the evaluation of this surface integral, evaluate the outer integral:

$$\int_{\mathscr{S}} z\, dS = \int_0^{2\pi} \frac{a^3}{2}\, d\theta = \left(\frac{a^3}{2}\right) 2\pi = \pi a^3.$$

♦

The result in Example 1 can be interpreted in terms of average value. The **average value** of $f(P)$ over a surface $\mathscr{S}$ is defined similarly to the average values of a function over a region in a plane or in space:

$$\text{Average value of } f(P) \text{ over a surface } \mathscr{S} = \frac{1}{\text{Area of } \mathscr{S}} \int_{\mathscr{S}} f(P)\, dS.$$

Observation 17.7.1:

Example 1 shows that the average value of z over the given hemisphere is

$$\frac{1}{\text{Area of } \mathscr{S}} \int_{\mathscr{S}} z\, dS = \frac{\pi a^3}{2\pi a^2} = \frac{a}{2}.$$

The average height above the equator is exactly half the radius.

We now turn our attention to developing our understanding of surface area.

Definition: *Surface Area*

The **surface area** of a surface $\mathscr{S}$ is

$$\text{(Surface) Area} = \int_{\mathscr{S}} 1\, dS.$$

CLARIFICATION: When we talk about the "area of a surface" we mean its surface area: $\int_{\mathscr{S}} 1\, dS$. The area of a plane region $\mathscr{R}$, $\int_{\mathscr{R}} 1\, dA$, is different (unless $\mathscr{S}$ is planar).

There is a simple relationship between the surface integrals and double integrals, as will be revealed next.

A General Technique

An integral over a curve, $\int_C f\, ds$, is evaluated by replacing ds with $(ds/dt)\, dt$ and evaluating $\int_a^b f\,(ds/dt)\, dt$, an integral over an interval. We will do something similar for an integral over a surface: We will replace a surface integral by a double integral over a set in a coordinate plane.

Formula 17.7.1: *Approximate Relationship Between dA and dS*

The idea is to replace a small patch on the surface $\mathscr{S}$ by its projection on a plane, say, the xy-plane. The area of the projection is not the same as the area of the patch. With the aid of Figure 17.7.3 the area of the shadow can be expressed in terms of the tilt of the patch.

The unit normal vector to the patch is $\mathbf{n}$. The (upward) unit normal vector to the shadow in the xy-plane is $\mathbf{k}$. The angle between $\mathbf{n}$ and $\mathbf{k}$ is the direction angle γ. Call the area of the patch dS, and the area of its projection dA. Then $dA \approx |\cos(\gamma)|\, dS$ or, if $\cos(\gamma)$ is not 0,

$$dS \approx \frac{dA}{|\cos(\gamma)|}. \qquad (17.7.1)$$

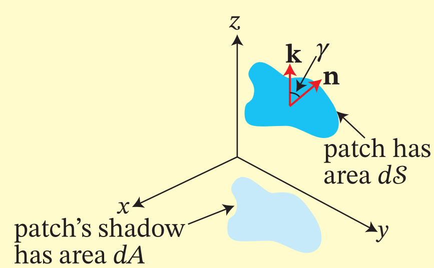


Figure 17.7.3

Direction angles and direction cosines were introduced in Section 14.4. See also Exercise 27.

For instance, if $\gamma = 0$, then $dA = dS$. If $\gamma = \pi/2$, then $dA = 0$ and dS is not defined. (In this case, project the small patches of $\mathscr{S}$ onto another coordinate plane and use the corresponding direction angle.) We use the absolute value of $\cos(\gamma)$, since γ could be larger than $\pi/2$.

Observation 17.7.2: *Basic Relationship between Surface Integrals and Double Integrals*

With the aid of (17.7.1), we replace a surface integral over a surface $\mathscr{S}$ with a double integral over its shadow in the xy-plane.

Let $\mathscr{S}$ be a surface that meets each line parallel to the z-axis at most once. Let f be a function whose domain includes $\mathscr{S}$. An approximating sum for $\int_{\mathscr{S}} f(P)\, dS$ is $\sum_{i=1}^n f(P_i)\Delta S_i$, where ΔS_i is the area of the small patch $\mathscr{S}_i$ in a typical partition of $\mathscr{S}$. The partition of $\mathscr{S}$ is shown in Figure 17.7.4.

Let $\mathscr{A}$ be the projection of $\mathscr{S}$ onto the xy-plane. The patch $\mathscr{S}_i$, with (surface) area ΔS_i, projects to $\mathscr{A}_i$, with area ΔA_i, and the point P_i on $\mathscr{S}_i$ projects down to Q_i in $\mathscr{A}_i$. Let γ_i be the direction angle between the unit normal vector to $\mathscr{S}$ at P_i and $\mathbf{k}$ (the unit normal vector to $\mathscr{A}$ at Q_i). Then $f(P_i)\Delta\mathscr{S}_i$ is approximately $(f(P_i)/|\cos(\gamma_i)|)\Delta A_i$. Thus an approximation of $\int_{\mathscr{S}} f(P)\, dS$ is

$$\sum_{i=1}^n \frac{f(P_i)}{|\cos(\gamma_i)|}\Delta A_i.$$

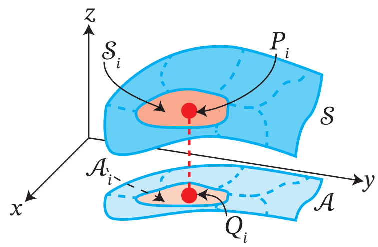


Figure 17.7.4

Taking the limit as the $\mathscr{A}_i$ are chosen smaller and smaller yields the following theorem.

Theorem 17.7.3: Evaluating a Surface Integral as a Double Integral

Assume $\mathscr{S}$ is a surface, $\mathscr{A}$ is its projection onto the xy-plane, and f is a function defined on $\mathscr{S}$. Assume that, for each point Q in $\mathscr{A}$, the line through Q parallel to the z-axis meets $\mathscr{S}$ in exactly one point, P. Define h on $\mathscr{A}$ by

$$h(Q) = f(P).$$

Then

$$\int_{\mathscr{S}} f(P)\, dS = \int_{\mathscr{A}} \frac{h(Q)}{|\cos(\gamma)|}\, dA.$$

*In this equation γ denotes the direction angle between the upward unit normal vector to the xy-plane, **k**, and **n**, a unit vector normal to the surface $\mathscr{S}$ at P. (See Figure 17.7.5.)*

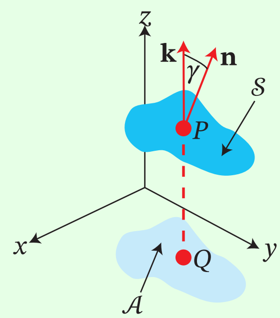


Figure 17.7.5

The key to being able to apply Theorem 17.7.3 is being able to compute the direction angle $\cos(\gamma)$.

Computing $\cos(\gamma)$

We find a vector perpendicular to the surface in order to compute $\cos(\gamma)$. If $\mathscr{S}$ is the level surface of $g(x, y, z)$, that is, $g(x, y, z) = c$ for some constant c, then for any point P on $\mathscr{S}$, that is, $g(P) = c$, the gradient $\nabla g(P)$ is a normal vector to $\mathscr{S}$ at P.

If the surface $\mathscr{S}$ is given as $z = f(x, y)$, it is a level surface of $g(x, y, z) = z - f(x, y)$. Theorem 17.7.4 provides formulas for $\cos(\gamma)$. However, it is unnecessary to memorize them. Just remember that a gradient provides a normal to a level surface.

Theorem 17.7.4: Two Formulas for $\cos(\gamma)$

Case 1: Level Surface. *If the surface $\mathscr{S}$ is part of the level surface $g(x, y, z) = c$, then*

$$|\cos(\gamma)| = \frac{|g_z|}{\sqrt{(g_x)^2 + (g_y)^2 + (g_z)^2}}. \tag{17.7.2}$$

Case 2: Graph of a Function. *If the surface $\mathscr{S}$ is given in the form $z = f(x, y)$, then*

$$|\cos(\gamma)| = \frac{1}{\sqrt{(f_x)^2 + (f_y)^2 + 1}}. \tag{17.7.3}$$

Proof of Theorem 17.7.4

Case 1: Level Surface. A normal vector to $\mathscr{S}$ at a point is provided by the gradient

$$\nabla g = g_x\mathbf{i} + g_y\mathbf{j} + g_z\mathbf{k}.$$

The cosine of the angle between $\mathbf{k}$ and ∇g is

$$\cos(\gamma) = \frac{\mathbf{k}\cdot\nabla g}{|\mathbf{k}|\,|\nabla g|} = \frac{\mathbf{k}\cdot\left(g_x\mathbf{i}+g_y\mathbf{j}+g_z\mathbf{k}\right)}{(1)\sqrt{\left(g_x\right)^2+\left(g_y\right)^2+\left(g_z\right)^2}}.$$

Therefore

$$|\cos(\gamma)| = \frac{|g_z|}{\sqrt{\left(g_x\right)^2+\left(g_y\right)^2+\left(g_z\right)^2}}.$$

Case 2: Graph of a Function. Rewrite $z = f(x, y)$ as $z - f(x, y) = 0$. The surface $z = f(x, y)$ is thus the level surface $g(x, y, z) = 0$, where $g(x, y, z) = z - f(x, y)$. Because $g_x = -f_x$, $g_y = -f_y$, and $g_z = 1$, Formula 17.7.2 for $\cos(\gamma)$ can be applied to yield

$$|\cos(\gamma)| = \frac{1}{\sqrt{\left(f_x\right)^2+\left(f_y\right)^2+1}}.$$

Observation 17.7.5: *Formulas for* $\cos(\alpha)$ *and* $\cos(\beta)$

Theorem 17.7.4 is stated for projections onto the xy-plane. Similar theorems hold for projections onto the xz- or yz-planes. The direction angle γ is then replaced by the corresponding direction angle, β or α, and the normal vector is dotted with $\mathbf{j}$ or $\mathbf{i}$. Draw a picture in each case; there is no point in trying to memorize the formulas.

EXAMPLE 2. Find the area of the part of the saddle $z = xy$ inside the cylinder $x^2 + y^2 = a^2$.

SOLUTION Let $\mathscr{S}$ be the part of the surface $z = xy$ inside the cylinder $x^2 + y^2 = a^2$. Then

$$\text{Area of } \mathscr{S} = \int_{\mathscr{S}} 1\, dS.$$

To evaluate this surface integral, first recognize that the projection of $\mathscr{S}$ onto the xy-plane is a disk of radius a and center $(0, 0)$.

Figure 17.7.6 shows the surface $z = xy$ shifted vertically to eliminate any visual overlapping with its projection, $\mathscr{A}$, onto the xy-plane. Then

$$\begin{aligned}\text{Area of } \mathscr{S} &= \int_{\mathscr{S}} 1\, dS\\ &= \int_{\mathscr{A}} \frac{1}{|\cos(\gamma)|}\, dA. \end{aligned} \tag{17.7.4}$$

To find an expression for $|\cos(\gamma))|$, we use (17.7.2). A normal vector to $\mathscr{S}$ is found by first writing $z = xy$ as $z - xy = 0$. This means $\mathscr{S}$ is a level surface of $g(x, y, z) = z - xy$, and a normal vector to $\mathscr{S}$ is

$$\nabla g = \frac{\partial g}{\partial x}\mathbf{i} + \frac{\partial g}{\partial y}\mathbf{j} + \frac{\partial g}{\partial z}\mathbf{k} = -y\mathbf{i} - x\mathbf{j} + \mathbf{k}.$$

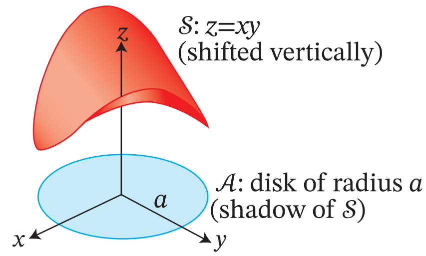


Figure 17.7.6

Then

$$\cos(\gamma) = \frac{\mathbf{k}\cdot\nabla g}{|\mathbf{k}|\ |\nabla g|} = \frac{\mathbf{k}\cdot(-y\mathbf{i} - x\mathbf{j} + \mathbf{k})}{\sqrt{y^2 + x^2 + 1}} = \frac{1}{\sqrt{y^2 + x^2 + 1}}.$$

By (17.7.4),

$$\text{Area of } \mathscr{S} = \int_{\mathscr{A}} \sqrt{y^2 + x^2 + 1}\, dA. \tag{17.7.5}$$

Now, the double integral in (17.7.5) is most naturally evaluated using polar coordinates:

$$\int_{\mathscr{A}} \sqrt{y^2 + x^2 + 1}\, dA = \int_0^{2\pi}\int_0^a \sqrt{r^2 + 1}\, r\, dr\, d\theta.$$

The inner integration, with respect to r, can be evaluated with the use of a substitution $u = r^1 + 1$, so $du = 2r\, dr$:

$$\begin{aligned}\int_0^a \sqrt{r^2 + 1}\, r\, dr &= \left.\frac{(r^2 + 1)^{3/2}}{3}\right|_0^a \\ &= \frac{(1 + a^2)^{3/2} - 1}{3}.\end{aligned}$$

Since this result, which is the integrand for the outer integral, does not depend on the polar angle, θ, the outer integral is particularly easy to evaluate:

$$\int_0^{2\pi} \frac{(1 + a^2)^{3/2} - 1}{3}\, d\theta = \frac{2\pi}{3}\left((1 + a^2)^{3/2} - 1\right).$$

The surface $z = xy$ inside the cylinder $x^2 + y^2 = a^2$ has (surface) area $2\pi\left((1 + a^2)^{3/2} - 1\right)/3$. ♦

A Geometric Application: Steradians

> Steradians comes from *stereo*, the Greek word for space, and *radians.*

Let $\mathscr{S}$ be a surface such that each ray from the point O meets $\mathscr{S}$ in at most one point. $\mathscr{S}$ subtends a solid angle at O. It consists of all rays from O that meet $\mathscr{S}$, as shown in Figure 17.7.7(a). To measure this solid angle introduce a sphere with center at O. Call its radius a. The solid angle intercepts a patch on the sphere. Call the area of that patch A. (See Figure 17.7.7(b).) The quotient A/a^2 is the **steradian measure** of the solid angle. For instance, take the case of a surface $\mathscr{S}$ that completely surrounds the point O. If the sphere has radius a, the area, A, of the patch determined by $\mathscr{S}$ is $4\pi a^2$, the surface area of the sphere. Thus the steradian measure of the solid angle subtended by $\mathscr{S}$ is $(4\pi a^2)/a^2 = 4\pi$. This is the analog of the fact that a closed convex curve subtends an angle of 2π radians at any point in the region it bounds.

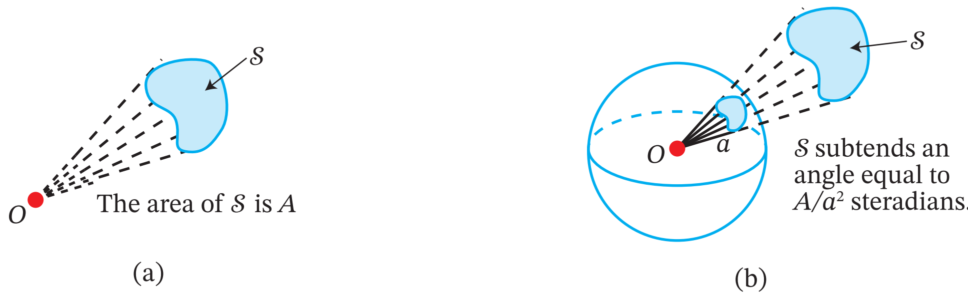


Figure 17.7.7

EXAMPLE 3. Find how large is the angle subtended by one face of a cube at its center.

SOLUTION Imagine a large sphere containing the cube and having center at the center of the cube. The entire surface of the cube subtends an angle of 4π steradians. Because there are six identical faces, each face subtends $4\pi/6 = 2\pi/3$ steradians, which is a little more than 2 steradians. ♦

In Section 15.4 the radians subtended by a curve C was expressed as a line integral over C, namely $\int_C (\widehat{\mathbf{r}}\cdot\mathbf{n})/r\,ds$. Almost identical reasoning shows that $\int_{\mathscr{S}} (\widehat{\mathbf{r}}\cdot\mathbf{n})/r^2\,dS$ equals the steradian measure of the angle subtended by a surface $\mathscr{S}$. (Recall that the angle between $\widehat{\mathbf{r}}$ and $\mathbf{n}$ is assumed to be acute or right, never obtuse.)

Exercise 26 outlines the reasoning leading to this definition.

The next example shows how to use the geometry of steradians to evaluate a surface integral.

EXAMPLE 4. One corner C of a cube of side b is at the origin of xyz-space. Find $\displaystyle\int_{\mathscr{S}} \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r^2}\,dS$ where $\mathscr{S}$ is one face of the cube that does not contain the origin.

SOLUTION The surface integral

$$\int_{\mathscr{S}} (\widehat{\mathbf{r}}\cdot\mathbf{n})/r^2\,dS$$

equals the steradian measure of the angle subtended by that face at C. Eight identical cubes of side b, all having C as a corner, fill up the space around C and form one large cube. The surface of the large cube consists of 24 congruent squares, each of which subtends the same angle at C. Because the origin is contained within the large cube, the angle subtended by it is 4π. Thus one face subtends $4\pi/24$ steradians at C. Consequently, the value of $\int_{\mathscr{S}} (\widehat{\mathbf{r}}\cdot\mathbf{n})/r^2\,dS$ is $4\pi/24 = \pi/6$, which is about 0.52. ♦

The following special case will be used in Section 18.5. It is the analog in space of the fact that in the plane $\oint_C (\widehat{\mathbf{r}}\cdot\mathbf{n})/r\,ds = 2\pi$ if the curve C encloses the origin.

Theorem 17.7.6: Closed Surfaces Subtend a Solid Angle of 4π Steradians

Let $\mathcal{O}$ be a point in the region bounded by the closed surface $\mathscr{S}$. Assume each ray from $\mathcal{O}$ meets $\mathscr{S}$ in exactly one point. Denote by $\mathbf{r}$ the position vector from $\mathcal{O}$ to that point, the length of $\mathbf{r}$ is r, and $\widehat{\mathbf{r}}$ is the unit vector in the same direction as $\mathbf{r}$. The unit exterior normal (that makes an acute or right angle with $\widehat{\mathbf{r}}$) is denoted as $\mathbf{n}$. Then

$$\int_{\mathscr{S}} \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r^2}\,dS = 4\pi. \qquad (17.7.6)$$

Observation 17.7.7: ***Emphasizing Theorem 17.7.6***

When $\mathscr{S}$ is a sphere of radius a and $\mathscr{O}$ is its center, it is easy to check that (17.7.6) is true. In this case we have $\hat{\mathbf{r}} = \mathbf{n}$, so $\hat{\mathbf{r}} \cdot \mathbf{n} = 1$. The integrand in (17.7.6) is $1/a^2$ and $\int_{\mathscr{S}} 1/a^2 \, dS = (1/a^2)4\pi a^2 = 4\pi$.

But, Theorem 17.7.6 says a lot more. In particular, it says that (17.7.6) holds when $\mathscr{S}$ is any sphere that contains the origin (even if it is not the sphere's center) or when $\mathscr{S}$ is any other surface that contains the origin.

Summary

After defining $\int_{\mathscr{S}} f(P)\, dS$, an integral over a surface, we showed how to compute it when the surface is part of a sphere or can be projected onto a coordinate plane. For a sphere of radius a, $\int_{\mathscr{S}} f(P)\, dS$ equals

$$\int_0^{2\pi}\int_0^{\pi} f(a,\theta,\phi)\, a^2 \sin(\phi)\, d\phi\, d\theta.$$

If each line parallel to the z-axis meets a surface $\mathscr{S}$ in at most one point an integral over $\mathscr{S}$ can be replaced by an integral over $\mathscr{A}$, the projection of $\mathscr{S}$ on the xy-plane:

$$\int_{\mathscr{S}} f(P)\, dS = \int_{\mathscr{A}} \frac{h(Q)}{|\cos(\gamma)|}\, dA.$$

To find $\cos(\gamma)$, use a gradient. If the surface is a level surface it can be represented as $g(x,y,z) = c$. If the surface is the graph of a function, $z = f(x,y)$, then it can be recognized as the level surface of $g(x,y,z) = z - f(x,y)$. In either case, $\nabla(g)$ is a normal vector to the surface.

We also defined the steradian measure of a solid angle and related it to an integral of the vector field $\hat{\mathbf{r}}/r^2$. In particular, for a surface $\mathscr{S}$ that encloses the origin, $\int_{\mathscr{S}} (\hat{\mathbf{r}}\cdot\mathbf{n})/r^2\, dS$ equals 4π, a fact that will be needed in the next chapter.

EXERCISES for Section 17.7

1. A small patch of a surface makes an angle of $\frac{\pi}{4}$ with the xy-plane. Its projection on that plane has area 0.05. Estimate the area of the patch.

2. A small patch of a surface makes an angle of 25° with the yz-plane. Its projection on that plane has area 0.03. Estimate the area of the patch.

3. Find the area of that part of the sphere of radius a that lies within a cone of vertex half-angle $\frac{\pi}{4}$ and vertex at the center of the sphere, as in Figure 17.7.8.

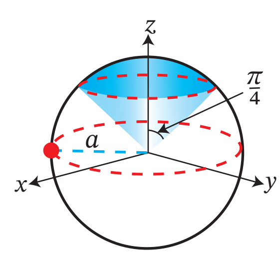


Figure 17.7.8

4. (a) Draw a diagram of the part of the plane $x + 2y + 3z = 12$ that lies inside the cylinder $x^2 + y^2 = 9$.
(b) Find that area using integration.
(c) Find that area using vectors.

5. (a) Draw a diagram of the part of the plane $z = x + 3y$ that lies inside the cylinder $r = 1 + \cos(\theta)$.
(b) Find its area.

6. Let $f(P)$ be the square of the distance from P to a diameter of a sphere of radius a. Find the average value of $f(P)$ for points on the sphere.

In Exercises 7 and 8 evaluate $\int_{\mathscr{S}} \mathbf{F}\cdot\mathbf{n}\, dS$ for the given sphere and vector field ($\mathbf{n}$ is the outward unit normal.)

7. The sphere $x^2+y^2+z^2=9$ and $\mathbf{F}=x^2\mathbf{i}+y^2\mathbf{j}+z^2\mathbf{k}$.

8. The sphere $x^2+y^2+z^2=1$ and $\mathbf{F}=x^3\mathbf{i}+y^2\mathbf{j}$.

---

9. Find the area of the part of the spherical surface $x^2+y^2+z^2=1$ that lies within the vertical cylinder erected on the circle $r=\cos\theta$ and above the xy-plane.

10. Find the area of that portion of the parabolic cylinder $z=\frac{1}{2}x^2$ between the three planes $y=0$, $y=x$, and $x=2$.

11. Evaluate $\int_{\mathscr{S}} x^2y\, dS$, where $\mathscr{S}$ is the portion in the first octant of a sphere with radius a and center at the origin, as follows: (a) Set up an integral using x and y as parameters. (b) Set up an integral using ϕ and θ as parameters. (c) Evaluate the easier of (a) and (b).

12. A triangle in the plane $z=x+y$ is directly above the triangle in the xy-plane whose vertices are $(1,2)$, $(3,4)$, and $(2,5)$. Find the area of (a) the triangle in the xy-plane and (b) the triangle in the plane $z=x+y$.

13. Let $\mathscr{S}$ be the triangle with vertices $(1,1,1)$, $(2,3,4)$, and $(3,4,5)$.
(a) Find the area of $\mathscr{S}$ using vectors. (b) Find the area of $\mathscr{S}$ using a surface integral.

14. Find the area of the portion of the cone $z^2=x^2+y^2$ that lies above one loop of the curve $r=\sqrt{\cos(2\theta)}$.

15. Let $\mathscr{S}$ be the triangle whose vertices are $(1,0,0)$, $(0,2,0)$, and $(0,0,3)$. Let $f(x,y,z)=3x+2y+2z$. Evaluate $\int_{\mathscr{S}} f(P)\, dS$.

16. An electric field radiates power at the rate of $\dfrac{k\sin^2(\phi)}{\rho^2}$ units per unit of area to the point $P=(\rho,\theta,\phi)$. Find the total power radiated to the sphere $\rho=a$.

17. A spherical surface of radius $2a$ has its center at the origin of a rectangular coordinate system. A circular cylinder of radius a has its axis parallel and at a distance a from the z-axis. Find the area of that part of the sphere that lies within the cylinder and is above the xy-plane.

A mass M is distributed on the surface $\mathscr{S}$. Let its density at P be $\sigma(P)$. The **moment of inertia** of the mass around the z-axis is defined as $\int_{\mathscr{S}} \left(x^2+y^2\right)\sigma(P)\, dS$. Exercises 18 and 19 concern this integral.

18. Find the moment of inertia around a diameter of a sphere of radius a on which mass is uniformly distributed.

19. Let a, b, and c be positive numbers. Find the moment of inertia about the z-axis of a homogeneous distribution of mass, M, of the triangle whose vertices are $(a,0,0)$, $(0,b,0)$, and $(0,0,c)$.

Exercises 20 and 21 involve calculations similar to those in Section 17.6 for a ball instead of its surface.

20. Let $\mathscr{S}$ be a spherical surface of radius a. Let A be a point at distance $b>a$ from the center of $\mathscr{S}$. For P in $\mathscr{S}$ let $f(P)$ be $\frac{1}{q}$, where q is the distance from P to A. Show that the average of $f(P)$ over $\mathscr{S}$ is $\frac{1}{b}$.

21. The data are the same as in Exercise 20 but $b<a$. Show that in this case the average of $\frac{1}{q}$ is $\frac{1}{a}$. The average does not depend on b in this case.

Exercises 22 to 26 deal with steradians. Answer Exercises 22 to 25 using only the definition and geometry.

22. A right circular cone has angle 2α radians at its vertex. What is the steradian measure of the solid angle its base subtends at the vertex (a) when α is less than $\frac{\pi}{2}$ and (b) when α approaches $\frac{\pi}{2}$ (from the left)?

23. Let $\mathscr{C}$ be a convex body bounded by the smooth surface $\mathscr{S}$. Show that if the origin is outside of $\mathscr{C}$, then the integral of $\frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r^2}$ over $\mathscr{S}$ is 0.

24. Let $\mathscr{C}$ be a convex body bounded by the smooth surface $\mathscr{S}$. (A surface is **smooth** if it has a continuous unit normal vector and no planar parts.) Let $\mathbf{n}$ denote the external normal to $\mathscr{S}$. Assume that the origin is on the surface $\mathscr{S}$. Use steradians to show that $\int_{\mathscr{S}} \frac{\mathbf{r}\cdot\mathbf{n}}{r^3}\,dS = 2\pi$.

25. (a) How many steradians are there in the solid angle subtended at a corner of a cube by its surface?
(b) Does this contradict Exercise 24?

26. This exercise shows that the steradian measure of the angle subtended by a surface $\mathscr{S}$ is measured by $\int_{\mathscr{S}} \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r^2}\,dS$.

(a) Draw a surface and a small patch $\mathscr{S}$ on it.
(b) Draw a sphere with center $\mathscr{O}$ that does not meet the patch.
(c) On the sphere draw the corresponding patch where rays from the origin to $\mathscr{S}$ meet the sphere.
(d) Following the approach used in Section 15.4 to write the radians subtended by a curve in terms of a line integral, complete the derivation of the steradian measure of a surface in terms of a surface integral.

Exercise 27 provides an alternate derivation of Formula 17.7.1.

27. Consider a region U in a plane tilted at some angle θ to a plane $\mathscr{P}$. The projection of U on $\mathscr{P}$ consists of all points where lines through U and perpendicular to $\mathscr{P}$ meet $\mathscr{P}$. The area of the projection of U onto $\mathscr{P}$ depends on the area of U and the angle between the plane of U and the plane $\mathscr{P}$.

Assume, as in Figure 17.7.9, that $0 < \theta < \frac{\pi}{2}$ and the line of intersection of the two planes is used as an axis. Relative to this axis, define $c(x)$ to be the length of the cross section of U for all x in the interval $[a, b]$.

(a) What is the length of the corresponding cross sections of V, the projection of U on $\mathscr{P}$?
(b) Show that (Area of V) = $\cos(\theta)$(Area of U).
(c) Explain why this result is also true when θ is zero.

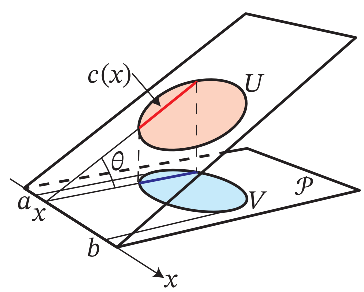


Figure 17.7.9

Spherical coordinates are also useful for integrating over a right circular cone as in Exercises 28 to 30. Place the origin at the vertex of the cone and the $\phi = 0$ ray (z-axis) along the axis of the cone, as shown in Figure 17.7.10(a). Let h be the height of the cone and α the vertex half-angle of the cone. So, $\tan(\alpha) = \frac{a}{h}$.

On the surface of the cone ϕ is constant, $\phi = \alpha$, but ρ and θ vary. A small patch on the surface of the cone corresponding to slight changes $d\theta$ and $d\rho$ has area approximately $(\rho\sin(\alpha)\,d\theta)\,d\rho = \rho\sin(\alpha)\,d\rho\,d\theta$. (Why?) So we may write

$$dS = \rho\sin(\alpha)\,d\rho\,d\theta.$$

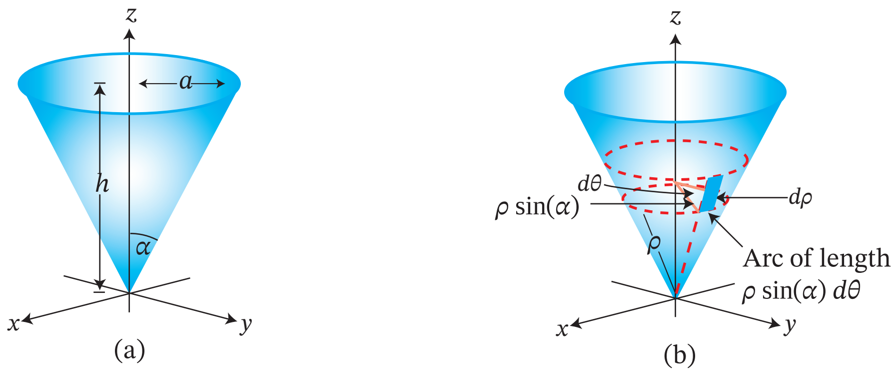


Figure 17.7.10

28. Find the average distance from points on the curved surface of a cone of radius a and height h to its axis.

29. Evaluate $\int_{\mathscr{S}} z^2\, dS$, where $\mathscr{S}$ is the entire surface of the cone shown in Figure 17.7.10(b), including its base.

30. Evaluate $\int_{\mathscr{S}} x^2\, dS$, where $\mathscr{S}$ is the curved surface of the right circular cone of radius 1 and height 1 with axis along the z-axis.

Integration over the curved surface of a right circular cylinder is easiest in cylindrical coordinates. Given the cylinder of radius a and axis on the z-axis, a patch on the cylinder corresponding to dz and $d\theta$ has area approximately $dS = a\, dz\, d\theta$. (Why?) Exercises 31 and 32 illustrate the use of these coordinates.

31. Let $\mathscr{S}$ be the entire surface of a solid cylinder of radius a and height h. For P in $\mathscr{S}$ let $f(P)$ be the square of the distance from P to one base. Find $\int_{\mathscr{S}} f(P)\, dS$. Include the two bases in the integration.

32. Let $\mathscr{S}$ be the curved part of the cylinder in Exercise 31. Let $f(P)$ be the square of the distance from P to a fixed diameter in a base. Find the average value of $f(P)$ for points in $\mathscr{S}$.

---

33. The areas of the projections of a small flat surface patch on the three coordinate planes are 0.01, 0.02, and 0.03. Is that enough information to find the area of the patch? If so, find it. If not, explain why not.

34. Let $\mathbf{F}$ describe the flow of a fluid in space. Then $\mathbf{F}(P) = \delta(P)\mathbf{v}(P)$ where $\delta(P)$ is the density of the fluid at P and $\mathbf{v}(P)$ is the velocity of the fluid at P. Making clear, large diagrams, explain why the rate at which the fluid is leaving the solid region enclosed by a surface $\mathscr{S}$ is $\int_{\mathscr{S}} \mathbf{F}\cdot\mathbf{n}\, dS$, where $\mathbf{n}$ denotes the unit outward normal to $\mathscr{S}$.

Exercises 35 and 36 are related. Exercise 36 is suggested by Exercise 20.

35. (a) Let g be a differentiable function such that $g\left(\frac{x+y}{2}\right) = \frac{g(x)+g(y)}{2}$ for any positive x and y. Show that $g(x) = kx + c$ for some k and c.

(b) Let f be a differentiable function such that $(x+y)f(x+y) + (x-y)f(x-y) = 2xf(x)$ for all positive x and y with $0 < y < x$. Deduce that there are constants k and c such that $f(x) = k + \frac{c}{x}$.

36. Let $d(P)$ be the distance from P to a point at a distance b from the center of a sphere of radius a, $a < b$. Let $f(x)$ have the property that for all a and b with $0 < a < b$, the average value of $f(d(P))$ on the sphere is equal to $f(b)$. Show that f must have the form $f(x) = k + \frac{c}{x}$.

---

37. If $\mathscr{S}$ is a smooth surface of a convex body, show that the volume of the solid bounded by $\mathscr{S}$ is the value of the surface integral $\int_{\mathscr{S}} z\cos(\gamma)\ dS$.

38. Assume $R(x,y,z)$ is a scalar function defined over a closed convex surface $\mathscr{S}$. Moreover, assume $\mathscr{A}$ is the projection of $\mathscr{S}$ on the xy-plane and the line through $(x,y,0)$ parallel to the z-axis meets $\mathscr{S}$ at (x,y,z_1) and (x,y,z_2), with $z_1 \le z_2$. As usual, γ is the direction angle between the (outward) normal vector to the surface and $\mathbf{k}$. (See Figure 17.7.11.)

Show that

$$\int_{\mathscr{S}} R(x,y,z)\cos(\gamma)\ dS = \int_{\mathscr{A}} \big(R(x,y,z_2) - R(x,y,z_1)\big)\ dA,$$

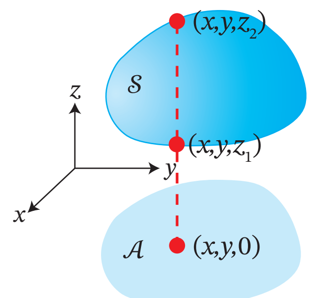


Figure 17.7.11

17.8 Moments, Centers of Mass, and Centroids

Now that we can integrate over plane regions, surfaces, and solid regions, we can define and calculate the center of mass of a physical object. Archimedes, the first person to study the center of mass, was interested in the stability of floating paraboloids. Today, the center of mass is important for naval architects, who need to design ships that do not tip over easily (so need a low center of mass) but can still navigate the world's major waterways and ports. And, a pole vaulter hopes that as they clear the bar their center of mass goes under the bar.

The Center of Mass

A small person on one side of a seesaw (which we regard as weightless) can balance a bigger person on the other side. For example, the two people in Figure 17.8.1 balance. Each person exerts a force on the seesaw, due to gravitational attraction, proportional to his mass. The small mass with the long lever arm balances the large mass with the small lever arm. Each person contributes the same tendency to turn but in opposite directions.

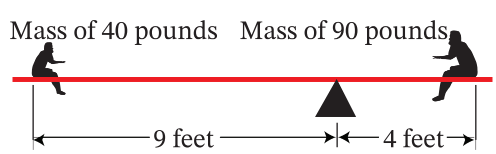


Figure 17.8.1

This tendency is called the **moment**:

$$\text{Moment} = (\text{Mass}) \cdot (\text{Lever arm}),$$

where the lever arm can be positive or negative. To be more precise, introduce on the seesaw an x-axis with its origin 0 at the fulcrum, the point on which the seesaw rests. Define the moment about 0 of a mass m located at the point x on the x-axis to be mx. Then the bigger body has a moment $(90)(4)$, while the smaller body has a moment $(40)(-9)$. The total moment of the lever-mass system is 0, and the masses balance.

If a mass m is located on a line with coordinate x, we define its moment about the point having coordinate k as the product $m(x-k)$.

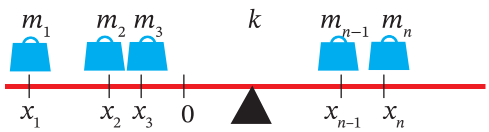


Figure 17.8.2

For several point masses $m_1, m_2, \ldots, m_n$, if mass m_i is located at x_i for $i = 1, 2, \ldots, n$, then $\sum_{i=1}^{n} m_i(x_i - k)$ is the total moment of the masses about the point k. If a fulcrum is placed at k, then the seesaw rotates clockwise if the total moment is greater that 0, rotates counterclockwise if it is less than 0, and is in **equilibrium** if the total moment is 0. See Figure 17.8.2.

To find where to place the fulcrum so that the entire system is in equilibrium, that is, it balances, find k such that

$$\sum_{i=1}^{n} m_i(x_i - k) = 0.$$

This can be manipulated into other forms, including

$$k\sum_{i=1}^{n} m_i = \sum_{i=1}^{n} m_i x_i.$$

From here it is straightforward to solve for k:

$$k = \frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i}.$$

This value of k is called the center of mass of the system of masses. It is the point about which all the masses balance. The x-coordinate of the center of mass is found by dividing the total moment about 0 by the total mass. It is usually denoted $\overline{x}$.

$\overline{x}$ is pronounced "x bar".

Formula 17.8.1: ***Center of Mass for a System of Point Masses***

The **center of mass** of a collection of n point masses at location $x_1, x_2, \ldots, x_n$ with masses $m_1, m_2, \ldots, m_n$ is

$$\overline{x} = \frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i}. \tag{17.8.1}$$

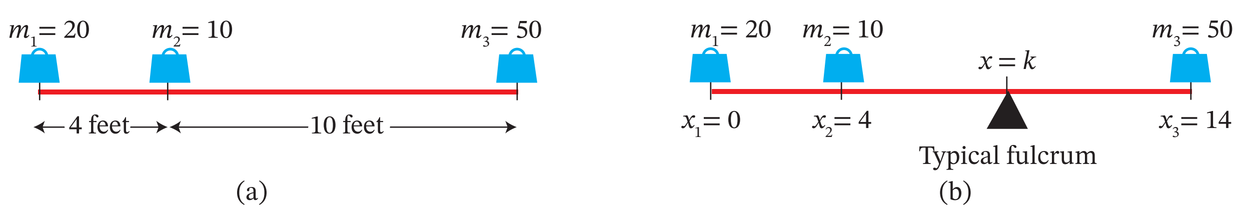


Figure 17.8.3

Finding the center of mass of a finite number of point masses involves only arithmetic. For example, suppose three masses are placed on a seesaw as in Figure 17.8.3(a). Introduce an x-axis with origin at mass $m_1 = 20$ pounds. Two additional masses are located at $x_2 = 4$ feet and $x_3 = 14$ feet with masses $m_2 = 10$ pounds and $m_3 = 50$ pounds, respectively. The total moment about $x = k$ is

$$M = 20(0 - k) + 10(4 - k) + 50(14 - k) = 740 - 80k.$$

This moment vanishes when $M = 0$, that is, when $k = 740/80 = 9.25$.

This is consistent with the formula for the center of mass:

$$\overline{x} = \frac{m_1x_1 + m_2x_2 + m_3x_3}{m_1 + m_2 + m_3} = \frac{0 + 40 + 700}{20 + 10 + 50} = \frac{740}{80} = 9.25.$$

The seesaw balances when the fulcrum is placed 9.25 feet to the right of mass m_1 ($k = \overline{x} = 9.25$). (See Figure 17.8.3(b).)

Calculus is needed to find the center of mass of a one-dimensional rod that occupies the interval $a \le x \le b$ with density $\lambda(x)$ at x. To apply the previous ideas, divide the rod into n pieces of width $\Delta x = (b-a)/n$. Let $x_i = a + i\Delta x$, $i = 0, \dots, n$. For the piece of the rod for $x_{i-1} \le x \le x_i$, select a point c_i in it. The mass of the piece is approximately $\lambda(c_i)\Delta x$. An approximation to the total moment about $x = k$ is

$$M \approx \sum_{i=1}^{n} \underbrace{\lambda(c_i)\Delta x}_{\text{mass}} \underbrace{(c_i - k)}_{\text{lever}}.$$

As n increases without bound these Riemann sums converge to a definite integral for the total moment

$$M = \int_a^b \lambda(x)(x-k)\, dx.$$

The total moment vanishes when

$$k = \frac{\int_a^b x\lambda(x)\, dx}{\int_a^b \lambda(x)\, dx}.$$

The denominator is the mass of the rod and the numerator is the rod's total moment.

To find the center of mass of a continuous distribution of matter in a plane region, we use a double integral. Let $\mathscr{R}$ be a region in the plane occupied by a thin piece of metal whose density at P is $\sigma(P)$. Let L be a line in the plane, as shown in Figure 17.8.4(a). We will find a formula for the unique line parallel to L, around which the mass in $\mathscr{R}$ balances.

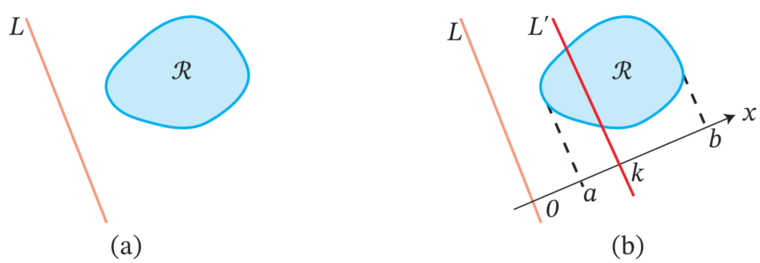


Figure 17.8.4

Let L' be any line parallel to L. We will compute the moment about L' and then see how to choose L' to make it 0. To compute the moment of $\mathscr{R}$ about L', introduce an x-axis perpendicular to L with its origin at its intersection with L. Assume that L' passes through the x-axis at the point $x = k$, as in Figure 17.8.4(b). In addition, assume that each line parallel to L meets $\mathscr{R}$ either in a line segment or at a point on the boundary of $\mathscr{R}$ or not at all. The lever arm of the mass distributed throughout $\mathscr{R}$ varies from point to point.

Partition $\mathscr{R}$ into n small regions $\mathscr{R}_1, \mathscr{R}_2, \dots, \mathscr{R}_n$. Let A_i be the area of $\mathscr{R}_i$. In each region the lever arm around L' varies only a little. If we pick a point P_1 in $\mathscr{R}_1$, P_2 in $\mathscr{R}_2, \dots, P_n$ in $\mathscr{R}_n$, and the x-coordinate of P_i is x_i, then

$$\underbrace{\sigma(P_i)A_i}_{\text{mass in } \mathscr{R}_i} \underbrace{(x_i - k)}_{\text{lever arm}}$$

is a local estimate of the turning tendency.

Thus

$$\sum_{i=1}^{n} \sigma\,(P_i) A_i\,(x_i - k) \tag{17.8.2}$$

would be a good estimate of the total turning tendency around L'. Taking the limit of (17.8.2) as all $\mathscr{R}_i$ are chosen smaller and smaller, we expect

$$\int_{\mathscr{R}} \sigma(P)\,dA\,(x-k) \tag{17.8.3}$$

to represent the turning tendency of the total mass around L'. The quantity (17.8.3) is called the **moment** of the mass distribution around L'.

EXAMPLE 1. Let $\mathscr{R}$ be the region under $y = x^2$ and above $[0,1]$ with the density $\sigma(x,y) = xy$. Find its moment around the line $x = \frac{1}{2}$.

Exercise 2 asks you for the details involved in evaluating each integral in this example.

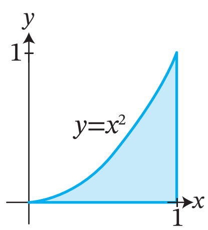


Figure 17.8.5

SOLUTION $\mathscr{R}$ is shown in Figure 17.8.5. The moment of $\mathscr{R}$ about $x = 1/2$ is given by (17.8.3):

$$\int_{\mathscr{R}} \left(x - \frac{1}{2}\right) xy\,dA. \tag{17.8.4}$$

One description of $\mathscr{R}$ is: $0 \le y \le x^2$, $0 \le x \le 1$. The double integral over $\mathscr{R}$ can be evaluated as the iterated integral:

$$\int_0^1 \int_0^{x^2} \left(x - \frac{1}{2}\right) xy\,dy\,dx.$$

The inner (y) integration evaluates to

$$\int_0^{x^2} \left(x - \frac{1}{2}\right) xy\,dy = \left(x - \frac{1}{2}\right) x \int_0^{x^2} y\,dy = \frac{1}{2}x^6 - \frac{1}{4}x^5.$$

Then, evaluating the outer (x) integration yields

$$\int_0^1 \left(\frac{1}{2}x^6 - \frac{1}{4}x^5\right) dx = \frac{5}{168}.$$

Since the total moment (17.8.4) is positive, the object would rotate clockwise around the line $x = 1/2$. ◆

Now that we have a way to find the moment around any line parallel to the y-axis we can find the line around which the moment is zero, the balancing line. We solve for k in the equation

$$\int_{\mathscr{R}} (x-k)\sigma(P)\,dA = 0.$$

Thus

$$\int_{\mathscr{R}} x\sigma(P)\,dA = k \int_{\mathscr{R}} \sigma(P)\,dA,$$

from which we find that

$$k = \frac{\displaystyle\int_{\mathscr{R}} x\sigma(P)\, dA}{\displaystyle\int_{\mathscr{R}} \sigma(P)\, dA}.$$

The denominator is the total mass. The numerator is the total moment. We can think of k as the average lever arm as weighted by the density.

There is therefore a unique balancing line parallel to the y-axis. Call its x-coordinate $\overline{x}$. Similarly, there is a unique balancing line parallel to the x-axis. Call its y-coordinate $\overline{y}$. The point $(\overline{x}, \overline{y})$ is called the center of mass of the region $\mathscr{R}$. We have:

Formula 17.8.2: *Center of Mass of Region $\mathscr{R}$ with Density σ*

The **center of mass** of a region $\mathscr{R}$ with density $\sigma(P)$ has coordinates $(\overline{x}, \overline{y})$ where

$$\overline{x} = \frac{\displaystyle\int_{\mathscr{R}} x\,\sigma(P)\, dA}{\displaystyle\int_{\mathscr{R}} \sigma(P)\, dA} \quad \text{and} \quad \overline{y} = \frac{\displaystyle\int_{\mathscr{R}} y\,\sigma(P)\, dA}{\displaystyle\int_{\mathscr{R}} \sigma(P)\, dA}.$$

The integral $\int_{\mathscr{R}} x\sigma(P)\, dA$ is called the **moment of $\mathscr{R}$ around the y-axis**, and is denoted M_y. Similarly, the **moment of $\mathscr{R}$ around the x-axis** is $M_x = \int_{\mathscr{R}} y\sigma(P)\, dA$ is .

If the density $\sigma(P)$ is constant everywhere in $\mathscr{R}$, then the two equations reduce to

$$\overline{x} = \frac{\displaystyle\int_{\mathscr{R}} x\, dA}{\displaystyle\int_{\mathscr{R}} dA} \quad \text{and} \quad \overline{y} = \frac{\displaystyle\int_{\mathscr{R}} y\, dA}{\displaystyle\int_{\mathscr{R}} dA}.$$

In this case the center of mass $\mathscr{R}$ is also called the centroid of the region, a purely geometric concept:

Formula 17.8.3: *Centroid of $\mathscr{R}$*

The **centroid** of the plane region $\mathscr{R}$ has the coordinates $(\overline{x}, \overline{y})$ where

$$\overline{x} = \frac{\displaystyle\int_{\mathscr{R}} x\, dA}{\displaystyle\int_{\mathscr{R}} dA} \quad \text{and} \quad \overline{y} = \frac{\displaystyle\int_{\mathscr{R}} y\, dA}{\displaystyle\int_{\mathscr{R}} dA}. \tag{17.8.5}$$

EXAMPLE 2. Find the center of mass of the region in Example 1.

SOLUTION Thinking back to Example 1, where the moment about $x = 1/2$ is positive, it appears as though we should expect the x-coordinate of the center of mass to be bigger than $1/2$.

To see if this is, in fact, the case, recall that the density at (x, y) in $\mathscr{R}$ is given by $\sigma = xy$. We compute three double integrals: the mass $\int_{\mathscr{R}} xy\, dA$ and the moments $M_y = \int_{\mathscr{R}} x(xy)\, dA$ and $M_x = \int_{\mathscr{R}} y(xy)\, dA$.

We find

$$M = \int_{\mathscr{R}} xy\,dA = \int_0^1 \int_0^{x^2} xy\,dy\,dx = \int_0^1 \frac{x^5}{2}\,dx = \frac{1}{12}.$$

Also

$$M_y = \int_{\mathscr{R}} x^2 y\,dA = \int_0^1 \int_0^{x^2} x^2 y\,dy\,dx = \int_0^1 \frac{x^6}{2}\,dx = \frac{1}{14}$$

and

$$M_x = \int_{\mathscr{R}} xy^2\,dA = \int_0^1 \int_0^{x^2} xy^2\,dy\,dx = \int_0^1 \frac{x^7}{3}\,dx = \frac{1}{24}.$$

Thus

$$\overline{x} = \frac{\frac{1}{14}}{\frac{1}{12}} = \frac{6}{7} \qquad \text{and} \qquad \overline{y} = \frac{\frac{1}{24}}{\frac{1}{12}} = \frac{1}{2}.$$

The fact that $\overline{x} = 6/7 > 1/2$ is consistent with the result of Example 1, and speculated at the outset of this example. ♦

Observation 17.8.1: ***An Important Point About an Important Point***

The center of mass $(\overline{x}, \overline{y})$ is found by first choosing an xy-coordinate system. What if we choose an $x'y'$-coordinate system at an angle to the xy-coordinate system? The center of mass computed in the $x'y'$-coordinate system, $(\overline{x'}, \overline{y'})$, is the same point as the center of mass, $(\overline{x}, \overline{y})$, computed in the xy-coordinate system? (See Figure 17.8.6.)

This observation is confirmed in Exercises 35 and 36 in Section 17.S.

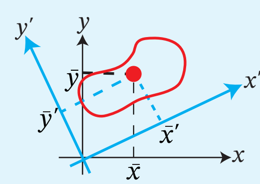


Figure 17.8.6

Shortcuts for Computing Centroids

Assume that f is a nonnegative function and let $\mathscr{R}$ be the region under $y = f(x)$ for x in $[a, b]$. Then, in computing the centroid of $\mathscr{R}$, we encounter the moment about the x-axis

$$M_x = \int_{\mathscr{R}} y\,dA.$$

Thus

$$M_x = \int_a^b \int_0^{f(x)} y\,dy\,dx = \int_a^b \frac{1}{2}(f(x))^2\,dx = \frac{1}{2}\int_a^b (f(x))^2\,dx.$$

By (17.8.5)

$$\overline{y} = \frac{\frac{1}{2}\int_a^b (f(x))^2\,dx}{\text{Area of } \mathscr{R}}. \tag{17.8.6}$$

EXAMPLE 3. Find the centroid of the semicircular region of radius a. (See Figure 17.8.7.)

SOLUTION By symmetry, $\bar{x} = 0$. To find $\bar{y}$, use (17.8.6). The function f is given by $f(x) = \sqrt{a^2 - x^2}$, an even function. The moment of $\mathscr{R}$ about the x-axis is

$$\int_{-a}^{a} \frac{\left(\sqrt{a^2-x^2}\right)^2}{2}\,dx = \int_{-a}^{a} \frac{a^2-x^2}{2}\,dx = 2\int_{0}^{a} \frac{a^2-x^2}{2}\,dx = \int_{0}^{a} (a^2-x^2)\,dx$$
$$= \left(a^2x - \frac{x^3}{3}\right)\Big|_0^a = \left(a^3 - \frac{a^3}{3}\right) - 0 = \frac{2}{3}a^3.$$

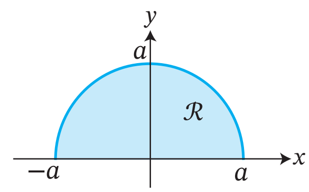


Figure 17.8.7

Thus

$$\bar{y} = \frac{\frac{2}{3}a^3}{\text{Area of } \mathscr{R}} = \frac{\frac{2}{3}a^3}{\frac{1}{2}\pi a^2} = \frac{4a}{3\pi}.$$

Since $4/(3\pi) \approx 0.42$ the centroid of $\mathscr{R}$ is at a height of about $0.42a$. ◆

Centers of Other Masses

We defined moment, center of mass, and centroid for masses situated in a plane. They generalize to masses distributed on a curve (such as a wire), a surface (such as a spherical surface), or in space (such as an ellipsoid).

For a curve, C, the mass has a linear density $\lambda(P)$. A short piece around P of length Δs would have mass approximately $\lambda(P)\Delta s$. Thus, the mass and moments of the curve would be

$$M = \int_C \lambda(P)\,ds, \qquad M_y = \int_C x\,\lambda(P)\,ds, \qquad \text{and} \qquad M_x = \int_C y\,\lambda(P)\,ds.$$

The definitions in the case of a surface $\mathscr{S}$ are obtained from the definitions for a plane region $\mathscr{R}$ by replacing $\mathscr{R}$ by $\mathscr{S}$ and dA by dS.

The definition for a solid object of density $\delta(P)$ occupying the region $\mathscr{R}$ is similar. We assume an xyz-coordinate system. The total mass is

$$M = \int_{\mathscr{R}} \delta(P)\,dV.$$

There are three moments, one relative to each of the coordinate planes, indicated by the subscripts:

$$M_{yz} = \int_{\mathscr{R}} x\,\delta(P)\,dV, \qquad M_{xz} = \int_{\mathscr{R}} y\,\delta(P)\,dV, \qquad M_{xy} = \int_{\mathscr{R}} z\,\delta(P)\,dV.$$

The center of mass is $(\bar{x}, \bar{y}, \bar{z})$, where

$$\bar{x} = \frac{\int_{\mathscr{R}} x\,\delta(P)\,dV}{M}, \qquad \bar{y} = \frac{\int_{\mathscr{R}} y\,\delta(P)\,dv}{M}, \qquad \bar{z} = \frac{\int_{\mathscr{R}} z\,\delta(P)\,dV}{M}.$$

Recall that when $\delta(P) = 1$ for all P in $\mathscr{R}$, then the mass is the same as the volume and the center of mass is called the centroid. Physically, the centroid is the average position of the object weighted by the density at each point.

EXAMPLE 4. Find the centroid of the top half of a solid ball of radius a shown in Figure 17.8.8.

Exercise 3 asks you for the details involved in evaluating each integral in this example.

SOLUTION This solid is a filled hemisphere of radius a. Place the origin of an xyz-coordinate system at the center of the circle of radius a that is the base of the hemisphere, as in Figure 17.8.8.

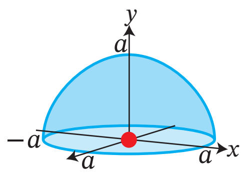


Figure 17.8.8

By symmetry, the centroid is on the z-axis. (If you spin the hemisphere about the z-axis you get the same hemisphere back, which must have the same centroid. If the centroid were not on the z-axis, you would get more than one centroid for the same object.) So $\overline{x} = \overline{y} = 0$. Calling the hemisphere $\mathscr{H}$, we have

$$\overline{z} = \frac{\int_{\mathscr{H}} z\, dV}{\text{Volume of } \mathscr{H}}.$$

The volume of the hemisphere is half that of a ball, $2\pi a^3/3$. To evaluate the moment $\int_{\mathscr{H}} z\, dV$, we use an iterated integral in spherical coordinates. Because $z = \rho\cos(\phi)$, we have:

$$\int_{\mathscr{H}} z\, dV = \int_0^{2\pi}\int_0^{\pi/2}\int_0^{a} (\rho\cos(\phi))\,\rho^2\sin(\phi)\, d\rho\, d\phi\, d\theta.$$

Straightforward computations reveal that

$$\int_{\mathscr{H}} z\, dV = \frac{\pi a^4}{4}$$

so that

$$\overline{z} = \frac{\frac{\pi a^4}{4}}{\frac{2}{3}\pi a^3} = \frac{3a}{8}.$$

The centroid is $(0, 0, 3a/8)$. ♦

EXAMPLE 5. Find the centroid of a homogeneous cone of height h and radius a.

Exercise 4 asks you for the details involved in evaluating each integral in this example.

SOLUTION As we saw for the hemisphere in Example 4, symmetry tells us the centroid lies on the axis of the cone.

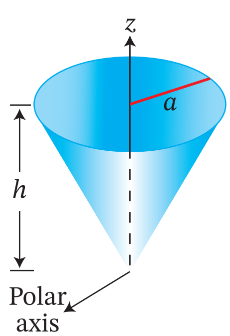


Figure 17.8.9

Introduce a spherical coordinate system with the origin at the vertex of the cone and with the axis of the cone lying on the ray $\phi = 0$, as in Figure 17.8.9. The vertex half-angle is $\arctan(a/h)$. The plane of the base of the cone is $z = h$ (in rectangular coordinates), or, in spherical coordinates, $\rho\cos(\phi) = h$. The cone's description is

$$0 \le \theta \le 2\pi, \qquad 0 \le \phi \le \arctan(a/h), \qquad 0 \le \rho \le h/\cos(\phi).$$

To find the centroid we compute $\int_{\mathscr{R}} z\, dV$ and divide the result by the volume of the cone, which is $\pi a^2 h/3$.

Now

$$\int_{\mathscr{R}} z\, dV = \int_0^{2\pi}\int_0^{\arctan(a/h)}\int_0^{h/\cos(\phi)} \rho\cos(\phi)\,(\rho^2\sin(\phi))\, d\rho\, d\phi\, d\theta.$$

For the inner integration ϕ and θ are constant:

$$\int_0^{h/\cos(\phi)} \rho\cos(\phi)\rho^2\sin(\phi)\, d\rho = \cos(\phi)\sin(\phi)\int_0^{h/\cos(\phi)} \rho^3\, d\rho = \frac{h^4\sin(\phi)}{4\cos^3(\phi)}.$$

The middle integration is

$$\int_0^{\arctan(a/h)} \frac{h^4 \sin(\phi)}{4\cos^3(\phi)}\, d\phi = \frac{h^4}{4} \int_0^{\arctan(a/h)} \frac{\sin(\phi)}{\cos^3(\phi)}\, d\phi = \frac{a^2h^2}{8}.$$

The outermost integral is

$$\int_0^{2\pi} \frac{a^2h^2}{8}\, d\theta = \frac{\pi a^2 h^2}{4}.$$

Thus,

$$\overline{z} = \frac{\int_{\mathscr{R}} z\, dV}{\text{Volume of } \mathscr{R}} = \frac{\frac{\pi a^2h^2}{4}}{\frac{\pi a^2 h}{3}} = \frac{3h}{4}.$$

The centroid of a cone is three-fourths of the way from the vertex to the base. ♦

Historical Note: Archimedes and Centroids

How did Archimedes find centroids? Integral calculus was not invented until 1684, some 1900 years after his death. In one approach he used axioms in the style of Euclid's geometry text, written a generation or two before him. These are the axioms:

1. The centroids of similar figures are similarly situated.
2. The centroid of a convex region lies within the region.
3. If an object is cut into two pieces, its centroid C lies on the line segment joining the centroids of the two pieces. Moreover, if the pieces are $\mathscr{R}$ with centroid A and $\mathscr{S}$ with centroid B, then $|CA|$ times the area of $\mathscr{R}$ equals $|CB|$ times the area of $\mathscr{S}$.

The book cited in Exercise 6 describes how Archimedes used these axioms to find the centroid of a triangle.

Summary

We defined the moment about a line and used it to define the center of mass for a plane distribution of mass. The moment of a mass about a line L measures the tendency of the mass to rotate about the line L. The center of mass for a region $\mathscr{R}$ in the xy-plane is the point in the region where the region balances.

The moment about the y-axis is

$$M_y = \int_{\mathscr{R}} x\,\sigma(P)\, dA.$$

The moment about the x-axis is

$$M_x = \int_{\mathscr{R}} y\,\sigma(P)\, dA.$$

The center of mass is $(\overline{x}, \overline{y})$ where

$$\overline{x} = \frac{M_y}{\text{Mass}}, \quad \overline{y} = \frac{M_x}{\text{Mass}}.$$

If the density is constant, we have a purely geometric concept. The centroid is $(\overline{x}, \overline{y})$ where

$$\overline{x} = \frac{1}{\text{Area of } \mathscr{R}} \int_{\mathscr{R}} x\ dA, \quad \text{and} \quad \overline{y} = \frac{1}{\text{Area of } \mathscr{R}} \int_{\mathscr{R}} y\ dA.$$

	curve (C)	solid ($\mathscr{R}$)
density	$\lambda(P)$	$\delta(P)$
M	$\int_C \lambda(P)\,ds$	$\int_{\mathscr{R}} \delta(P)\,dV$
M_{yz}	$\int_C x\,\lambda(P)\,ds$	$\int_{\mathscr{R}} x\,\delta(P)\,dV$
M_{xz}	$\int_C y\,\lambda(P)\,ds$	$\int_{\mathscr{R}} y\,\delta(P)\,dV$
M_{xy}	$\int_C z\,\lambda(P)\,ds$	$\int_{\mathscr{R}} z\,\delta(P)\,dV$

Table 17.8.1

In three dimensions, the corresponding definitions for curves and solids are summarized in Table 17.8.1.

EXERCISES for Section 17.8

1. (a) How would you define the center of mass of a curve? Call its linear density $\lambda(P)$.
(b) Find the center of mass of a semicircle of radius a.

In Exercises 2 to 4 give a description of the region $\mathscr{R}$ and carry out the integrations in

2. Example 1.
3. Example 4.
4. Example 5.

5. Example 4 showed that the centroid of a hemisphere is at a distance less than half the radius from the sphere's center. Why is that to be expected?

6. Find the centroid of a solid paraboloid of revolution. This is the region above $z = x^2 + y^2$ and below the plane $z = c$. Archimedes found the centroid without calculus and used the result to analyze the equilibrium of a floating paraboloid. (If it is slightly tilted, will it come back to the vertical or topple over?) The following book explains how he did this 2200 years ago, and more.

REFERENCE: S. Stein, *Archimedes: What Did He Do Besides Cry Eureka?*, Mathematical Assoc. of America, 1999.

7. Using cylindrical coordinates, find $\overline{z}$ for the region below the paraboloid $z = x^2 + y^2$ and above the disk bounded by $r = 2\cos(\theta)$ in the $r\theta$-plane. Include a drawing of the region.

8. Find the z-coordinate, $\overline{z}$, of the centroid of the part of the saddle $z = xy$ that lies above the portion of the disk bounded by the circle $x^2 + y^2 = a^2$ in the first quadrant.

In Exercises 9 to 16 find the centroid of $\mathscr{R}$. (Exercises 13 to 16 require integral tables or techniques of Chapter 8.)

9. $\mathscr{R}$ is bounded by $y = x^2$ and $y = 4$.

10. $\mathscr{R}$ is bounded by $y = x^4$ and $y = 1$.

11. $\mathscr{R}$ is bounded by $y = 4x - x^2$ and the x-axis.

12. $\mathscr{R}$ is bounded by $y = x$, $x + y = 1$, and the x-axis.

13. $\mathscr{R}$ is the region bounded by $y = e^x$ and the x-axis, between the lines $x = 1$ and $x = 2$.

14. $\mathscr{R}$ is the region bounded by $y = \sin(2x)$ and the x-axis, between the lines $x = 0$ and $x = \frac{\pi}{2}$.

15. $\mathscr{R}$ is the region bounded by $y = \sqrt{1+x}$ and the x-axis, between the lines $x = 0$ and $x = 3$.

16. $\mathscr{R}$ is the region bounded by $y = \ln(x)$ and the x-axis between the lines $x = 1$ and $x = e$.

In Exercises 17 to 24 find the center of mass of the lamina.

17. The triangle with vertices $(0,0)$, $(1,0)$, $(0,1)$; density at (x, y) is $x + y$.

18. The triangle with vertices $(0,0)$, $(2,0)$, $(1,1)$; density at (x, y) is y.

19. The square with vertices $(0,0)$, $(1,0)$, $(1,1)$, $(0,1)$; density at (x, y) is $y\arctan(x)$.

20. The finite region bounded by $y = 1 + x$ and $y = 2^x$; density at (x, y) is $x + y$.

21. The triangle with vertices $(0,0)$, $(1,2)$, $(1,3)$; density at (x, y) is xy.

22. The finite region bounded by $y = x^2$, the x-axis, and $x = 2$; density at (x, y) is e^x.

23. The finite region bounded by $y = x^2$ and $y = x + 6$, situated to the right of the y-axis; density at (x, y) is $2x$.

24. The trapezoid with vertices $(0,0)$, $(3,0)$, $(2,1)$, $(0,1)$; density at (x, y) is $\sin(x)$.

---

25. In a letter of 1680 Leibniz wrote:

> Huygens, as soon as he had published his book on the pendulum, gave me a copy of it; and at that time I was quite ignorant of Cartesian algebra and also of the method of indivisibles, indeed I did not know the correct definition of the center of gravity. For, when by chance I spoke of it to Huygens, I let him know that I thought that a straight line drawn through the center of gravity always cut a figure into two equal parts; since that clearly happened in the case of a square, or a circle, an ellipse, and other figures that have a center of magnitude. I imagined that it was the same for all other figures. Huygens laughed when he heard this, and told me that nothing was further from the truth.
>
> REFERENCE: C.H. Edwards, *The Historical Development of the Calculus*, Springer-Verlag, New York, 1979, p. 239.

Give an example showing that Huygens was right.

NOTE: Leibniz' center of gravity is really the center of mass.

26. Cut an irregular shape out of cardboard and find three balancing lines for it experimentally. Are they concurrent? That is, do they pass through a common point?

27. Let f and g be continuous functions such that $f(x) \geq g(x) \geq 0$ for x in $[a, b]$. Let $\mathscr{R}$ be the region above $[a, b]$ that is bounded by the curves $y = f(x)$ and $y = g(x)$. Assume the density of $\mathscr{R}$ is 1.

(a) Set up a definite integral in terms of f and g for the moment of $\mathscr{R}$ about the y-axis.

(b) Set up a definite integral with respect to x in terms of f and g for the moment of $\mathscr{R}$ about the x-axis.

In Exercises 28 to 31 find (a) the moment of $\mathscr{R}$ about the y-axis, (b) the moment of $\mathscr{R}$ about the x-axis, (c) the area of $\mathscr{R}$, (d) $\bar{x}$, and (e) $\bar{y}$. Assume the density is 1. See Exercise 27.

28. $\mathscr{R}$ is bounded by the curves $y = x^2$, $y = x^3$.

29. $\mathscr{R}$ is bounded by $y = x$, $y = 2x$, $x = 1$, and $x = 2$.

30. $\mathscr{R}$ is bounded by the curves $y = 3^x$ and $y = 2^x$ between $x = 1$ and $x = 2$.

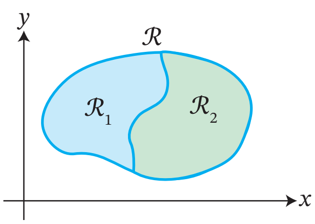
31. $\mathscr{R}$ is bounded by the curves $y = x - 1$ and $y = \ln(x)$, between $x = 1$ and $x = e$.

---

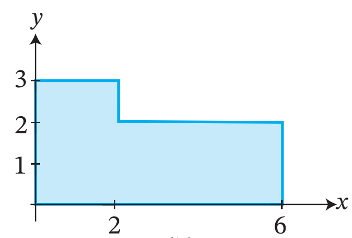
32. If $\mathscr{R}$ is the region below $y = f(x)$ and above $[a, b]$, show that $\bar{x} = \dfrac{1}{\text{Area of } \mathscr{R}} \displaystyle\int_a^b x f(x)\, dx$.

Exercises 33 to 35 are related.

33. A planar distribution of matter occupies the region $\mathscr{R}$. It is cut into two pieces, occupying regions $\mathscr{R}_1$ and $\mathscr{R}_2$, as in Figure 17.8.10(a). The part in $\mathscr{R}_1$ has mass M_1 and center of mass $(\bar{x}_1, \bar{y}_1)$. The part in $\mathscr{R}_2$ has mass M_2 and center of mass $(\bar{x}_2, \bar{y}_2)$. Find the center of mass $(\bar{x}, \bar{y})$ of the entire mass. Express it in terms of M_1, M_2, $\bar{x}_1$, $\bar{x}_2$, $\bar{y}_1$, and $\bar{y}_2$.



(a)



(b)

Figure 17.8.10

34. Let $\mathscr{R}$, $\mathscr{R}_1$, and $\mathscr{R}_2$ be as in Exercise 33. Show that the center of mass of $\mathscr{R}$ lies on the line segment joining the centers of mass of $\mathscr{R}_1$ and $\mathscr{R}_2$.

35. Use the formula in Exercise 33 to find the center of mass of the homogeneous lamina shown in Figure 17.8.10(b).

Exercise 36 shows that the medians of a triangle meet at the centroid of the triangle. (A **median** of a triangle is a line that passes through a vertex and the midpoint of the opposite edge.)

36. Let $\mathscr{R}$ be a triangle with vertices A, B, and C. Let M denote the midpoint of the side AB.

Introduce an xy-coordinate system such that the origin is at A and B lies on the x-axis, as in Figure 17.8.11.

(a) Compute $(\bar{x}, \bar{y})$.

(b) Find the equation of the median through C and M.

(c) Verify that the centroid lies on the median computed in (b).

(d) Use physical intuition to explain why the centroid lies on each median.

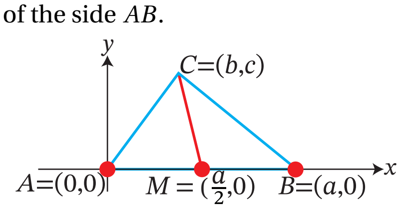


Figure 17.8.11

In Exercises 37 and 38 find $\bar{z}$ for the given surface.

37. The portion of the paraboloid $2z = x^2 + y^2$ below the plane $z = 9$.

38. The portion of the plane $x + 2y + 3z = 6$ above the triangle in the xy-plane with vertices $(0,0)$, $(4,0)$, and $(0,1)$.

Exercises 39 and 40 are related, and involve the objects shown in Figure 17.8.12. Figure 17.8.12(a) is a semicircular wire of radius a, Figure 17.8.12(b) is the top half of the surface of a ball of radius a, and Figure 17.8.12(c) the top half of a ball of radius a.

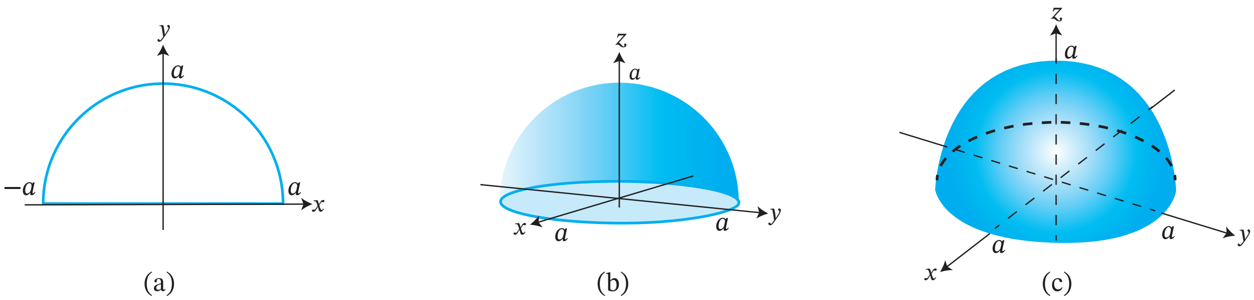


Figure 17.8.12

39. Which of the objects shown in Figure 17.8.12 has the highest centroid? Give your opinion, with reason.

40. Using calculus, determine the highest centroid in Exercise 39.

---

41. The corners of a homogeneous triangular piece of metal are $(0,0)$, $(1,0)$, and $(0,2)$.
(a) Is the line $y = \frac{11x}{5}$ a balancing line? (b) If not, if the metal rests on the line which way would it rotate?

Definition: *Section of a region*

Let $\mathscr{R}$ be a convex set in the plane. A **section** of $\mathscr{R}$ is a part of $\mathscr{R}$ that is bounded by a chord and part of the boundary, as shown in Figure 17.8.13.

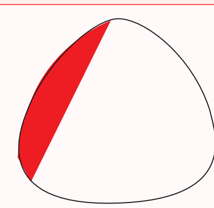


Figure 17.8.13

42. If $\mathscr{R}$ is a convex set in the plane, show that different sections have different centroids.

43. Do you think every point in $\mathscr{R}$ that is not on the boundary is the centroid of some section of $\mathscr{R}$? Why? See Exercise 42.

44. Archimedes (287-212 B.C.) investigated the centroid of a section of a parabola. A section of the parabola $y = x^2$ is shown in Figure 17.8.14. M is the midpoint of the chord and N is the point on the parabola directly below M.

He showed without calculus that the centroid is on the line MN, three-fifths of the way from N to M. Obtain his result with calculus.

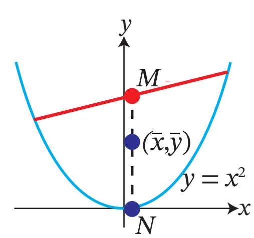


Figure 17.8.14

45. Is every point in the region bounded by the parabola the centroid of some section? See Exercise 44.

46. The plane $z = c$ in Exercise 6 is perpendicular to the axis of the paraboloid. Archimedes was also interested in the case when the plane is not perpendicular to the axis. Find the centroid of the region below the tilted plane $z = ky$ and above the paraboloid $z = x^2 + y^2$.

Exercises 47 to 49 concern Pappus's Theorem, Theorem 17.8.2, which relates the volume of a solid of revolution to the centroid of the plane region $\mathscr{R}$ that is revolved to form the solid.

Theorem 17.8.2: Pappus's Theorem

Let $\mathscr{R}$ be a region in the plane and L a line in the plane that does not cross $\mathscr{R}$, though it can touch $\mathscr{R}$ at its border. Then the volume of the solid formed by revolving $\mathscr{R}$ about L is equal to the product

$$\left(\text{Distance the centroid of } \mathscr{R} \text{ is rotated}\right) \cdot \left(\text{Area of } \mathscr{R}\right).$$

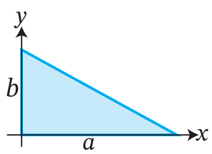


Figure 17.8.15

47. (a) Prove Pappus's Theorem.
(b) Use Pappus's Theorem to find the volume of the torus formed by revolving a disk of radius 3 inches about a line in the plane of the disk and 5 inches from its center.

48. Use Pappus's Theorem to find the centroid of the half disk $\mathscr{R}$ of radius a.

49. Use Pappus's Theorem to find the centroid of the right triangle in Figure 17.8.15.

50. This exercise concerns hydrostatic pressure.

Section 7.6 introduced water pressure on a planar surfaces.

(a) Show that the force of water against a submerged, vertical plane surface occupying the plane region $\mathscr{R}$ equals the pressure at the centroid of $\mathscr{R}$ times the area of $\mathscr{R}$.
(b) Is the assertion in (a) correct if $\mathscr{R}$ is not vertical?

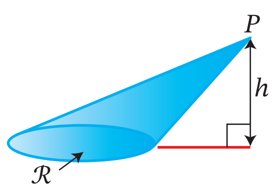


Figure 17.8.16

51. Let $\mathscr{R}$ be a region in a plane and P a point a distance $h > 0$ from the plane. P and $\mathscr{R}$ determine a cone with base $\mathscr{R}$ and vertex P, as shown in Figure 17.8.16. Let the area of $\mathscr{R}$ be A. What can be said about the distance from the centroid of the cone to the plane of $\mathscr{R}$?
(a) What is the distance in the case of a right circular cone?
(b) Experiment with another cone with a convenient base of your choice.
(c) Make a conjecture.
(d) Explain why it is true.

52. Let f and g be two continuous functions such that $f(x) \geq g(x) \geq 0$ for x in $[0,1]$. Let $\mathscr{R}$ be the region under $y = f(x)$ and above $[0,1]$ and let $\mathscr{R}^*$ be the region under $y = g(x)$ and above $[0,1]$.

CONTRIBUTED BY: Jeff Lichtman

(a) Do you think the centroid of $\mathscr{R}$ is at least as high as the centroid of $\mathscr{R}^*$? (Give your opinion, without any supporting calculations.)
(b) Let $g(x) = x$. Define $f(x)$ to be $1/3$ for $0 \leq x \leq 1/3$ and to be x if $1/3 \leq x \leq 1$. (f is continuous.) Find $\overline{y}$ for $\mathscr{R}$ and also for $\mathscr{R}^*$. (Which is larger?)
(c) Let a be a constant, $0 \leq a \leq 1$. Let $f(x) = a$ for $0 \leq x \leq a$, and let $f(x) = x$ for $a \leq x \leq 1$. Find $\overline{y}$ for $\mathscr{R}$.
(d) Show that the number a for which $\overline{y}$ defined in (c) is a minimum is a root of $x^3 + 3x - 1 = 0$.
(e) Show that the equation in (d) has only one real root q.
(f) Find q to four decimal places.
(g) Show that $\overline{y} = q$ with $a = q$.

53. Let $\mathscr{S}$ be the part of the paraboloid $z = x^2 + y^2$ below the plane $z = a^2$.
   (a) Set up a double integral in the xy-plane for the moment of $\mathscr{S}$ about the xy-plane.
   (b) Express it as an iterated integral in polar coordinates.
   (c) Evaluate the integral.
   (d) Find the centroid of $\mathscr{S}$.

17.9 Magnification and Multiple Integrals

The magnification of a mapping was first discussed in Section 16.9. In this section the magnification of a mapping is used to evaluate multiple integrals by replacing the domain of integration by a simpler domain. This is similar to the way we replaced an integral over a curve with an integral over an interval (replacing ds by $(ds/dx)\ dx$, in Section 9.4) and an integral over a surface by an integral over a plane region (replacing dS by $(1/|\cos(\gamma)|)dA$ in Section 17.7).

Magnification Enters the Integral

Assume that F is a mapping from a region $\mathscr{R}$ in uv-space to a region $\mathscr{S}$ in xy-space and $f(P)$ is a scalar function defined on $\mathscr{S}$. We will express the multiple integral $\int_{\mathscr{S}} f(P)\ dS$ as a multiple integral over $\mathscr{R}$. If $\mathscr{R}$ is simpler than $\mathscr{S}$, it may be easier to compute the integral over $\mathscr{R}$ than the integral over $\mathscr{S}$.

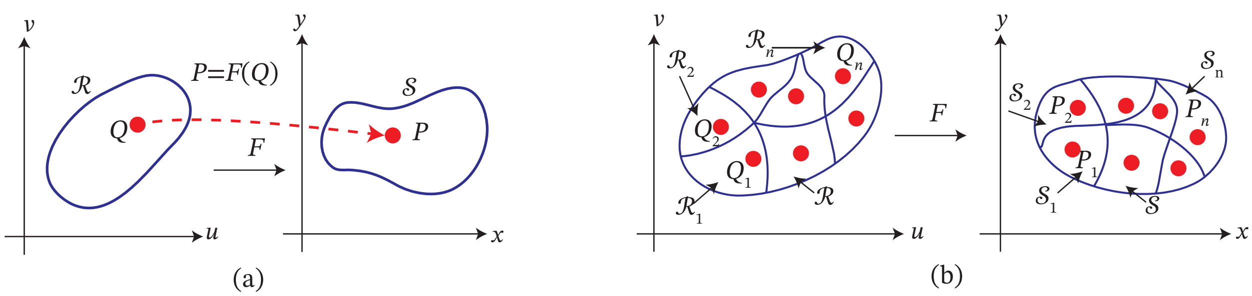


Figure 17.9.1

For a point P in $\mathscr{S}$, let Q be the point in $\mathscr{R}$ such that $F(Q) = P$, as shown in Figure 17.9.1(a). We can form an approximating sum for $\int_{\mathscr{S}} f(P)\ dA_{x,y}$ indirectly, as follows:

Partition $\mathscr{R}$ into n small patches $\mathscr{R}_1, \mathscr{R}_2, \dots, \mathscr{R}_n$. Then $\mathscr{S}_1 = F(\mathscr{R}_1)$, $\mathscr{S}_2 = F(\mathscr{R}_1)$, $\dots$, $\mathscr{S}_n = F(\mathscr{R}_n)$ is a partition of $\mathscr{S}$. Pick points Q_1 in $\mathscr{R}_1$, Q_2 in $\mathscr{R}_2$, $\dots$, Q_n in $\mathscr{R}_n$. Let $P_1 = F(Q_1)$, $P_2 = F(Q_2)$, $\dots$, $P_n = F(Q_n)$, as shown in Figure 17.9.1(b). Let the area of $\mathscr{S}_i$ be S_i.

Then
$$\sum_{i=1}^{n} f(P_i)S_i = \sum_{i=1}^{m} f(F(Q_i))S_i \tag{17.9.1}$$
is an approximation of $\int_{\mathscr{S}} f(P)\ dA_{u,v}$.

Suppose the area of $\mathscr{R}_i$ is A_i and $M_F(Q)$ is the magnification of F at Q. Because $M_F(Q_i)A_i$ is an approximation of S_i, in view of (17.9.1),
$$\sum_{i=1}^{n} f(F(Q_i))M_F(Q_i)A_i \tag{17.9.2}$$
is an approximation of $\int_{\mathscr{S}} f(P)\ dA_{x,y}$. But (17.9.2) is also an approximation of
$$\int_{\mathscr{R}} f(F(Q))\ M_F(Q)\ dA_{u,v}.$$
Taking limits as all $\mathscr{R}_i$ are chosen smaller and smaller leads to the following formula for changing domains in a multiple integral.

Formula 17.9.1: *Change of Domain in Multiple Integrals*

Assume that F is a mapping from a region $\mathcal{R}$ in uv-space to a region $\mathcal{S}$ in xy-space, as shown in Figure 17.9.1(a). When $f(P)$ be a scalar function defined on $\mathcal{S}$ and $M_F(Q)$ is the magnification of F for each Q in $\mathcal{R}$, then

$$\int_{\mathcal{S}} f(P)\, dA_{x,y} = \int_{\mathcal{R}} f(F(Q))\, M_F(Q)\, dA_{u,v}. \tag{17.9.3}$$

Equation (17.9.3) says we can replace an integral over $\mathcal{S}$ by an integral over $\mathcal{R}$. The equation remains valid even if F is not one-to-one on the boundary of $\mathcal{R}$. The notation in (17.9.3) is precise but forbidding. In shorthand, it is summarized by the equation $\boxed{dA_{x,y} = M_F(u,v)\, dA_{u,v}}$.

Applying the Idea

EXAMPLE 1. Let $\mathcal{S}$ be the parallelogram bounded by the lines $x+y=1$, $x+y=4$, $y-2x=2$, and $y-2x=3$. Evaluate $\int_{\mathcal{S}} x^2\, dA$.

SOLUTION The set $\mathcal{S}$ is shown in Figure 17.9.2(a).

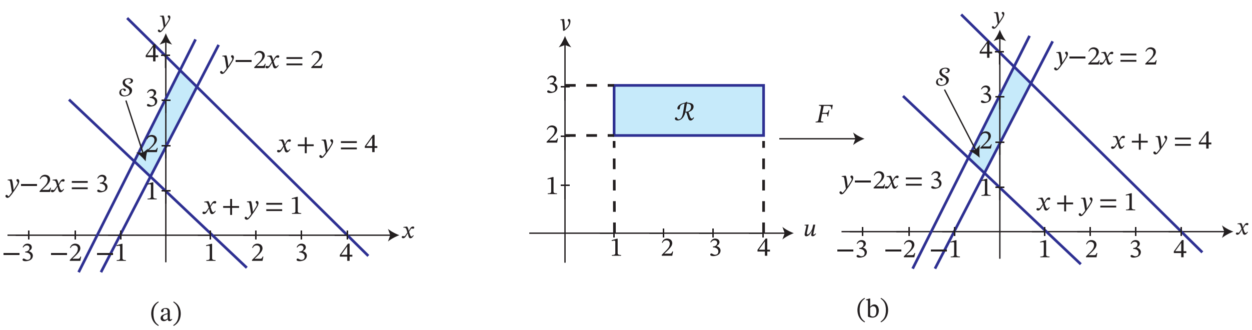


Figure 17.9.2

Evaluating $\int_{\mathcal{S}} x^2\, dA$ by iterated integrals would require breaking $\mathcal{S}$ into two triangles and a parallelogram. Instead, let us change the domain.

Define $u = x+y$ and $v = y-2x$. Then the four sides of $\mathcal{S}$ are $u=1$, $v=2$, $u=4$, and $v=3$. Moreover, solving for x and y in terms of u and v reveals that $x=(u-v)/3$ and $y=(2u+v)/3$. Or, in the language of mappings, $\mathcal{S}$ is the image of the rectangle $\mathcal{R}$ in uv-space described by $1\le u\le 4$, $2\le v\le 3$ by the mapping

$$F(u,v) = \left(\frac{u-v}{3}, \frac{2u+v}{3}\right),$$

as shown in Figure 17.9.2(b). If $F(u,v)=(x,y)$, we have $x=(u-v)/3$ and $y=(2u+v)/3$.

The magnification of a mapping is the absolute value of the Jacobian of the mapping. (Recall that the Jacobian of a mapping was defined, in Section 16.9, in terms of partial derivatives.) In this case, the Jacobian of F is

$$\det\begin{pmatrix} \frac{\partial}{\partial u}\left(\frac{u-v}{3}\right) & \frac{\partial}{\partial u}\left(\frac{2u+v}{3}\right) \\ \frac{\partial}{\partial v}\left(\frac{u-v}{3}\right) & \frac{\partial}{\partial v}\left(\frac{2u+v}{3}\right)\end{pmatrix} = \det\begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{-1}{3} & \frac{1}{3}\end{pmatrix} = \frac{1}{9}+\frac{2}{9} = \frac{1}{3}.$$

Since this determinant is positive, it is the magnification of this mapping. Then, using (17.9.3),

$$\int_{\mathscr{S}} x^2 \, dA_{x,y} = \int_{\mathscr{R}} \left(\frac{u-v}{3}\right)^2 \left(\frac{1}{3}\right) \, dA_{u,v}.$$

The latter integral is easily evaluated as an iterated integral in which u and v are the variables:

$$\int_{\mathscr{R}} \left(\frac{u-v}{3}\right)^2 \left(\frac{1}{3}\right) \, dA_{u,v} = \frac{1}{27} \int_{\mathscr{R}} (u-v)^2 \, dA_{u,v} = \frac{1}{27} \int_1^4 \int_2^3 (u-v)^2 \, dv \, du.$$

The inner integration, with respect to v, is

$$\int_2^3 (u-v)^2 \, dv = \left.\frac{-(u-v)^3}{3}\right|_{v=2}^{v=3} = \frac{-(u-3)^3}{3} - \frac{-(u-2)^3}{3} = \frac{-(u-3)^3}{3} + \frac{(u-2)^3}{3}.$$

The outer integration, with respect to u, is

$$\int_1^4 \left(\frac{-(u-3)^3}{3} + \frac{(u-2)^3}{3}\right) du = \left.\left(\frac{-(u-3)^4}{12} + \frac{(u-2)^4}{12}\right)\right|_{u=1}^{u=4} = \left(\frac{-(1^4)}{12} + \frac{2^4}{12}\right) - \left(\frac{-(-2)^4}{12} + \frac{(-1)^4}{12}\right) = \frac{30}{12} = \frac{5}{2}.$$

Thus $\int_{\mathscr{S}} x^2 \, dA = (1/27) \cdot (5/2) = 5/54$. ♦

EXAMPLE 2. Consider the region $\mathscr{S}$ in the first quadrant of the xy-plane bounded by the circles of radii 1 and 2 with centers at the origin and by the lines $y = x$ and $y = \frac{x}{\sqrt{3}}$. Find $\int_{\mathscr{S}} (x^2 + y^2) \, dA$.

SOLUTION The region $\mathscr{S}$ appeared in Examples 2 and 3 of Section 16.9. It is the image by the mapping $F(u, v) = (u\cos(v), u\sin(v))$ of the rectangle $\mathscr{R}$ in the uv-plane described by $1 \le u \le 2$, $\pi/6 \le v \le \pi/4$. Then

$$\int_{\mathscr{S}} (x^2 + y^2) \, dA_{x,y} = \int_{\mathscr{R}} \left((u\cos(v))^2 + (u\sin(v))^2\right) M_F(u, v) \, dA_{u,v}.$$

The magnification of F is

$$M_F(u, v) = \left|\det\begin{pmatrix} \frac{\partial}{\partial u}(u\cos(v)) & \frac{\partial}{\partial u}(u\sin(v)) \\ \frac{\partial}{\partial v}(u\cos(v)) & \frac{\partial}{\partial v}(u\sin(v)) \end{pmatrix}\right| = \left|\det\begin{pmatrix} \cos(v) & \sin(v) \\ -u\sin(v) & u\cos(v) \end{pmatrix}\right| = |u|.$$

Because $u > 0$ and $(u\cos(v))^2 + (u\sin(v))^2 = u^2$, we have

$$\int_{\mathscr{S}} (x^2 + y^2) \, dA_{x,y} = \int_{\mathscr{R}} \underbrace{u^2}_{\text{integrand}} \cdot \underbrace{|u|}_{\text{magnification of } F} \, dA_{u,v} = \int_{\mathscr{R}} u^3 \, dA_{u,v}.$$

Finally,

$$\int_{\mathscr{R}} u^3 \, dA_{u,v} = \int_{\pi/6}^{\pi/4} \int_1^2 u^3 \, du \, dv = \int_{\pi/6}^{\pi/4} \left.\frac{u^4}{4}\right|_{u=1}^{u=2} dv = \int_{\pi/6}^{\pi/4} \frac{15}{4} \, dv = \frac{15}{4}\left(\frac{\pi}{4} - \frac{\pi}{6}\right) = \frac{5}{16}\pi.$$

♦

In this example the magnification introduced the extra factor u in the integrand. But u and v are the same as the r and θ of polar coordinates. The extra factor r was introduced in Section 17.3 because the area of a small patch is $r \, \Delta r \, \Delta\theta$, not simply $\Delta r \, \Delta\theta$. Now we see that the factor r can also be viewed as the magnification of a certain mapping.

Observation 17.9.1: *Magnification of Mappings to Cylindrical and Spherical Coordinates*

Exercise 17 in Section 16.9 develops the magnification of a mapping from uvw-space to xyz-space. One can check that the magnification associated with cylindrical coordinates (r, θ, z) where

$$x = r\cos(\theta), \qquad y = r\sin(\theta), \qquad z = z$$

is r, which then must be introduced as a factor in the integrand.

The magnification associated with spherical coordinates (ρ, θ, ϕ), where

$$x = \rho\sin(\phi)\cos(\theta), \qquad y = \rho\sin(\phi)\sin(\theta), \qquad z = \rho\cos(\phi),$$

is $\rho^2\sin(\phi)$, which is consistent with what we found in Section 17.6 by considering the volume of a small patch corresponding to changes $\Delta\rho$, $\Delta\phi$, $\Delta\theta$ in the coordinates.

Summary

A mapping F from $\mathscr{R}$ to $\mathscr{S}$ enables us to replace an integral over $\mathscr{S}$ by an integral over $\mathscr{R}$:

$$\int\limits_{\mathscr{S}} f(P)\, dA_{x,y} = \int\limits_{\mathscr{R}} f(F(Q))\, M_F(Q)\, dA_{u,v}.$$

The magnification of F appears because it tells by how much F magnifies the area of a small patch in $\mathscr{R}$. The factors that we introduced in earlier sections into integrands, r for polar and cylindrical coordinates and $\rho^2\sin(\phi)$ for spherical coordinates, are instances of magnifications.

EXERCISES for Section 17.9

1. State the relation between an integral over $\mathscr{R}$ and an integral over its image $\mathscr{S}$ by a mapping F, $\mathscr{S} = F(\mathscr{R})$. Use as few symbols as you can.

2. State, using as few symbols as you can, why the magnification of F appears in the integrand on the right-hand side of $\int\limits_{\mathscr{S}} f(P)\, dA_{x,y} = \int\limits_{\mathscr{R}} f(F(Q)) M_F(Q)\, dA_{u,v}$. Start your explanation from an approximating sum.

3. Denote by $\mathscr{S}$ the elliptical region $\frac{x^2}{25} + \frac{y^2}{16} \le 1$.

(a) Sketch $\mathscr{S}$.

(b) Find a mapping F from the disk $\mathscr{R}$ described by $u^2 + v^2 \le 1$ to $\mathscr{S}$.

(c) Use it to evaluate $\int\limits_{\mathscr{S}} \sin^2(x)\, dA$.

(d) Do enough of the direct calculation of $\int\limits_{\mathscr{S}} \sin^2(x)\, dA$ without using a mapping to see that it is more complicated than the method in (c).

4. (a) Find a linear mapping F from the uv-plane to the xy-plane such that the image of the square with vertices $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$ is the parallelogram $\mathscr{S}$ with vertices $(0,0)$, $(3,2)$, $(5,-1)$, and $(2,-3)$.

(b) Use it to evaluate $\int\limits_{\mathscr{S}} xy\, dA$.

5. (a) Use the mapping in Example 1 to find the area of the parallelogram in that example.

(b) Find the area of the parallelogram by using the cross product.

6. The magnification of a mapping F at (u, v) is $u^2 v$. Find the area of the image of the rectangle $1 \le u \le 2$, $3 \le v \le 5$.

In Exercises 7 to 10, construct a mapping F from some region $\mathscr{R}$ to $\mathscr{S}$, and use it to evaluate the given integral.

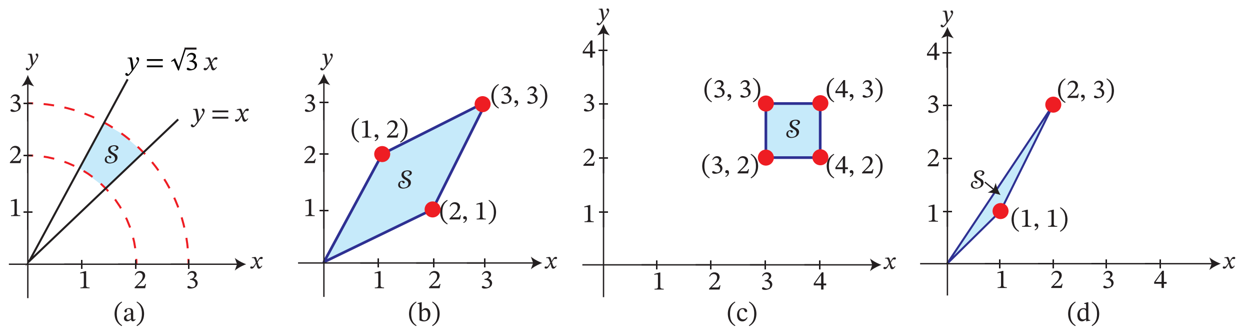


Figure 17.9.3

7. $\int_{\mathscr{S}} x^2\, dA$. (See Figure 17.9.3(a).)

8. $\int_{\mathscr{S}} (x+y)\, dA$. (See Figure 17.9.3(b).)

9. $\int_{\mathscr{S}} xy\, dA$. (See Figure 17.9.3(c).)

10. $\int_{\mathscr{S}} \cos(x)\, dA$. (See Figure 17.9.3(d).)

11. It is plausible that if the local magnification of a mapping has the constant value k, then it would magnify all areas by that factor. However, there is nothing in the definition of the local magnification that assures us that this is so. After all, the definition involves a limit of quotients that need not equal k. However, a mapping that has a constant local magnification k does magnify all areas by the factor k. Justify this claim.

12. Show that an integral of the form $\int_a^b f(x)\, dx$ can be replaced by an integral of the form $\int_0^1 g(u)\, du$.

13. Mappings $F(u,v) = (au+bv+e, cu+dv+f)$, with $ad-bc \neq 0$, are **affine mappings**. Note that $F(0,0) = (e,f)$. The linear mappings are the affine mappings that map the origin in the uv-plane to the origin in the xy-plane. Show that if $\mathscr{S}$ is a triangle in the xy-plane there is an affine mapping from the triangle $\mathscr{R}$ in the uv-plane whose vertices are $(0,0)$, $(1,0)$, and $(0,1)$ onto $\mathscr{S}$. That implies that an integral over a triangle can be replaced by an integral over the fixed triangle $\mathscr{R}$.

14. **Sam:** I can even use a mapping to get rid of improper integrals.
Jane: Another of your tricks.
Sam: Say I had $\int_0^\infty e^{-x^2}\, dx$. The mapping $x = \tan(u)$ sends the interval $\left[0, \frac{\pi}{2}\right)$ onto the infinite interval $[0,\infty)$. Since $dx = \sec^2(u)\, du$, it follows that that integral equals $\int_0^{\pi/2} e^{-\tan^2(u)} \sec^2(u)\, du$.
Jane: Very impressive. Surely something is wrong.
Is Sam's claim correct for a change?

15. Let F be described by $x = u^2 - v^2$, $y = uv$.
(a) Find the magnification of F.
(b) Let $\mathscr{R}$ be the square with vertices $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$ and let $\mathscr{S}$ be $F(R)$. Sketch $\mathscr{S}$.
(c) Evaluate $\int_{\mathscr{S}} y^2\, dA$ using the magnification of F found in (a).

16. (a) Sketch the set $\mathscr{S}$ in xyz-space given by $x^2 + \frac{y^2}{9} + \frac{z^2}{16} \le 1$.

(b) Find its volume by using a mapping from a ball $\mathscr{R}$ of radius 1, whose volume is $\frac{4\pi}{3}$.

17. This exercise uses the magnification to find the formula for the area of a surface of revolution. In Section 9.5 we developed such a formula. Let f be a positive function defined for x in $[a,b]$. The graph of f is rotated around the x-axis to produce a surface of revolution, $\mathscr{S}$.

(a) Sketch $\mathscr{S}$.
(b) Show on the sketch that each point (x,y,z) on $\mathscr{S}$ is determined by x and an angle θ, $0 \le \theta \le 2\pi$.
(c) Express (x,y,z) in terms of x and θ.
(d) Part (c) describes a mapping from the rectangle $a \le x \le b$, $0 \le \theta \le 2\pi$. Find its magnification.
(e) Use the magnification to show that the area of $\mathscr{S}$ is $2\pi \int_a^b \sqrt{1 + (f'(x))^2}\, dx$.

We have concentrated on using mappings to simplify the domain of integration. They may also simplify an integrand, as the next two exercises show.

18. Consider $\int_{\mathscr{S}} \exp\left(\frac{y-x}{x+y}\right) dA$ where $\mathscr{S}$ is the triangle bounded by the lines $x = 0$, $y = 0$, and $x + y = 1$. The substitution $u = x + y$, $v = y - x$ simplifies the integrand.

(a) Sketch $\mathscr{S}$ and the set $\mathscr{R}$ in uv-space that is the image of $\mathscr{S}$ under the mapping $u = x + y$, $v = y - x$.
(b) Find the mapping F from $\mathscr{R}$ to $\mathscr{S}$ that is the inverse of the mapping in (a).
(c) Use F to evaluate the integral. (Choose the iterated integral wisely.)

19. Evaluate $\int_{\mathscr{S}} (x+y)^2 \cos^2(x-y)\, dA$, where $\mathscr{S}$ is the square whose vertices have rectangular coordinates $\left(\frac{\pi}{2}, 0\right)$, $\left(\pi, \frac{\pi}{2}\right)$, $\left(\frac{\pi}{2}, \pi\right)$, and $\left(0, \frac{\pi}{2}\right)$.

17.S Chapter Summary

This chapter generalized the notion of a definite integral over an interval to integrals over plane sets, surfaces, and solids. The definitions of multiple integrals are almost the same, the integral of $f(P)$ over a set being the limit of sums of the form $\sum f(P_i)\, A_i$, $\sum f(P_i)\, S_i$, or $\sum f(P_i)\, V_i$ for integrals over plane sets, surfaces, or solids, respectively.

Three different interpretations of double integrals are given in Table 17.S.1. The only difference is the interpretation of the integrand.

Integral	Integrand	Interpretation
$\int_{\mathscr{R}} 1\, dA$	1	Area of $\mathscr{R}$
$\int_{\mathscr{R}} \sigma(P)\, dA$	$\sigma(P)$ = density (mass per unit area)	Mass of solid
$\int_{\mathscr{R}} c(P)\, dA$	$c(P)$ = length of cross section of solid	Volume of solid

Table 17.S.1

Average value extends easily to functions of several variables. For instance, if $f(P)$ is defined on a plane region $\mathscr{R}$, the average value of f over $\mathscr{R}$ is defined as

$$\frac{1}{\text{Area of } \mathscr{R}} \int_{\mathscr{R}} f(P)\, dA.$$

Mass, first moment, center of mass, and moment of inertia were also discussed in this chapter. Table 17.S.2 summarizes these ideas for (two-dimensional) planar regions in the xy-plane and Table 17.S.3 summarizes the same ideas for (three-dimensional) solid regions in space.

Some multiple integrals (also known as double or triple integrals) can be calculated by repeated integrations over intervals, that is, as iterated integrals. This requires a description of the region in a coordinate system and replaces dA or dV by an expression based on the area or volume of a small patch swept out by small changes in the coordinates, as recorded in Table 17.S.4.

An integral over a surface $\mathscr{S}$, $\int_{\mathscr{S}} f(P)\, dA$, can be replaced by an integral over the projection of $\mathscr{S}$ onto a plane $\mathscr{R}$, replacing dS by $dA/|\cos(\gamma)|$ where γ is the (typically varying) angle between a normal to $\mathscr{S}$ and a normal to $\mathscr{R}$.

The final section showed the role of the magnification of a mapping when substituting an integral over one set by an integral over another. The magnification of a mapping is the factor that is inserted into an integrand if the computation uses coordinates other than rectangular: r for polar coordinates and $\rho^2 \sin(\phi)$ for spherical coordinates. That is, if F is a mapping from $\mathscr{R}$ to $\mathscr{S}$, $\int_{\mathscr{S}} f(P)\, dA_{x,y}$ equals $\int_{\mathscr{R}} f(F(Q)) M_F(Q)\, dA_{u,v}$ where M_F is the magnification of the mapping F. The equation holds whether $\mathscr{R}$ and $\mathscr{S}$ are solids, surfaces, or intervals. The last case was called integration by substitution in Section 8.2.

Formula	Significance
$\int_{\mathscr{R}} 1\, dA$	Area of $\mathscr{R}$
$\frac{1}{\text{Area}(\mathscr{R})} \int_{\mathscr{R}} f(P)\, dA$	Average value of f over $\mathscr{R}$
$M = \int_{\mathscr{R}} \sigma(P)\, dA$	Total mass of $\mathscr{R}$ ($\sigma(P)$ denotes density for a plane region)
$M_x = \int_{\mathscr{R}} y\, \sigma(P)\, dA$ $M_y = \int_{\mathscr{R}} x\, \sigma(P)\, dA$	(First) Moments about x- and y-axes, respectively ($\sigma(P)$ denotes density for a plane region)
$\left(\frac{M_y}{M}, \frac{M_x}{M}\right)$	Center of mass, $(\overline{x}, \overline{y})$ (Centroid, when $\sigma(P) = 1$)
$I_x = \int_{\mathscr{R}} y^2\, \sigma(P)\, dA$ $I_y = \int_{\mathscr{R}} x^2\, \sigma(P)\, dA$	Moment of inertia (second moments) about x- and y-axes ($\sigma(P)$ denotes density for a plane region)
$\int_{\mathscr{R}} r(P)\, \sigma(P)\, dA$	(First) Moment about a line L in the plane ($r(P)$ is the distance from P to the fixed line L)
$\int_{\mathscr{R}} r(P)^2\, \sigma(P)\, dA$	Moment of inertia (second moment) around a line L for a plane region ($r(P)$ is the distance from P to the fixed line L)

Table 17.S.2

Formula	Significance
$\displaystyle\int_{\mathscr{R}} 1\, dV$	Volume of $\mathscr{R}$
$\displaystyle\frac{1}{\text{Volume}(\mathscr{R})}\int_{\mathscr{R}} f(P)\, dV$	Average value of f over $\mathscr{R}$
$\displaystyle M = \int_{\mathscr{R}} \delta(P)\, dV$	Total mass of $\mathscr{R}$ ($\delta(P)$ denotes density for a solid region)
$\displaystyle M_{xy} = \int_{\mathscr{R}} z\,\delta(P)\, dV$ $\displaystyle M_{xz} = \int_{\mathscr{R}} y\,\delta(P)\, dV$ $\displaystyle M_{yz} = \int_{\mathscr{R}} x\,\delta(P)\, dV$	(First) Moment relative to xy-plane, xz-plane, and yz-plane, respectively ($\delta(P)$ denotes density for a solid region)
$\displaystyle\left(\frac{M_{yz}}{M}, \frac{M_{xz}}{M}, \frac{M_{xy}}{M}\right)$	Center of mass of solid, $(\overline{x}, \overline{y}, \overline{z})$ (Centroid, when $\delta(P) = 1$.)
$\displaystyle\int_{\mathscr{R}} r(P)\,\delta(P)\, dV$	(First) Moment about a line L for a solid region ($r(P)$ is the distance from P to the fixed line L)
$\displaystyle\int_{\mathscr{R}} r(P)\,\delta(P)\, dV$	(First) Moment relative to a plane ($r(P)$ is the distance from P to a fixed plane)
$\displaystyle\int_{\mathscr{R}} r(P)^2\,\delta(P)\, dV$	Moment of inertia (second moment) around a line L for a solid region ($r(P)$ is the distance from P to the fixed line L)

Table 17.S.3

Coordinate System	Substitution
Rectangular (2-D)	$dA = dx\, dy$
Rectangular (3-D)	$dV = dx\, dy\, dz$
Polar	$dA = r\, dr\, d\theta$
Cylindrical	$dV = r\, dr\, d\theta\, dz$
Cylindrical (surface)	$dS = r\, d\theta\, dz$
Spherical	$dV = \rho^2 \sin(\phi)\, d\phi\, d\rho\, d\theta$
Spherical (surface)	$dS = \rho^2 \sin(\phi)\, d\phi\, d\theta$

Table 17.S.4

EXERCISES for Section 17.S

1. In each of the following conversions from one integral to another, what is the "local magnification" that is involved?

(a) Using substitution to replace an integral over $[a, b]$ by an integral over $[c, d]$. (Section 8.2)

(b) Replacing an integral over a curve by an integral over an interval. (Section 15.3)

(c) Replacing an integral over a surface that is the graph of $z = f(x, y)$ by an integral over its projection on the xy-plane. (Section 17.7)

(d) Replacing a double integral in rectangular coordinates by an equivalent iterated integral in polar coordinates. (Section 17.3)

(e) Using spherical coordinates to integrate over a ball or solid cone. (Section 17.6)

(f) Using spherical coordinates to integrate over the surface of a sphere or cone. (Section 17.7)

2. The temperature at (x, y) at time t is $T(x, y, t) = e^{-tx}\sin(x+3y)$. Let $f(t)$ be the average temperature in the rectangle $0 \le x \le \pi$, $0 \le y \le \pi/2$ at time t. Find $\frac{df}{dt}$.

3. Let f be a function such that $f(-x, y) = -f(x, y)$.

(a) Give some examples of such functions.

(b) For what type regions $\mathscr{R}$ in the xy-plane is $\int\limits_{\mathscr{R}} f(x, y)\, dA$ certainly equal to 0?

4. Find $\int\limits_{\mathscr{R}} (2x^3y^2+7)\, dA$ where $\mathscr{R}$ is the square with vertices $(1,1)$, $(-1,1)$, $(-1,-1)$, and $(1,-1)$. Do this with as little work as possible.

5. Let $f(x, y)$ be a continuous function. Define $g(x)$ to be $\int\limits_{\mathscr{R}} f(P)\, dA$, where $\mathscr{R}$ is the rectangle with vertices $(3,0)$, $(3,5)$, $(x,0)$, and $(x,5)$, $x > 3$. Express $\frac{dg}{dx}$ as a suitable integral.

6. Let $\mathscr{R}$ be a plane lamina in the shape of the region bounded by the graph of the function $r = 2a\sin(\theta)$ $(a > 0)$. If the variable density of the lamina is given by $\sigma(r, \theta) = \sin(\theta)$, find the center of mass of $\mathscr{R}$.

In Exercises 7 and 8, use iterated integrals in polar coordinates to find the point.

7. The centroid of the region within the cardioid $r = 1+\cos(\theta)$.

8. The centroid of the region within the leaf of $r = \cos^3(\theta)$ that lies along the polar axis.

In Exercises 9 to 12 find the moment of inertia of a homogeneous lamina of mass M of the given shape, around the given line.

9. A disk of radius a, about the line perpendicular to it through its center.

10. A disk of radius a, about a line perpendicular to it through a point on the circumference.

11. A disk of radius a, about a diameter.

12. A disk of radius a, about a tangent.

---

13. (a) In a diagram much larger than Figure 16.9.4 in Section 16.9, show $\mathscr{C}$ and the parallelogram that approximates it. Include the vectors $\frac{\partial \mathbf{r}}{\partial u}\Delta u$ and $\frac{\partial \mathbf{r}}{\partial v}\Delta v$.

(b) Why are the vectors in (a) tangent to the curves that meet at $F(u_0, v_0)$?

14. Let $F(u,v,w) = (u\sin(v)\cos(w), u\sin(v)\sin(w), u\cos(v))$. Let $\mathscr{R}$ in uvw-space be described by $1 \le u \le 2$, $0 \le v \le \pi/4$, $0 \le w \le \pi/2$. (a) Sketch $\mathscr{R}$. (b) Sketch $F(\mathscr{R})$. (c) Find the magnification of F.

15. Let $\mathscr{S}$ be the spherical surface with radius a and center at the origin. We want to find $\int_{\mathscr{S}} (xz + y^2)\,dS$.

(a) Why is $\int_{\mathscr{S}} xz\,dS = 0$?

(b) Why is $\int_{\mathscr{S}} x^2\,dS = \int_{\mathscr{S}} y^2\,dS = \int_{\mathscr{S}} z^2\,dS$?

(c) Why is $\int_{\mathscr{S}} y^2\,dS = \int_{\mathscr{S}} (a^2/3)\,dS$?

(d) Show that $\int_{\mathscr{S}} (xz + y^2)\,dS = 4\pi a^4/3$.

16. Let a be a positive number and $\mathscr{R}$ the region bounded by $y = x^a$, the x-axis, and the line $x = 1$.

It is true that the centroid lies in $\mathscr{R}$ for all positive values of a, but the proof is more difficult.

(a) Show that the centroid of $\mathscr{R}$ is $\left(\frac{a+1}{a+2}, \frac{a+1}{4a+2}\right)$.

(b) Find $\lim_{a\to\infty} \overline{x}$ and $\lim_{a\to\infty} \overline{y}$.

(c) Show that the centroid of $\mathscr{R}$ lies in $\mathscr{R}$ for all large values of a.

17. Define the **moment of a curve in the xy-plane around the x-axis** to be $\int_0^L y\,ds$, where L is the length of the curve and s is arc length. The **moment of the curve around the y-axis** is defined as $\int_0^L x\,ds$. The **centroid** of the curve, $(\overline{x}, \overline{y})$, is defined by setting $\overline{x} = \dfrac{\int_0^L x\,ds}{\text{Length of curve}}$ and $\overline{y} = \dfrac{\int_0^L y\,ds}{\text{Length of curve}}$. Find the centroid of the top half of the circle $x^2 + y^2 = a^2$.

18. Show that the area of the surface obtained by revolving about the x-axis a curve that lies above it is equal to the length of the curve times the distance that the centroid of the curve moves.

NOTE: See Exercise 17.

19. Use Exercise 18 to find the surface area of the torus formed by revolving a circle of radius a around a line a distance b from its center, $b \ge a$.

20. Use Exercise 18 to find the area of the curved part of a cone of radius a and height h.

21. Let $f(P)$ and $g(P)$ be continuous functions defined on a plane region $\mathscr{R}$.

(a) Show that $\left(\int_{\mathscr{R}} f(P)g(P)\,dA\right)^2 \le \left(\int_{\mathscr{R}} f(P)^2\,dA\right)\left(\int_{\mathscr{R}} g(P)^2\,dA\right)$.

(b) Show that if equality occurs in the inequality in (a), then either f is a nonzero constant times g or at least one of f and g is the zero function. (Assume $\mathscr{R}$ is the closure of an open set, that is, for each point P on the boundary of $\mathscr{R}$ there is a curve consisting of interior points of $\mathscr{R}$ except that the terminal point is P.)

22. A solid region $\mathscr{R}$ is bounded below by the xy-plane, above by the surface $z = f(P)$, and the sides by the surface of a cylinder, as shown in Figure 17.S.1. The volume of $\mathscr{R}$ is V. If V is fixed, show that the top surface that minimizes the height of the centroid of $\mathscr{R}$ is a horizontal plane. CONTRIBUTED BY: G. D. Chakerian

NOTE: Water in a glass illustrates this, for nature minimizes the height of the centroid of the water.

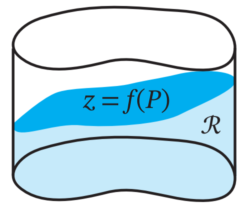


Figure 17.S.1

23. Find the average distance from points in a disk of radius a to the center of the disk.
- (a) Set up the pertinent definite integral in rectangular coordinates.
- (b) Set it up in polar coordinates.
- (c) Evaluate the easier integral in (a) and (b).

24. Find the average distance from points in a square of side a to the center of the square.
- (a) Set up the pertinent definite integral in rectangular coordinates.
- (b) Set it up in polar coordinates.
- (c) Evaluate the easier integral in (a) and (b).

25. Find the average distance from points in a ball of radius a to the center of the ball.
- (a) Set up the pertinent definite integral in rectangular coordinates.
- (b) Set it up in spherical coordinates.
- (c) Evaluate the easier integral in (a) and (b).

26. Find the average distance from points in a cube of side a to the center of the cube.
- (a) Set up the pertinent definite integral in rectangular coordinates.
- (b) Set it up in polar coordinates.
- (c) Evaluate the easier integral in (a) and (b).

Exercises 27 and 28 refer to the distance from a point to a curve, that is, the shortest distance from the given point to a point on the curve.

27. Find the average distance from points in a square of side a to the border of the square.
- (a) Set up the pertinent definite integral in rectangular coordinates.
- (b) Set it up in polar coordinates.
- (c) Evaluate the easier integral in (a) and (b).

28. Find the average distance from the points in a disk of radius a to the circular border.
- (a) Before doing any calculations, decide whether the average distance is greater than $a/2$ or less than $a/2$. Explain how you made this decision.
- (b) Carry out the calculation using a convenient coordinate system.

---

29. (a) Show that a region of diameter d can always fit into a disk of diameter $2d$.
(b) Can it always fit into a disk of diameter d?

30. If a region has diameter d, (a) how small can its area be? (b) Show that its area is less than or equal to $\pi d^2/2$.

31. Let A and B be two points in the xy-plane. A curve C (in the xy-plane) consists of all points P such that the sum of the distances from P to A and P to B is constant, say $2a$. The distance from P to A is a function of arc length on C. Find the average of that distance.

Exercises 32 to 34 concern the moment of inertia. Note that if the object is homogeneous and has mass M and volume V, its density $\delta(P)$ is M/V.

32. A homogeneous rectangular solid box has mass M and sides of lengths a, b, and c. Find its moment of inertia about an edge of length a.

33. A rectangular homogeneous box of mass M has dimensions a, b, and c. Show that the moment of inertia of the box about a line through its center and parallel to the sides of length a is $M(b^2+c^2)/12$.

34. A right solid circular cone has altitude h, radius a, constant density, and mass M.

(a) Why is its moment of inertia about its axis less than Ma^2?

(b) Show that its moment of inertia about its axis is $3Ma^2/10$.

Exercises 35 and 36 imply that the centroid of a region $\mathscr{R}$ does not depend on the particular choice of xy-axes used to define it. That means that the centroid is an intrinsic geometric property of $\mathscr{R}$. More generally, the center of mass does not depend on the choice of axes; it is an intrinsic property of the distribution of the mass.

35. Matter in a plane region $\mathscr{R}$ has density $\sigma(P)$. Relative to the xy-axes its center of mass is $P = (\overline{x}, \overline{y})$. Introduce a second rectangular coordinate system with $x'y'$-axes parallel to the original system. The $x'y'$-axes are a translation of the xy-axes. The origin of the $x'y'$-axes is at the point (h, k) relative to the xy-axes. The center of mass is $Q = (\overline{x'}, \overline{y'})$ when computed using the $x'y'$-axes. Show that $P = Q$. This shows that the center of mass does not depend on the choice of axes, as long as one set of axes is a translation of the other set.

36. Assume that the center of mass is $(0,0)$ when computed relative to the xy-axes. These axes are rotated around $(0,0)$ by an angle θ to produce $x'y'$-axes.

(a) Show that the center of mass computed with the $x'y'$-axes is the same as the center of mass computed with the xy-axes.

(b) From (a) and the preceding Exercise, show that the center of mass does not depend on the particular coordinate system chosen.

Exercises 37 and 38 are related. Both concern a decreasing function $z = g(y)$ that depends only on y such that $g(1) = 0$. Let $\mathscr{R}$ be the solid of revolution formed by revolving about the z-axis the region $\mathscr{R}_0$ in the yz-plane bounded by $y = 0$, $z = 0$, and $z = g(y)$.

37. Show that $\int_{\mathscr{R}} z\,dV = \int_0^1 \pi y \left(g(y)\right)^2\,dy$ and $\int_{\mathscr{R}} z\,dV = \int_0^{g(0)} \pi z \left(g^{-1}(z)\right)^2\,dz$.

38. (a) Show that the z-coordinates of the centroid of the solid $\mathscr{R}$ and of the centroid of the plane region $\mathscr{R}_0$ that was revolved to to form $\mathscr{R}$ are given by $\dfrac{\int_0^1 \frac{x}{2}\left(g(x)\right)^2\,dx}{\int_0^1 x g(x)\,dx}$ and $\dfrac{\int_0^1 \frac{1}{2}\left(g(x)\right)^2\,dx}{\int_0^1 g(x)\,dx}$, respectively.

(b) By considering $\int_0^1\int_0^1 g(x)g(y)(x-y)(g(x)-g(y))\,dx\,dy$, show that the centroid of the solid of revolution is below that of the plane region.

---

39. A solid of varying density $\delta(P)$ occupies the region $\mathscr{R}$ in space. Let L_1 be a line through its center of mass and L_2 a line parallel to L_1 and at a distance r from it. Let I_1 be the moment of inertia of the solid around L_1 and I_2 the moment of inertia around L_2.

(a) Show that $I_2 = I_1 + r^2 M$ where M is the mass of the solid.

(b) Which choice of r leads to the smallest value of I_2?

(c) How could the center of mass be defined in terms of moments of inertia?

40. **Sam:** I can make things clearer and cut the book by two pages.

Jane: How?

Sam: Those guys make separate definitions for integrals over line segments, curves, plane regions, solid regions, and surfaces.

Jane: You don't like the definitions?

Sam: They're the same definition over and over.

Jane: So?

Sam: I'd make one definition back in Chapter 6 to do them all wholesale.

Jane: Impossible.

Sam: Look, each type involves a measure of size, length for intervals and curves, area for plane and surface regions, and volume for solid ones. I'll just write $m(\mathscr{R})$ for the measure of the region $\mathscr{R}$. I'm done. All those sums involving partitions have the same form: $\sum f(P_i)m(\mathscr{R}_i)$ Then I take the limit as the little $\mathscr{R}_i$ get small. That gets all types in one blow.

Jane: But we won't know how to compute them.

Sam: I'll add suspense. I'll promise that later material in Chapters 6, 15, and 17 will show how to compute them.

(a) Write out Sam's definition.

(b) Should Sam's definition appear in Chapter 6? (Explain.)

Calculus is Everywhere # 24
Solving the Wave Equation

Unlike every other CIE, this one does not use any ideas introduced in the current chapter. Instead, it brings together several fundamental ideas introduced in earlier chapters to finish the discussion started in CIE 23 at the end of Chapter 16.

The CIE 23 (Wave in a Rope) in Chapter 16 introduced the partial differential equation known as the wave equation:

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 y}{\partial t^2}. \tag{C.24.1}$$

We will solve it to find y as a function of x and t. First, we solve some simpler equations, which will help us solve equation (C.24.1).

EXAMPLE 1. Let $u(x, y)$ satisfy $\frac{\partial u}{\partial x} = 0$. Find the form of $u(x, y)$.

SOLUTION Since $\partial u/\partial x$ is 0, for a fixed value of y, $u(x, y)$ is constant. Thus, $u(x, y)$ depends only on y, and can be written in the form $h(y)$ for some function h of a single variable.

Conversely, any function $u(x, y)$ that can be written in the form $h(y)$ has the property that $\partial u/\partial x = 0$. ◆

EXAMPLE 2. Let $u(x, y)$ satisfy $\frac{\partial^2 u}{\partial x \partial y} = 0$. Find the form of $u(x, y)$.

SOLUTION We know that

$$\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial y}\right) = \frac{\partial^2 u}{\partial x \partial y} = 0.$$

By Example 1, $u_y = h(y)$ for some function $h(y)$.

When this equation is integrated over the interval $0 \le y \le b$ and for any x, the fundamental theorem of calculus can be applied to yield:

$$u(x, b) - u(x, 0) = \int_0^b \frac{\partial u}{\partial y}\,dy = \int_0^b h(y)\,dy.$$

Let H be an antiderivative of h. Then

$$u(x, b) - u(x, 0) = H(b) - H(0).$$

Replacing b by y shows that

$$u(x, y) = u(x, 0) + H(y) - H(0).$$

So $u(x, y)$ is a sum of a function of x and a function of y,

$$u(x, y) = f(x) + g(y). \tag{C.24.2}$$

A quick calculation shows that any function of this form satisfies $u_{xy} = 0$. ◆

We will solve the wave equation (C.24.1) by using a change of variables that transforms it into the one solved in Example 2.

Let c be a positive constant. The new variables are

$$p = x + ct \qquad \text{and} \qquad q = x - ct.$$

Solving these for x and t we obtain

$$x = \frac{1}{2}(p + q) \qquad \text{and} \qquad t = \frac{1}{2c}(p - q).$$

Now, apply the chain rule, where y is a function of p and q and p and q are functions of x and t, as indicated in Figure C.24.1. Thus $y(x,t) = u(p,q)$.

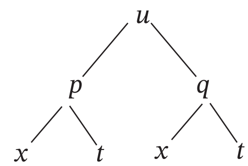


Figure C.24.1

Because

$$\frac{\partial p}{\partial x} = 1, \qquad \frac{\partial p}{\partial t} = c, \qquad \frac{\partial q}{\partial x} = 1, \qquad \text{and} \qquad \frac{\partial q}{\partial t} = -c,$$

we have

$$\frac{\partial y}{\partial x} = \frac{\partial u}{\partial p}\frac{\partial p}{\partial x} + \frac{\partial u}{\partial q}\frac{\partial q}{\partial x} = \frac{\partial u}{\partial p} + \frac{\partial u}{\partial q}.$$

Then

$$\begin{aligned}\frac{\partial^2 y}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial p} + \frac{\partial u}{\partial q}\right)\\ &= \frac{\partial}{\partial p}\left(\frac{\partial u}{\partial p} + \frac{\partial u}{\partial q}\right)\frac{\partial p}{\partial x} + \frac{\partial}{\partial q}\left(\frac{\partial u}{\partial p} + \frac{\partial u}{\partial q}\right)\frac{\partial q}{\partial x}\\ &= \left(\frac{\partial^2 u}{\partial p^2} + \frac{\partial^2 u}{\partial p \partial q}\right)\cdot 1 + \left(\frac{\partial^2 u}{\partial q \partial p} + \frac{\partial^2 u}{\partial q^2}\right)\cdot 1.\end{aligned}$$

Thus

$$\frac{\partial^2 y}{\partial x^2} = \frac{\partial^2 u}{\partial p^2} + 2\frac{\partial^2 u}{\partial p \partial q} + \frac{\partial^2 u}{\partial q^2}. \tag{C.24.3}$$

A similar calculation shows that

$$\frac{\partial^2 y}{\partial t^2} = c^2\left(\frac{\partial^2 u}{\partial p^2} - 2\frac{\partial^2 u}{\partial p \partial q} + \frac{\partial^2 u}{\partial q^2}\right). \tag{C.24.4}$$

Substituting (C.24.3) and (C.24.4) in (C.24.1) leads to

$$\frac{\partial^2 u}{\partial p^2} + 2\frac{\partial^2 u}{\partial p \partial q} + \frac{\partial^2 u}{\partial q^2} = \frac{1}{c^2}\left(c^2\right)\left(\frac{\partial^2 u}{\partial p^2} - 2\frac{\partial^2 u}{\partial p \partial q} + \frac{\partial^2 u}{\partial q^2}\right),$$

which reduces to

$$4\frac{\partial^2 u}{\partial p \partial q} = 0.$$

By Example 2, there are function $f(p)$ and $g(q)$ such that

$$u(p,q) = f(p) + g(q)$$

which, since $y(x,t) = u(p,q)$, can be written as

$$y(x,t) = f(x+ct) + g(x-ct). \tag{C.24.5}$$

The expression (C.24.5) is the most general solution of the wave equation (C.24.1).

QUESTION: What does a solution (C.24.5) look like? What does the parameter c tell us?

To answer these questions, suppose

$$y(x,t) = g(x-ct). \tag{C.24.6}$$

Here t represents time. For each t, $y(x,t) = g(x-ct)$ is a function of x and we can graph it in the xy-plane. For $t = 0$, (C.24.6) becomes

$$y(x,0) = g(x).$$

That is the graph of $y = g(x)$, whatever g is, as shown in Figure C.24.2(a).

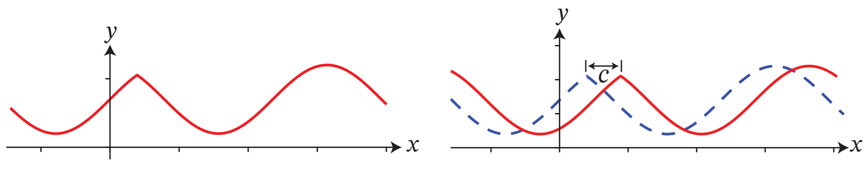


Figure C.24.2

One unit of time later, when $t = 1$,

$$y = y(x,1) = g(x - c \cdot 1) = g(x - c).$$

The value of $y(x,1)$ is the same as the value of g at $x - c$, c units to the left of x. So the graph at $t = 1$ is the graph of g in Figure C.24.2(a) shifted to the right c units, as in Figure C.24.2(b).

As t increases, the initial wave shown in Figure C.24.2(a) moves to the right at the constant speed, c. Thus c tells us the wave's velocity. That will play a role in Maxwell's prediction that electromagnetic waves travel at the speed of light, as we will see in CIE 26 at the end of Chapter 18.

EXERCISES for CIE C.24

1. We interpreted $y(x,t) = g(x - ct)$ as the description of a wave moving with speed c to the right. What is the corresponding interpretation of $y(x,t) = f(x + ct)$.

2. Which functions $u(x,y)$ have both $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$ equal to 0 for all x and y?

3. Let $u(x,y)$ satisfy $\frac{\partial^2 u}{\partial x^2} = 0$. Find the form of $u(x,y)$.

4. Show that any function of the form $u(x,y) = f(x) + g(y)$ satisfies $u_{xy} = 0$.

5. Verify that any function of the form $u(x,y) = f(x) + g(y)$ satisfies the wave equation.

6. Modify the steps leading to (C.24.3) to put together a verification of (C.24.4).

7. Let k be a positive constant.

(a) What are the solutions to the equation $\frac{\partial^2 y}{\partial x^2} = k\frac{\partial^2 y}{\partial t^2}$?

(b) What is the speed of the waves?

Chapter 18

Theorems of Green and Stokes

Imagine a fluid or gas moving through space. Its density may vary from point to point. Also its velocity vector may vary from point to point. Figure 18.0.1 illustrates four examples. The diagrams show flows in the plane because they are easier to visualize than flows in space. The plots in Figure 18.0.1 resemble the slope fields of Section 3.6, but instead of short segments it shows vectors, which may be short or long, and point in a definite direction. Two commonly asked questions direction fields are:

- For a fixed region, is the amount of fluid in the region increasing, decreasing, or not changing?
- At a point, does the field create a tendency for the fluid to rotate? If we put a little propeller in the fluid would it turn? If so, in which direction, and how fast or slow?

Questions like these arise in several realistic contexts, including fluid flow, electromagnetism, thermodynamics, and gravity. Techniques for answering them, and more, will be developed in this chapter.

ASSUMPTION: Throughout this chapter, unless specified otherwise, it will be assumed that all partial derivatives of the first and second order exist and are continuous.

18.1 Conservative Vector Fields

In Section 15.3 we defined integrals of the form

$$\int_C (P\,dx + Q\,dy + R\,dz) \tag{18.1.1}$$

where P, Q, and R are scalar functions of x, y, and z and C is a curve in space. Similarly, in the xy-plane, when P and Q are scalar functions of x and y, we have

$$\int_C (P\,dx + Q\,dy).$$

Instead of three scalar fields, P, Q, and R, think of a single vector-valued function $\mathbf{F}(x,y,z) = P(x,y,z)\mathbf{i} + Q(x,y,z)\mathbf{j} + R(x,y,z)\mathbf{k}$, called a **vector field**. In contrast to a scalar field that produces a single number, a vector field produces a vector (typically either two- or three-dimensional).

In Chapter 15 the formal vector $d\mathbf{r} = dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k}$ was introduced as a way to rewrite (18.1.1) as

$$\int_C \mathbf{F}\cdot d\mathbf{r}.$$

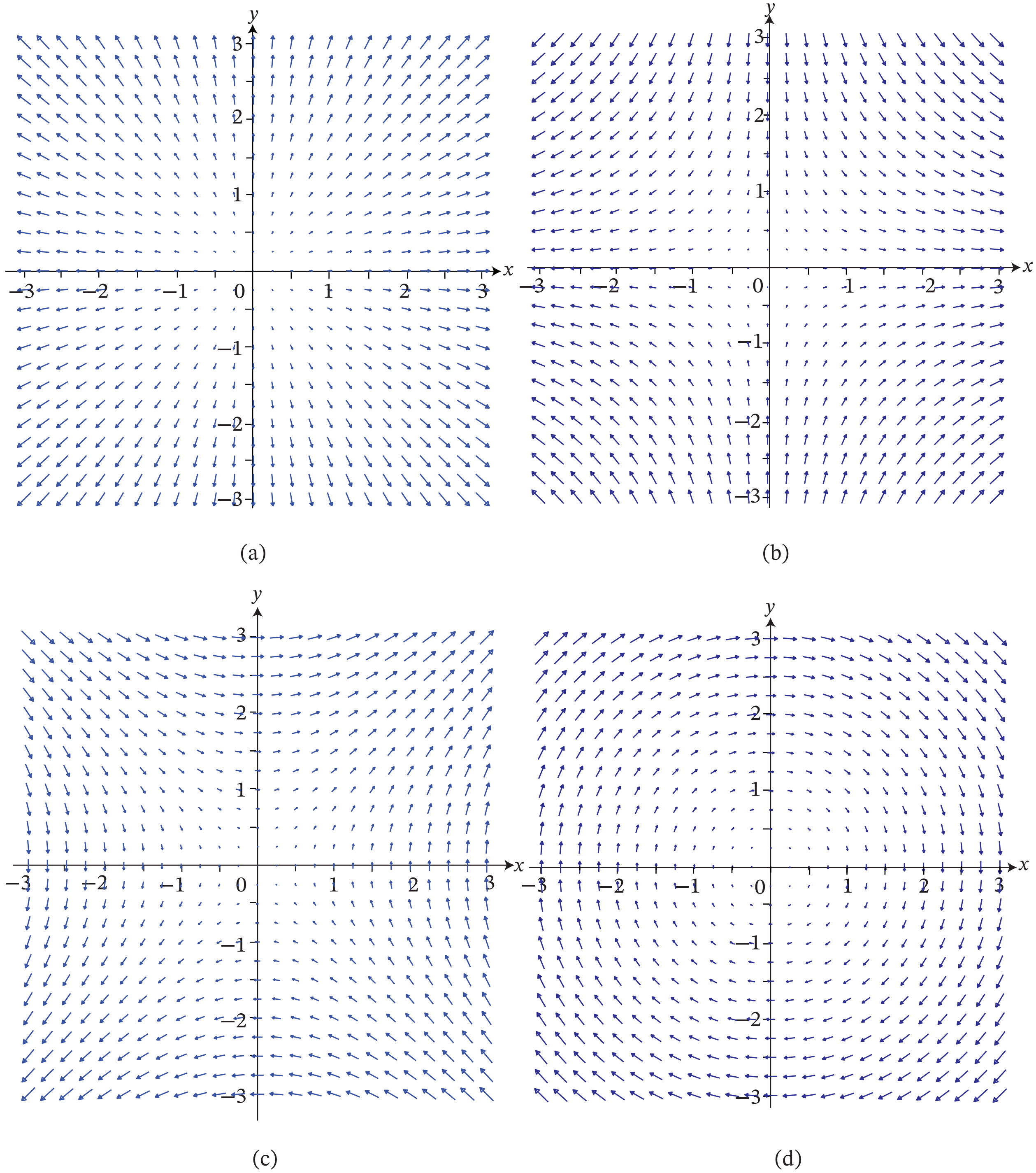


Figure 18.0.1

The vector notation is compact, is the same in the plane and in space, and emphasizes the idea of a vector field. More important, it frees us from referring to any specific coordinate system. The longer notations

$$\int_C (P\,dx + Q\,dy + R\,dz) \qquad \text{and} \qquad \int_C (P(x,y,z)\,dx + Q(x,y,z)\,dy + R(x,y,z)\,dz)$$

are used to prove theorems and to carry out calculations.

Understanding Conservative Vector Fields

In Section 15.3 we made this definition:

Definition: *Conservative Vector Field*

A vector field **F** defined in a plane or spatial region is called **conservative** if

$$\int_{C_1} \mathbf{F}\cdot d\mathbf{r} = \int_{C_2} \mathbf{F}\cdot d\mathbf{r}$$

whenever C_1 and C_2 are any two curves in the region with the same initial and terminal points.

Our goal is to find a more convenient characterization of a conservative field. The first step is Theorem 18.1.1, an equivalent definition of conservative that involves the line integral along a simple closed curve. We review the definitions of simple and closed, then state the theorem, and give its proof.

Recall from Section 15.3 that a closed curve is a curve that begins and ends at the same point, forming a loop. It is simple if it passes through no point more than once other than its start and finish points. A curve that starts at one point and ends at a different point is simple if it never intersects itself. Figure 18.1.1 shows some curves that are simple ((a) and (c)) and some that are not ((b) and (d)).

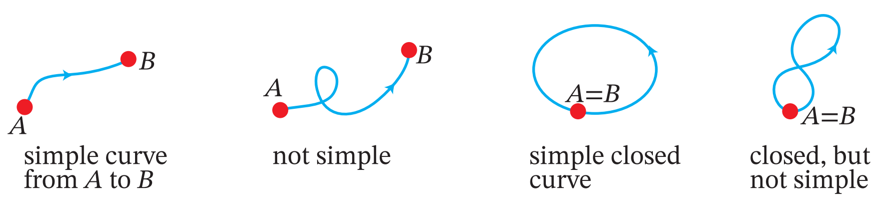


Figure 18.1.1

Theorem 18.1.1: Characterization of a Conservative Vector Field

A vector field **F** *is conservative if and only if* $\oint_C \mathbf{F}\cdot d\mathbf{r} = 0$ *for every simple closed curve in the region where* **F** *is defined.*

Proof of Theorem 18.1.1

Assume that **F** is conservative and let C be a simple closed curve that starts and ends at the point A. Pick a point B on the curve and break C into two curves: C_1 from A to B and C_1^* from B to A, as in Figure 18.1.2(a). Then, $C = C_1 + C_1^*$ is a closed curve.

Let C_2 be the curve C_1^* traversed in the opposite direction, from A to B. Then, since **F** is conservative, and since C_1 and C_2 both start at A and end at B,

$$\begin{aligned}
\oint_C \mathbf{F}\cdot d\mathbf{r} &= \int_{C_1} \mathbf{F}\cdot d\mathbf{r} + \int_{C_1^*} \mathbf{F}\, d\mathbf{r} \\
&= \int_{C_1} \mathbf{F}\cdot d\mathbf{r} - \int_{C_2} \mathbf{F}\cdot d\mathbf{r} \qquad (\text{ note the sign change }) \\
&= 0 \qquad\qquad\qquad\qquad\qquad\quad (\text{ because } \mathbf{F} \text{ is conservative }).
\end{aligned}$$

(a) (b)

Figure 18.1.2

To prove the converse, start with the assumption that $\oint_C \mathbf{F}\cdot d\mathbf{r} = 0$ for every simple closed curve C. The goal is to show that $\int_{C_1} \mathbf{F}\cdot d\mathbf{r} = \int_{C_2} \mathbf{F}\cdot d\mathbf{r}$ for all curves C_1 and C_2 with the same initial and terminal points.

Let curves C_1 and C_2 share the same initial and terminal points, as in Figure 18.1.2(b). Recall that $-C_2$ is the curve C_2 traversed in the opposite direction. Then the composite curve $C = C_1 + (-C_2)$ is a closed curve. But, C is a simple closed curve only if C_1 and C_2 overlap only at their endpoints, A and B. In this case, because C is a simple closed curve and $\mathbf{F}$ is assumed to be conservative,

$$0 = \oint_C \mathbf{F}\cdot d\mathbf{r} = \int_{C_1} \mathbf{F}\cdot d\mathbf{r} + \int_{-C_2} \mathbf{F}\cdot d\mathbf{r} = \int_{C_1} \mathbf{F}\cdot d\mathbf{r} - \int_{C_2} \mathbf{F}\cdot d\mathbf{r}.$$

Consequently,

$$\int_{C_1} \mathbf{F}\cdot d\mathbf{r} = \int_{C_2} \mathbf{F}\cdot d\mathbf{r}.$$

While the converse is still true when the curves intersect elsewhere, the curve $C = C_1 + (-C_2)$ is not simple. A different proof is needed to handle this case; it is left for another course.

Every Gradient Field is Conservative

Knowing whether a vector field is conservative is important in the study of gravity, electromagnetism, and thermodynamics. In the rest of this section we describe ways to determine whether a vector field $\mathbf{F}$ is conservative.

It is impossible to evaluate $\oint_C \mathbf{F}\cdot d\mathbf{r}$ for every simple closed curve and see if it is always 0 because there are infinitely many of them. The first practical test involves gradients and makes use of Theorem 18.1.2.

A vector field that is the gradient of a scalar field is conservative. That is the substance of Theorem 18.1.2. It says that the circulation of a gradient field of a scalar function f along a curve is the difference in values of f at the end points.

The fundamental theorem of calculus asserts that $\int_a^b f'(x)\,dx = f(b) - f(a)$. Theorem 18.1.2 asserts that $\int_C \nabla f \cdot d\mathbf{r} = f(B) - f(A)$, where f is a function of two or three variables and C is a curve from A to B. Because of its resemblance to the fundamental theorem of calculus, it is sometimes called the **Fundamental Theorem of Vector Fields**.

Theorem 18.1.2: Fundamental Theorem of Vector Fields

Let f be a scalar field defined in a region $\mathcal{R}$ in the plane or in space. Then the gradient field $\mathbf{F} = \nabla f$ is conservative. For any points A and B in $\mathcal{R}$ and any curve C in $\mathcal{R}$ from A to B,

$$\int_C \nabla f \cdot d\mathbf{r} = f(B) - f(A).$$

Proof of the Fundamental Theorem of Vector Fields (Theorem 18.1.2)
For simplicity take the plane case. Let C be given by the parameterization $\mathbf{r} = \mathbf{G}(t)$ for t in $[a,b]$. In particular, assume $\mathbf{G}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}$. Then

$$\int_C \nabla f \cdot d\mathbf{r} = \int_C \left(\frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy\right) = \int_a^b \left(\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}\right)dt.$$

If we define H by $H(t) = f(x(t), y(t))$, then the chain rule asserts that

$$\frac{dH}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Thus, by the fundamental theorem of calculus,

$$\int_a^b \left(\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}\right)dt = \int_a^b \frac{dH}{dt}\,dt = H(b) - H(a).$$

Because $H(b) = f(x(b), y(b)) = f(B)$, and $H(a) = f(x(a), y(a)) = f(A)$ we have

$$\int_C \nabla f \cdot d\mathbf{r} = f(B) - f(A), \tag{18.1.2}$$

and the theorem is proved. ●

In differential form Theorem 18.1.2 reads:

- If f is defined on the xy-plane, and C starts at A and ends at B,

$$\int_C \left(\frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy\right) = f(B) - f(A) \tag{18.1.3}$$

- If f is defined in space, then,

$$\int_C \left(\frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy + \frac{\partial f}{\partial z}\,dz\right) = f(B) - f(A). \tag{18.1.4}$$

One vector equation (18.1.2) covers both (18.1.3) and (18.1.4). This illustrates an advantage of vector notation.

Observation 18.1.3: *Evaluating Certain Line Integrals*

Some line integrals are easy to evaluate. By the fundamental theorem of vector fields: $\int_C \nabla f \cdot d\mathbf{r} = f(B) - f(A)$, where A and B are the initial and terminal points of the curve C.

EXAMPLE 1. Let $f(x,y,z) = \dfrac{1}{\sqrt{x^2+y^2+z^2}}$, which is defined everywhere except at the origin. (a) Find the gradient field $\mathbf{F} = \nabla f$ and (b) compute $\displaystyle\int_C \mathbf{F}\cdot d\mathbf{r}$ where C is a smooth curve from $(1,2,2)$ to $(3,4,0)$.

SOLUTION

(a) Straightforward computations show that

$$\frac{\partial f}{\partial x} = \frac{-x}{(x^2+y^2+z^2)^{3/2}}, \qquad \frac{\partial f}{\partial y} = \frac{-y}{(x^2+y^2+z^2)^{3/2}}, \quad \text{and} \quad \frac{\partial f}{\partial z} = \frac{-z}{(x^2+y^2+z^2)^{3/2}}.$$

So

$$\nabla f = \frac{-x\mathbf{i} - y\mathbf{j} - z\mathbf{k}}{(x^2+y^2+z^2)^{3/2}}. \tag{18.1.5}$$

If we let $\mathbf{r}(x,y,z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, $r = |\mathbf{r}|$, and $\widehat{\mathbf{r}} = \mathbf{r}/r$, then (18.1.5) can be written as

$$\mathbf{F} = \nabla f = \frac{-\mathbf{r}}{r^3} = \frac{-\widehat{\mathbf{r}}}{r^2}.$$

(b) For any smooth curve C from $(1,2,2)$ to $(3,4,0)$,

$$\begin{aligned}
\int_C \mathbf{F}\cdot d\mathbf{r} &= \int_C \nabla f \cdot d\mathbf{r} && (\text{ by definition, } \mathbf{F} = \nabla f\text{ })\\
&= f(3,4,0) - f(1,2,2) && (\text{ fundamental theorem of vector fields })\\
&= \frac{1}{\sqrt{3^2+4^2+0^2}} - \frac{1}{\sqrt{1^2+2^2+2^2}} && (\text{ subtract values of } f \text{ at endpoints })\\
&= \frac{1}{5} - \frac{1}{3} = -\frac{2}{15}.
\end{aligned}$$

◆

For a constant k, a vector field, $\mathbf{F}(\mathbf{r}) = k\widehat{\mathbf{r}}/r^2$, is called an **inverse-square central field**. They play an important role in gravity and electromagnetism. In Example 1, $|\nabla f| = |-\mathbf{r}|/r^3 = r/r^3 = 1/r^2$ and $f(x,y,z) = 1/r$. In the study of gravity, $\nabla f(x,y,z) = -\widehat{\mathbf{r}}/r^2$ measures gravitational attraction, and $f(x,y,z) = 1/r$ measures potential.

EXAMPLE 2. Evaluate $\displaystyle\oint_C (y\,dx + x\,dy)$ around a closed curve C taken counterclockwise.

SOLUTION In Section 15.3 it was shown that if the area enclosed by a curve C is a, and if C is swept out counterclockwise then $\oint_C x\,dy = a$ and $\oint_C y\,dx = -a$. Thus,

$$\oint_C (y\,dx + x\,dy) = -a + a = 0.$$

A second solution uses Theorem 18.1.2. The gradient of xy is

$$\nabla(xy) = \frac{\partial(xy)}{\partial x}\mathbf{i} + \frac{\partial(xy)}{\partial y}\mathbf{j} = y\mathbf{i} + x\mathbf{j}.$$

Hence, by Theorem 18.1.2, if the endpoints of C are A and B

$$\oint_C (y\,dx + x\,dy) = \oint_C \nabla(xy)\cdot d\mathbf{r} = xy\Big|_A^B.$$

Because C is a closed curve, $A = B$, and so the integral is 0. ◆

A differential form $P(x,y,z)\ dx + Q(x,y,z)\ dy + R(x,y,z)\ dz$ is called **exact** if there is a scalar function f such that $P(x,y,z) = \partial f/\partial x$, $Q(x,y,z) = \partial f/\partial y$, and $R(x,y,z) = \partial f/\partial z$. Then the expression takes the form

$$\frac{\partial f}{\partial x}\ dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z}\ dz.$$

That is the same thing as saying that the vector field $\mathbf{F} = P(x,y,z)\mathbf{i} + Q(x,y,z)\mathbf{j} + R(x,y,z)\mathbf{k}$ is a gradient field: $\mathbf{F} = \nabla f$.

If F is Conservative Must It Be a Gradient Field?

The proof of the next theorem, Theorem 18.1.4l, is similar to the proof of the second part of the fundamental theorem of calculus, Theorem 6.4.6. Recall that FTC II states that every continuous function has an antiderivative.

SUGGESTION: It would be helpful to review the proof of FTC II (page 332) before reading the proof of Theorem 18.1.4.

If **F** is conservative, is it necessarily the gradient of some scalar function? The answer is yes. That is the substance of the next theorem. First we introduce some terminology about regions.

Recall that a region $\mathcal{R}$ in the plane is open if for each point P in $\mathcal{R}$ there is a disk with center at P that lies entirely in $\mathcal{R}$.

An open region in space is defined similarly, with "disk" replaced by "ball."

Before proceeding, we need one more technical definition: A region $\mathcal{R}$ is **arcwise-connected** if every two points in it can be joined by a curve that lies completely in $\mathcal{R}$. For practical purposes, an arcwise-connected region has only one piece.

Theorem 18.1.4

Let **F** *be a conservative vector field defined in an arcwise-connected open region* $\mathcal{R}$ *in the plane (or in space). If* **F** *is also continuous on* $\mathcal{R}$*, then there is a scalar function* f *defined there such that* $\mathbf{F} = \nabla f$.

Proof of Theorem 18.1.4

Suppose $\mathbf{F} = P(x,y)\mathbf{i} + Q(x,y)\mathbf{j}$. (If **F** is defined in space then **F** has three components, but the proof is similar.) Define the scalar function f as follows. Let (a,b) and (x,y) be points in $\mathcal{R}$. Select a curve C in $\mathcal{R}$ that starts at (a,b) and ends at (x,y). Define $f(x,y)$ to be $\int_C \mathbf{F}\cdot d\mathbf{r}$. Since **F** is conservative, $f(x,y)$ depends only on the point (x,y) and not on the choice of C. (See Figure 18.1.3.)

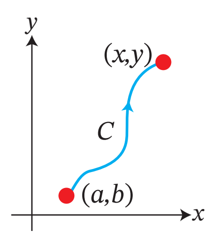


Figure 18.1.3

All that remains is to show that $\nabla f = \mathbf{F}$; that is, $\partial f/\partial x = P$ and $\partial f/\partial y = Q$. We will go through the details for the first case, $\partial f/\partial x = P$. The other is similar (see Exercise 21).

Let (x_0,y_0) be a point in $\mathcal{R}$ and form the difference quotient whose limit is $\partial f/\partial x(x_0,y_0)$, namely,

$$\frac{f(x_0+h,y_0) - f(x_0,y_0)}{h},$$

for h small enough so that (x_0+h,y_0) is also in the region.

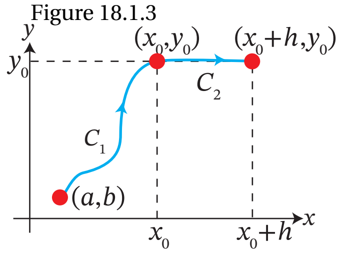


Figure 18.1.4

Let C_1 be a curve from (a,b) to (x_0,y_0) and let C_2 be the straight path from (x_0,y_0) to (x_0+h,y_0). (See Figure 18.1.4.) Let C be the curve from $(0,0)$ to (x_0+h,y_0) formed by taking C_1 first and continuing on C_2. Then

$$f(x_0,y_0) = \int_{C_1} \mathbf{F}\cdot d\mathbf{r},$$

and

$$f(x_0+h,y_0) = \int_C \mathbf{F}\cdot d\mathbf{r} = \int_{C_1} \mathbf{F}\cdot d\mathbf{r} + \int_{C_2} \mathbf{F}\cdot d\mathbf{r}.$$

Thus

$$\frac{f(x_0+h,y_0)-f(x_0,y_0)}{h}=\frac{\int\limits_{C_2}\mathbf{F}\cdot d\mathbf{r}}{h}=\frac{\int\limits_{C_2}(P(x,y)\,dx+Q(x,y)\,dy)}{h}.$$

On C_2, y is constant, $y=y_0$, so $dy=0$. Thus $\int_{C_2}Q(x,y)\,dy=0$. Also,

$$\int\limits_{C_2}P(x,y)\,dx=\int\limits_{x_0}^{x_0+h}P(x,y_0)\,dx.$$

By the mean value theorem for definite integrals (see Section 6.3), and the continuity of P on $\mathscr{R}$, there is a number x^* between x_0 and x_0+h such that

$$\int\limits_{x_0}^{x_0+h}P(x,y_0)\,dx=P(x^*,y_0)h.$$

Hence

$$\begin{aligned}\frac{\partial f}{\partial x}(x_0,y_0)&=\lim_{h\to0}\frac{f(x_0+h,y_0)-f(x_0,y_0)}{h}\\&=\lim_{h\to0}\frac{1}{h}\int\limits_{x_0}^{x_0+h}P(x,y_0)\,dx\\&=\lim_{h\to0}P(x^*,y_0)=P(x_0,y_0).\end{aligned}$$

Consequently,

$$\frac{\partial f}{\partial x}(x_0,y_0)=P(x_0,y_0),$$

as was to be shown. Similarly (see Exercise 21), we can show that

$$\frac{\partial f}{\partial y}(x_0,y_0)=Q(x_0,y_0).$$

Observation 18.1.5: *Different Ways to Characterize a Conservative Field*

For a vector field $\mathbf{F}$ defined throughout some region in the plane (or space) the following four statements are equivalent

(a) the vector field $\mathbf{F}$ is conservative,

(b) $\mathbf{F}$ can be written as ∇f for some function f,

(c) $\oint\limits_C\mathbf{F}\cdot d\mathbf{r}=0$ for all closed curves C, and

(d) $\int\limits_C\mathbf{F}\cdot d\mathbf{r}$ depends only on the endpoints of the curve C.

We used property (d) as the definition.

Almost A Test For Being Conservative

Now we give a simple way to tell that a vector field $\mathbf{F}$ is not conservative. This test is simpler to apply than attempting to find a line integral $\int_C\mathbf{F}\cdot d\mathbf{r}$ that is not 0.

The test depends on the equality of the two mixed second-order partial derivatives:

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right). \tag{18.1.6}$$

If $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ is a gradient there is a scalar function f such that

$$\frac{\partial f}{\partial x} = P, \qquad \frac{\partial f}{\partial y} = Q, \qquad \frac{\partial f}{\partial z} = R.$$

Then, using the equality of mixed partial derivatives, (18.1.6),

$$\frac{\partial P}{\partial y} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial Q}{\partial x}.$$

So,

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}.$$

Similarly we find

$$\frac{\partial Q}{\partial z} = \frac{\partial R}{\partial y} \qquad \text{and} \qquad \frac{\partial P}{\partial z} = \frac{\partial R}{\partial x}.$$

These findings prove the following theorem.

Theorem 18.1.6: Partial Test for a Conservative Field

If the vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ *is conservative, then*

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 0, \qquad \frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} = 0, \qquad \frac{\partial R}{\partial x} - \frac{\partial P}{\partial z} = 0. \tag{18.1.7}$$

IMPORTANT FACT: *If at least one of the equations in* (18.1.7) *does not hold, then* $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ *is not conservative and* $P\,dx + Q\,dy + R\,dz$ *is not exact.*

EXAMPLE 3. Show that $\cos(y)\,dx + \sin(xy)\,dy + \ln(1+x)\,dz$ is not exact.

SOLUTION Checking (18.1.7) we compute

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = \frac{\partial(\sin(xy))}{\partial x} - \frac{\partial(\cos(y))}{\partial y} = y\cos(xy) + \sin(y),$$

which is not 0. There is no need to check the remaining equations in (18.1.7) before concluding that the expression $\cos(y)\,dx + \sin(xy)\,dy + \ln(1+x)\,dz$ is not exact. Thus the vector field $\cos(y)\mathbf{i} + \sin(xy)\mathbf{j} + \ln(1+x)\mathbf{k}$ is not a gradient field and hence not conservative. ◆

To restate (18.1.7) as a vector equation introduce the 3×3 determinant

$$\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{pmatrix} \tag{18.1.8}$$

Expanding this determinant as though its entries were numbers, we get

$$\mathbf{i}\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) - \mathbf{j}\left(\frac{\partial R}{\partial x} - \frac{\partial P}{\partial z}\right) + \mathbf{k}\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right). \tag{18.1.9}$$

If the scalar equations in (18.1.7) hold, then (18.1.9) is the $\mathbf{0}$-vector. This vector consisting of this combination of partial derivatives is given a name.

Definition: *Curl of a Vector Field*

The **curl** of the vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ is the vector field given by (18.1.8) or (18.1.9). It is denoted $\mathbf{curl\,F}$.

$$\mathbf{curl\,F}(x,y,z) = \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ P & Q & R \end{pmatrix} = \mathbf{i}\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) - \mathbf{j}\left(\frac{\partial R}{\partial x} - \frac{\partial P}{\partial z}\right) + \mathbf{k}\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right).$$

The determinant (18.1.8) is like the one for the cross product of two vectors. For this reason, it is also denoted $\nabla \times \mathbf{F}$ (read as "del cross F"). That is easier to write than (18.1.8) or (18.1.9), which refer to components.

The vector field $\mathbf{curl\,F}$ is called curl because if $\mathbf{F}$ describes a fluid flow, then $\mathbf{curl\,F}$ describes the tendency of the fluid to rotate and form whirlpools — that is, to curl. (The history behind the origin of the name "curl" can be found in the historical note in Section 18.4 (see page 1068).)

The definition of curl in space also applies to a vector field in the plane, $\mathbf{F} = P(x,y)\mathbf{i} + Q(x,y)\mathbf{j}$. Writing $\mathbf{F}$ as $P(x,y)\mathbf{i} + Q(x,y)\mathbf{j} + 0\mathbf{k}$ and observing that $\partial Q/\partial z = 0$ and $\partial P/\partial z = 0$, we find that

$$\nabla \times \mathbf{F} = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k}.$$

EXAMPLE 4. Compute the curl of $\mathbf{F} = xyz\mathbf{i} + x^2\mathbf{j} - xy\mathbf{k}$.

SOLUTION The curl of $\mathbf{F}$ is given by

$$\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ xyz & x^2 & -xy, \end{pmatrix}$$

which is evaluated in the following steps:

$$\begin{aligned} &\left(\frac{\partial}{\partial y}(-xy) - \frac{\partial}{\partial z}(x^2)\right)\mathbf{i} - \left(\frac{\partial}{\partial x}(-xy) - \frac{\partial}{\partial z}(xyz)\right)\mathbf{j} + \left(\frac{\partial}{\partial x}(x^2) - \frac{\partial}{\partial y}(xyz)\right)\mathbf{k} \\ &= (-x-0)\mathbf{i} - (-y-xy)\mathbf{j} + (2x - xz)\mathbf{k} \\ &= -x\mathbf{i} + (y+xy)\mathbf{j} + (2x-xz)\mathbf{k}. \end{aligned}$$

◆

From (18.1.7), for vector fields in space or in the xy-plane we have a new characterization of conservative vector fields.

Theorem 18.1.7:

If $\mathbf{F}$ is a conservative vector field, then $\nabla \times \mathbf{F} = \mathbf{0}$.

The Converse of Theorem 18.1.7 Is False

The converse of Theorem 18.1.7 is not true. There are vector fields $\mathbf{F}$ whose curls are $\mathbf{0}$ that are not conservative. Example 5 provides one such $\mathbf{F}$ in the xy-plane. Its curl is $\mathbf{0}$ but it is not conservative.

EXAMPLE 5. Let $\mathbf{F} = \dfrac{-y\mathbf{i}}{x^2+y^2} + \dfrac{x\mathbf{j}}{x^2+y^2}$. Show that (a) $\nabla \times \mathbf{F} = \mathbf{0}$, but (b) $\mathbf{F}$ is not conservative.

SOLUTION

(a) We compute

$$\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{-y}{x^2+y^2}i & \frac{x}{x^2+y^2}i & 0 \end{pmatrix},$$

which equals

$$\left(\frac{\partial(0)}{\partial y} - \frac{\partial}{\partial z}\left(\frac{x}{x^2+y^2}\right)\right)\mathbf{i} - \left(\frac{\partial(0)}{\partial x} - \frac{\partial}{\partial z}\left(\frac{-y}{x^2+y^2}\right)\right)\mathbf{j} + \left(\frac{\partial}{\partial x}\left(\frac{x}{x^2+y^2}\right) - \frac{\partial}{\partial y}\left(\frac{-y}{x^2+y^2}\right)\right)\mathbf{k}.$$

The **i** and **j** components are clearly 0, and a computation shows that the **k** component is

$$\begin{aligned}\frac{\partial}{\partial x}\left(\frac{x}{x^2+y^2}\right) - \frac{\partial}{\partial y}\left(\frac{-y}{x^2+y^2}\right) &= \frac{(x^2+y^2)(1) - x(2x)}{(x^2+y^2)^2} - \frac{(x^2+y^2)(-1) - (-y)(2y)}{(x^2+y^2)^2} \\ &= \frac{y^2-x^2}{(x^2+y^2)^2} - \frac{y^2-x^2}{(x^2+y^2)^2} = 0.\end{aligned}$$

Thus the curl of **F** is **0**.

(b) To show that **F** is not conservative, it suffices to exhibit a closed curve C such that $\oint_C \mathbf{F}\cdot d\mathbf{r}$ is not 0. One such C is the unit circle with center at the origin oriented counterclockwise. A parameterization of C is

$$x = \cos(\theta), \qquad y = \sin(\theta), \qquad 0 \le \theta \le 2\pi.$$

On it $x^2 + y^2 = 1$. Figure 18.1.5 shows a few values of **F** at points on C.
It appears that $\oint_C \mathbf{F}\cdot d\mathbf{r}$, which measures circulation, is positive, not 0. Its exact value is $\oint_C \mathbf{F}\cdot d\mathbf{r}$:

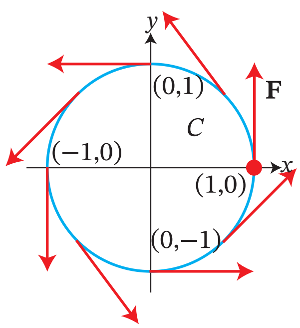


Figure 18.1.5

$$\begin{aligned}\oint_C \mathbf{F}\cdot d\mathbf{r} &= \oint_C \left(\frac{-y\,dx}{x^2+y^2} + \frac{x\,dy}{x^2+y^2}\right) \\ &= \int_0^{2\pi} (-\sin(\theta)\ d(\cos(\theta)) + \cos(\theta)\ d(\sin(\theta))) \\ &= \int_0^{2\pi} (\sin^2(\theta) + \cos^2(\theta))\ d\theta = \int_0^{2\pi} d\theta = 2\pi.\end{aligned}$$

This establishes (b): **F** is not conservative. ◆

FINAL POINT: The curl of **F** being **0** is not enough to assure us that a vector field **F** is conservative. An extra condition must be satisfied by **F**. This condition concerns the domain of **F**. This extra assumption will be developed for planar fields in Section 18.2 and for spatial fields in Section 18.6. Only then will we have a complete test for determining whether a vector field is conservative.

Summary

We showed that a vector field being conservative is equivalent to its being the gradient of a scalar field. Then we defined the curl of a vector field. If the field is denoted **F**, the curl of **F** is a new vector field denoted **curl F** or $\nabla \times \mathbf{F}$. If **F** is conservative, then $\nabla \times \mathbf{F}$ is **0**. However, if the curl of **F** is **0**, it does not follow that **F** is conservative. An extra assumption on the domain of **F** must be added. That assumption will be described later in this chapter.

EXERCISES for Section 18.1

In Exercises 1 to 4 answer true or false, then explain your answer.

1. If $\mathbf{F}$ is conservative, then $\nabla \times \mathbf{F} = \mathbf{0}$.

2. If $\nabla \times \mathbf{F} = \mathbf{0}$, then $\mathbf{F}$ is conservative.

3. If $\mathbf{F}$ is a gradient field, then $\nabla \times \mathbf{F} = \mathbf{0}$.

4. If $\nabla \times \mathbf{F} = \mathbf{0}$, then $\mathbf{F}$ is a gradient field.

---

5. (a) Use the fact that a gradient, ∇f, is conservative to show that its curl is $\mathbf{0}$.
(b) Compute $\nabla \times \nabla f$ in terms of components to show that the curl of a gradient is $\mathbf{0}$.

6. Using information in this section, describe various ways of showing a vector field $\mathbf{F}$ is not conservative.

7. Using information in this section, describe various ways of showing a vector field $\mathbf{F}$ is conservative.

8. Decide if the sets are open, closed, neither open nor closed, or both open and closed.

(a) unit disk with its boundary
(b) unit disk without any of its boundary points
(c) the x-axis
(d) the xy-plane
(e) the xy-plane with the x-axis removed
(f) a square with its edges and corners
(g) a square with its edges but with its corners removed
(h) a square with none of its edges

9. In Example 1 we computed a line integral by using the fact that the vector field $\mathbf{F} = \dfrac{-x\mathbf{i} - y\mathbf{j} - z\mathbf{k}}{(x^2+y^2+z^2)^{3/2}}$ is a gradient field. Choose a specific curve C from $(1,2,2)$ to $(3,4,0)$ and compute $\int_C \mathbf{F} \cdot d\mathbf{r}$ by parameterizing C.

10. Let $\mathbf{F} = y\cos(x)\mathbf{i} + (\sin(x) + 2y)\mathbf{j}$.
(a) Show that $\mathbf{curl}\,\mathbf{F}$ is $\mathbf{0}$ and $\mathbf{F}$ is defined in an arcwise-connected region of the plane.
(b) Construct a function f whose gradient is $\mathbf{F}$.

11. Let $f(x,y,z) = e^{3x}\ln(z+y^2)$. Compute $\displaystyle\int_C \nabla f \cdot d\mathbf{r}$ where C is the straight path from $(1,1,1)$ to $(4,3,1)$.

12. We obtained the first equation in (18.1.7). Derive the other two.

13. Find the curl of $\mathbf{F}(x,y,z) = e^{x^2} yz\mathbf{i} + x^3\cos^2(3y)\mathbf{j} + (1+x^6)\mathbf{k}$.

14. Find the curl of $\mathbf{F}(x,y) = \tan^2(3x)\mathbf{i} + e^{3x}\ln(1+x^2)\mathbf{j}$.

15. (a) Using theorems of this section, explain why the curl of a gradient is $\mathbf{0}$, that is, $\mathbf{curl}(\nabla f) = \nabla \times \nabla f = \mathbf{0}$ for a scalar function $f(x,y,z)$.
(b) By a computation using components, show that $\mathbf{curl}(\nabla f) = \mathbf{0}$ for the scalar function $f(x,y,z)$.

16. Let $f(x,y) = \cos(x+y)$. Evaluate $\int_C \nabla f \cdot d\mathbf{r}$, where C is the part of the parabola $y = x^2$ from $(0,0)$ to $(2,4)$.

17. In Example 5 we computed $\displaystyle\oint_C \mathbf{F} \cdot d\mathbf{r}$, where $\mathbf{F} = \dfrac{-y\mathbf{i} + x\mathbf{j}}{x^2+y^2}$ and C is the unit circle with center at the origin. Compute the integral when C is the circle of radius 5 with center at the origin.

18. Let $\mathbf{F}$ and $\mathbf{G}$ be conservative fields defined throughout the xy-plane. Is $\mathbf{F} + \mathbf{G}$ necessarily conservative?

19. Show that $\mathbf{curl}(f\mathbf{F}) = \nabla f \times \mathbf{F} + f\,\mathbf{curl}\,\mathbf{F}$.

20. If $\mathbf{F}$ and $\mathbf{G}$ are conservative, is $\mathbf{F} \times \mathbf{G}$?

21. The partial proof of Theorem 18.1.4 given in this section showed that $\dfrac{\partial f}{\partial x} = P$. To complete the proof of this theorem, prove that $\dfrac{\partial f}{\partial y} = Q$, thereby completing the proof that $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} = \nabla f$.

22. Assume that $\mathbf{F}(x, y)$ is conservative. Let C_1 be the straight path from $(0,0,0)$ to $(1,0,0)$ and C_2 the straight path from $(1,0,0)$ to $(1,1,1)$. If $\int_{C_1} \mathbf{F}\cdot d\mathbf{r} = 3$ and $\int_{C_2} \mathbf{F}\cdot d\mathbf{r} = 4$, what can be said about $\int_C \mathbf{F}\cdot d\mathbf{r}$ where C is the straight path from $(0,0,0)$ to $(1,1,1)$?

23. Let $\mathbf{F}(x, y)$ be the field $\mathbf{F}(x, y) = g\left(\sqrt{x^2+y^2}\right)\dfrac{x\mathbf{i}+y\mathbf{j}}{\sqrt{x^2+y^2}}$ where g is a scalar function. If we denote $x\mathbf{i}+y\mathbf{j}$ as $\mathbf{r}$, then $\mathbf{F}(x, y) = g(r)\widehat{\mathbf{r}}$, where $r = |\mathbf{r}|$ and $\widehat{\mathbf{r}} = |\mathbf{r}|/r$. Show that $\displaystyle\oint_{ABCDA} \mathbf{F}\cdot d\mathbf{r} = 0$ for any path $ABCDA$ of the form shown in Figure 18.1.6. The path consists of two circular arcs and parts of two rays from the origin.

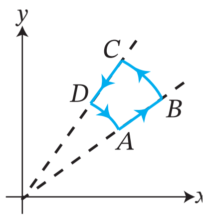


Figure 18.1.6

24. In view of the previous exercise, we may expect $\mathbf{F}(x, y) = g(\sqrt{x^2+y^2})\dfrac{x\mathbf{i}+y\mathbf{j}}{\sqrt{x^2+y^2}}$ to be conservative. Show that it is by showing that $\mathbf{F}$ is the gradient of $G(x, y) = H(\sqrt{x^2+y^2})$, where H is an antiderivative of g, that is, $H' = g$.

25. The domain of a vector field $\mathbf{F}$ is all of the xy-plane. Assume that there are two points A and B such that $\displaystyle\int_C \mathbf{F}\cdot d\mathbf{r}$ is the same for all curves C from A to B. Deduce that $\mathbf{F}$ is conservative.

26. A gas at temperature T_0 and pressure P_0 is brought to temperature $T_1 > T_0$ and pressure $P_1 > P_0$. The work done in this process is given by the line integral in the TP-plane $\displaystyle\int_C \left(\frac{RT}{P}\,dP - R\,dT\right)$, where R is a constant and C is the curve that records the various combinations of T and P during the process.

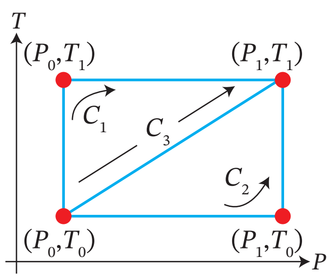


Figure 18.1.7

Evaluate the integral over the paths shown in Figure 18.1.7.

(a) The curve C_1 keeps pressure constant at P_0 while the temperature is raised from T_0 to T_1 and then the temperature is kept constant at T_1 while the pressure is raised from P_0 to P_1.

(b) The curve C_2 keeps the temperature constant at T_0 while the pressure is raised from P_0 to P_1 and then the temperature is raised from T_0 to T_1 while the pressure is kept constant at P_1.

(c) For C_3 the pressure and temperature are raised so that the path from (P_0, T_0) to (P_1, T_1) is straight.

Observation 18.1.8:

Because the integrals are path dependent, the differential expression $\dfrac{RT}{P}dP - R\,dT$ defines a thermodynamic quantity that depends on the process, not only on the state. That is, the vector field $\dfrac{RT}{P}\mathbf{i} - R\mathbf{j}$ is not conservative.

27. Assume that $\mathbf{F}(x, y)$ is defined throughout the xy-plane and that $\displaystyle\oint_C \mathbf{F}(x, y)\cdot d\mathbf{r} = 0$ for every closed curve that can fit inside some disk of diameter 0.01. Show that $\mathbf{F}$ is conservative.

28. We used the mean value theorem for definite integrals to prove that $\displaystyle\lim_{h\to 0}\frac{1}{h}\int_{x_0}^{x_0+h} P(x, y_0)\,dx$ equals $P(x_0, y_0)$. Find a proof of this result that uses a part of the fundamental theorem of calculus.

18.2 Green's Theorem and Circulation

In this section we discuss a connection between an integral of a vector field over a closed curve C in the plane to an integral of a related scalar function over the region $\mathscr{R}$ in the plane whose boundary is C. We will also see what this means in terms of the circulation of a vector field. In the next section we explore a similar relationship related to flux across a curve.

Statement of Green's Theorem

We begin by stating Green's theorem. Then we will see several of its applications. The proof is at the end of the next section.

Theorem 18.2.1: Green's Theorem

Let C be a simple, closed counterclockwise curve in the xy-plane, bounding a region $\mathscr{R}$. Let P and Q be scalar functions defined at least on an open set containing $\mathscr{R}$. Assume P and Q have continuous first-order partial derivatives. Then

$$\oint_C (P\,dx + Q\,dy) = \int_{\mathscr{R}} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\,dA.$$

Since P and Q are independent of each other, Green's theorem consists of two theorems:

$$\oint_C P\,dx = -\int_{\mathscr{R}} \frac{\partial P}{\partial y}\,dA \qquad \text{and} \qquad \oint_C Q\,dy = \int_{\mathscr{R}} \frac{\partial Q}{\partial x}\,dA. \tag{18.2.1}$$

EXAMPLE 1. In Section 15.3 we showed that if the counterclockwise curve C bounds a region $\mathscr{R}$, then $\oint_C y\,dx$ is the negative of the area of $\mathscr{R}$. Obtain this result with the aid of Green's theorem.

SOLUTION Let $P(x,y) = y$, and $Q(x,y) = 0$. Then Green's theorem says that

$$\begin{aligned}
\oint_C y\,dx &= \oint_C (y\,dx + 0\,dy) \\
&= \int_{\mathscr{R}} \left(\frac{\partial 0}{\partial x} - \frac{\partial y}{\partial y}\right)\,dA \qquad (\text{Green's theorem, with } P = 0 \text{ and } Q = y) \\
&= -\int_{\mathscr{R}} 1\,dA \\
&= -(\text{Area of } \mathscr{R}).
\end{aligned}$$

◆

Green's Theorem and Circulation

What does Green's theorem say about a vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$?

First of all, $\oint_C (P\,dx + Q\,dy)$ now becomes simply $\oint_C \mathbf{F}\cdot d\mathbf{r}$, which is recognized as the circulation of $\mathbf{F}$ around the closed curve C.

Next, notice that the right-hand side of Green's theorem looks a bit like the curl of a vector field in the plane. To be specific, we compute the curl of **F**:

$$\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ P(x,y) & Q(x,y) & 0 \end{pmatrix} = 0\mathbf{i} - 0\mathbf{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k}.$$

Thus the curl of **F** equals the vector function

$$\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k}. \tag{18.2.2}$$

To obtain the (scalar) integrand on the right-hand side of (18.2.2), we "dot (18.2.2) with **k**,"

$$\left(\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k}\right)\cdot\mathbf{k} = \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}.$$

We can now express Green's theorem in a vector form. In particular, circulation around a closed curve can be expressed in terms of a double integral of the component of the curl perpendicular to the surface over a region.

Theorem 18.2.2: Green's Theorem in Vector Notation

If the counterclockwise closed curve C bounds the region $\mathscr{R}$ in the xy-plane, then

$$\oint_C \mathbf{F}\cdot d\mathbf{r} = \int_{\mathscr{R}} (\nabla\times\mathbf{F})\cdot\mathbf{k}\, dA.$$

Recall that if **F** describes the flow of a fluid in the xy-plane, then $\oint_C \mathbf{F}\cdot d\mathbf{r}$ represents its circulation, or tendency to form whirlpools. This theorem tells us that the magnitude of the curl of **F** represents the tendency of the fluid to rotate. If the curl of **F** is **0** everywhere, then **F** is called **irrotational** — there is no rotational tendency.

This form of Green's theorem provides an easy way to show that a vector field **F** is conservative. It uses the idea of a simply connected region. Informally "a simply connected region in the xy-plane comes in one piece and has no holes." More precisely, an arcwise-connected region $\mathscr{R}$ in the plane or in space is **simply connected** if each closed curve in $\mathscr{R}$ can be shrunk gradually to a point while remaining in $\mathscr{R}$. Figure 18.2.1 shows two regions in the plane. The region in Figure 18.2.1(a) is simply connected, while the region in Figure 18.2.1(b) the right is not simply connected.

Figure 18.2.1

For instance, the xy-plane is simply connected. So is the xy-plane without its positive x-axis. However, the xy-plane without the origin is not simply connected, because a circular path around the origin cannot be shrunk to a point while staying within the region.

The corresponding cases in three-dimensional space are quite different.

If, as in Figure 18.2.2(a), the origin is removed from xyz-space, what is left is simply connected. However, if we remove the z-axis, as in Figure 18.2.2(b), what is left is not simply connected.

Now we can state an easy way to tell whether a vector field is conservative.

Space without origin.
Simply connected
(a)

Space without z axis.
Not simply connected
(b)

Figure 18.2.2

Theorem 18.2.3: Test for a Conservative Vector Field in the Plane

If a vector field $\mathbf{F}$ *is defined in a simply connected region in the* xy*-plane and* $\nabla\times\mathbf{F}=\mathbf{0}$ *throughout that region, then* $\mathbf{F}$ *is conservative.*

Proof of Theorem 18.2.3

Let C be any simple closed curve in the region and $\mathscr{R}$ the region it bounds. We wish to prove that the circulation of $\mathbf{F}$ around C is $\mathbf{0}$. We have

$$\oint_C \mathbf{F}\cdot d\mathbf{r} = \int_{\mathscr{R}} (\mathbf{curl}\,\mathbf{F})\cdot\mathbf{k}\; dA.$$

Since $\mathbf{curl}\,\mathbf{F}$ is $\mathbf{0}$ throughout $\mathscr{R}$, it follows that $\oint_C \mathbf{F}\cdot d\mathbf{r}=0$. ●

Observation 18.2.4: *Final Remarks on Example 5 in Section 18.1*

In Example 5 in Section 18.1 there is a vector field whose curl is $\mathbf{0}$ but it is not conservative. In view of the theorem just proved, its domain must not be simply connected. Indeed, the domain of the vector field in that example is the xy-plane without the origin.

EXAMPLE 2. Let $\mathbf{F}(x,y,z)=e^x y\mathbf{i}+(e^x+2y)\mathbf{j}$. Show that $\mathbf{F}$ is conservative. Exhibit a scalar function f whose gradient is $\mathbf{F}$.

SOLUTION Since $\mathbf{F}$ is defined throughout the xy-plane, a simply connected region, a straightforward calculation shows that $\nabla\times\mathbf{F}=\mathbf{0}$. Theorem 18.2.3 tells us that $\mathbf{F}$ is conservative.

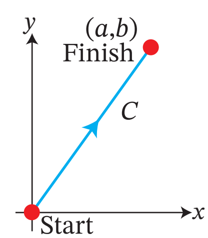


Figure 18.2.3

By Section 18.1, we know that there is a scalar function f such that $\nabla f=\mathbf{F}$. There are several ways to find f. We show one of these methods. Additional approaches are pursued in Exercises 9 to 11.

The approach chosen here is suggested by the construction in the proof of Theorem 18.1.4. For a point (a,b), define $f(a,b)$ to equal $\int_C \mathbf{F}\cdot d\mathbf{r}$, where C is any curve from $(0,0)$ to (a,b). We call the arbitrary point (a,b) instead of (x,y) in order to have x and y available to describe an integrand. Any curve with the prescribed endpoints will do. For simplicity, choose C to be the curve that goes from $(0,0)$ to (a,b) in a straight line. (See Figure 18.2.3.)

When a is not zero, we can use x as a parameter and write this segment as: $x=t$, $y=(b/a)t$ for $0\le t\le a$. (If $a=0$, we would use y as a parameter.) Then, defining $f(a,b)$ comes down to evaluating a line integral, using the parameterization of C, applying integration by parts, and using the fundamental theorem of calculus, as follows:

$$\begin{aligned}
f(a,b) &= \int\limits_C \left(e^x y\, dx + (e^x + 2y)\, dy\right) \\
&= \int\limits_0^a \left(e^t \frac{bt}{a}\, dt + \left(e^t + \frac{2bt}{a}\right) \frac{b}{a}\, dt\right) \\
&= \frac{b}{a} \int\limits_0^a \left(te^t + e^t + \frac{2bt}{a}\right) dt \\
&= \frac{b}{a} \left((t-1)e^t + e^t + \frac{bt^2}{a}\right)\Bigg|_0^a \\
&= \frac{a}{a} \left(te^t + \frac{bt^2}{a}\right)\Bigg|_0^a \\
&= be^a + b^2.
\end{aligned}$$

Since $f(a,b) = be^a + b^2$, we see that $f(x,y) = ye^x + y^2$ is the desired function.

One could check this by showing that the gradient of f is indeed $ye^x\mathbf{i} + (e^x + 2y)\mathbf{j}$. Other suitable potential functions f are $ye^x + y^2 + k$ for any constant k. ♦

The next example uses the cancellation principle, which is based on the fact that the sum of two line integrals in opposite directions on a curve is zero. This idea is used here to develop the two-curve version of Green's theorem; it will be used several more times before the end of this chapter.

EXAMPLE 3. Figure 18.2.4(a) shows two closed counterclockwise curves C_1 and C_2 that enclose a ring-shaped region $\mathscr{R}$ in which $\nabla \times \mathbf{F}$ is $\mathbf{0}$. Show that the circulation of $\mathbf{F}$ over C_1 equals the circulation of $\mathbf{F}$ over C_2.

(a) (b)

Figure 18.2.4

SOLUTION Cut $\mathscr{R}$ into two regions, each bounded by a simple curve, to which we can apply Theorem 18.2.3. Assume C_3 is the boundary of one of the regions and C_4 bounds the other, with the usual counterclockwise orientation. On the cuts, C_3 and C_4 go in opposite directions. On the outer curve C_3 and C_4 have the same orientation as C_1. On the inner curve they have the opposite orientation of C_2. (See Figure 18.2.4(b).) Thus

$$\int\limits_{C_3} \mathbf{F} \cdot dr + \int\limits_{C_4} \mathbf{F} \cdot dr = \int\limits_{C_1} \mathbf{F} \cdot dr - \int\limits_{C_2} \mathbf{F} \cdot dr. \tag{18.2.3}$$

By Theorem 18.2.3 the two integrals on the left-hand side of (18.2.3) are 0. Thus

$$\int\limits_{C_1} \mathbf{F} \cdot d\mathbf{r} = \int\limits_{C_2} \mathbf{F} \cdot d\mathbf{r} \tag{18.2.4}$$

♦

The ideas introduce in Example 3 are the basis for the following "two-curve" variation of Green's theorem.

Corollary 18.2.5: Changing the Curve when curl $\mathbf{F} = 0$

Assume two nonoverlapping curves C_1 and C_2 lie in a region where **curl F** *is* **0** *and form the border of a ring. Then, if C_1 and C_2 both have the same orientation,*

$$\oint_{C_1} \mathbf{F}\cdot d\mathbf{r} = \oint_{C_2} \mathbf{F}\cdot d\mathbf{r}.$$

Observation 18.2.6: *What Corollary 18.2.5 Really Says*

Corollary 18.2.5 says "moving a closed curve within a region of zero-curl does not change the circulation".

The next Example illustrates this point.

EXAMPLE 4. Let $\mathbf{F} = \dfrac{-y\mathbf{i} + x\mathbf{j}}{x^2+y^2}$ and C be the closed counterclockwise curve bounding the square whose vertices are $(-2,-2)$, $(2,-2)$, $(2,2)$, and $(-2,2)$. Evaluate the circulation of $\mathbf{F}$ around C as easily as possible.

SOLUTION This vector field appeared in Example 5 of Section 18.1. Since its curl is $\mathbf{0}$ at all points except the origin, where $\mathbf{F}$ is not defined, we may use the two-curve version of Green's theorem. Thus $\oint_C \mathbf{F}\cdot d\mathbf{r}$ equals the circulation of $\mathbf{F}$ over the unit circle in Example 5 of Section 18.1, hence equals 2π. ♦

NOTE: This is a lot easier than integrating $\mathbf{F}$ directly over each of the four edges of the square.

Graphing $\nabla\times\mathbf{F}$ when $\mathbf{F}$ is a Planar Vector Field

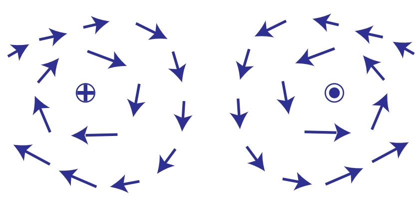


Figure 18.2.5

The curl of the planar vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$ has the form, $\nabla\times\mathbf{F} = z(x,y)\mathbf{k}$. If $z(x,y)$ is positive, the curl points directly up from the page. Indicate this by the symbol $\odot$, which suggests the point of an arrow or the nose of a rocket. If $z(x,y)$ is negative, the curl points down from the page. To show this, use the symbol $\oplus$, which suggests the feathers of an arrow or the fins of a rocket. These are standard notations in physics. Figure 18.2.5 illustrates their use.

Summary

Two main ideas in this section are:

- expressing Green's theorem in terms of scalar functions: $\displaystyle\oint_C (P\,dx + Q\,dy) = \int_{\mathscr{R}} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\,dA$
- generalizing it into a statement about the circulation of a vector field: $\displaystyle\oint_C \mathbf{F}\cdot d\mathbf{r} = \int_{\mathscr{R}} (\nabla\times\mathbf{F})\cdot\mathbf{k}\,dA$.

In both cases the curve C is both closed and oriented counterclockwise, and $\mathscr{R}$ is the enclosed region.

We used these ideas to obtain the following test for a conservative vector field in the plane (Theorem 18.2.3): If the curl of $\mathbf{F}$ is $\mathbf{0}$ and if the domain of $\mathbf{F}$ is simply connected, then $\mathbf{F}$ is conservative.

Also, in a region in which $\nabla\times\mathbf{F} = \mathbf{0}$, the value of $\oint_C \mathbf{F}\cdot d\mathbf{r}$ does not change as you gradually change C to other curves in the region.

EXERCISES for Section 18.2

1. State Green's theorem in words, using no mathematical symbols.

2. State the two-curve form of Green's theorem in words, using no mathematical symbols.

In Exercises 3 through 6 verify Green's theorem for the given functions P and Q and curve C.

3. $P = xy$, $Q = y^2$ and C is the border of the square whose vertices are $(0,0)$, $(1,0)$, $(1,1)$ and $(0,1)$.

4. $P = x^2$, $Q = 0$ and C is the boundary of the unit circle with center $(0,0)$.

5. $P = e^y$, $Q = e^x$ and C is the triangle with vertices $(0,0)$, $(1,0)$, and $(0,1)$.

6. $P = \sin(y)$, $Q = 0$ and C is the boundary of the portion of the unit disk with center $(0,0)$ in the first quadrant.

---

7. Figure 18.2.6 shows a vector field for a fluid flow $\mathbf{F}$. At the indicated points A, B, C, and D tell when the curl of $\mathbf{F}$ is pointed up, down, or is $\mathbf{0}$. (Use the $\odot$ and $\oplus$ notation.)

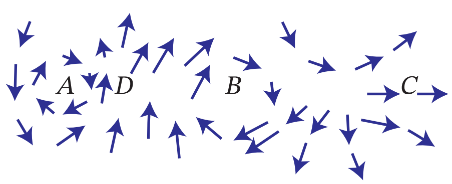


Figure 18.2.6

8. Assume that $\mathbf{F}$ describes a fluid flow. Let P be a point in the domain of $\mathbf{F}$ and C a small circular path around P. Assume the curl of $\mathbf{F}$ points upward.

(a) In what direction is the fluid tending to turn near P, clockwise or counterclockwise?
(b) If C is oriented clockwise, is $\oint_C \mathbf{F} \cdot d\mathbf{r}$ positive or negative?

Exercises 9 to 11 explore alternate ways to find a scalar function f such that $\nabla f = \mathbf{F}$.

9. In Example 2 we constructed a function f by using a straight path from $(0,0)$ to (a,b). Instead, construct f by using a path that consists of two line segments, the first from $(0,0)$ to $(a,0)$, and the second, from $(a,0)$ to (a,b).

10. In Example 2 we constructed a function f by using a straight path from $(0,0)$ to (a,b). Instead, construct f by using a path that consists of two line segments, the first from $(0,0)$ to $(0,b)$, and the second from $(0,b)$ to (a,b).

11. Another way to construct a potential function f for a vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$ is to work directly with the requirement that $\nabla f = \mathbf{F}$. That is, with the equations $\frac{\partial f}{\partial x} = P(x,y)$ and $\frac{\partial f}{\partial y} = Q(x,y)$. The following steps outline how to use this approach for the vector field of Example 2.

(a) Integrate $\frac{\partial f}{\partial x} = e^x y$ with respect to x to conclude that $f(x,y) = e^x y + C(y)$. Note that the "constant of integration" can be any function of y, which we call $C(y)$. (Why?)
(b) Next, differentiate the result found in (a) with respect to y. This gives two formulas for $\frac{\partial f}{\partial y}$: $e^x + C'(y)$ and $e^x + 2y$. Use these facts to explain why $C'(y) = 2y$.
(c) Find the function C that solves the equation found in (b).
(d) Combine the results of (a) and (c) to obtain the general form for a potential function for this vector field.

In Exercises 12 through 15:

(a) Check that $\mathbf{F}$ is conservative in the given domain.

(b) Construct f such that $\nabla f = \mathbf{F}$, using integrals on curves.

(c) Construct f such that $\nabla f = \mathbf{F}$, using antiderivatives, as in Exercise 11.

12. $\mathbf{F}(x,y) = 3x^2y\mathbf{i} + x^3\mathbf{j}$, domain the xy-plane

13. $\mathbf{F}(x,y) = y\cos(xy)\mathbf{i} + (x\cos(xy) + 2y)\mathbf{j}$, domain the xy-plane

14. $\mathbf{F}(x,y) = \left(ye^{xy} + \dfrac{1}{x}\right)\mathbf{i} + xe^{xy}\mathbf{j}$, domain all (x,y) with $x > 0$

15. $\mathbf{F}(x,y) = \dfrac{2y\ln(x)}{x}\mathbf{i} + (\ln(x))^2\mathbf{j}$, domain all (x,y) with $x > 0$

---

16. Verify Green's theorem when $\mathbf{F}(x,y) = x\mathbf{i} + y\mathbf{j}$ and $\mathscr{R}$ is the disk of radius a with center at the origin.

17. In Example 1 Green's theorem was used to show that $-\oint_C y\,dx$ is the area that C encloses. Use Green's theorem to show that $\oint_C x\,dy$ also equals this area. (This result was obtained in Section 15.3 without Green's theorem.)

18. Let $\mathscr{A}$ be a plane region with boundary C a simple closed curve swept out counterclockwise. Use Green's theorem to show that the area of $\mathscr{A}$ equals $\dfrac{1}{2}\oint_C (-y\,dx + x\,dy)$.

19. (a) Use Exercise 18 to find the area of the region bounded by the line $y = x$ and the curve given parametrically as $x = t^6 + t^4$, $y = t^3 + t$ for t in $[0,1]$.

(b) Repeat (a) using the formula for the area of a region found in Exercise 17.

(c) Repeat (a) using the formula for the area of a region found in Example 1.

(d) Your answers to (a), (b), and (c) should, of course, be the same. What is different is the amount of work needed to setup and to evaluate the resulting definite integral. Which of the three formulas is easiest to apply in this case?

20. Assume that $\mathbf{curl}\,\mathbf{F}$ at $(0,0)$ is $-3\mathbf{k}$. Let C sweep out counterclockwise the circle of radius a, center at $(0,0)$. When a is small, estimate the circulation $\oint_C \mathbf{F}\cdot d\mathbf{r}$.

In Exercises 21 to Exercises 26, determine the domain of the vector field $\mathbf{F}$ and whether $\mathbf{F}$ is conservative on that domain.

21. $\mathbf{F}(x,y) = x\mathbf{i} - y\mathbf{j}$

22. $\mathbf{F}(x,y) = \dfrac{x\mathbf{i} - y\mathbf{j}}{x^2 + y^2}$

23. $\mathbf{F}(x,y) = 3\mathbf{i} + 4\mathbf{j}$

24. $\mathbf{F}(x,y) = \left(6xy - y^3\right)\mathbf{i} + \left(4y + 3x^2 - 3xy^2\right)\mathbf{j}$

25. $\mathbf{F}(x,y) = \dfrac{y\mathbf{i} - x\mathbf{j}}{1 + x^2y^2}$

26. $\mathbf{F}(x,y) = \dfrac{x\mathbf{i} + y\mathbf{j}}{x^2 + y^2}$

---

27. Let $\mathbf{F}(x,y) = y^2\mathbf{i}$.

(a) Sketch the vector field $\mathbf{F}$.

(b) Without computing it, predict where $(\nabla\times\mathbf{F})\cdot\mathbf{k}$ is positive, negative, or zero.

(c) Compute $(\nabla\times\mathbf{F})\cdot\mathbf{k}$.

(d) What would happen if you dipped a paddle wheel with small blades into the flow. Assume you keep its axis parallel to $\mathbf{k}$.

28. Figure 18.2.7 shows a fluid flow **F**. All the vectors are parallel, but their magnitudes increase from bottom to top. A small simple curve C is placed in the flow.

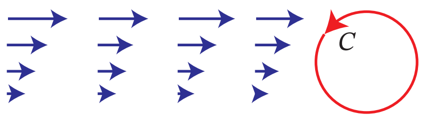


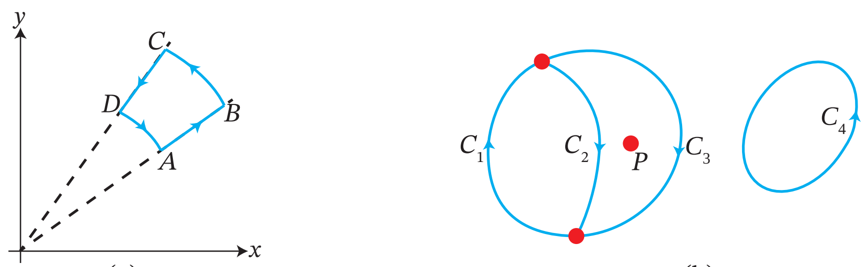
Figure 18.2.7

(a) Assume C has a counterclockwise orientation. Is the circulation around C positive, negative, or 0? Justify your opinion.

(b) A paddle wheel with small blades is dipped into the flow with its axis parallel to **k**. Would the paddle wheel rotate? If so, which way?

29. Check that the curl of the vector field in Example 2 is $\mathbf{0}$, as asserted.

30. Explain in words, without explicit calculations, why the circulation of the field $f(r)\widehat{\mathbf{r}}$ around the curve $ABCDA$ in Figure 18.2.8(a) is zero. As usual, f is a scalar function, $r = |\mathbf{r}|$, and $\widehat{\mathbf{r}} = \frac{\mathbf{r}}{r}$.



(a) (b)

Figure 18.2.8

Figure 18.2.8(b) shows four curves C_1, C_2, C_3, and C_4 and a point P where the vector field **F** is not defined. Assume $\nabla \times \mathbf{F} = \mathbf{0}$ and $\int\limits_{C_1} \mathbf{F}\cdot d\mathbf{r} = 5$. In Exercises 31 to 36 what, if anything, can be said about $\int\limits_{C} \mathbf{F}\cdot d\mathbf{r}$ for the following curves?

31. $C = C_1 + C_2$ (C_1 followed by C_2.)

32. $C = C_1 + C_3$

33. $C = C_2 + (-C_3)$

34. $C = C_2$

35. $C = C_3$

36. $C = C_4$

In Exercises 37 to 40 show that the vector field **F** is conservative and then construct a scalar function of which it is the gradient. Use the method in Example 2.

37. $\mathbf{F}(x, y) = 2xy\mathbf{i} + x^2\mathbf{j}$

38. $\mathbf{F}(x, y) = \sin(y)\mathbf{i} + (x\cos(y) + 3)\mathbf{j}$

39. $\mathbf{F}(x, y) = (y+1)\mathbf{i} + (x+1)\mathbf{j}$

40. $\mathbf{F}(x, y) = 3y\sin^2(xy)\cos(xy)\mathbf{i} + (1 + 3x\sin^2(xy)\cos(xy))\mathbf{j}$

---

41. Show that (a) $3x^2y\,dx + x^3\,dy$ is exact and (b) $3xy\,dx + x^2\,dy$ is not exact.

42. Show that the differential form $\dfrac{x\,dx + y\,dy}{x^2+y^2}$ is exact and find a function f such that df equals the given expression. (That is, determine f such that $\nabla f \cdot d\mathbf{r}$ agrees with the given differential form.)

43. Use Exercise 18 to obtain the formula for area in polar coordinates: Area $= \dfrac{1}{2}\int\limits_{\alpha}^{\beta} r^2\,d\theta$.

44. Figure 18.2.9(a) shows the direction of the curl of a vector field at three points. Draw a nonzero vector field compatible with these values.

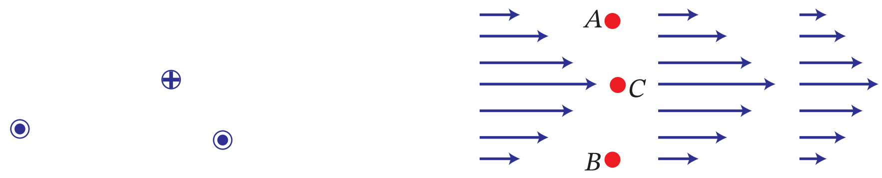


Figure 18.2.9

45. Figure 18.2.9(b) shows a vector field in the xy-plane. A paddle wheel with small blades is inserted into the flow with its axis parallel to **k**. Will the paddle wheel turn if it is inserted (a) at A? (b) at B? (c) at C? If so, in which direction?

46. A curve is given parametrically by $x = t(1-t^2)$, $y = t^2(1-t^3)$, for t in $[0,1]$.

(a) Sketch the points corresponding to $t = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$, and 1, and use them to sketch the curve.

(b) Let $\mathscr{R}$ be the region enclosed by the curve. What difficulty arises when you try to compute the area of $\mathscr{R}$ by a definite integral involving vertical or horizontal cross sections?

(c) Three different formula for the area of a region in terms of a line integral along the boundary of the region were obtained in Example 1 and Exercises 17 and 18. Use one of these formulas to find the area of $\mathscr{R}$.

47. Repeat Exercise 46 for $x = \sin(\pi t)$ and $y = t - t^2$, for t in $[0,1]$.

48. Assume that the circulation of **F** along every circle in the xy-plane is 0. Is **F** conservative?

49. Assume that you know that Green's theorem is true for every triangle $\mathscr{R}$.

(a) Deduce that Green's theorem therefore holds for quadrilaterals.

(b) Deduce that Green's theorem holds for polygons.

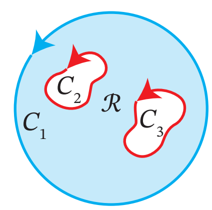


Figure 18.2.10

50. Assume that $\nabla \times \mathbf{F} = \mathbf{0}$ in the region $\mathscr{R}$ bounded by an exterior curve C_1 and two interior curves C_2 and C_3, as in Figure 18.2.10. Show that $\int_{C_1} \mathbf{F}\cdot d\mathbf{r} = \int_{C_2} \mathbf{F}\cdot d\mathbf{r} + \int_{C_3} \mathbf{F}\cdot d\mathbf{r}$.

All three curves are closed and have a counterclockwise orientation.

18.3 Green's Theorem, Flux, and Divergence

In the previous section we introduced Green's theorem and applied it to obtain information about circulation and curl:

$$\oint_C \mathbf{F}\cdot d\mathbf{r} = \int_{\mathscr{R}} (\nabla \times \mathbf{F})\cdot \mathbf{k}\, dA. \qquad (18.3.1)$$

Since $\mathbf{F}\cdot d\mathbf{r}$ is an abbreviation for $(\mathbf{F}\cdot\mathbf{T})\,ds$, we see that this result concerned line integrals of the tangential component of **F**, $\mathbf{F}\cdot\mathbf{T}$, along the curve C. Now we will translate Green's theorem to obtain corresponding information about the line integral of the normal component of a vector field **F**, $\mathbf{F}\cdot\mathbf{n}$, along the curve C. Thus Green's theorem also provides information about the flow of the vector field **F** across a closed curve C.

Green's Theorem Expressed in Terms of Flux

Let $\mathbf{F} = M\mathbf{i} + N\mathbf{j}$ and C be a counterclockwise closed curve. At a point on a closed curve the **unit exterior normal vector** (or **unit outward normal vector**) **n** is perpendicular to the curve and points outward from the region enclosed by the curve. To compute $\mathbf{F}\cdot\mathbf{n}$ in terms of M and N, we first express **n** in terms of **i** and **j**.

The vector

$$\mathbf{T} = \frac{dx}{ds}\mathbf{i} + \frac{dy}{ds}\mathbf{j}$$

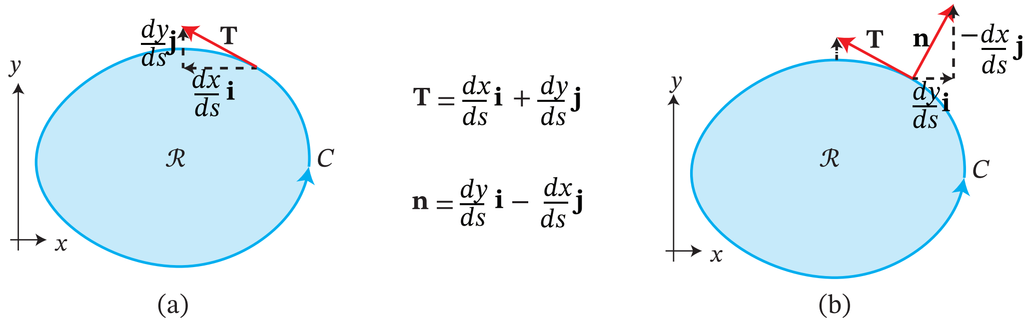


Figure 18.3.1

is tangent to the curve, has length 1, and points in the direction in which the curve is swept out, as shown in Figure 18.3.1(a).

Figure 18.3.1(b) shows the exterior unit normal $\mathbf{n}$ has its x-component equal to the y-component of $\mathbf{T}$ and its y-component equal to the negative of the x-component of $\mathbf{T}$. Thus

$$\mathbf{n} = \frac{dy}{ds}\mathbf{i} - \frac{dx}{ds}\mathbf{j}.$$

Consequently, if $\mathbf{F} = M\mathbf{i} + N\mathbf{j}$, then

$$\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \oint_C (M\mathbf{i}+N\mathbf{j})\cdot\left(\frac{dy}{ds}\mathbf{i} - \frac{dx}{ds}\mathbf{j}\right)\,ds = \oint_C \left(M\frac{dy}{ds} - N\frac{dx}{ds}\right)\,ds$$
$$= \oint_C (M\,dy - N\,dx) = \oint_C (-N\,dx + M\,dy). \tag{18.3.2}$$

In (18.3.2), $-N$ plays the role of P and M plays the role of Q in Green's theorem. Since Green's theorem states that

$$\oint_C (P\,dx + Q\,dy) = \int_{\mathcal{R}} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\,dA$$

we have

$$\oint_C (-N\,dx + M\,dy) = \int_{\mathcal{R}} \left(\frac{\partial M}{\partial x} - \frac{\partial(-N)}{\partial y}\right)\,dA$$

or, if $\mathbf{F} = M\mathbf{i} + N\mathbf{j}$, then

$$\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \int_{\mathcal{R}} \left(\frac{\partial M}{\partial x} + \frac{\partial N}{\partial y}\right)\,dA.$$

We have just proven the following theorem about flux in the plane.

Theorem 18.3.1: Green's Theorem Expressed in Terms of Flux

If $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$, *then*

$$\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \int_{\mathcal{R}} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right)\,dA$$

where C *is the boundary of the region* $\mathcal{R}$ *(with a counterclockwise orientation).*

The expression

$$\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y},$$

the sum of two partial derivatives, is called the **divergence** of $\mathbf{F} = P\mathbf{i}+Q\mathbf{j}$. It is written div $\mathbf{F}$ or $\nabla\cdot\mathbf{F}$. The latter notation is suggested by the symbolic dot product

$$\nabla\cdot\mathbf{F} = \left(\frac{\partial}{\partial x}\mathbf{i}+\frac{\partial}{\partial y}\mathbf{j}\right)\cdot(P\mathbf{i}+Q\mathbf{j}) = \frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}.$$

The symbols $\nabla\cdot\mathbf{F}$ are pronounced "del dot eff". Theorem 18.3.1 is called the **divergence theorem in the plane**. The next theorem is the same as Theorem 18.3.1 with the exception that it is written without explicit reference to components.

Theorem 18.3.2: Divergence Theorem in the Plane (Vector Form)

Let $\mathbf{F}$ *be a vector field in the* xy*-plane. Then*

$$\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \int_{\mathscr{R}} \operatorname{div}\mathbf{F}\,dA$$

where the closed curve C *is the boundary of the planar region* $\mathscr{R}$.

EXAMPLE 1. Compute the divergence of (a) $\mathbf{F} = e^{xy}\mathbf{i}+\arctan(3x)\mathbf{j}$ and (b) $\mathbf{F} = -x^2\mathbf{i}+2xy\mathbf{j}$.

SOLUTION (a) $\frac{\partial}{\partial x}e^{xy}+\frac{\partial}{\partial y}\arctan(3x) = ye^{xy}+0 = ye^{xy}$ and (b) $\frac{\partial}{\partial x}(-x^2)+\frac{\partial}{\partial y}(2xy) = -2x+2x = 0.$ ♦

The double integral of the divergence of a vector field $\mathbf{F}$ describing fluid flow over a region thus describes the amount of flow across the border of the region. It tells how rapidly the fluid is leaving (diverging) or entering (converging) the region. Hence the name "divergence".

In the next section we will be using the divergence of a vector field defined in space, $\mathbf{F} = P\mathbf{i}+Q\mathbf{j}+R\mathbf{k}$, where P, Q, and R are functions of x, y, and z. It is not a surprise that, in this case, the divergence of $\mathbf{F}$ is defined as the sum of three partial derivatives:

$$\nabla\cdot\mathbf{F} = \frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}.$$

It will play a role in measuring flux across a surface.

EXAMPLE 2. Verify that $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$ equals $\int_{\mathscr{R}} \nabla\cdot\mathbf{F}\,dA$, when $\mathbf{F}(x,y) = x\mathbf{i}+y\mathbf{j}$, $\mathscr{R}$ is the disk of radius a and center at the origin and C is the boundary curve of $\mathscr{R}$.

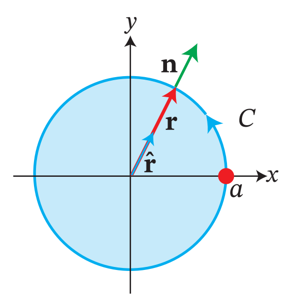


Figure 18.3.2

SOLUTION We compute $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$, where C is the circle bounding $\mathscr{R}$. See Figure 18.3.2. Since C is a circle centered at $(0,0)$, the unit exterior normal $\mathbf{n}$ is $\hat{\mathbf{r}}$:

$$\mathbf{n} = \hat{\mathbf{r}} = \frac{x\mathbf{i}+y\mathbf{j}}{|x\mathbf{i}+y\mathbf{j}|} = \frac{x\mathbf{i}+y\mathbf{j}}{a}.$$

Thus

$$\begin{aligned}\oint_C \mathbf{F}\cdot\mathbf{n}\,ds &= \oint_C (x\mathbf{i}+y\mathbf{j})\cdot\left(\frac{x\mathbf{i}+y\mathbf{j}}{a}\right)ds = \oint_C \frac{x^2+y^2}{a}\,ds\\ &= \oint_C \frac{a^2}{a}\,ds = a\oint_C ds = a(2\pi a) = 2\pi a^2. \qquad (18.3.3)\end{aligned}$$

Next we compute $\int_{\mathscr{R}} \left(\partial P/\partial x + \partial Q/\partial y\right)\, dA$. Since $P = x$ and $Q = y$, $\partial P/\partial x + \partial Q/\partial y = 1+1 = 2$. Then

$$\int_{\mathscr{R}} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) dA = \int_{\mathscr{R}} 2\, dA,$$

which is twice the area of the disk R, and hence is $2\pi a^2$. This agrees with (18.3.3), confirming Theorem 18.3.1. ◆

As the next example shows, a double integral can provide a way to compute the flux $\oint_C \mathbf{F}\cdot\mathbf{n}\, ds$.

EXAMPLE 3. Let $\mathbf{F} = x^2\mathbf{i} + xy\mathbf{j}$. Evaluate $\oint_C \mathbf{F}\cdot\mathbf{n}\, ds$ over the curve that bounds the quadrilateral with vertices $(1,1)$, $(3,1)$, $(3,4)$, and $(1,2)$.

SOLUTION The shaded region in Figure 18.3.3 is $\mathscr{R}$; its boundary is C. The line integral could be evaluated directly, but would require parameterizing each of the four edges of C. With Green's theorem we can instead evaluate an integral over a single plane region. (Even though the unit normal vector $\mathbf{n}$ is not defined at the four vertices of C, Green's theorem applies to this problem.)

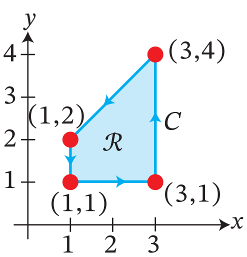


Figure 18.3.3

We begin by applying Green's theorem (Theorem 18.3.1) to find the double integral that is equivalent to the original line integral:

$$\begin{aligned}\oint_C \mathbf{F}\cdot\mathbf{n}\, ds &= \int_{\mathscr{R}} \nabla\cdot\mathbf{F}\, dA = \int_{\mathscr{R}} \left(\frac{\partial(x^2)}{\partial x} + \frac{\partial(xy)}{\partial y}\right) dA \\ &= \int_{\mathscr{R}} (2x+x)\, dA = \int_{\mathscr{R}} 3x\, dA.\end{aligned}$$

Then, to evaluate the double integral we express it as an iterated integral:

$$\int_{\mathscr{R}} 3x\, dA = \int_1^3 \int_1^{y(x)} 3x\, dy\, dx,$$

where $y(x)$ is determined by the equation of the line that provides the top edge of $\mathscr{R}$. The line through $(1,2)$ and $(3,4)$ has the equation $y = x+1$. Therefore,

$$\int_{\mathscr{R}} 3x\, dA = \int_1^3 \int_1^{x+1} 3x\, dy\, dx.$$

Evaluation of this iterated integral begins with the inner integration:

$$\int_1^{x+1} 3x\, dy = 3xy\Big|_{y=1}^{y=x+1} = 3x(x+1) - 3x = 3x^2.$$

The outer integration is equally easy to evaluate:

$$\int_1^3 3x^2\, dx = x^3\Big|_1^3 = 27 - 1 = 26.$$

In conclusion, $\oint_C \mathbf{F}\cdot\mathbf{n}\, ds = 26$. ◆

This same line integral appears in Exercise 15 where it is to be evaluated using parameterizations of the four segments that make up the boundary curve C. Examples like this one provide an opportunity to see the benefits of Green's theorem.

A Local View of $\operatorname{div}\mathbf{F}$

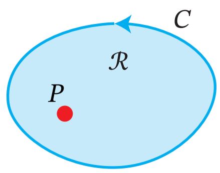


Figure 18.3.4

We have presented a global view of $\operatorname{div}\mathbf{F}$, integrating it over a region $\mathscr{R}$ to get the total divergence across the boundary of $\mathscr{R}$. There is a way of viewing $\operatorname{div}\mathbf{F}$ locally. It uses an extension of the permanence property of Section 2.5 to the plane and to space.

Let P be a point in the plane and $\mathbf{F}$ a vector field describing fluid flow. Choose a small region $\mathscr{R}$ around P, and let C be its boundary. See Figure 18.3.4. Then the net flow out of R is

$$\oint_C \mathbf{F}\cdot\mathbf{n}\,ds.$$

By Green's theorem, the net flow is also

$$\int_{\mathscr{R}} \operatorname{div}\mathbf{F}\,dA.$$

Since $\operatorname{div}\mathbf{F}$ is continuous and $\mathscr{R}$ is small, $\operatorname{div}\mathbf{F}$ is almost constant throughout $\mathscr{R}$, staying close to the divergence of $\mathbf{F}$ at P. Thus

$$\int_{\mathscr{R}} \operatorname{div}\mathbf{F}\,dA \approx (\operatorname{div}\mathbf{F}(P))\ (\text{Area of }\mathscr{R}).$$

or, equivalently,

$$\frac{\text{Net flow out of }\mathscr{R}}{\text{Area of }\mathscr{R}} \approx \operatorname{div}\mathbf{F}(P). \tag{18.3.4}$$

This means that the divergence of $\mathbf{F}$ at P, $\operatorname{div}\mathbf{F}(P)$, is a measure of the rate at which fluid tends to leave a small region around P, hence another reason for the name "divergence." If $\operatorname{div}\mathbf{F}$ is positive, fluid near P tends to get less dense (diverge). If $\operatorname{div}\mathbf{F}$ is negative, fluid near P tends to accumulate (converge). Physicists also refer to $\operatorname{div}\mathbf{F}$ as "flux density," for if it is multiplied by the area of a small region around it, the product approximates the flux out of the region.

Estimate (18.3.4) suggests another definition of $\operatorname{div}\mathbf{F}$ at P. This definition involves a limit as the diameters of regions $\mathscr{R}$ containing P approach 0. Rather than writing $\lim_{\operatorname{diam}\mathscr{R}\to 0}$ we will write $\lim_{\mathscr{R}\to P}$.

Definition: *Local Definition of* $\operatorname{div}\mathbf{F}(P)$

The divergence of the vector field $\mathbf{F}$ at a point P in the domain of $\mathbf{F}$ is

$$\operatorname{div}\mathbf{F}(P) = \lim_{\mathscr{R}\to P} \frac{1}{\text{Area of }\mathscr{R}} \oint_C \mathbf{F}\cdot\mathbf{n}\,ds$$

where each $\mathscr{R}$ is a region enclosing the point P and whose boundary C is a simple closed curve.

Observation 18.3.3: *Making Sense of the Local Definition of* $\operatorname{div}\mathbf{F}(P)$

The local definition of $\operatorname{div}\mathbf{F}(P)$ also appeals to our physical intuition. We began by defining $\operatorname{div}\mathbf{F}$ mathematically, as $\partial P/\partial x + \partial Q/\partial y$. We now see its physical meaning, which is independent of a coordinate system. The coordinate-free definition is needed in Section 18.9.

EXAMPLE 4. Estimate the flux of $\mathbf{F}$ across a small circle C of radius a if $\operatorname{div}\mathbf{F}$ at the center of the circle is 3.

SOLUTION The flux of $\mathbf{F}$ across C is $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$, which equals $\int_{\mathscr{R}} \operatorname{div}\mathbf{F}\,dA$, where $\mathscr{R}$ is the disk that C bounds. Since $\operatorname{div}\mathbf{F}$ is continuous, it changes little in a small enough disk, and we treat it as almost constant. Then

$$\int_{\mathscr{R}} \operatorname{div}\mathbf{F}\,dA \approx (3)(\text{Area of }\mathscr{R}) = 3(\pi a^2) = 3\pi a^2.$$

◆

As Steve Whitaker of the chemical engineering department at the University of California at Davis has observed, "The concepts that one must understand to prove a theorem are frequently the concepts one must understand to apply the theorem."

The following proof of Green's theorem does more than show that Green's theorem is true. It has been known for over 150 years, and no one has said it is false — not even Sam. Studying a proof reinforces and extends one's understanding of several fundamental ideas.

Proof of Green's Theorem

In the proof we will use the concepts of a double integral, an iterated integral, a line integral, and the fundamental theorem of calculus. The reasoning provides a review of four basic ideas.

We will prove that $\oint_C P\,dx = -\int_{\mathscr{R}} \partial P/\partial y\, dA$. A similar proof confirms that $\oint_C Q\,dy = \int_{\mathscr{R}} \partial Q/\partial x\, dA$. (See Exercise 38.)

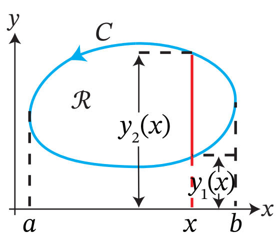


Figure 18.3.5

To avoid distracting details we assume that $\mathscr{R}$ is **strictly convex**. It has no dents and its border has no straight line segments. The ideas in the proof show up clearly in this special case. Under these assumptions $\mathscr{R}$ can be described as $a \le x \le b$, $y_1(x) \le y \le y_2(x)$ with $y_1(a) = y_2(a)$ and $y_1(b) = y_2(b)$, as shown in Figure 18.3.5. We will express $\int_{\mathscr{R}} \partial P/\partial y\, dA$ and $\oint_C P\,dx$ as definite integrals over the interval $[a,b]$. Exercise 37 asks for a proof in a slightly more general case. Proofs of Green's theorem for more general regions are typically discussed in a more advanced course such as real analysis or vector analysis.

First, the description of $\mathscr{R}$ allows the double integral to be written as the following iterated integral with y as the inner variable:

$$\int_{\mathscr{R}} \frac{\partial P}{\partial y}\, dA = \int_a^b \int_{y_1(x)}^{y_2(x)} \frac{\partial P}{\partial y}\, dy\, dx. \tag{18.3.5}$$

The fundamental theorem of calculus allows us to rewrite the inner integral in (18.3.5) as

$$\int_{y_1(x)}^{y_2(x)} \frac{\partial P}{\partial y}\, dy = P(x, y_2(x)) - P(x, y_1(x)).$$

Thus

$$\int_{\mathscr{R}} \frac{\partial P}{\partial y}\, dA = \int_a^b \big(P(x, y_2(x)) - P(x, y_1(x))\big)\, dx. \tag{18.3.6}$$

To express $\oint_C P\,dx$ as an integral over $[a,b]$, break C into two paths, one along the bottom part of $\mathscr{R}$, described by $y = y_1(x)$, the other along the top part of $\mathscr{R}$, described by $y = y_2(x)$. Denote the bottom path C_1 and the top path C_2. (See Figure 18.3.6.)

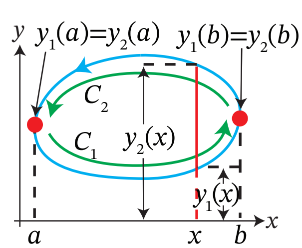


Figure 18.3.6

Then

$$\oint_C P\,dx = \int_{C_1} P\,dx + \int_{C_2} P\,dx. \tag{18.3.7}$$

The line integrals along C_1 and C_2 in (18.3.7) can be expressed as

$$\int_{C_1} P\,dx = \int_{C_1} P(x, y_1(x))\,dx = \int_a^b P(x, y_1(x))\,dx,$$

and

$$\int_{C_2} P\,dx = \int_{C_2} P(x, y_2(x))\,dx = \int_b^a P(x, y_2(x))\,dx = -\int_a^b P(x, y_2(x))\,dx.$$

Thus, by (18.3.7),

$$\oint_C P\,dx = \int_a^b P(x, y_1(x))\,dx - \int_a^b P(x, y_2(x))\,dx$$
$$= \int_a^b \left(P(x, y_1(x)) - P(x, y_2(x))\right)dx.$$

As this is the negative of the right side of (18.3.6), this completes the proof that

$$\oint_C P\,dx = -\int_{\mathscr{R}} \frac{\partial P}{\partial y}\,dA,$$

in the special case that $\mathscr{R}$ is strictly convex. •

The requirement that $\mathscr{R}$ is strictly convex is relaxed, a little, in Exercise 37.

Summary

We introduced the divergence of a vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$, namely the scalar field $\operatorname{div}\mathbf{F} = \nabla\cdot\mathbf{F} = \partial P/\partial x + \partial Q/\partial y$. Vector fields whose divergences are always zero are called divergence-free or incompressible.

We translated Green's theorem into a theorem about the flux of a vector field in the xy-plane,

$$\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \int_{\mathscr{R}} \operatorname{div}\mathbf{F}\,dA.$$

It says that the integral of the normal component of $\mathbf{F}$ around a simple closed curve equals the integral of the divergence of $\mathbf{F}$ over the region which the curve bounds.

From this it follows that

$$\operatorname{div}\mathbf{F}(P) = \lim_{\mathscr{R}\to P} \frac{1}{\text{Area of }\mathscr{R}} \oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \lim_{\mathscr{R}\to P} \frac{\text{Flux across } C}{\text{Area of }\mathscr{R}}$$

where C is the boundary of the region $\mathscr{R}$, which contains P.

The section concluded with a proof of Green's theorem.

EXERCISES for Section 18.3

1. State the divergence form of Green's theorem in symbols for a vector field $\mathbf{H}$.
2. State the divergence form of Green's theorem in words, using no symbols.

In Exercises 3 to 6 compute the divergence of

3. $\mathbf{F} = x^3y\mathbf{i} + x^2y^3\mathbf{j}$
4. $\mathbf{F} = \arctan(3xy)\mathbf{i} + e^{y/x}\mathbf{j}$
5. $\mathbf{F} = \ln(x+y)\mathbf{i} + xy\left(\arcsin(y)\right)^2\mathbf{j}$
6. $\mathbf{F} = y\sqrt{1+x^2}\mathbf{i} + \ln\left((x+1)^3(\sin(y))^{3/5}\right)e^{x+y})\mathbf{j}$

In Exercises 7 to 10 compute $\int_{\mathscr{R}} \operatorname{div} \mathbf{F}\, dA$ and $\oint_C \mathbf{F}\cdot\mathbf{n}\, ds$ and check that they are equal.

7. $\mathbf{F} = 3x\mathbf{i} + 2y\mathbf{j}$, and $\mathscr{R}$ is the disk of radius 1 with center $(0,0)$.

8. $\mathbf{F} = 5y^3\mathbf{i} - 6x^2\mathbf{j}$, and $\mathscr{R}$ is the disk of radius 2 with center $(0,0)$.

9. $\mathbf{F} = xy\mathbf{i} + x^2y\mathbf{j}$, and $\mathscr{R}$ is the rectangle with vertices $(0,0)$, $(a,0)$ (a,b) and $(0,b)$, where a, $b > 0$.

10. $\mathbf{F} = \cos(x+y)\mathbf{i} + \sin(x+y)\mathbf{j}$, and $\mathscr{R}$ is the triangle with vertices $(0,0)$, $(a,0)$ and (a,b), where a, $b > 0$.

In Exercises 11 to 14 use Green's theorem expressed in terms of divergence to evaluate $\oint_C \mathbf{F}\cdot\mathbf{n}\, ds$, where C is the boundary of $\mathscr{R}$.

11. $\mathbf{F} = e^x\sin(y)\mathbf{i} + e^{2x}\cos(y)\mathbf{j}$, and $\mathscr{R}$ is the rectangle with vertices $(0,0)$, $(1,0)$, $\left(0,\frac{\pi}{2}\right)$, and $\left(1,\frac{\pi}{2}\right)$.

12. $\mathbf{F} = y\tan(x)\mathbf{i} + y^2\mathbf{j}$, and $\mathscr{R}$ is the square with vertices $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$.

13. $\mathbf{F} = 2x^3y\mathbf{i} - 3x^2y^2\mathbf{j}$, and $\mathscr{R}$ is the triangle with vertices $(0,1)$, $(3,4)$, and $(2,7)$.

14. $\mathbf{F} = \frac{-1}{xy^2}\mathbf{i} + \frac{1}{x^2y}\mathbf{j}$, and $\mathscr{R}$ is the triangle with vertices $(1,1)$, $(2,2)$, and $(1,2)$.

---

15. In Example 3 we found $\oint_C \mathbf{F}\cdot\mathbf{n}\, ds$ by evaluating a double integral. Determine its value directly.

16. Let $\mathbf{F}(x,y) = \mathbf{i}$, a constant field.
(a) Evaluate directly the flux of $\mathbf{F}$ around the triangular path, $(0,0)$ to $(1,0)$, to $(0,1)$, back to $(0,0)$.
(b) Use the divergence of $\mathbf{F}$ to evaluate the flux in (a).

17. Assume a is a small number, $\mathscr{R}$ is the square with vertices (a,a), $(-a,a)$, $(-a,-a)$, and $(a,-a)$, and C is the boundary of $\mathscr{R}$, oriented counterclockwise. If the divergence of $\mathbf{F}$ at the origin is 3, find an estimate of $\oint_C \mathbf{F}\cdot\mathbf{n}\, ds$.

18. Assume $|\mathbf{F}(P)| \le 4$ for points P on a curve of length L that bounds a region $\mathscr{R}$ of area A. What can be said about the integral $\int_{\mathscr{R}} \nabla\cdot\mathbf{F}\, dA$?

19. Verify the divergence form of Green's theorem when $\mathbf{F} = 3x\mathbf{i} + 4y\mathbf{j}$ and C is the square whose vertices are $(2,0)$, $(5,0)$, $(5,3)$, and $(2,3)$.

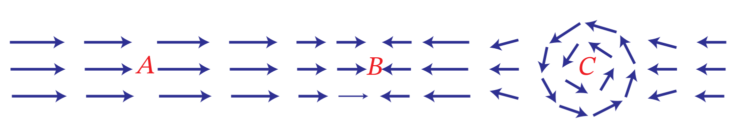


Figure 18.3.7

20. Figure 18.3.7 shows the flow $\mathbf{F}$ of a fluid. Decide whether $\nabla\cdot\mathbf{F}$ is positive, negative, or zero at A, B, and C.

Definition: *Divergence-Free, or Incompresible, Vector Field*

A vector field $\mathbf{F}$ is said to be **divergence-free** or **incompressible** when $\nabla\cdot\mathbf{F} = 0$ at every point in the field.

TERMINOLOGY: Motivation for this terminology is found in Section 18.5.

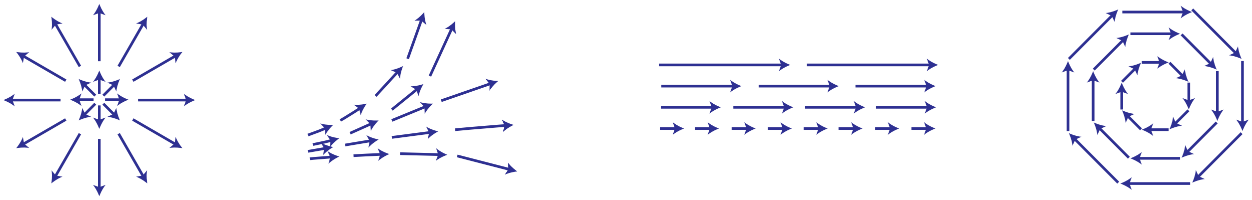


Figure 18.3.8

21. Figure 18.3.8 shows four vector fields. Two are divergence-free and two are not. Decide which two are not, copy them onto a sheet of drawing paper, and sketch a closed curve C for which $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$ is not 0.

22. For a scalar field f and a vector field $\mathbf{F}$,

(a) is the curl of the gradient of f always $\mathbf{0}$?

(b) is the divergence of the gradient of f always 0?

(c) is the divergence of the curl of $\mathbf{F}$ always 0?

(d) is the gradient of the divergence of $\mathbf{F}$ always $\mathbf{0}$?

23. Assume $\operatorname{div}\mathbf{F}$ at $(0.1, 0.1)$ is 3. Estimate $\displaystyle\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$, where C is the curve around the square whose vertices are $(0,0)$, $(0.2,0)$, $(0.2,0.2)$, and $(0,0.2)$.

24. Use Green's theorem to find the area of the region bounded by the line $y = x$ and the curve $x = t^6 + t^4$, $y = t^3 + t$ for t in $[0,1]$.

NOTE: See Exercise 19.

25. Let f be a scalar function. Let $\mathscr{R}$ be a convex region and C its boundary taken counterclockwise. Show that

$$\int_{\mathscr{R}} \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) dA = \oint_C \left(\frac{\partial f}{\partial x}\,dy - \frac{\partial f}{\partial y}\,dx\right).$$

26. Let $\mathbf{F}$ be the vector field whose formula in polar coordinates is $\mathbf{F}(r,\theta) = r^n\hat{\mathbf{r}}$, where $\mathbf{r} = x\mathbf{i} + y\mathbf{j}$, $r = |\mathbf{r}|$, and $\hat{\mathbf{r}} = \mathbf{r}/r$. Show that the divergence of $\mathbf{F}$ is $(n+1)r^{n-1}$.

27. Assume that $\mathbf{F}$ is defined everywhere in the xy-plane except at the origin and that the divergence of $\mathbf{F}$ is identically 0. Let C_1 and C_2 be two counterclockwise simple curves circling the origin. They may intersect. Show that $\displaystyle\oint_{C_1} \mathbf{F}\cdot\mathbf{n}\,ds = \oint_{C_2} \mathbf{F}\cdot\mathbf{n}\,ds$.

MAIN POINT Exercise 27 shows that if the divergence of $\mathbf{F}$ is 0, a line integral of $\mathbf{F}$ over a complicated curve can be replaced by a line integral of $\mathbf{F}$ over a simpler curve.

28. A region $\mathscr{R}$ with a hole is bounded by two oriented curves C_1 and C_2, as in Figure 18.3.9, which includes outward-pointing unit normal vectors (relative to the origin, but not to $\mathscr{R}$).

(a) Show that $\displaystyle\oint_{C_2} \mathbf{F}\cdot\mathbf{n}\,ds - \oint_{C_1} \mathbf{F}\cdot\mathbf{n}\,ds = \int_{\mathscr{R}} \nabla\cdot\mathbf{F}\,dA$.

(b) If $\nabla\cdot\mathbf{F} = 0$ in $\mathscr{R}$, show that $\displaystyle\oint_{C_1} \mathbf{F}\cdot\mathbf{n}\,ds = \oint_{C_2} \mathbf{F}\cdot\mathbf{n}\,ds$.

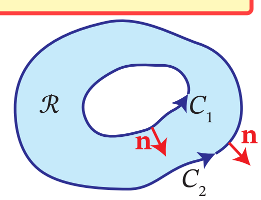


Figure 18.3.9

29. Let $\mathbf{F}$ be a vector field in the xy-plane whose flux across any rectangle is 0. Show that its flux across the curves in Figure 18.3.10(a) and (b) is also 0.

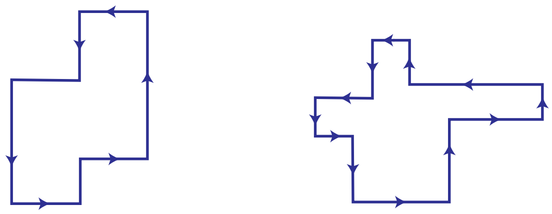


Figure 18.3.10

30. The line integral for flux is $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$, and for circulation it is $\oint_C \mathbf{F}\cdot\mathbf{T}\,ds$. Why is the first integral independent of the orientation of the curve but the second one is dependent on the orientation?

31. The field $\mathbf{F}$ is defined throughout the xy-plane. If the flux of $\mathbf{F}$ across every circle is 0, must the flux of $\mathbf{F}$ across every square be 0? Explain.

32. Let $\mathbf{F}(x, y)$ describe a fluid flow. Assume $\nabla\cdot\mathbf{F}$ is never 0 in a certain region $\mathscr{R}$. Show that none of the streamlines in the region forms a loop within $\mathscr{R}$.

33. Assume $\mathscr{R}$ is a region in the xy-plane bounded by the closed curve C and $f(x, y)$ is defined on the xy-plane. Show that $\int_{\mathscr{R}} \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial x^2}\right)\,dA = \oint_C D_{\mathbf{n}}(f)\,ds$ where $D_{\mathbf{n}}(f)$ is the directional derivative of f in the direction of the unit vector $\mathbf{n}$.

34. Evaluate $\oint_C \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r}\,ds$ where C is the closed curve consisting of the top half of the unit circle centered at the origin, the graph of $y = -2x - 2$ in the third quadrant, and the graph of $y = 2x^2 - 2$ in the fourth quadrant.

35. Assume $\mathbf{F} = \frac{\widehat{\mathbf{r}}}{|\mathbf{r}|}$ in the xy-plane, C is the circle of radius a and center $(0,0)$, with unit exterior normal $\mathbf{n}$.

(a) Evaluate $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$ without using Green's theorem.

(b) Define C' to be the circle of radius 3 and center $(4,0)$. Evaluate $\oint_{C'} \mathbf{F}\cdot\mathbf{n}\,ds$, doing as little work as possible.

36. (a) Draw enough vectors for the field $\mathbf{F}(x, y) = \frac{x\mathbf{i} + y\mathbf{j}}{x^2 + y^2}$ to show what it looks like.

(b) Compute $\nabla\cdot\mathbf{F}$.

(c) Does your sketch in (a) agree with what you found for $\nabla\cdot\mathbf{F}$ in (b)? If not, redraw the vector field.

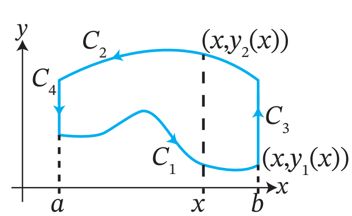


Figure 18.3.11

37. We proved that $\oint_C P\,dx = -\int_{\mathscr{R}} \frac{\partial P}{\partial y}\,dA$ in a special case, namely that $\mathscr{R}$ is strictly convex. Prove it in this more general case, in which we assume less about the region $\mathscr{R}$. Assume that $\mathscr{R}$ has the description $a \le x \le b$, $y_1(x) \le y \le y_2(x)$. Figure 18.3.11 shows such a region, which need not be convex. The curved path C breaks up into four paths, two of which are straight (or may be empty).

38. We proved the first part of (18.2.1), namely that $\oint_C P\,dx = -\int_{\mathscr{R}} \frac{\partial P}{\partial y}\,dA$. Now, prove the second part of (18.2.1). In particular, show that $\oint_C Q\,dy = \int_{\mathscr{R}} \frac{\partial Q}{\partial x}\,dA$.

18.4 Central Vector Fields

Central vector fields are a special but important type of vector field that appear in the study of gravity and the attraction or repulsion of electric charges. They radiate from a point mass or point charge. They provide a way to deal with "action at a distance." One particle acts on another indirectly, through the vector field it creates.

Introduction to Central Vector Fields

A **central vector field** is a continuous vector field defined everywhere in the plane (or in space) except, perhaps, at a point $\mathcal{O}$, with the properties:

(i) Each vector points towards (or away from) $\mathcal{O}$.

(ii) The magnitudes of all vectors at a given distance from $\mathcal{O}$ are equal.

The point $\mathcal{O}$ is called the center, or **pole**, of the field. A central vector field is also called **radially symmetric**. There are various ways to think of a central vector field. For one in the plane, the vectors at points on a circle with center $\mathcal{O}$ are perpendicular to the circle and have the same length, as shown in Figure 18.4.1(a) and (b).

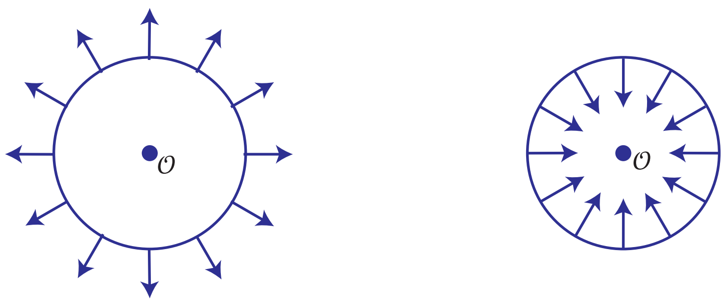


Figure 18.4.1

The same holds for central vector fields in space, with "circle" replaced by "sphere."

In the plane, the flux of a vector field across a closed curve C is $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$. Likewise, in space, the **flux** of a vector field $\mathbf{F}$ across a closed surface $\mathcal{S}$ is $\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\,dS$.

The formula for a central vector field has a simple form. Let the field be $\mathbf{F}$ and let P be any point other than $\mathcal{O}$. Denote the vector $\overrightarrow{\mathcal{O}P}$ by $\mathbf{r}$, its magnitude by r, and the unit vector $\mathbf{r}/r$ by $\hat{\mathbf{r}}$. Then there is a scalar function f, defined for all positive numbers, such that

$$\mathbf{F}(P) = f(r)\hat{\mathbf{r}}.$$

The magnitude of $\mathbf{F}(P)$ is $|f(r)|$. If $f(r)$ is positive, $\mathbf{F}(P)$ points away from $\mathcal{O}$. If $f(r)$ is negative, $\mathbf{F}(P)$ points toward $\mathcal{O}$.

If the domain of a central field $\mathbf{F}$ is part of the xy-plane we can write $\mathbf{F}(P)$ as $\mathbf{F}(x,y)$ or $\mathbf{F}(\mathbf{r})$, where $P=(x,y)$ and $\mathbf{r}=\overrightarrow{\mathcal{O}P}$.

The astute reader will recall that we encountered the two-dimensional central field $\mathbf{F}(\mathbf{r}) = \hat{\mathbf{r}}/r$ in Section 15.4 and the three-dimensional inverse-square central field $\mathbf{F}(\mathbf{r}) = \hat{\mathbf{r}}/r^2$ in Section 18.1.

The vector field $\mathbf{F}(\mathbf{r}) = (1/r)\hat{\mathbf{r}}$ can also be written as

$$\mathbf{F}(\mathbf{r}) = \frac{\mathbf{r}}{r^2}. \tag{18.4.1}$$

Its magnitude is not inversely proportional to the square of r because the magnitude of $\mathbf{r}/r^2$ is $r/r^2 = 1/r$, the reciprocal of the first power of r.

EXAMPLE 1. Evaluate the flux $\displaystyle\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$ for the central field $\mathbf{F}(x,y) = f(r)\hat{\mathbf{r}}$, where $\mathbf{r} = x\mathbf{i}+y\mathbf{j}$, over the closed curve shown in Figure 18.4.2. We have $a<b$ and the path goes from $A=(a,0)$ to $B=(b,0)$ to $C=(0,b)$ to $D=(0,a)$ and ends at $A=(a,0)$.

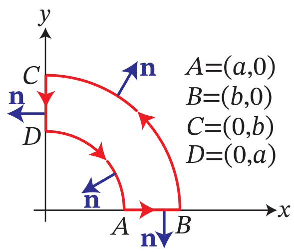


Figure 18.4.2

SOLUTION On the paths from A to B and from C to D the exterior normal, $\mathbf{n}$, is perpendicular to $\mathbf{F}$, so $\mathbf{F}\cdot\mathbf{n}=0$, and these integrands contribute nothing to the integral.

On the arc BC, $\mathbf{F}$ equals $f(b)\widehat{\mathbf{r}}$. There $\widehat{\mathbf{r}}=\mathbf{n}$, so $\mathbf{F}\cdot\mathbf{n}=f(b)$ since $\mathbf{r}\cdot\mathbf{n}=1$. The length of arc BC is $(2\pi b)/4=\pi b/2$. Thus

$$\int_B^C \mathbf{F}\cdot\mathbf{n}\,ds=\int_B^C f(b)\,ds=f(b)\int_B^C ds=\frac{\pi}{2}bf(b).$$

On the arc DA, $\widehat{\mathbf{r}}=-\mathbf{n}$. A similar calculation shows that

$$\int_D^C \mathbf{F}\cdot\mathbf{n}\,ds=-\frac{\pi}{2}af(a).$$

Combining these results we conclude that

$$\oint_{ABCDA}\mathbf{F}\cdot\mathbf{n}\,ds=0+\frac{\pi}{2}bf(b)+0-\frac{\pi}{2}af(a)=\frac{\pi}{2}(bf(b)-af(a)).$$

♦

A central field $f(r)\widehat{\mathbf{r}}$ has zero flux across all paths of the special type shown in Figure 18.4.2 only when

$$bf(b)-af(a)=0$$

for positive a and b. In particular, with $a=1$ and $b=r$,

$$rf(r)-1f(1)=0 \qquad \text{or} \qquad f(r)=\frac{f(1)}{r}.$$

Thus $f(r)$ is inversely proportional to r and there is a constant c such that

$$f(r)=\frac{c}{r}.$$

If $f(r)$ is not of the form c/r, the vector field $\mathbf{F}(x,y)=f(r)\widehat{\mathbf{r}}$ does not have zero flux across these paths. In Exercise 5 you compute the divergence of $(c/r)\widehat{\mathbf{r}}$ and show that it is zero.

Observation 18.4.1: *Incompressible Central Fields in the Plane*

The only divergence-free vector fields in the plane are those in which the magnitude is inversely proportional to the distance from their pole:

$$\mathbf{F}(\mathbf{r})=\frac{c}{r}\widehat{\mathbf{r}}.$$

RECALL: An incompressible, or divergence-free, vector field is one whose divergence is zero.

Knowing that the central field $\mathbf{F}=\widehat{\mathbf{r}}/r$ has zero divergence helps us to evaluate some line integrals of the form $\oint_C(\widehat{\mathbf{r}}\cdot\mathbf{n})/r\,ds$, as the next example shows.

EXAMPLE 2. Let $\mathbf{F}(\mathbf{r}) = \frac{\widehat{\mathbf{r}}}{r}$. Evaluate $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$ where C is the counterclockwise circle of radius 1 with center $(2,0)$, as shown in Figure 18.4.3.

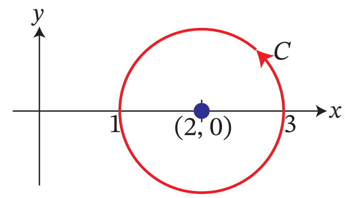


Figure 18.4.3

SOLUTION By Green's theorem, the line integral of the normal component of $\mathscr{F}$ along C equals the integral of the divergence of $\mathscr{F}$ over $\mathscr{R}$:

$$\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \int_{\mathscr{R}} \nabla\cdot\mathbf{F}\,dA. \tag{18.4.2}$$

Exercise 5 shows that the field $\mathbf{F}$ has divergence zero throughout the region $\mathscr{R}$ that C bounds. As a result, the right side of (18.4.2) is 0. Therefore $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = 0$, which is consistent with Observation 18.4.1. ♦

The next example involves a curve that surrounds a point where the vector field $\mathbf{F} = \widehat{\mathbf{r}}/r$ is not defined.

EXAMPLE 3. Let C be a simple closed curve enclosing the origin. Evaluate $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$, where $\mathbf{F} = \widehat{\mathbf{r}}/r$.

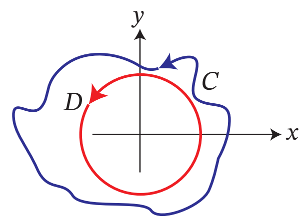


Figure 18.4.4

SOLUTION Figure 18.4.4 shows C and a small circle D centered at the origin and in the region that C bounds. Without a formula for C, we can not compute $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$ directly. However, since the divergence of $\mathbf{F}$ is 0 throughout the region bounded by C and D, we have, by the two-curve case of Green's theorem, Corollary 18.2.5,

$$\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \oint_D \mathbf{F}\cdot\mathbf{n}\,ds. \tag{18.4.3}$$

The integral on the right-hand side of (18.4.3) can be computed directly. To do so, let the radius of D be a. Then for points P on D, $\mathbf{F}(P) = \widehat{\mathbf{r}}/a$. Because $\widehat{\mathbf{r}}$ and $\mathbf{n}$ are the same unit vector, $\widehat{\mathbf{r}}\cdot\mathbf{n} = 1$. Thus

$$\oint_D \mathbf{F}\cdot\mathbf{n}\,ds = \oint_D \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{a}\,ds = \oint_D \frac{1}{a}\,ds = \frac{1}{a}2\pi a = 2\pi.$$

Hence $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = 2\pi$. ♦

The conclusion to Example 3 should not be a surprise, for the integral equals the number of radians that C subtends, as was shown in Section 15.4.

Central Vector Fields in Space

A central vector field in space with center at the origin has the form $\mathbf{F}(x,y,z) = \mathbf{F}(\mathbf{r}) = f(r)\widehat{\mathbf{r}}$. We show that if the flux of $\mathbf{F}$ over surfaces described below is zero then $f(r)$ must be inversely proportional to the square of r.

The surface $\mathscr{S}$ shown in Figure 18.4.5 consists of the parts of two spherical surfaces, one of radius a, the other of radius b, $a < b$, located in the octant where the coordinates are all positive, together with parts of the coordinate planes between the two spheres.

Let $\mathscr{R}$ be the region bounded by $\mathscr{S}$. On its three flat sides $\mathbf{F}$ is perpendicular to the exterior normal. On the outer sphere $\mathbf{F}(x,y,z)\cdot\mathbf{n} = f(b)$. On the inner sphere $\mathbf{F}(x,y,z)\cdot\mathbf{n} = -f(a)$. Thus, since the surface area of a sphere of radius r is $4\pi r^2$, the flux of $\mathbf{F}$ over the surface $\mathscr{S}$ is

$$\begin{aligned}\oint_{\mathscr{S}} \mathbf{F}\cdot\mathbf{n}\,dS &= f(b)\left(\frac{1}{8}\right)(4\pi b^2) - f(a)\left(\frac{1}{8}\right)(4\pi a^2)\\ &= \frac{\pi}{2}\left(f(b)b^2 - f(a)a^2\right).\end{aligned}$$

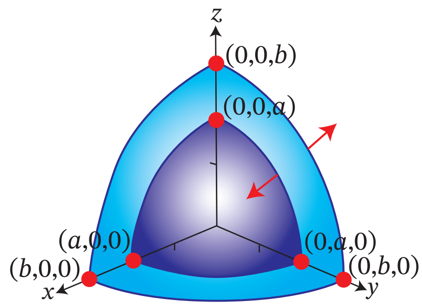


Figure 18.4.5

So, $f(b)b^2 - f(a)a^2 = 0$ for all positive value of a and b with $b > a$.

Similar to what we did in Example 1 to find divergence-free central vector fields in the plane, consider the situation with $b = r$ and $a = 1$: $f(r)\,r^2 - f(1)1^2$. It follows that there is a constant c such that

$$f(r) = \frac{c}{r^2}.$$

The magnitude must be proportional to the inverse square of r.

Compare with Example 1. See Exercise 16.

Observation 18.4.2: *Incompressible Central Fields in Space*

The only central vector fields in space with zero divergence are those in which the magnitude is inversely proportional to the square of the distance from their pole:

$$\mathbf{F}(\mathbf{r}) = \frac{c}{r^2}\widehat{\mathbf{r}}.$$

In physics books the integral $\int_{\mathscr{S}} \widehat{\mathbf{r}}\cdot\mathbf{n}/r^2\,dS$ is also written as

$$\int_{\mathscr{S}} \frac{\mathbf{r}\cdot\mathbf{n}}{r^3}\,dS, \quad \int_{\mathscr{S}} \frac{\widehat{\mathbf{r}}\cdot d\mathbf{S}}{r^2}, \quad \int_{\mathscr{S}} \frac{\mathbf{r}\cdot d\mathbf{S}}{r^3}, \quad \text{or} \quad \int_{\mathscr{S}} \frac{\cos(\mathbf{r},\mathbf{n})}{r^2}\,dS.$$

NOTATION: Recall, from Section 14.2, that $\cos(\mathbf{r},\mathbf{n})$ denotes the cosine of the angle between $\mathbf{r}$ and $\mathbf{n}$.

The symbol $d\mathbf{S}$ is introduced to represent $\mathbf{n}\,dS$. Figure 18.4.6 shows a small patch on the surface, dS together with an exterior normal unit vector $\mathbf{n}$.

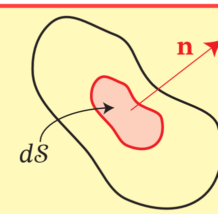


Figure 18.4.6

Summary

This section investigated central vector fields: $\mathbf{F}(P) = f(r)\widehat{\mathbf{r}}$. Central vector fields are radially symmetric (for every point P, the vector $\mathbf{F}(P)$ points towards or away from the pole of the vector field. The function $f(r)$ is the magnitude of the vector field.

In the plane the only divergence-free central fields are of the form $(c/r)\widehat{\mathbf{r}}$ where c is a constant, and are called inverse-first-power fields.

In space the only incompressible central fields are of the form $(c/r^2)\widehat{\mathbf{r}}$, an inverse-square field.

EXERCISES for Section 18.4

1. Define a central field in words, using no symbols.

2. Define a central field with center at $\mathcal{O}$ in symbols.

3. Give an example of a nonzero central field in the plane that (a) is compressible and (b) has zero divergence.

4. Give an example of a nonzero central field in space that (a) has nonzero divergence and (b) is divergence-free.

5. Define $\mathbf{F}(x, y)$ to be an inverse-first-power central field in the plane, that is $\mathbf{F}(x, y) = \frac{c}{r}\widehat{\mathbf{r}}$, where $\mathbf{r} = x\mathbf{i} + y\mathbf{j}$. Compute the divergence of $\mathbf{F}$.

6. Define $\mathbf{F}(\mathbf{r}) = \frac{\widehat{\mathbf{r}}}{r}$. Evaluate the flux $\oint\limits_C \mathbf{F}\cdot\mathbf{n}\,ds$ as simply as you can for the ellipses in Figure 18.4.7.

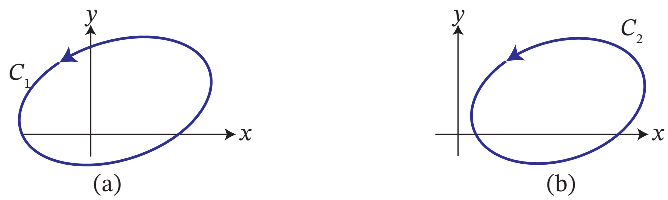


Figure 18.4.7

7. Figure 18.4.8 shows a cube with one corner at the origin and edges along the positive axes. Use steradian measure of a solid angle to evaluate the integral of $\frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r^2}$ over (a) the square $EFGH$, (b) the square $ABCD$ (see Exercise 8), and (c) the entire surface of the cube.

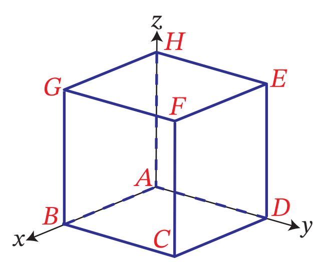


Figure 18.4.8

8. (a) Why is the integral in Exercise 7(b) improper?
(b) How could one express the integral in Exercise 7(b) as a limit of multiple integrals that are not improper?
(c) Use (b) to evaluate the integral in Exercise 7(b).

9. Evaluate the flux of $\mathbf{F}(\mathbf{r}) = \frac{\widehat{\mathbf{r}}}{r^3}$ over the sphere of radius 2 with center at the origin.

Steradians were introduced in Section 17.7. The next few problems involve steradians, including a few problems where you are asked to complete a computation both with and without using steradians.

10. In Example 2, where C is the counterclockwise circle of radius 1 with center $(2,0)$, the integral $\oint\limits_C \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r}\,ds$ is 0. How would you explain this in terms of subtended angles and steradians?

11. A rectangular solid is bounded by the planes $x = \pm a$, $y = \pm b$, and $z = \pm c$. Evaluate $\int\limits_{\mathcal{S}} \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r^2}\,d\mathcal{S}$ (a) by interpreting it in terms of the size of a subtended solid angle and (b) directly, without using steradians.

12. Consider the surface integral $\int\limits_{\mathcal{S}} \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r^2}\,dS$ where $\mathcal{S}$ is the triangle whose vertices are $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$. Evaluate $\int\limits_{\mathcal{S}} \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r^2}\,dS$ (a) by using steradians and (b) directly, without using steradians.

---

13. Define $\mathbf{F}(x,y,z) = \dfrac{x\mathbf{i}+y\mathbf{j}+0\mathbf{k}}{x^2+y^2}$. Note that $\mathbf{F}$ is a vector field in space. (a) What is the domain of $\mathbf{F}$? (b) Sketch $\mathbf{F}(1,1,0)$ and $\mathbf{F}(1,1,2)$ with tails at the given points. (c) Show $\mathbf{F}$ is not a central field. (d) Show its divergence is 0.

Exercises 14 to 18 reinforce and extend our understanding about central fields in different dimensions.

14. Consider a planar central field, $\mathbf{F}(x,y) = \dfrac{g\left(\sqrt{x^2+y^2}\right)}{\sqrt{x^2+y^2}}\left(x\mathbf{i}+y\mathbf{j}\right)$ for some scalar function g. Show that $\nabla\times\mathbf{F}$ is $\mathbf{0}$.

15. (This continues Exercise 14.) Show that a planar central field $\mathbf{F}$ is a gradient field; to be specific, $\mathbf{F} = \nabla G\left(\sqrt{x^2+y^2}\right)$ where G is an antiderivative of g.

16. Carry out the computation to show that the only central fields in space that have zero divergence have the form $\mathbf{F}(\mathbf{r}) = \dfrac{c}{r^2}\widehat{\mathbf{r}}$ if the origin of the coordinates is at the center of the field.

17. Show that the curl of a central vector field in space is $\mathbf{0}$.

18. If we worked in four-dimensional space instead of the two-dimensional plane or three-dimensional space, which central fields do you think would have zero divergence? Carry out a calculation to confirm your conjecture.

---

19. Define $\mathbf{F} = \dfrac{\widehat{\mathbf{r}}}{r^2}$ and S to be the surface of the lopsided pyramid with square base, whose vertices are $(0,0,0)$, $(1,1,0)$, $(0,1,0)$, $(0,1,1)$, and $(1,1,1)$. (a) Sketch the pyramid. (b) What is the integral of $\mathbf{F}\cdot\mathbf{n}$ over the square base? (c) What is the integral of $\mathbf{F}\cdot\mathbf{n}$ over each of the remaining four faces? (d) Evaluate $\displaystyle\oint_S \mathbf{F}\cdot\mathbf{n}\,dS$.

Exercises 20 and 21 are related.

20. Assume C is the circle $x^2+y^2=4$ in the xy-plane. For each point Q in the plane consider the central field with center Q, $\mathbf{F}(P) = \dfrac{\overrightarrow{PQ}}{|PQ|^2}$. Its magnitude is inversely proportional to the first power of the distance from P to Q. Evaluate the flux of $\mathbf{F}$ across C when (a) $Q=(0,0)$, the origin, (b) $Q=(1,0)$, (c) Q is a point on C, and (d) $Q=(3,5)$.

21. Assume $\mathscr{S}$ is the sphere $x^2+y^2+z^2=4$ in space. For each point Q in space consider the central field with center Q, $\mathbf{F}(P) = \dfrac{\overrightarrow{PQ}}{|PQ|^3}$. Its magnitude is inversely proportional to the square of the distance from P to Q. Evaluate the flux of $\mathbf{F}$ across $\mathscr{S}$ when (a) $Q=(0,0,0)$, the origin, (b) $Q=(1,0,0)$, (c) Q is a point on $\mathscr{S}$, and (d) $Q=(3,5,7)$.

---

22. Assume C is the circle of radius 2 and center at $(0,0)$ and $\mathbf{F}$ is the central field in the plane with center at $(1,0)$ and magnitude inversely proportional to the first power of the distance to $(1,0)$: $\mathbf{F}(x,y) = \dfrac{(x-1)\mathbf{i}+y\mathbf{j}}{|(x-1)\mathbf{i}+y\mathbf{j}|^2}$.

(a) By thinking in terms of subtended angle, evaluate the flux across C: $\displaystyle\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$.

(b) Evaluate the flux across C by carrying out the integration directly.

23. This exercise gives a geometric way to see why a central force is conservative. Consider $\mathbf{F}(x, y) = f(r)\hat{\mathbf{r}}$. Figure 18.4.9 shows $\mathbf{F}(x, y)$ and a short vector $\overrightarrow{d\mathbf{r}}$ and two circles.

(a) Why is $\mathbf{F}(x, y)\cdot d\mathbf{r}$ approximately $f(r)\ dr$, where dr is the difference in the radii of the two circles?

(b) Assume C is a curve from A to B, where $A = (a, \alpha)$ and $B = (b, \beta)$ in polar coordinates. Why is $\int\limits_C \mathbf{F}\cdot d\mathbf{r} = \int\limits_a^b f(r)\ dr$?

(c) Why is $\mathbf{F}$ conservative?

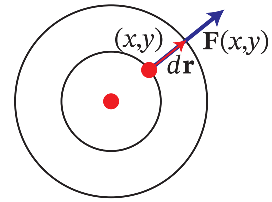


Figure 18.4.9

Radiation, light, or sound radiates uniformly in all directions from a point source at a distance h from a plane. Half of the signal hits the plane. In applications it is important to know what fraction of that half hits a disk located in the plane. There are tables that list that fraction as a function of h and the distance the center of the disk is from the point in the plane closest to the source (with the radius of the disk taken as 1).

Exercises 24 to 27 concern the special case where the center of the disk is on the line through the source perpendicular to the plane of the disk.

24. Consider the disk with radius a. Define $s(a, h)$ to be the steradian measure of the angle subtended by the disk. Explain why the fraction of interest equals $\frac{s(a,h)}{2\pi}$.

25. Recall, from Exercise 32 in Section 7.4, that the surface area of the portion of a sphere between two parallel planes that intersect the sphere is proportional to the distance between the planes. Use this information to show that $s(a, h) = 2\pi\left(1 - \frac{h}{\sqrt{a^2+h^2}}\right)$.

26. Use the integral for the flux of the field $\frac{\hat{\mathbf{r}}}{r^2}$ to find $s(a, h)$.

27. For $h = 0.8$ and $a = 1$ a table lists $s(a, h) = 2.35811$.

(a) Does that agree with the formula obtained in Exercise 25?

(b) What fraction of the radiation that strikes the plane hits the disk?

(c) The table lists only disks of radius 1. How would you use the table if $h = 3$ and $a = 2$?

---

28. Show that the derivative of $\frac{1}{3}\tan^3(x) - \tan(x) + x$ is $\tan^4(x)$.

29. Use integration by parts to show that $\int \tan^n(x)\ dx = \frac{\tan^{n-1}(x)}{n-1} - \int \tan^{n-2}(x)\ dx$.

30. Formula 32 in the Table of Integrals (in Appendix A) is $\int \frac{dx}{x(ax+b)} = \frac{1}{b}\ln\left|\frac{x}{ax+b}\right|$.

(a) Use a partial fraction expansion to evaluate the antiderivative.

(b) Use differentiation to check that the formula is correct.

31. Repeat Exercise 30 for formula 33 in the table of antiderivatives: $\int \frac{dx}{x^2(ax+b)} = \frac{-1}{bx} + \frac{a}{b^2}\ln\left|\frac{ax+b}{x}\right|$.

32. Show that $x\arccos(x) - \sqrt{1-x^2}$ is an antiderivative of $\arccos(x)$.

33. Find $\int \arctan(x)\ dx$.

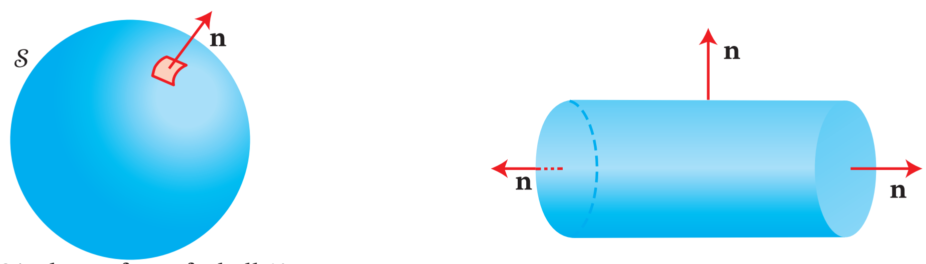
34. (a) Find $\int xe^{ax}\ dx$., (b) Use integration by parts to show that $\int x^m e^{ax}\ dx = \frac{x^m e^{ax}}{a} - \frac{m}{a}\int x^{m-1}e^{ax}\ dx$., and (c) Verify the equation in (b) by differentiating the right-hand side..

18.5 Divergence Theorem in Space (Gauss's Theorem)

In Sections 18.2 and 18.3 we developed Green's theorem and applied it in two forms for a vector field $\mathbf{F}$ in the plane. One concerned the line integral of the tangential component of $\mathbf{F}$, $\oint_C \mathbf{F}\cdot\mathbf{T}\,ds$, also written as $\oint_C \mathbf{F}\cdot d\mathbf{r}$. The other concerned the integral of the normal component of $\mathbf{F}$, $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$. In this section we develop the **divergence theorem**, an extension of the second form from the plane to space. Another name for the divergence theorem is **Gauss's Theorem**. The extension of the first form to space is the subject of Section 18.6. In Section 18.7 the divergence theorem will be applied to electromagnetism.

The Divergence Theorem

A region $\mathcal{V}$ in space is bounded by a surface $\mathcal{S}$. For instance, for many problems in electromagnetism $\mathcal{V}$ might be a ball and $\mathcal{S}$ its surface, a sphere (see Figure 18.5.1(a)). In other cases, $\mathcal{V}$ is a right circular cylinder and $\mathcal{S}$ is its surface, which consists of two disks and its curved side (see Figure 18.5.1(b)).



$\mathcal{S}$ is the surface of a ball $\mathcal{V}$
(a)

(b)

Figure 18.5.1

Both figures show unit exterior normal vectors, that is, they are perpendicular to the surface. The divergence theorem relates an integral over the surface to an integral over the region it bounds. It is assumed that all surfaces of interest have a continuous exterior normal vector (such as a sphere) or are made up of a finite number of such surfaces (such as the surface of a cylinder or cube). That there are curves or points where the exterior normal is not defined complicates some of the proofs, but does not affect the validity of the basic result.

Theorem 18.5.1: Divergence Theorem — One-Surface Case

Let $\mathcal{V}$ be the region in space bounded by the surface $\mathcal{S}$. Let $\mathbf{n}$ denote the exterior unit normal vector of $\mathcal{V}$ along the boundary $\mathcal{S}$. Then, for any vector field $\mathbf{F}$ defined on $\mathcal{V}$,

$$\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\,dS = \int_{\mathcal{V}} \nabla\cdot\mathbf{F}\,dV.$$

In words, the integral of the normal component of $\mathbf{F}$ over a surface equals the integral of the divergence of $\mathbf{F}$ over the region the surface bounds.

The integral $\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\,dS$ is called the **flux** of the field $\mathbf{F}$ across the surface $\mathcal{S}$.

Direction cosines were defined in Section 14.4.

If $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ and $\cos(\alpha)$, $\cos(\beta)$, and $\cos(\gamma)$ are the direction cosines of the exterior normal vector, then the divergence theorem reads

$$\int_{\mathcal{S}} (P\mathbf{i}+Q\mathbf{j}+R\mathbf{k})\cdot\left(\cos(\alpha)\mathbf{i}+\cos(\beta)\mathbf{j}+\cos(\gamma)\mathbf{k}\right)\,dS = \int_{\mathcal{V}} \left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}\right)dV.$$

Evaluating the dot product puts the divergence theorem in the form

$$\int_{\mathcal{S}} (P\cos(\alpha)+Q\cos(\beta)+R\cos(\gamma))\,dS = \int_{\mathcal{V}} \left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}\right)dV.$$

When the divergence theorem is expressed in this form, we see that it amounts to three scalar theorems:

$$\int_{\mathscr{S}} P\cos(\alpha)\,dS = \int_{\mathscr{V}} \frac{\partial P}{\partial x}\,dV, \quad \int_{\mathscr{S}} Q\cos(\beta)\,dS = \int_{\mathscr{V}} \frac{\partial Q}{\partial y}\,dV, \quad \text{and} \quad \int_{\mathscr{S}} R\cos(\gamma)\,dS = \int_{\mathscr{V}} \frac{\partial R}{\partial z}\,dV. \tag{18.5.1}$$

Establishing these equations will prove the divergence theorem. We delay the proof to the end of this section, after we have shown how the divergence theorem is applied.

The divergence theorem also is the basis for this coordinate-free description of divergence. Its motivation is like the motivation for its analog in the plane (see the local definition of div $\mathbf{F}(P)$ on page 1042).

Definition: *Local Definition of* div $\mathbf{F}(P)$

The divergence of a vector field $\mathbf{F}$ at a point P in the domain of $\mathbf{F}$ is

$$\operatorname{div}\mathbf{F}(P) = \lim_{\mathscr{V}\to P} \frac{1}{\text{Volume of } \mathscr{V}} \int_{\mathscr{S}} \mathbf{F}\cdot\mathbf{n}\,d\mathscr{S},$$

where each $\mathscr{V}$ is a region enclosing the point P and whose boundary is the closed surface $\mathscr{S}$.

Two-Surface Version of the Divergence Theorem

The divergence theorem also holds if the solid region has holes in it. Then the boundary consists of several separate closed surfaces. The most important case is when there is just one hole and hence an inner surface $\mathscr{S}_1$ and an outer surface $\mathscr{S}_2$, as shown in Figure 18.5.2.

Theorem 18.5.2: Divergence Theorem — Two-Surface Case

If $\mathscr{V}$ is a region in space bounded by the surfaces $\mathscr{S}_1$ and $\mathscr{S}_2$, $\mathbf{n}$ denote the exterior normal vector along the boundary, and $\mathbf{F}$ is a vector field defined on $\mathscr{V}$, then

$$\int_{\mathscr{S}_1} \mathbf{F}\cdot\mathbf{n}\,dS + \int_{\mathscr{S}_2} \mathbf{F}\cdot\mathbf{n}\,dS = \int_{\mathscr{V}} \operatorname{div}\mathbf{F}\,dV.$$

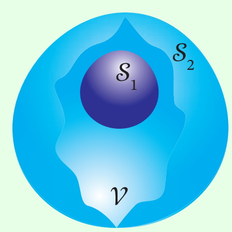


Figure 18.5.2

EMPHASIS: *The vectors* $\mathbf{n}$ *in Theorem 18.5.2 are unit normal vectors that point into the* exterior *of* $\mathscr{V}$.

The importance of this form of the divergence theorem is that it allows us to conclude that if div $\mathbf{F}$ is 0 on the solid region $\mathscr{V}$ the two surfaces bound, then the fluxes across the surfaces are the same.

Corollary 18.5.3: Changing the Surface when div $\mathbf{F} = 0$

Let $\mathscr{S}_1$ and $\mathscr{S}_2$ be two closed surfaces that form the boundary of the region $\mathscr{V}$. Let $\mathbf{F}$ be a vector field defined on $\mathscr{V}$ such that the divergence of $\mathbf{F}$ is 0 throughout $\mathscr{V}$. Then

$$\int_{\mathscr{S}_1} \mathbf{F}\cdot\mathbf{n}\,dS = \int_{\mathscr{S}_2} \mathbf{F}\cdot\mathbf{n}\,dS \tag{18.5.2}$$

where $\mathbf{n}$ *is the unit outward normal vector for each surface.*

REFLECTION: Take a minute to compare Corollaries 18.2.5 and 18.5.3. While the former concerns circulation around a closed curve and the latter concerns flux across a closed surface, they are similar in that they identify conditions where the curve or surface can be changed without changing the value of the corresponding integral. (See also Example 3) in Section 18.2.)

The proof of Theorem 18.5.2 is omitted because it does not add anything more to our understanding of the divergence theorem. As a first application of Corollary 18.5.3, Example 1 illustrates how, since the divergence of $\mathbf{F}$ is 0, the integral of $\mathbf{F}\cdot\mathbf{n}$ over one surface can be replaced by an integral of $\mathbf{F}\cdot\mathbf{n}$ over a more convenient surface.

EXAMPLE 1. Consider $\mathbf{F}(\mathbf{r}) = \dfrac{\widehat{\mathbf{r}}}{r^2}$, the inverse-square vector field with center at the origin. Let $\mathcal{S}$ be a convex surface that encloses the origin. Find $\displaystyle\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\, dS$, the flux of $\mathbf{F}$ over the surface $\mathcal{S}$.

SOLUTION Recall from Observation 18.4.2 that $\nabla\cdot\mathbf{F} = 0$ for all $\mathbf{r} \neq \mathbf{0}$. (See also Exercise 11). So Corollary 18.5.3 applies and we can use (18.5.2). Choose $\mathcal{S}_1 = \mathcal{S}$ and $\mathcal{S}_2$ to be a sphere with center at the origin and radius a small enough to ensure $\mathcal{S}_2$ is completely enclosed by $\mathcal{S}$. Then, since $\nabla\cdot\mathbf{F} = 0$ in the region between $\mathcal{S}_1$ and $\mathcal{S}_2$, (18.5.2) gives

$$\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\, dS = \int_{\mathcal{S}_2} \mathbf{F}\cdot\mathbf{n}\, dS.$$

Because $\left(\widehat{\mathbf{r}}/r^2\right)\cdot\mathbf{n} = (\widehat{\mathbf{r}}\cdot\widehat{\mathbf{r}})/r^2 = 1/a^2$ on the sphere $\mathcal{S}_2$, we have

$$\int_{\mathcal{S}_2} \mathbf{F}\cdot\mathbf{n}\, dS = \int_{\mathcal{S}_2} \frac{1}{a^2}\, dS = \frac{1}{a^2}\int_{\mathcal{S}_1} dS = \frac{1}{a^2}4\pi a^2 = 4\pi.$$

Thus $\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\, dS$ evaluates to 4π. ◆

Example 1 agrees with the fact, first observed in Section 17.7, that a convex surface that encloses the origin subtends an angle of 4π steradians at any point in the region it bounds.

A **uniform** or **constant** vector field is a vector field whose vectors are identical. We use one in the next example.

EXAMPLE 2. Verify the divergence theorem for the constant field $\mathbf{F}(x, y, z) = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$ and the surface $\mathcal{S}$ of the cube whose sides have length 5 in Figure 18.5.3.

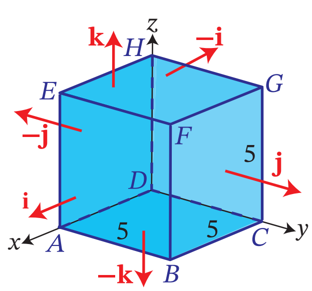


Figure 18.5.3

SOLUTION To find $\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\, dS$ we consider the integral of $\mathbf{F}\cdot\mathbf{n}$ over each of its six faces. (Observe that each face is a square with side length 5, so its area is 25.)

On the bottom face, $ABCD$, the unit exterior normal vector is $-\mathbf{k}$. Thus

$$\mathbf{F}\cdot\mathbf{n} = (2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k})\cdot(-\mathbf{k}) = -4.$$

So

$$\int_{ABCD} \mathbf{F}\cdot\mathbf{n}\, dS = \int_{ABCD} (-4)\, dS = -4\int_{ABCD} dS = (-4)(25) = -100.$$

The integral over the top face involves the exterior unit normal vector $\mathbf{k}$ instead of $-\mathbf{k}$. Then $\int_{EFGH} \mathbf{F}\cdot\mathbf{n}\, dS = 100$. The sum of the two integrals is 0. Similar computations show that the flux of $\mathbf{F}$ over the entire surface is 0.

The divergence theorem says that the flux equals $\int_{\mathcal{V}} \operatorname{div}\mathbf{F}\, dV$, where $\mathcal{V}$ is the solid cube. Now, $\operatorname{div}\mathbf{F} = \partial(2)/\partial x + \partial(3)/\partial y + \partial(4)/\partial z = 0 + 0 + 0 = 0$. So the integral of $\operatorname{div}\mathbf{F}$ over $\mathcal{V}$ is 0, verifying the divergence theorem. ◆

Example 2 actually shows the flux of any uniform vector field across a closed surface is zero. The proof of this statement is merely a simple application of Theorem 18.5.1. Once you believe $\operatorname{div}\mathbf{F} = 0$, there is no need to even attempt to parameterize the surface $\mathcal{S}$.

Why $\operatorname{div}\mathbf{F}$ is Called the Divergence

Let $\mathbf{F}(x, y, z)$ be the vector field describing the flow for a gas. That is, $\mathbf{F}(x, y, z)$ is the product of the density of the gas at (x, y, z) and the velocity vector of the gas there.

The integral $\int_{\mathscr{S}} \mathbf{F}\cdot\mathbf{n}\, dS$ over a closed surface $\mathscr{S}$ represents the tendency of the gas to leave the region $\mathcal{V}$ that $\mathscr{S}$ bounds. If the flux is positive the gas is tending to escape or diverge. If negative, the effect is for the amount of gas in $\mathcal{V}$ to increase and become denser.

Let $\rho(x, y, z, t)$ be the density of the gas at time t at the point (x, y, z), with units mass per unit volume. Then $\int_{\mathcal{V}} \rho\ dV$ is the total mass of gas in $\mathcal{V}$ at a given time. The rate at which the mass in $\mathcal{V}$ changes is given by the derivative

$$\frac{d}{dt}\int\limits_{\mathcal{V}} \rho\ dV.$$

When this derivative is positive the amount of gas in $\mathcal{V}$ is increasing.

If ρ is sufficiently well-behaved we may differentiate past the integral sign. Therefore

$$\frac{d}{dt}\int\limits_{\mathcal{V}} \rho\ dV = \int\limits_{\mathcal{V}} \frac{\partial\rho}{\partial t}\ dV$$

represents the rate at which the amount of gas in $\mathcal{V}$ changes. Because $\int_{\mathscr{S}} \mathbf{F}\cdot\mathbf{n}\ dS$ is the rate at which gas escapes from $\mathcal{V}$,

$$\int\limits_{\mathscr{S}} \mathbf{F}\cdot\mathbf{n}\ dS = -\int\limits_{\mathcal{V}} \frac{\partial\rho}{\partial t}\ dV. \tag{18.5.3}$$

But by the divergence theorem, $\int_{\mathscr{S}} \mathbf{F}\cdot\mathbf{n}\ dS = \int_{\mathcal{V}} \nabla\cdot\mathbf{F}\ dV$, and so

$$\int\limits_{\mathcal{V}} \nabla\cdot\mathbf{F}\ dV = -\int\limits_{\mathcal{V}} \frac{\partial\rho}{\partial t}\ dV$$

or

$$\int\limits_{\mathcal{V}} \left(\nabla\cdot\mathbf{F} + \frac{\partial\rho}{\partial t}\right)\ dV = 0. \tag{18.5.4}$$

From this it is possible to conclude that $\nabla\cdot\mathbf{F} + \partial\rho/\partial t = 0$. Justification for this conclusion, known as the **equation of continuity**, is given next.

Recall from Section 6.3 that the zero-integral principle says: If a continuous function f on an interval $[a, b]$ has the property that $\int_c^d f(x)\ dx = 0$ for every subinterval $[c, d]$ then $f(x) = 0$ on $[a, b]$. The following extension of the zero-integral principle to three-dimensional space is proved in Exercise 28.

Theorem 18.5.4: Zero-Integral Principle in Space

Let $\mathcal{V}$ be a region in space bounded by a surface, and let f be a continuous function on $\mathcal{V}$. Assume that for every region $\mathscr{R}$ in $\mathcal{V}$, $\int_{\mathscr{R}} f(P)\ dS = 0$. Then $f(P) = 0$ for all P in $\mathcal{V}$.

Equation (18.5.4) holds not just for the solid $\mathcal{V}$ but for any solid region within $\mathcal{V}$. By the zero-integral principle in space, the integrand must be zero throughout $\mathcal{V}$, and we conclude that

$$\nabla\cdot\mathbf{F} = -\frac{\partial p}{\partial t}.$$

This tells us that $\operatorname{div}\mathbf{F}$ at a point P represents the rate gas is getting more dense (heavier) or less dense (lighter) at the point P. That is why $\operatorname{div}\mathbf{F}$ is called the divergence of $\mathbf{F}$. Where $\operatorname{div}\mathbf{F}$ is positive, the gas is dissipating (diverging). Where $\operatorname{div}\mathbf{F}$ is negative, the gas is collecting. Similar reasoning is the basis for referring to a vector field for which the divergence is zero as **incompressible**.

We conclude this section with a proof of the divergence theorem.

Proof of the Divergence Theorem (Theorem 18.5.1)

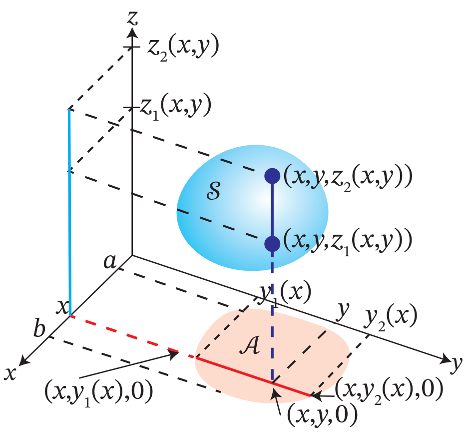


Figure 18.5.4

We prove Theorem 18.5.1 for the special case that each line parallel to an axis meets the surface $\mathscr{S}$ in at most two points and $\mathscr{S}$ encloses a solid region, $\mathcal{V}$, that is convex. We prove the third equation in (18.5.1).

We wish to show that

$$\int_{\mathscr{S}} R\cos(\gamma)\, dS = \int_{\mathcal{V}} \frac{\partial R}{\partial z}\, dV. \tag{18.5.5}$$

Let $\mathscr{A}$ be the projection of $\mathscr{S}$ on the xy-plane. Its description is

$$a \le x \le b, \qquad y_1(x) \le y \le y_2(x).$$

The description of $\mathcal{V}$ is then (see Figure 18.5.4)

$$a \le x \le b, \qquad y_1(x) \le y \le y_2(x), \qquad z_1(x,y) \le z \le z_2(x,y).$$

The right-hand side of (18.5.5) can be evaluated as follows:

$$\int_{\mathcal{V}} \frac{\partial R}{\partial z}\, dV = \int_a^b \int_{y_1(x)}^{y_2(x)} \int_{z_1(x,y)}^{z_2(x,y)} \frac{\partial R}{\partial z}\, dz\, dy\, dx.$$

The first integration gives, by the fundamental theorem of calculus,

$$\int_{z_1(x,y)}^{z_2(x,y)} \frac{\partial R}{\partial z}\, dz = R(x,y,z_2(x,y)) - R(x,y,z_1(x,y)).$$

We have, therefore,

$$\int_{\mathcal{V}} \frac{\partial R}{\partial z}\, dV = \int_a^b \int_{y_1(x)}^{y_2(x)} \big(R(x,y,z_2(x,y)) - R(x,y,z_1(x,y))\big)\, dy\, dx,$$

hence

$$\int_{\mathcal{V}} \frac{\partial R}{\partial z}\, dV = \int_{\mathscr{A}} \big(R(x,y,z_2(x,y)) - R(x,y,z_1(x,y))\big)\, dA. \tag{18.5.6}$$

Next, observe that the right-hand side of (18.5.6) is actually the difference of two integrals over $\mathscr{A}$. In fact, each of these double integrals is a surface integral over a different part of the surface $\mathscr{S}$.

First, $\int_{\mathscr{A}} R(x,y,z_2(x,y))\, dA$ involves the top part of $\mathscr{S}$, where $\cos(\gamma)$ is positive; call this part S_1. Then, recalling (17.7.1), we have

$$\int_{\mathscr{A}} R(x,y,z_2(x,y))\, dA = \int_{\mathscr{S}_1} R(x,y,z)\cos(\gamma)\, dS. \tag{18.5.7}$$

On the bottom part of $\mathscr{S}$, which we call $\mathscr{S}_2$, $\cos(\gamma)$ is negative because the angle between $\mathbf{k}$ and the outward unit normal vector is between $\pi/2$ and π. Thus

$$\int_{\mathscr{A}} R(x,y,z_1(x,y))\, dA = -\int_{\mathscr{S}_2} R(x,y,z)\cos(\gamma)\, dS. \tag{18.5.8}$$

The surfaces $\mathscr{S}_1$ and $\mathscr{S}_2$ do not quite cover all of $\mathscr{S}$. Let $\mathscr{S}_3$ denote the points on $\mathscr{S}$ where $\cos(\gamma) = 0$. Then

$$\int_{\mathscr{S}_3} R(x,y,z)\cos(\gamma)\, dS = 0. \tag{18.5.9}$$

Now, since every point on $\mathscr{S}$ is on exactly one of $\mathscr{S}_1$, $\mathscr{S}_2$, or $\mathscr{S}_3$,

$$\begin{aligned}\int_{\mathcal{V}} \frac{\partial R}{\partial z}\,dV &= \int_{\mathscr{A}} R(x,y,z_2(x,y))\,dA - \int_{\mathscr{A}} R(x,y,z_1(x,y))\,dA\\ &= \int_{\mathscr{S}_1} R(x,y,z)\cos(\gamma)\,dS + \int_{\mathscr{S}_2} R(x,y,z)\cos(\gamma)\,dS + \int_{\mathscr{S}_3} R(x,y,z)\cos(\gamma)\,dS\\ &= \int_{\mathscr{S}} R(x,y,z)\cos(\gamma)\,dS.\end{aligned}$$

This establishes (18.5.5). The remainder of the proof of the divergence theorem uses similar approaches to derive the other two parts of (18.5.1). (See Exercises 33 and 34.) •

Summary

We stated the divergence theorem for a single surface and for two surfaces. Also known as Gauss's theorem, it lets us calculate the flux of a vector field across a surface in terms of an integral of its divergence over the region bounded by the surface.

The divergence theorem is especially useful for fields that are incompressible (divergence-free), such as the inverse-square field in space, $\widehat{\mathbf{r}}/r^2$. The flux across a surface of such a field depends on whether its center is inside or outside the surface. If the center is at Q and the field is of the form $c\overrightarrow{QP}/|QP|^3$, its flux across a surface not enclosing Q is 0. If the surface encloses Q, its flux is $4\pi c$. This consequence of the divergence theorem can also be explained in terms of solid angles.

EXERCISES for Section 18.5

1. State the divergence theorem in symbols.
2. State the divergence theorem using only words.
3. Explain, using no symbols, the meaning of $\nabla\cdot\mathbf{F}$ at a point P. (Use the coordinate-free definition of $\nabla\cdot\mathbf{F}$.)
4. What is the two-surface version of Gauss's theorem?
5. Why is there a minus sign in (18.5.3)?

Exercises 6 to 6 ask you to verify the divergence theorem for a given vector field $\mathbf{F}$ and surface $\mathscr{S}$.

6. $\mathbf{F}(x,y,z) = x\mathbf{i} + y\mathbf{j} + 0\mathbf{k}$ and $\mathscr{S}_1$ is the sphere with center $(0,0,0)$ and radius 3.
7. $\mathbf{F}(x,y,z) = x\mathbf{i}$ and $\mathscr{S}$ is the cube with vertices $(0,0,0)$, $(2,0,0)$, $(2,2,0)$, $(0,2,0)$, $(0,0,2)$, $(2,0,2)$, $(2,2,2)$, and $(0,2,2)$.
8. $\mathbf{F} = 4\mathbf{i} - 3\mathbf{j} + \pi\mathbf{k}$ and $\mathscr{S}$ is the tetrahedron whose four vertices are $(0,0,0)$, $(1,0,0)$, and $(0,1,0)$, and $(0,0,1)$.
9. $\mathbf{F}(x,y,z) = (x^2+y^2+z^2)(x\mathbf{i}+y\mathbf{j}+z\mathbf{k})$ where the surfaces are spheres of radii 2 and 3, both centered at the origin.

---

10. Let $\mathbf{F} = 2x\mathbf{i} + 3y\mathbf{j} + (5z+6x)\mathbf{k}$, and let $\mathbf{G} = (2x+4z^2)\mathbf{i} + (3y+5x)\mathbf{j} + 5z\mathbf{k}$. Show that $\int_{\mathscr{S}} \mathbf{F}\cdot\mathbf{n}\,dS = \int_{\mathscr{S}} \mathbf{G}\cdot\mathbf{n}\,dS$, where $\mathscr{S}$ is a surface bounding a region in space.
11. Show that $\mathbf{F} = \dfrac{\widehat{\mathbf{r}}}{r^2}$, for all $\mathbf{r} \neq \mathbf{0}$, is an incompressible vector field.

In Exercises 12 to 19 use the divergence theorem to evaluate the given integrals.

12. Let $\mathcal{V}$ be the solid region bounded by the xy-plane and the paraboloid $z = 9 - x^2 - y^2$. Evaluate $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$ where $\mathbf{F} = y^3\mathbf{i} + z^3\mathbf{j} + x^3\mathbf{k}$ and $\mathcal{S}$ is the boundary of $\mathcal{V}$.

13. Evaluate $\int_{\mathcal{V}} \nabla \cdot \mathbf{F}\, dV$ for $\mathbf{F} = \sqrt{x^2 + y^2 + z^2}(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})$ and $\mathcal{V}$ the ball of radius 2 and center at $(0,0,0)$.

14. Evaluate $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$, where $\mathbf{F} = z\sqrt{x^2 + z^2}\mathbf{i} + (y+3)\mathbf{j} - x\sqrt{x^2 + z^2}\mathbf{k}$ and $\mathcal{S}$ is the boundary of the solid region between $z = x^2 + y^2$ and the plane $z = 4x$.

15. Evaluate $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$, where $\mathbf{F} = x\mathbf{i} + (3y + z)\mathbf{j} + (4x + 2z)\mathbf{k}$ and $\mathcal{S}$ is the surface of the cube bounded by the planes $x = 1$, $x = 3$, $y = 2$, $y = 4$, $z = 3$, and $z = 5$.

16. Evaluate $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$, where $\mathbf{F} = 4xz\mathbf{i} - y^2\mathbf{j} + yz\mathbf{k}$ and $\mathcal{S}$ is the surface of the cube bounded by the planes $x = 0$, $x = 1$, $y = 0$, $y = 1$, $z = 0$, and $z = 1$, with the face corresponding to $x = 1$ removed.

17. Evaluate $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$, where $\mathbf{F} = x\mathbf{i} + y\mathbf{j} + 2x\mathbf{k}$ and $\mathcal{S}$ is the boundary of the tetrahedron with vertices $(1,2,3)$, $(1,0,1)$ $(2,1,4)$, and $(1,3,5)$.

18. Let $\mathcal{S}$ be a surface of area S that bounds a region $\mathcal{V}$ of volume V. Assume that $\mathbf{F}$ is defined throughout $\mathcal{V}$ with $|\mathbf{F}(P)| \le 5$ for all points P on the surface $\mathcal{S}$. What can be said about $\int_{\mathcal{V}} \nabla \cdot \mathbf{F}\, dV$?

19. Evaluate $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$, where $\mathbf{F} = x^3\mathbf{i} + y^3\mathbf{j} + z^3\mathbf{k}$ and $\mathcal{S}$ is the sphere of radius a and center $(0,0,0)$.

In Exercises 20 to 23 evaluate $\int_{\mathcal{S}} \frac{\hat{\mathbf{r}}}{r^2} \cdot \mathbf{n}\, dS$.

20. $\mathcal{S}$ is the sphere of radius 2 and center $(5,3,1)$.

21. $\mathcal{S}$ is the sphere of radius 3 and center $(1,0,1)$.

22. $\mathcal{S}$ is the surface of the box bounded by the planes $x = -1$, $x = 2$, $y = 2$, $y = 3$, $z = -1$, and $z = 6$.

23. $\mathcal{S}$ is the surface of the box bounded by the planes $x = -1$, $x = 2$, $y = -1$, $y = 3$, $z = -1$, and $z = 4$.

---

24. Assume that the flux of $\mathbf{F}$ across every sphere is 0. Must the flux of $\mathbf{F}$ across the surface of every cube be 0 also?

25. If $\mathbf{F}$ is always tangent to a surface $\mathcal{S}$ what can be said about the integral of $\nabla \cdot \mathbf{F}$ over the region that $\mathcal{S}$ bounds?

26. Let $\mathbf{F}(\mathbf{r}) = f(r)\hat{\mathbf{r}}$ be a central vector field in space that has zero divergence. Show that $f(r)$ has the form $f(r) = a/r^2$ for some constant a. To do this, consider the flux of $\mathbf{F}$ across the surface of the region bounded by a cone with vertex at the origin and two spheres centered at the origin, see Figure 18.5.5.

27. Assume **F** is defined everywhere except at the origin and is divergence-free and $\mathscr{S}_1$ and $\mathscr{S}_2$ are two closed surfaces that enclose the origin. Explain why $\int_{\mathscr{S}_1} \mathbf{F}\cdot\mathbf{n}\, dS = \int_{\mathscr{S}_2} \mathbf{F}\cdot\mathbf{n}\, dS$. NOTE: The two surfaces may intersect.

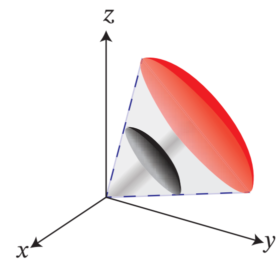


Figure 18.5.5

28. Provide the details for the proof of the zero-integral principle in space. Treat the assumptions $f(P) > 0$ and $f(P) < 0$ separately.

29. Show that the flux of an inverse-square central field $c\frac{\widehat{\mathbf{r}}}{r^2}$ across any closed surface that bounds a region that does not contain the origin is zero.

30. (a) Show that the proof in the text of the divergence theorem applies to a tetrahedron.

(b) Deduce that if the divergence theorem holds for a tetrahedron then it holds for any polyhedron.

NOTE: You can use without proof the fact that a polyhedron can be cut into tetrahedra.

31. Exercise 26 showed, by considering a particular type of surface, that the only central fields with zero divergence are the inverse-square fields. Show this, instead, by computing the divergence of $\mathbf{F}(x, y, z) = f(r)\widehat{\mathbf{r}}$, where $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ and $r = |\mathbf{r}|$.

32. Assume **F** is defined everywhere in space except at the origin, is divergence-free, and $\lim_{r\to\infty} r^2\mathbf{F}(\mathbf{r}) = \mathbf{0}$, where $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ and $r = |\mathbf{r}|$. What can be said about $\int_{\mathscr{S}} \mathbf{F}\cdot\mathbf{n}\, dS$, where $\mathscr{S}$ is the sphere of radius 2 centered at $(0,0,0)$?

We proved one part of the divergence theorem (Theorem 18.5.1). Exercises 33 and 34 concern the other two parts.

33. Prove that $\int_{\mathscr{S}} Q\cos(\beta)\, dS = \int_{\mathscr{V}} \frac{\partial Q}{\partial y}\, dV$.

34. Prove that $\int_{\mathscr{S}} P\cos(\alpha)\, dS = \int_{\mathscr{V}} \frac{\partial P}{\partial x}\, dV$.

18.6 Stokes' Theorem

In Section 18.2 we learned that Green's theorem can be written as

$$\oint_C \mathbf{F}\cdot d\mathbf{r} = \int_{\mathscr{R}} (\mathbf{curl}\,\mathbf{F})\cdot\mathbf{k}\, dA,$$

where C is traversed counterclockwise and bounds the region $\mathscr{R}$ in the xy-plane. Stokes' theorem in this section extends this to closed curves C in space that bound a surface $\mathscr{S}$, as in Figure 18.6.1. It asserts that

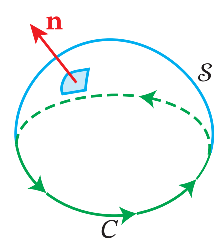


Figure 18.6.1

$$\oint_C \mathbf{F}\cdot d\mathbf{r} = \int_{\mathscr{S}} (\mathbf{curl}\,\mathbf{F})\cdot\mathbf{n}\, d\mathscr{S}$$

for any vector field **F** that has (piecewise) continuous first-order partial derivatives on an open set in space that contains the surface $\mathscr{S}$.

As usual, the vector **n** is a unit normal vector to the surface $\mathscr{S}$. However, there are two unit normal vectors at each point on the surface. We will soon describe how to decide which one to use, or if it is not possible to define **n** in this manner. When a suitable **n** can be identified, the choice depends on the orientation of C; such surfaces will be called **orientable**.

In words, Stokes' theorem says that the circulation of a vector field around a closed curve is equal to the integral of the normal component of the curl of the field over any orientable surface that the curve bounds. The key to finding a more precise statement of Stokes' theorem is coming up with a clear and consistent definition of the unit normal vector to a surface.

Choosing the Unit Normal Vector, n

In order to state Stokes' theorem precisely, we must describe what kind of surface $\mathcal{S}$ is permitted and which of the two normal vectors **n** to choose.

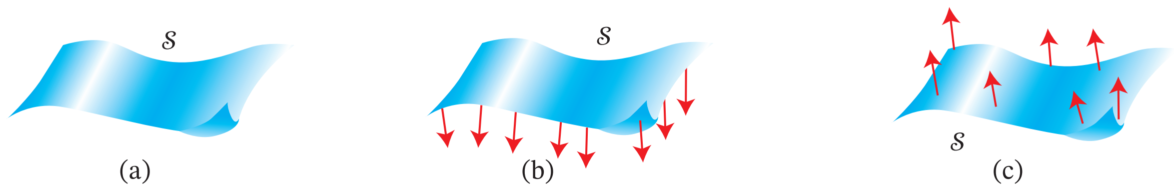


Figure 18.6.2

For the surfaces $\mathcal{S}$ that we consider it is possible to assign at each point on $\mathcal{S}$ a unit normal **n** in a (piecewise) continuous manner. On the surface shown in Figure 18.6.2(a), there are two ways to do this, shown in Figures 18.6.2(b) and (c). The piecewise continuous assumption means that Stokes' theorem applies to cubes, cylinders, and other surfaces with corners and smooth edges.

For the **Möbius band** shown in Figure 18.6.3 it is impossible to choose a continuously varying outward unit normal vector. Start at the cyan dot with the red normal vector marked with a "1" and move the normal vector continuously along the surface to choices 2, 3, and so on. Note that each normal vector is based on the curve that is in the middle of the band. One lap around the band is completed when you return to the initial point (cyan dot) on the surface. But, even though the normal vectors have been selected in a consistent (and continuous) manner, by the time you return to the initial (cyan) point on the surface, the normal vector identified as "9" is pointing in the opposite direction.

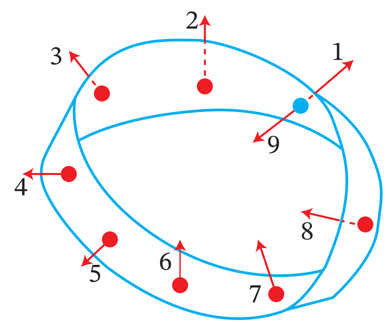


Figure 18.6.3

A smooth surface for which a continuous choice can be made is called **orientable** or **two-sided**. The Möbius band is not orientable because it is smooth but one-sided. The definition of orientable can be extended to nonsmooth surfaces, like the cube, that can be approximated by smooth, orientable surfaces. For our current purposes it is important to understand that Stokes' theorem holds for orientable surfaces, which include (because discontinuities between faces of a surface can be ignored), for instance, any part of the surface of a convex body, such as a ball, cube, or cylinder.

Right-Hand Rule for Choosing n

Let $\mathcal{S}$ be an orientable surface that is bounded by a parameterized curve C so that the curve is swept out in a definite direction. If the surface is flat (part of a plane) or almost flat, we can use the right-hand rule to choose **n**: The direction of **n** should match the direction the thumb of the right hand points when the fingers of the right hand curl in the direction of C and the thumb and palm are perpendicular to the plane or to the surface. Figure 18.6.4 illustrates the choice of **n**. For instance, in Figure 18.6.4(a), since C is counterclockwise in the xy-plane and $\mathcal{S}$ is the plane region bound by C, the right-hand rule picks out **k** as the normal vector on C, and so the normal vectors on the rest of the surface $\mathcal{S}$ will also point upwards. When, as in Figure 18.6.4)b, C has a clockwise orientation, the right-hand rule requires −**k** as the appropriate normal vector on C and so the normal vectors at points on $\mathcal{S}$ must also point downwards. If the surface is not sufficiently "flat" it can be difficult to make sense of "clockwise" and "counterclockwise". The proper choice of the outward unit normal **n** will be made by the following extension:

Definition: *Right-Hand Rule for Choosing the Normal Vector to a Surface*

Imagine walking along the curve C in the direction of its orientation but standing perpendicular to the surface and on the side of the surface such that *nearby points* on the surface are on your *left*. Then choose the normal **n** to be the one that points towards your head.

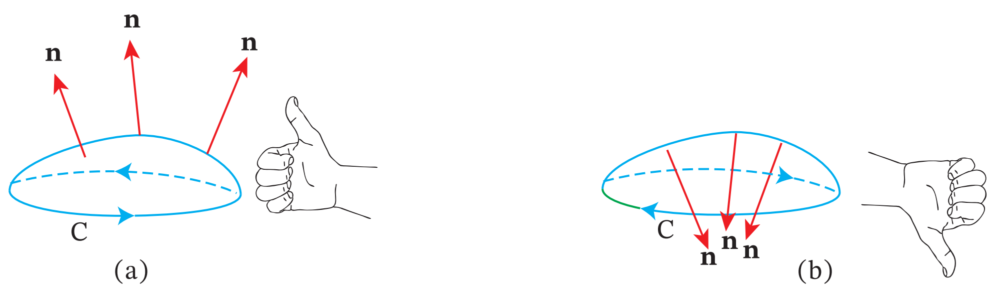


Figure 18.6.4

Figure 18.6.5 illustrates the right-hand rule for a surface that consists of five faces of a cube with a boundary curve oriented as shown. Note that the front face of the cube is not part of $\mathscr{S}$.

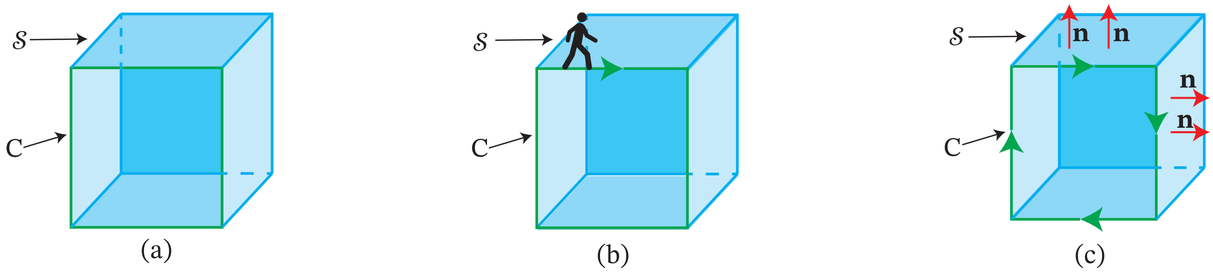


Figure 18.6.5

We are now prepared to provide a precise statement of Stokes' theorem.

Theorem 18.6.1: Stokes' theorem

Let $\mathscr{S}$ be an orientable surface bounded by the parameterized curve C. At each point of $\mathscr{S}$ let $\mathbf{n}$ be the unit normal chosen by the right-hand rule. Let $\mathbf{F}$ be a vector field defined on some region in space including $\mathscr{S}$. Then

$$\oint_C \mathbf{F} \cdot d\mathbf{r} = \int_{\mathscr{S}} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS.$$

Some Applications of Stokes' Theorem

While each surface has a unique and well-defined boundary, that boundary is the boundary for many surfaces. This fact creates the opportunity to use Stokes' theorem to replace $\int_{\mathscr{S}} (\mathbf{curl}\,\mathbf{F}) \cdot \mathbf{n}\, dS$ by a similar integral over a surface that might be simpler than $\mathscr{S}$. That is the substance of the following observation.

Observation 18.6.2: *Replacing One Surface by Another*

Let $\mathscr{S}_1$ and $\mathscr{S}_2$ be two surfaces with the same boundary, C. Assume the common boundary is a simple closed curve and the normal vectors on $\mathscr{S}_1$ and $\mathscr{S}_2$ are consistent with the the orientation of C according to the right-hand rule. Moreover, assume the vector field $\mathbf{F}$ is defined (and has continuous first-order partial derivatives) on an open set that contains both $\mathscr{S}_1$ and $\mathscr{S}_2$. Then

$$\int_{\mathscr{S}_1} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS = \int_{\mathscr{S}_2} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS. \tag{18.6.1}$$

This is true because both of the surface integrals in (18.6.1) are equal to $\oint_C \mathbf{F} \cdot d\mathbf{r}$.

The utility of (18.6.1) is that it provides a way to replace an integral of **curl F** $\cdot$ **n** over a complicated surface by an integral over a simpler surface. Often, the simpler surface will be part of a plane or a a hemisphere.

EXAMPLE 1. Let $\mathbf{F} = xe^z\mathbf{i} + (x + xz)\mathbf{j} + 3e^z\mathbf{k}$ and let $\mathcal{S}$ be the top half of the sphere $x^2 + y^2 + z^2 = 1$. Find $\int_{\mathcal{S}} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS$, where $\mathbf{n}$ is the outward unit normal vector to $\mathcal{S}$.

SOLUTION Let $\mathcal{S}^*$ be the flat base of the hemisphere. Then, as shown in Figure 18.6.6, $\mathcal{S}$ and $\mathcal{S}^*$ have the same boundary curve, C. Note that using the outward normal vector on $\mathcal{S}$ dictates the counter-clockwise orientation on C. For this to be the correct orientation of the boundary curve to $\mathcal{S}^*$, the normal vector on $\mathcal{S}^*$ must be upward, that is, $\mathbf{n} = \mathbf{k}$, not $-\mathbf{k}$.

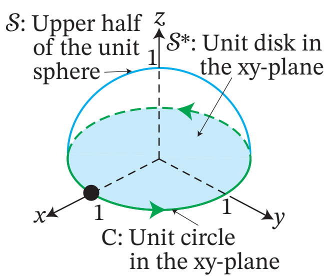


Figure 18.6.6

Then, by (18.6.1),

$$\int_{\mathcal{S}} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS = \int_{\mathcal{S}^*} (\nabla \times \mathbf{F}) \cdot \mathbf{k}\, dS.$$

A calculation shows that

$$\nabla \times \mathbf{F} = -x\mathbf{i} + xe^z\mathbf{j} + (z+1)\mathbf{k},$$

hence $(\nabla \times \mathbf{F}) \cdot \mathbf{k} = z + 1$. Because $z = 0$ on $\mathcal{S}^*$,

$$\int_{\mathcal{S}^*} (\nabla \times \mathbf{F}) \cdot \mathbf{k}\, dS = \int_{\mathcal{S}^*} dS = \pi.$$

Thus the original integral over $\mathcal{S}$ is also π. ♦

Just as there are two-curve versions of Green's theorem and of the divergence theorem, there is a two-curve version of Stokes' theorem.

Theorem 18.6.3: Stokes' Theorem for a Surface Bounded by Two Closed Curves

Let $\mathcal{S}$ be an orientable surface whose boundary consists of the two closed curves C_1 and C_2. Give C_1 an orientation. Orient $\mathcal{S}$ consistent with the right-hand rule as applied to C_1. Give C_2 the same orientation as C_1. (If C_2 is moved on $\mathcal{S}$ to C_1, the orientations will agree.) Then

$$\oint_{C_1} \mathbf{F} \cdot d\mathbf{r} - \oint_{C_2} \mathbf{F} \cdot d\mathbf{r} = \int_{\mathcal{S}} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS. \tag{18.6.2}$$

Proof of Theorem 18.6.3

Figure 18.6.7(a) shows the typical situation. Note the orientation of the boundary curves C_1 and C_2.

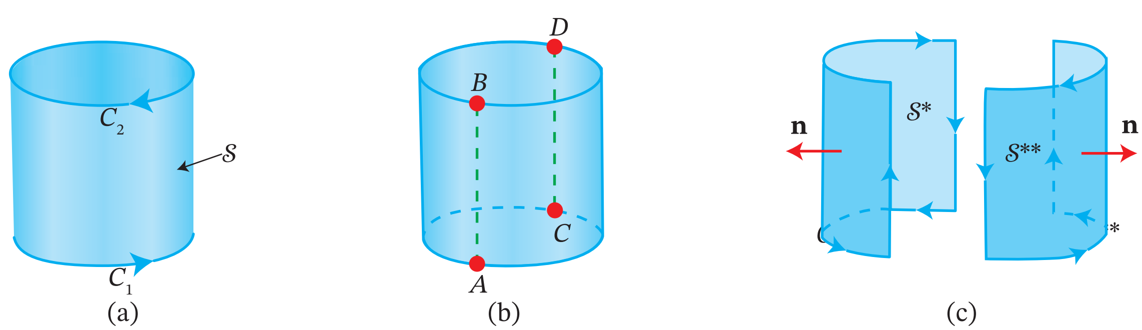


Figure 18.6.7

We will obtain (18.6.2) from Stokes' theorem with the aid of the cancellation principle (first encountered in Section 18.2). Introduce curves AB and CD on $\mathscr{S}$, cutting $\mathscr{S}$ into two surfaces, $\mathscr{S}^*$ and $\mathscr{S}^{**}$.

Let C^* be the curve that bounds $\mathscr{S}^*$, oriented so that where it overlaps C_1 it has the same orientation as C_1. Let C^{**} be the curve that bounds $\mathscr{S}^{**}$, again oriented to match C_1 on any shared boundary curves. See Figure 18.6.7(c). Notice that while on portions of C^* and C^{**} that are also on C_1 the orientations are the same, the orientations are opposite on portions of C^* and C^{**} that are also on C_2.

By Stokes' theorem,

$$\oint_{C^*} \mathbf{F}\cdot d\mathbf{r} = \int_{\mathscr{S}^*} (\mathbf{curl}\,\mathbf{F})\cdot \mathbf{n}\, dS \tag{18.6.3}$$

and

$$\oint_{C^{**}} \mathbf{F}\cdot d\mathbf{r} = \int_{\mathscr{S}^{**}} (\mathbf{curl}\,\mathbf{F})\cdot \mathbf{n}\, dS. \tag{18.6.4}$$

Adding (18.6.3) and (18.6.4) and using the cancellation principle gives

$$\begin{aligned}\int_{\mathscr{S}} (\mathbf{curl}\,\mathbf{F})\cdot \mathbf{n}\, dS &= \int_{\mathscr{S}^*} (\mathbf{curl}\,\mathbf{F})\cdot \mathbf{n}\, dS + \int_{\mathscr{S}^{**}} (\mathbf{curl}\,\mathbf{F})\cdot \mathbf{n}\, dS\\ &= \oint_{C^*} \mathbf{F}\cdot d\mathbf{r} + \oint_{C^{**}} \mathbf{F}\cdot d\mathbf{r}\\ &= \oint_{C_1} \mathbf{F}\cdot d\mathbf{r} - \oint_{C_2} \mathbf{F}\cdot d\mathbf{r}.\end{aligned}$$

●

In practice, it is most common to apply (18.6.2) when the vector field $\mathbf{F}$ is irrotational, that is $\mathbf{curl}\,\mathbf{F} = \mathbf{0}$. This is so important we state it explicitly:

Corollary 18.6.4: Replacing One Curve by Another when curl F = 0

Assume C_1 and C_2 are two closed curves that together form the boundary of an orientable surface $\mathscr{S}$ and the orientations of C_1 and C_2 are compatible with the orientation of $\mathscr{S}$. If the irrotational vector field $\mathbf{F}$ is defined and has continuous first-order partial derivatives on an open set that contains $\mathscr{S}$, then

$$\oint_{C_1} \mathbf{F}\cdot d\mathbf{r} = \oint_{C_2} \mathbf{F}\cdot d\mathbf{r}. \tag{18.6.5}$$

Equation (18.6.5) follows directly from (18.6.2) since $\int_{\mathscr{S}} (\mathbf{curl}\,\mathbf{F})\cdot \mathbf{n}\, dS = 0$.

EXAMPLE 2. Assume that $\mathbf{F}$ is irrotational and defined everywhere except on the z-axis. The curves are as given in Figure 18.6.8. Assume that $\oint_{C_1} \mathbf{F}\cdot d\mathbf{r} = 3$. (a) Find $\oint_{C_2} \mathbf{F}\cdot d\mathbf{r}$. (b) Find $\oint_{C_3} \mathbf{F}\cdot d\mathbf{r}$. (c) Find $\oint_{C_4} \mathbf{F}\cdot d\mathbf{r}$.

SOLUTION

(a) Notice that every surface with boundary curve C_2 must must include at least one point on the z-axis, where $\nabla\times\mathbf{F}$ is not defined. As a result, Stokes' theorem cannot be applied directly to C_2. But, because there is an orientable surface $\mathscr{S}$ with boundary C_1 and C_2, and C_1 and C_2 have the same orientation, by (18.6.5), $\oint_{C_2} \mathbf{F}\cdot d\mathbf{r} = \oint_{C_1} \mathbf{F}\cdot d\mathbf{r} = 3$.

(b) Similarly, since the orientation of C_3 is opposite the orientation of C_2, (18.6.5) tells us that $\oint_{-C_3} \mathbf{F}\cdot d\mathbf{r} = \oint_{C_1} \mathbf{F}\cdot d\mathbf{r} = 3$. So, $\oint_{C_3} \mathbf{F}\cdot d\mathbf{r} = -\oint_{-C_3} \mathbf{F}\cdot d\mathbf{r} = -3$.

(c) By Stokes' theorem, (18.6.1), with $\mathscr{S}$ any orientable surface with boundary C_4 that avoids the z-axis, $\oint_{C_4} \mathbf{F}\cdot d\mathbf{r} = \int_{\mathscr{S}} (\nabla\times\mathbf{F})\cdot\mathbf{n}\, dS = 0$.

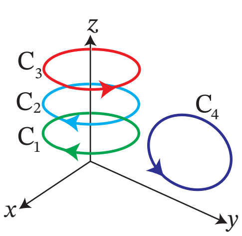


Figure 18.6.8

◆

Curl and Conservative Fields

In Section 18.2 we saw that if $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$ is defined on a simply connected region in the xy-plane and if $\mathbf{curl}\,\mathbf{F} = \mathbf{0}$, then $\mathbf{F}$ is conservative. Now that we have Stokes' theorem, this can be extended to a field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ defined on a simply connected region in space.

Theorem 18.6.5: Test for a Conservative Vector Field in Space

Let $\mathbf{F}$ *be defined on a simply connected region in space. If* $\mathbf{curl}\,\mathbf{F} = \mathbf{0}$*, then* $\mathbf{F}$ *is conservative.*

Proof of Theorem 18.6.5

We provide only a sketch of the proof. Let C be a simple closed curve situated in the simply connected region. To avoid topological complexities, we assume that it bounds an orientable surface $\mathcal{S}$. To show that $\oint_C \mathbf{F}\cdot d\mathbf{r} = 0$, we use the same argument as in Section 18.2, except that the curve C and surface $\mathcal{S}$ are no longer restricted to the xy-plane:

$$\oint_C \mathbf{F}\cdot d\mathbf{r} = \int_{\mathcal{S}} (\nabla\times\mathbf{F})\cdot\mathbf{n}\,dS = \int_{\mathcal{S}} 0\,dS = 0.$$

It follows from Theorem 18.6.5 that a central field $\mathbf{F}$ that is defined throughout space with the possible exception of its center is conservative. Why? First, its curl is $\mathbf{0}$ (see Exercises 14 and 17 in Section 18.4). Second, its domain is simply connected.

Exercise 23 of Section 18.4 presents a geometric argument that shows why a central field is conservative.

Why Curl is Called Curl

Let $\mathbf{F}$ be a vector field describing the flow of a fluid, as in Section 18.1. Stokes' theorem will give a physical interpretation of $\mathbf{curl}\,\mathbf{F}$.

Let P be a fixed point in space. Imagine a small circular disk $\mathcal{S}$ with center P. Let C be the boundary of $\mathcal{S}$ oriented in such a way that C and $\mathbf{n}$ fit the right-hand rule. (See Figure 18.6.9(a).)

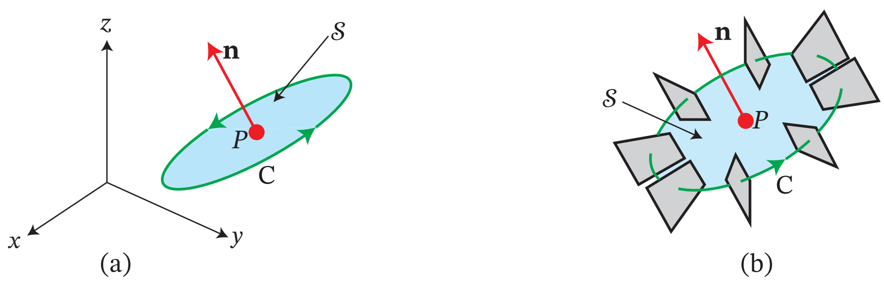


Figure 18.6.9

Now examine the two sides of

$$\int_{\mathcal{S}} (\mathbf{curl}\,\mathbf{F})\cdot\mathbf{n}\,dS = \oint_C \mathbf{F}\cdot\mathbf{T}\,ds. \tag{18.6.6}$$

The right side measures the tendency of the fluid to move along C (rather than, say, perpendicular to it.) Thus $\oint_C \mathbf{F}\cdot\mathbf{T}\,ds$ might be thought of as the circulation or whirling tendency of the fluid along C. For each tilt of the small disk $\mathcal{S}$ at P, or, equivalently, each choice of unit normal vector $\mathbf{n}$, the line integral $\oint_C \mathbf{F}\cdot\mathbf{T}\,ds$ measures a circulation. It records the tendency of a paddle wheel at P with axis along $\mathbf{n}$ to rotate. (See Figure 18.6.9(b).)

On the left side of (18.6.6), if $\mathcal{S}$ is small, the integrand is almost constant and the integral is approximately

$$\mathbf{curl}\,\mathbf{F}(P)\cdot\mathbf{n}\cdot(\text{Area of }\mathcal{S}). \tag{18.6.7}$$

Keeping the center of $\mathcal{S}$ at P, vary the vector $\mathbf{n}$ by tilting $\mathcal{S}$. For which choice of $\mathbf{n}$ will (18.6.7) be largest? Of course, (8.4.3) is largest when $\mathbf{n}$ has the same direction as the fixed vector $\mathbf{curl}\,\mathbf{F}(P)$. With that choice of $\mathbf{n}$, (18.6.7) becomes

$$|\mathbf{curl}\,\mathbf{F}(P)|(\text{Area of } \mathscr{S}).$$

Thus a paddle wheel placed in the fluid at P rotates most quickly when its axis is in the direction of **curl F** in that position. The magnitude of $\mathbf{curl}\,\mathbf{F}(P)$ is a measure of how fast the paddle wheel rotates when placed at P. Thus **curl F** records the direction and magnitude of maximum circulation at a given point.

Historical Note: The Origin of the Term curl

In a letter to the mathematician Peter Guthrie Tait written on November 7, 1870, Maxwell offered some names for $\nabla \times \mathbf{F}$:

> Here are some rough-hewn names. Will you like a good Divinity shape their ends properly so as to make them stick? ...
>
> The vector part $\nabla \times \mathbf{F}$ I would call the twist of the vector function. Here the word twist has nothing to do with a screw or helix. The word *turn* ... would be better than twist, for twist suggests a screw. Twirl is free from the screw motion and is sufficiently racy. Perhaps it is too dynamical for pure mathematicians, so for Cayley's sake I might say Curl (after the fashion of Scroll.)

His last suggestion, curl, has stuck; we still use it today.

A Vector Definition of Curl

In Section 18.1 **curl F** was defined in terms of the partial derivatives of the components of **F**. By Stokes' theorem, **curl F** is related to the circulation $\oint_C \mathbf{F}\cdot d\mathbf{r}$. We exploit this to obtain a new view of **curl F**, free of coordinates.

Let P be a point in space. For a unit vector $\mathbf{n}$, $\mathscr{S}$ denotes a region containing P situated in the plane through P perpendicular to $\mathbf{n}$. Let C be the boundary of $\mathscr{S}$ oriented by the right-hand rule. Then, by the reasoning just used,

$$(\mathbf{curl}\,\mathbf{F}(P)\cdot\mathbf{n})\,(\text{Area of } \mathscr{S}) \approx \oint\limits_C \mathbf{F}\cdot d\mathbf{r}.$$

Thus

$$\mathbf{curl}\,\mathbf{F}(P)\cdot\mathbf{n} = \lim_{\mathscr{S}\to P} \frac{\oint_C \mathbf{F}\cdot d\mathbf{r}}{\text{Area of } \mathscr{S}}. \tag{18.6.8}$$

This gives meaning to the component of $\mathbf{curl}\,\mathbf{F}(P)$ in a direction $\mathbf{n}$. The magnitude and direction of **curl F** at P can be described in terms of **F**, without looking at the components of **F**. The left-hand side of (18.6.8) is maximized when $\mathbf{n}$ is the unit vector in the direction of $\mathbf{curl}\,\mathbf{F}(P)$. For this choice of $\mathbf{n}$ the dot product equals the magnitude of $\mathbf{curl}\,\mathbf{F}(P)$.

Definition: *Coordinate-Free Definition of* $\mathbf{curl}\,\mathbf{F}(P)$

The magnitude of $\mathbf{curl}\,\mathbf{F}(P)$ is the maximum value of

$$\lim_{\mathscr{S}\to P} \frac{1}{\text{Area of } \mathscr{S}} \oint\limits_C \mathbf{F}\cdot d\mathbf{r} \tag{18.6.9}$$

for all unit vectors $\mathbf{n}$. The regions $\mathscr{S}$ are in the plane through P perpendicular to $\mathbf{n}$. The direction of $\mathbf{curl}\,\mathbf{F}(P)$ is given by the vector $\mathbf{n}$ that maximizes (18.6.9).

EXAMPLE 3. Let **F** be a vector field such that at the origin $\mathbf{curl}\,\mathbf{F} = 2\mathbf{i}+4\mathbf{j}+4\mathbf{k}$. Estimate $\oint\limits_C \mathbf{F}\cdot d\mathbf{r}$ if C encloses a disk of radius 0.01 in the xy-plane with center $(0,0,0)$ and C is swept out clockwise. (See Figure 18.6.10.)

SOLUTION Let $\mathscr{S}$ be the disk whose boundary is C. Choose the normal to $\mathscr{S}$ that is consistent with the orientation of C and the right-hand rule; that choice is $\mathbf{n} = -\mathbf{k}$. Thus

$$(\mathbf{curl\,F}) \cdot (-\mathbf{k}) \approx \frac{\oint_C \mathbf{F} \cdot d\mathbf{r}}{\text{Area of } \mathscr{S}}.$$

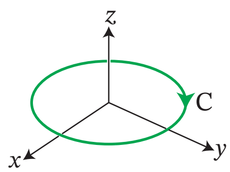


Figure 18.6.10

Because the area of $\mathscr{S}$ is $\pi(0.01)^2$ and $\mathbf{curl\,F} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$, we find

$$(2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}) \cdot (-\mathbf{k}) \approx \frac{\oint_C \mathbf{F} \cdot d\mathbf{r}}{\pi(0.01)^2}.$$

From this it follows that

$$\oint_C \mathbf{F} \cdot d\mathbf{r} \approx -4\pi(0.01)^2 = -0.0004\pi.$$

♦

Proof of Stokes' Theorem

We include the proof of Theorem 18.6.1 because it reviews several basic ideas that must be remembered and verified each time we want to utilize Stokes' theorem. The proof uses Green's theorem, the normal to a surface $z = f(x, y)$, and the expression for an integral over a surface as an integral over its projection on a plane.

Proof of Stokes' Theorem (Theorem 18.6.1)

Assume that the surface $\mathscr{S}$ meets each line parallel to an axis in at most one point. That is, there are one-to-one projections of $\mathscr{S}$ onto each of the three coordinate planes.

Write $\mathbf{F}(x, y, z)$ as $P(x, y, z)\mathbf{i} + Q(x, y, z)\mathbf{j} + R(x, y, z)\mathbf{k}$, or $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$. If the equation for $\mathscr{S}$ is written as $f(x, y) - z = 0$, a unit normal for $\mathscr{S}$ is found to be

$$\mathbf{n} = \frac{-\frac{\partial f}{\partial x}\mathbf{i} - \frac{\partial f}{\partial y}\mathbf{j} + \mathbf{k}}{\sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1}}.$$

(Since the $\mathbf{k}$-component of $\mathbf{n}$ is positive, it is the correct normal, given by the right-hand rule.) Let C^* be the projection of C on the xy-plane, swept out counterclockwise.

A computation shows that Stokes' theorem, expressed in components, reads

See Exercise 36.

$$\oint_C P\,dx + Q\,dy + R\,dz = \int_{\mathscr{S}} \left(\frac{\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\left(-\frac{\partial f}{\partial x}\right) - \left(\frac{\partial R}{\partial x} - \frac{\partial P}{\partial z}\right)\left(-\frac{\partial f}{\partial y}\right) + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)(1)}{\sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1}} \right) dS.$$

This reduces to three separate equations, one for P, one for Q, and one for R. We will establish the result for P, namely

$$\oint_C P\,dx = \int_{\mathscr{S}} \left(\frac{\frac{\partial P}{\partial z}\left(-\frac{\partial f}{\partial y}\right) - \frac{\partial P}{\partial y}(1)}{\sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1}} \right) dS. \tag{18.6.10}$$

The line integral of P around the closed curve, $\oint_C P\,dx$, can be translated into a line integral around C^*, the projection of C onto the xy-plane with a counterclockwise orientation:

$$\oint_C P\,dx = \oint_{C^*} P(x,y,f(x,y))\,dx. \tag{18.6.11}$$

Applying Green's theorem, and then the chain rule, to the right-hand side of (18.6.11) yields

$$\oint_{C^*} P(x,y,f(x,y))\,dx = \int_{\mathscr{S}^*} -\frac{\partial}{\partial y}P(x,y,f(x,y))\,dA = \int_{\mathscr{S}^*} -\left(\frac{\partial P}{\partial y}+\frac{\partial P}{\partial z}\frac{\partial f}{\partial y}\right)dA. \tag{18.6.12}$$

Lastly, to convert the surface integral over $\mathscr{S}^*$ in (18.6.12) to an equivalent surface integral over $\mathscr{S}$, we know that dA can be replaced by $dS/\sqrt{(\partial f/\partial x)^2+(\partial f/\partial y)^2+1}$. Thus,

$$\int_{\mathscr{S}^*} -\left(\frac{\partial P}{\partial y}+\frac{\partial P}{\partial z}\frac{\partial f}{\partial y}\right)dA = \int_{\mathscr{S}} -\left(\frac{\partial P}{\partial y}+\frac{\partial P}{\partial z}\frac{\partial f}{\partial y}\right)\frac{dS}{\sqrt{\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial y}\right)^2+1}}. \tag{18.6.13}$$

When (18.6.11), (18.6.12), and (18.6.13) are combined, they provide a proof of (18.6.10).

The proofs of the other two parts of Stokes' theorem are the subject of Exercises 40 and 41. Note that one of these is very similar to the one just given, but the other one requires using the projection of C onto a different coordinate plane (but is otherwise similar to the proofs for the other components).

We assumed that $\mathscr{S}$ has a special form, meeting lines parallel to an axis just once. More general surfaces, such as the surface of a sphere or a polyhedron, can be partitioned into pieces of this type. The proof that Stokes' theorem holds in these general cases is covered in more advanced courses.

Summary

Stokes' theorem relates the circulation of a vector field over a closed curve C to a surface integral over a surface $\mathscr{S}$ with boundary C. The integrand over the surface is the component of the curl of the field perpendicular to the surface,

$$\oint_C \mathbf{F}\cdot d\mathbf{r} = \int_{\mathscr{S}} (\mathbf{curl\,F})\cdot\mathbf{n}\,dS.$$

The normal $\mathbf{n}$ to $\mathscr{S}$ is determined by the right-hand rule.

Stokes published his theorem in 1854 (without proof, for it appeared as a question on a Cambridge University examination). By 1870 it was in common use. It is the most recent of the three theorems discussed in this chapter, for Green published his theorem in 1828 and Gauss published the divergence theorem in 1839.

Sections 18.7 and 18.9 will show how Stokes' theorem is applied in the theory of electromagnetism. Section 18.8 shows how vector-valued functions, and the differential operators, **grad**, div, and **curl**, can be converted to other coordinate systems.

EXERCISES for Section 18.6

1. State Stokes' theorem using mathematical symbols.

2. State Stokes' theorem in words, using no mathematical symbols.

3. For curves in the xy-plane we spoke of clockwise and counterclockwise orientations. Why is this distinction not made for curves in space?

4. Show that if $F(r)$ is an antiderivative of $f(r)$, then $f(r)\hat{\mathbf{r}}$ is the gradient of $F(r)$.

Observation 18.6.6: *Every Radial Function, $f(r)\hat{\mathbf{r}}$ is Conservative*

From Exercise 4 we conclude that every radial function $f(r)\hat{\mathbf{r}}$ is conservative.

5. (a) Use the fact that a gradient, ∇f, is conservative to show that its curl is $\mathbf{0}$.
(b) Compute $\nabla \times \nabla f$ in terms of components to show that the curl of a gradient is $\mathbf{0}$.

In Exercises 6 to 9 verify Stokes' theorem for the given field $\mathbf{F}$ and surface $\mathcal{S}$.

6. $\mathbf{F} = xy^2\mathbf{i} + y^3\mathbf{j} + y^2z\mathbf{k}$, $\mathcal{S}$ is the top half of the sphere $x^2 + y^2 + z^2 = 1$.

7. $\mathbf{F} = y\mathbf{i} + xz\mathbf{j} + x^2\mathbf{k}$, $\mathcal{S}$ is the triangle with vertices $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$.

8. $\mathbf{F} = y^5\mathbf{i} + x^3\mathbf{j} + z^4\mathbf{k}$, $\mathcal{S}$ is the portion of $z = x^2 + y^2$ below the plane $z = 1$.

9. $\mathbf{F} = -y\mathbf{i} + x\mathbf{j} + z\mathbf{k}$, $\mathcal{S}$ is the portion of the cylinder $z = x^2$ inside the cylinder $x^2 + y^2 = 4$.

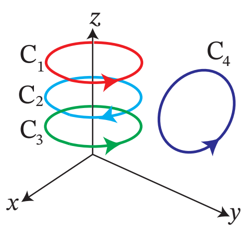


Figure 18.6.11

10. Assume that $\mathbf{F}$ is defined everywhere except on the z-axis and is irrotational. The curves C_1, C_2, C_3, and C_4 are shown in Figure 18.6.11. What, if anything, can be said about (a) $\oint_{C_1} \mathbf{F}\cdot d\mathbf{r}$, (b) $\oint_{C_2} \mathbf{F}\cdot d\mathbf{r}$, (c) $\oint_{C_3} \mathbf{F}\cdot d\mathbf{r}$, and (d) $\oint_{C_4} \mathbf{F}\cdot d\mathbf{r}$.

11. Evaluate $\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\ dS$, where $\mathbf{F}(x,y,z) = x\mathbf{i} - y\mathbf{j}$ and $\mathcal{S}$ is the surface of the cube bounded by the three coordinate planes and the planes $x = 1$, $y = 1$, $z = 1$, exclusive of the surface in the plane $x = 1$. Let $\mathbf{n}$ be outward from the cube.

12. Using Stokes' theorem (Theorem 18.6.1), evaluate $\int_{\mathcal{S}} (\nabla\times\mathbf{F})\cdot\mathbf{n}\ dS$, where $\mathbf{F} = (x^2 + y - 4)\mathbf{i} + 3xy\mathbf{j} + (2xz + z^2)\mathbf{k}$, and $\mathcal{S}$ is the portion of the surface $z = 4 - (x^2 + y^2)$ above the xy-plane. Let $\mathbf{n}$ be the upward normal.

In Exercises 13 to 16 use Stokes' theorem (Theorem 18.6.1) to evaluate $\oint_C \mathbf{F}\cdot d\mathbf{r}$ for the given field $\mathbf{F}$ and curve C. Assume that C is oriented counterclockwise when viewed from above.

13. $\mathbf{F}(x,y,z) = \sin(xy)\mathbf{i}$, C is the intersection of the plane $x + y + z = 1$ and the cylinder $x^2 + y^2 = 1$.

14. $\mathbf{F}(x,y,z) = e^x\mathbf{j}$, C is the triangle with vertices $(2,0,0)$, $(0,3,0)$, and $(0,0,4)$.

15. $\mathbf{F}(x,y,z) = xy\mathbf{k}$, C is the intersection of the plane $z = y$ with the cylinder $x^2 - 2x + y^2 = 0$.

16. $\mathbf{F}(x,y,z) = \cos(x+z)\mathbf{j}$, C is the boundary of the rectangle with vertices $(1,0,0)$, $(1,1,1)$, $(0,1,1)$, and $(0,0,0)$.

---

17. The surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$ the top and bottom halves, respectively, of a sphere of radius a in space. The vector field $\mathbf{F}$ is defined and has continuous first-order partial derivatives on an open set that contains the sphere and let $\mathbf{n}$ denote an exterior normal to the sphere. What relation, if any, is there between $\int_{\mathcal{S}_1} (\nabla\times\mathbf{F})\cdot\mathbf{n}\ dS$ and $\int_{\mathcal{S}_2} (\nabla\times\mathbf{F})\cdot\mathbf{n}\ dS$?

18. The vector field $\mathbf{F}$ is defined throughout space such that $\mathbf{F}(P)$ is perpendicular to the curve C at each point P on C, the boundary of a surface $\mathcal{S}$. What can one conclude about $\int_{\mathcal{S}} (\nabla\times\mathbf{F})\cdot\mathbf{n}\ dS$?

19. Which central fields, $\mathbf{F}$, have $\mathbf{curl}\,\mathbf{F} = \mathbf{0}$?

20. Closed curves C_1 and C_2 in the xy-plane encircle the origin and are similarly oriented, as in Figure 18.6.12(a). Vector field $\mathbf{F}$ is defined and has continuous first-order partial derivatives on an open set that contains the plane except at the origin. Assume that $\nabla \times \mathbf{F} = \mathbf{0}$.

(a) Must $\oint_{C_1} \mathbf{F} \cdot d\mathbf{r} = 0$? (b) What, if any, relation exists between $\oint_{C_1} \mathbf{F} \cdot d\mathbf{r}$ and $\oint_{C_2} \mathbf{F} \cdot d\mathbf{r}$?

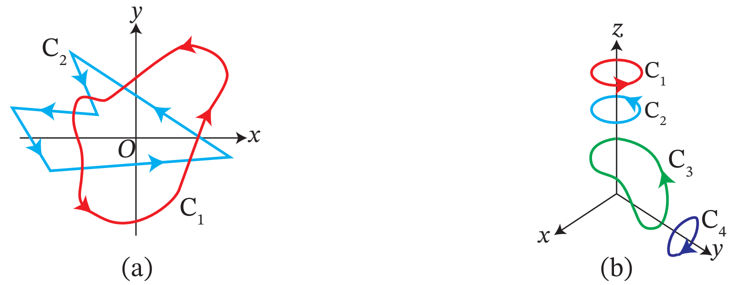


Figure 18.6.12

21. Vector field $\mathbf{F}$ is defined everywhere in space except on the z-axis. Assume that $\mathbf{F}$ is irrotational and $\oint_{C_1} \mathbf{F} \cdot d\mathbf{r} = 7$. See Figure 18.6.12(b). What, if anything, can be said about (a) $\oint_{C_2} \mathbf{F} \cdot d\mathbf{r}$, (b) $\oint_{C_3} \mathbf{F} \cdot d\mathbf{r}$, and (c) $\oint_{C_4} \mathbf{F} \cdot d\mathbf{r}$?

In Exercises 22 and 23 the solid $\mathcal{V}$ is bounded by $z = x + 2$, $x^2 + y^2 = 1$, and $z = 0$ and $\mathbf{F} = y\mathbf{i} + xz\mathbf{j} + (x + 2y)\mathbf{k}$. Define the surface $\mathcal{S}_1$ to be the portion of the plane $z = x + 2$ that lies within the cylinder $x^2 + y^2 = 1$. Define the surface $\mathcal{S}_2$ to be the curved sides of $\mathcal{V}$ together with the base of $\mathcal{V}$. And, denote by C the closed curve that is the boundary of $\mathcal{S}_1$, with a counterclockwise orientation as viewed from above.

22. Use Stokes' theorem for $\mathcal{S}_1$ to evaluate $\oint_C \mathbf{F} \cdot d\mathbf{r}$.

23. Use Stokes' theorem for $\mathcal{S}_2$ to evaluate $\oint_C \mathbf{F} \cdot d\mathbf{r}$.

---

24. Verify Stokes' theorem for the special case when $\mathbf{F}$ has the form ∇f, that is, when $\mathbf{F}$ is a gradient field.

25. Denote by $\mathcal{S}$ the surface of a convex solid, $\mathcal{V}$. Assume the vector field $\mathbf{F}$ is defined throughout $\mathcal{V}$. Show that $\int_{\mathcal{S}} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS = 0$ by

(a) using the divergence theorem, and

(b) drawing a closed curve on C on $\mathcal{S}$ and using Stokes' theorem on the parts into which C divides $\mathcal{S}$.

26. Evaluate $\oint_C \mathbf{F} \cdot d\mathbf{r}$ if $\mathbf{F}(x, y, z) = \dfrac{-y\mathbf{i} + x\mathbf{j}}{x^2 + y^2}$ and C is the intersection of the plane $z = 2x + 2y$ and the paraboloid $z = 2x^2 + 3y^2$ oriented counterclockwise as viewed from above.

27. Assume $\mathbf{F}(x, y)$ is a vector field defined everywhere in the plane except at the origin, $(0, 0)$, and that $\nabla \times \mathbf{F} = \mathbf{0}$. Also assume C_1 is the circle $x^2 + y^2 = 1$ oriented counterclockwise; C_2 is the circle $x^2 + y^2 = 4$ oriented clockwise, C_3 is the circle $(x - 2)^2 + y^2 = 1$ oriented counterclockwise, and C_4 is the circle $(x - 1)^2 + y^2 = 9$ oriented clockwise. If $\oint_{C_1} \mathbf{F} \cdot d\mathbf{r}$ is 5, evaluate (a) $\oint_{C_2} \mathbf{F} \cdot d\mathbf{r}$, (b) $\oint_{C_3} \mathbf{F} \cdot d\mathbf{r}$, and (c) $\oint_{C_4} \mathbf{F} \cdot d\mathbf{r}$.

28. Define the vector field $\mathbf{F}(x, y, z) = \dfrac{\mathbf{r}}{r^a}$, where $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ and a is a real number.

(a) Show that $\nabla \times \mathbf{F} = \mathbf{0}$. (b) Show that $\mathbf{F}$ is conservative. (c) Exhibit a scalar function f such that $\mathbf{F} = \nabla f$.

Exercises 29 to 31 are related.

29. Assume that $\mathbf{G}$ is the curl of a vector field $\mathbf{F}$, $\mathbf{G} = \nabla \times \mathbf{F}$. Let $\mathcal{S}$ be a surface that bounds a solid region $\mathcal{V}$. Let C be a closed curve on the surface $\mathcal{S}$ breaking $\mathcal{S}$ into two pieces $\mathcal{S}_1$ and $\mathcal{S}_2$. Show that $\int\limits_{\mathcal{S}_1} \mathbf{G}\cdot\mathbf{n}\, dS = -\int\limits_{\mathcal{S}_2} \mathbf{G}\cdot\mathbf{n}\, dS$.

30. Using the divergence theorem, show that $\int\limits_{\mathcal{S}} \mathbf{G}\cdot\mathbf{n}\, dS = 0$.

31. Using Stokes' theorem, show that $\int\limits_{\mathcal{S}} \mathbf{G}\cdot\mathbf{n}\, dS = 0$.

32. The curve C is formed by the intersection of the plane $z = x$ and the paraboloid $z = x^2 + y^2$. Orient C to be counterclockwise when viewed from above. Evaluate $\oint\limits_C (xyz\, dx + x^2\, dy + xz\, dz)$.

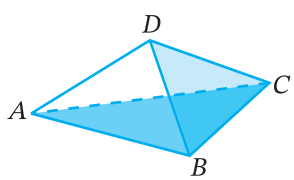


Figure 18.6.13

33. The vector field $\mathbf{F}$ is defined throughout space and has continuous divergence and curl.

(a) For which $\mathbf{F}$ is $\int\limits_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\, dS = 0$ for all spheres $\mathcal{S}$?

(b) For which $\mathbf{F}$ is $\oint\limits_C \mathbf{F}\cdot d\mathbf{r} = 0$ for all circles C?

34. Assume that Stokes' theorem (Theorem 18.6.1) is true for triangles. Deduce that it holds for the surface $\mathcal{S}$ in Figure 18.6.13, consisting of the three triangles DAB, DBC, DCA with boundary curve $ABCA$.

35. **SAM:** Theorem 18.6.3 must be wrong.
JANE: Always questioning authority.
SAM: In the assumptions C_1 and C_2 play the same role. But they play different roles in the conclusion. There's a negative sign in front of the integral over C_2. That does not seem correct.
JANE: Maybe you are right, for once.

Is Sam right? If not, what is his error?

Exercises 36 to 41 concern the proof of Stokes' theorem (Theorem 18.6.1).

36. Carry out the calculations in the proof that translated Stokes' theorem into an equation involving P, Q, and R.

37. Draw a picture of the surfaces $\mathcal{S}$ and $\mathcal{S}^*$ and the curves C and C^* that appear in the proof of Stokes' theorem.

38. List the major ideas involved in the proof of Stokes' theorem, giving an explanation for each.

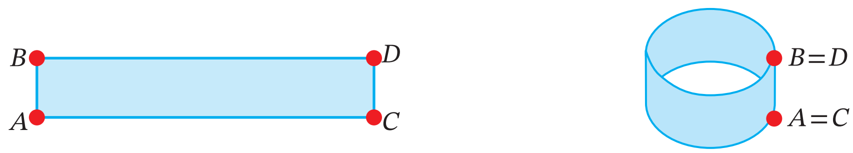
39. In the proof of Stokes' theorem (Theorem 18.6.1) we used a normal $\mathbf{n}$. Show that it is the correct one, compatible with the counterclockwise orientation of C^*.

40. (a) State Stokes' theorem for $\int\limits_C Q\, dy$. (b) Prove Stokes' theorem for $\int\limits_C Q\, dy$.

41. (a) State Stokes' theorem for $\int\limits_C R\, dz$. (b) Prove Stokes' theorem for $\int\limits_C R\, dz$.

Exercises 42 and 43 concern an orientable surface encountered in daily life and Exercises 44 and 45 are devoted to a nonorientable surface known as a Möbius band, named after August Möbius, who discovered it in 1858. These exercises contrast the two concepts; afterwards we describe why the bands are called orientable and nonorientable.

42. (a) Make two strips of paper at least 20" long and about 2" wide. (This can be done by cutting a piece of $8\frac{1}{2}$" × 11" paper into 4 strips, each just over 2" wide and taping 2 of them end-to-end resulting in a strip about 20" long. The second strip is used in Exercise 44.) Label the four corners of each strip as in Figure 18.6.14(a).



(a) (b)

Figure 18.6.14

(b) Form one strip into a band by wrapping edge CD to edge AB, matching corners C and A and corners D with B; then tape the ends to form a band. It resembles a bracelet or belt, as shown if Figure 18.6.14(b).

(c) Using a pencil or match stick to represent a unit normal vector, form a outward-pointing normal vector from the band. Then move the normal vector continuously around the band until it returns to the starting point. Note that during this process the normal vector never points inward.

(d) Repeat (c), but choose the inward-pointing normal vector. When moving around the band, does it ever point outward?

(e) At each point in the band there are two normal vectors. Is it impossible to start with the outward normal vector, continuously move the normal vector along the curve, and end up with the inward normal vector?

43. Cut the band created in Exercise 42 midway between its two edges. What is the result?

44. This problem uses the second 20" × 2" strip of paper constructed in Exercise 42(a).

(a) Form a second strip into a band by twisting the strip 180° so that corners C and B are matched and so are corners D and A; then tape the ends to form a new band, called a Möbius band. When you put on a belt with a half twist in it you may have unintentionally been wearing a Möbius band.

(b) As in Exercise 42(c), choose a normal vector to the band and move it continuously around the band. When it returns to the initial point it points in the direction opposite that of the initial choice. For these reasons we say "The Möbius band is a **one-sided surface**." and "The Möbius band is **nonorientable**."

45. (a) Cut the Möbius band in Exercise 44 midway along the paper. What is the result?

(b) Make another Möbius band and cut it staying about one-third of the way from one edge to the other edge. What is the result?

Observation 18.6.7: *The Relationship between* n *and Orientation*

Now that you have completed Exercises 42 to 45, you might be wondering exactly how the normal vector relates to orientation. The short answer is: by the right-hand rule.

Draw a small circle on the surface with its center at the base of a normal. Orient the circle so that the normal vector is pointed in the direction given by the right-hand rule. Then we could use the oriented circle instead of the normal vector, for it contains the same information as the normal vector. An orientable surface is then one in which we cannot continuously move an oriented circle on the surface so that it returns to the same point with the opposite "spin."

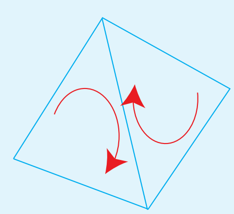


Figure 18.6.15

When an orientable surface is cut into triangles it is possible to orient each triangle so that on an edge shared by two adjacent triangles the orientations induced by the two triangles are opposites, as shown in Figure 18.6.15.

18.7 Connections Between the Electric Field and $\widehat{\mathbf{r}}/|\mathbf{r}|^2$

This section develops one of the four equations that describe the phenomena of electricity and magnetism. It depends on properties of the central inverse-square field in three-dimensional space.

The Mathematics

The central inverse-square field centered at a point C is,

$$\mathbf{F}(P) = \frac{\overrightarrow{CP}}{|CP|^3} \tag{18.7.1}$$

at any point P other than C. The discussion utilizes the following two properties of the central inverse-square field.

(i) If the center, C, of the field is inside the region bounded by a surface $\mathscr{S}$, the flux across $\mathscr{S}$ is 4π.

(ii) If the center, C, of the field is outside the region bounded by $\mathscr{S}$, the flux across $\mathscr{S}$ is 0.

That is,

$$\int_{\mathscr{S}} \frac{\overrightarrow{CP}\cdot\mathbf{n}}{|CP|^3}\,dS = \begin{cases} 4\pi & \text{if } C \text{ is contained within } \mathscr{S} \\ 0 & \text{if } C \text{ is outside } \mathscr{S}. \end{cases}$$

The divergence theorem is used to prove these results, with the proof of the first property using the two-surface version. Both cases depend on the fact that the central inverse-square field has zero divergence everywhere except at its center. The two properties also can be obtained with the aid of the steradian measure of solid angles.

The concept of the integral of a vector field is also needed. The definition is similar to the definition of the definite integral in Section 6.2. Let $\mathbf{F}(P)$ be a continuous vector field defined on a solid region $\mathscr{R}$. Break $\mathscr{R}$ into regions $\mathscr{R}_1, \mathscr{R}_2, \dots, \mathscr{R}_n$ and choose a point P_i in $\mathscr{R}_i$, $1 \le i \le n$. Let the volume of $\mathscr{R}_i$ be V_i. The sums $\sum_{i=1}^n \mathbf{F}(P_i)V_i$ have a limit as all $\mathscr{R}_i$ are chosen smaller and smaller. This limit, denoted $\int_{\mathscr{R}} \mathbf{F}(P)\,dV$, is called the **integral of F over** $\mathscr{R}$. It can be computed componentwise. For example, if $\mathbf{F} = F_1\mathbf{i} + F_2\mathbf{j} + F_3\mathbf{k}$ then $\int_{\mathscr{R}} \mathbf{F}(P)\,dV = \int_{\mathscr{R}} F_1\,dV\mathbf{i} + \int_{\mathscr{R}} F_2\,dV\mathbf{j} + \int_{\mathscr{R}} F_3\,dV\mathbf{k}$. Similar definitions hold for vector fields defined on surfaces or curves.

The Physics

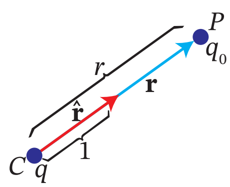


Figure 18.7.1

We make some assumptions about the fundamental electrical charges, electrons and protons. A proton has a positive charge and an electron has a negative charge of equal absolute value. Two like charges exert a force of repulsion on each other and unlike charges attract each other.

Denote the locations of charges q and q_0 by C and P, respectively. Define $\mathbf{r}$ to be the vector from C to P, as in Figure 18.7.1; so $r = |\mathbf{r}|$ is the distance between the points C and P. The unit vector in the direction of $\mathbf{r}$ is $\widehat{\mathbf{r}} = \mathbf{r}/r$.

If both q and q_0 are protons or both are electrons, the force pushes them farther apart. If one is a proton and the other is an electron, the force draws them closer. The magnitude of the force is inversely proportional to r^2.

Assume that q is positive. The magnitude of the force it exerts on charge q_0 is proportional to q and q_0, and it is also inversely proportional to r^2. For some constant k, the magnitude of the force is

$$k\frac{q\,q_0}{r^2}.$$

The force is directed along the vector $\mathbf{r}$. If q_0 is also positive, it is in the same direction as $\mathbf{r}$. If q_0 is negative, it is in the direction of $-\mathbf{r}$. So we have the vector equation

$$\mathbf{F} = k\frac{q\,q_0}{r^2}\widehat{\mathbf{r}} \tag{18.7.2}$$

where k is positive.

To simplify expressions that appear later, it is customary to write k as $1/(4\pi\epsilon_0)$, where $\epsilon_0 \approx 8.8542\times 10^{-12}$ Farad/m is known as the **permittivity of free space** or **vacuum permittivity**. (Additional information about ϵ_0 can be found in CIE 26, How Maxwell Did It, at the end of this chapter.) Then (18.7.2) is written

$$\mathbf{F}(P) = \frac{q\, q_0}{4\pi\epsilon_0 r^2}\widehat{\mathbf{r}}. \tag{18.7.3}$$

Equation (18.7.3) is also known as **Coulomb's law**.

To physicists, a charge q produces an associated vector field $\mathbf{E}$. This vector field, called the **electrostatic field**, exerts a force on other charges.

A positive charge q at point C creates a central inverse-square vector field $\mathbf{E}$ with center at C. It is defined everywhere except at C. Its value at a point P is

$$\mathbf{E}(P) = \frac{q\,\widehat{\mathbf{r}}}{4\pi\epsilon_0 r^2}$$

where $\mathbf{r} = \overrightarrow{CP}$, as in Figure 18.7.1.

While this introduction is admittedly brief, we conclude by emphasizing the following points:

1. Typical inverse-square vector fields with $q > 0$ and $q < 0$ are shown in Figures 18.7.2(a) and (b), respectively.
2. The value of $\mathbf{E}$ depends only on q and the vector from C to P.
3. The force $\mathbf{F}$ exerted by charge q on charge q_0 at P is obtained by multiplying $\mathbf{E}$ by q_0: $\mathbf{F} = q_0\mathbf{E}$.
4. The field $\mathbf{E}$ can be calculated, in principle, by putting a charge q_0 at P, observing the force $\mathbf{F}$, and then dividing $\mathbf{F}$ by q_0.
5. The field $\mathbf{E}$ enables the charge q to act at a distance on other charges. It plays the role of a rubber band or a spring.

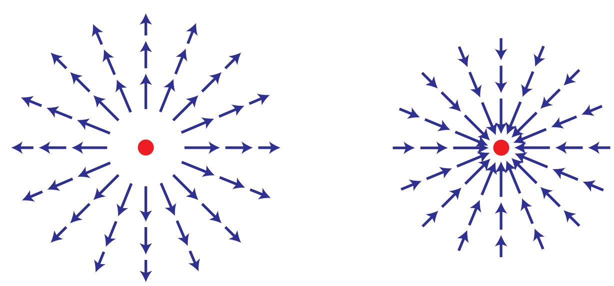

(a) (b)

Figure 18.7.2

The Electrostatic Field Due to a Distribution of Charge

Electrons and protons usually do not exist in isolation. Charge may be distributed on a line, a curve, a surface, or in space.

The amount of charge typically varies from point to point through a region $\mathscr{R}$ in space. Denote the **charge density** at P by $\delta(P)$. Like the density of mass, it is defined as a limit. Let $\mathcal{V}(r)$ be the ball of radius r and center at P. Then we have the definition

$$\delta(P) = \lim_{r\to 0^+} \frac{\text{Charge in } \mathcal{V}(r)}{\text{Volume of } \mathcal{V}(r)}.$$

The charge in the region $\mathcal{V}(r)$ is approximately the volume of $\mathcal{V}(r)$ times $\delta(P)$. We will be interested only in uniform charges, where the density is constant (within the region $\mathcal{V}(r)$ with the value δ. Thus the charge in a region of volume V is $\delta\ V$. The field due to a uniform charge distributed in a region $\mathscr{R}$ is the sum of the fields due to the individual point charges. To estimate this field due to a distribution of charge, partition $\mathscr{R}$ into small regions $\mathscr{R}_1, \mathscr{R}_2, \dots, \mathscr{R}_n$ and choose a point P_i in $\mathscr{R}_i$, $i = 1, 2\dots, n$. The volume of $\mathscr{R}_i$ is V_i. The charge in $\mathscr{R}_i$ is $\delta\ V_i$, where δ is the charge density within $\mathscr{R}$. Figure 18.7.3 shows the contribution to the field at a point P.

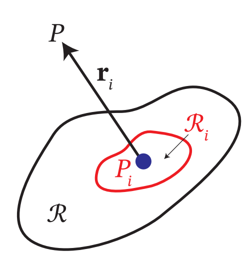


Figure 18.7.3

Let $\mathbf{r}_i$ be the vector from P_i to P, and $r_i = |\mathbf{r}_i|$. Then the field due to the charge in $\mathscr{R}_i$ is approximately

$$\frac{\delta\ V_i\,\widehat{\mathbf{r}}_i}{4\pi\epsilon_0\ r_i^2} = \frac{\delta\,\widehat{\mathbf{r}}_i}{4\pi\epsilon_0\ r_i^2} V_i.$$

As an estimate of the total field due to all n charges, we have the sum

$$\sum_{i=1}^{n} \frac{\delta\,\widehat{\mathbf{r}}_i}{4\pi\epsilon_0 r_i^2}\, V_i. \tag{18.7.4}$$

Taking the limit as all the regions $\mathcal{R}_i$ are chosen smaller, we find the electric field at a point P to be

$$\mathbf{E}(P) = \int\limits_{\mathcal{R}} \frac{\delta\,\widehat{\mathbf{r}}}{4\pi\epsilon_0 r^2}\, dV = \frac{\delta}{4\pi\epsilon_0} \int\limits_{\mathcal{R}} \frac{\widehat{\mathbf{r}}}{r^2}\, dV.$$

If, instead of being distributed throughout a solid region, the charge is on a surface $\mathcal{S}$ with uniform surface density σ, then

$$\mathbf{E}(P) = \frac{\sigma}{4\pi\epsilon_0} \int\limits_{\mathcal{S}} \frac{\widehat{\mathbf{r}}}{r^2}\, dS.$$

And, if the charge lies on a line or a curve C (such as a wire), with uniform density λ, then

$$\mathbf{E}(P) = \frac{\lambda}{4\pi\epsilon_0} \int\limits_{C} \frac{\widehat{\mathbf{r}}}{r^2}\, ds.$$

Though they are not central fields they are divergence-free. This is to be expected because the sums (18.7.4) whose limit is the integral are divergence-free fields.

The above discussion depends critically on the fact that there is no charge inside the region bounded by $\mathcal{S}$.

The Flux Across a Surface Due to a Distribution of Charge

A surface $\mathcal{S}$ bounds a solid region $\mathcal{V}$. The field $\mathbf{E}$ due to a point charge q creates a flux across $\mathcal{S}$. If the point charge is inside the solid region bounded by $\mathcal{S}$, the flux is

$$\frac{q}{4\pi\epsilon_0} 4\pi = \frac{q}{\epsilon_0}.$$

If the charge is outside $\mathcal{S}$, then the flux is 0.

A distribution of charge also creates a flux across $\mathcal{S}$. Assume the charge is distributed in space with a density $\delta(P)$ at point P. The charge in a small region around P of volume dV is approximately $\delta(P)\, dV$. The flux across $\mathcal{S}$ caused by the charge in this small region is $\delta(P)\, dV/\epsilon_0$. If the charge occupies a region $\mathcal{R}$ the flux it creates is then expressed by an integral. To evaluate this integral break $\mathcal{R}$ into two regions, $\mathcal{R}_1$ inside $\mathcal{S}$ and $\mathcal{R}_2$ outside $\mathcal{S}$.

The flux due to the charge in $\mathcal{R}_1$ is

$$\int\limits_{\mathcal{R}_1} \frac{\delta(P)\, dV}{\epsilon_0} = \frac{1}{\epsilon_0} \int\limits_{\mathcal{R}_1} \delta(P)\, dV$$

where $\int_{\mathcal{R}_1} \delta(P)\, dV$ is the charge in $\mathcal{R}_1$.

The charge in $\mathcal{R}_2$ creates a flux of 0 across $\mathcal{S}$. This brings us to one of the four fundamental equations of electromagnetism.

Theorem 18.7.1: Gauss's Law for Electricity

A charge distributed on a curve, a surface, or a solid region induces a flux across a closed surface. This flux equals Q/ϵ_0, where Q is the charge enclosed within the surface. The charge outside the surface creates no flux across the surface.

CAUTION: Do not confuse Gauss's Law for Electricity and Gauss's theorem (the divergence theorem), introduced in Section 18.5.

Applying Gauss's Law to Find E

Example 1 shows that Gauss's law provides a way to determine the electrostatic field **E** associated with a distribution of charge. But, it must be noted that this approach works only in special cases with lots of symmetry. The remainder of this section obtains the same field directly, without the aid of Gauss's law. The contrast illustrates the power of Gauss's law for electricity.

EXAMPLE 1. A positive charge Q is distributed uniformly on a sphere of radius a. Use Gauss's law for electricity to find the electrostatic field **E** at a point B a distance $b > a$ from the center of the sphere.

SOLUTION Figure 18.7.4(a) shows the sphere and a hypothetical vector **v** representing the electrostatic field at B due to the charge on the sphere.

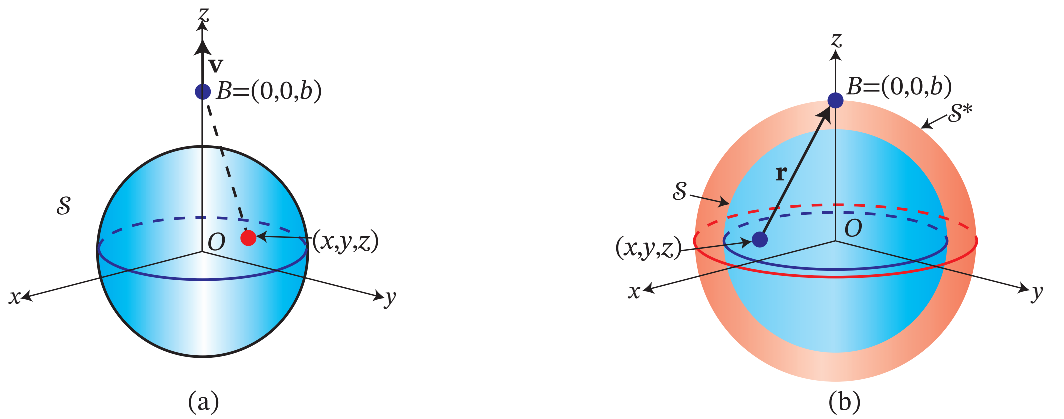


Figure 18.7.4

When the sphere with its charge is spun around the axis OB we get the same situation as when we started. Therefore the vector **v** must be parallel to $\overrightarrow{OB}$. By the symmetry of the sphere the magnitude of **v** depends only on the distance b from B to O. Call this magnitude $f(b)$. All that remains is to find $f(b)$.

To do this, introduce another sphere $\mathscr{S}^*$, with center O and radius b, as in Figure 18.7.4(b). The flux of **E** across $\mathscr{S}^*$ is $\int_{\mathscr{S}^*} \mathbf{E} \cdot \mathbf{n}\, dS$.

But $\mathbf{E} \cdot \mathbf{n}$ is just $f(b)$ since **E** and **n** are parallel and $\mathbf{E}(P)$ has magnitude $f(b)$ for all points P on S^*. Thus

$$\int_{\mathscr{S}^*} \mathbf{E} \cdot \mathbf{n}\, dS = \int_{\mathscr{S}^*} f(b)\, dS = f(b) \int_{\mathscr{S}^*} dS = f(b) 4\pi b^2.$$

By Gauss's law $Q/\epsilon_0 = f(b)\, 4\pi b^2$. Therefore $f(b) = Q/(4\pi\epsilon_0 b^2)$ for $b > a$. ♦

The same result is obtained if the entire charge Q were at the center of the sphere. The same technique shows that for the charge in Example 1 the field inside the sphere is **0**.

Let $f(r)$ be the magnitude of **E** at a distance r from the center of the sphere. For $r > a$, $f(r) = Q/(4\pi\epsilon_0 r^2)$ and for $0 < r < a$, $f(r) = 0$. Exercise 13 concerns $f(a)$. The graph of f is shown in Figure 18.7.5.

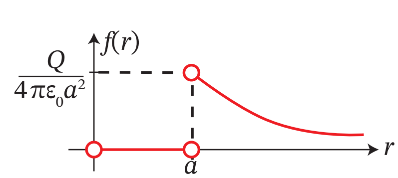


Figure 18.7.5

Finding E without Gauss's Law

To appreciate the power of Gauss's law for electricity we repeat Example 1 using explicit evaluation of the surface integral to determine the electric field in Example 1 directly.

EXAMPLE 2. A positive charge Q is distributed uniformly on a sphere of radius a. Find the electrostatic field $\mathbf{E}$ at a point B located a distance $b > a$ from the center of the sphere by evaluating the surface integral as an iterated integral. NOTE: Details omitted here will be supplied in Exercise 15.

SOLUTION We use an iterated integral to evaluate

$$\mathbf{E}(B) = \frac{\sigma}{4\pi\epsilon_0} \int\limits_{\mathcal{S}} \frac{\widehat{\mathbf{r}}}{r^2}\, dS. \tag{18.7.5}$$

Since the charge is uniform over a region with area $4\pi a^2$, $\sigma = Q/(4\pi a^2)$.

Place a rectangular coordinate system with its origin at the center of the sphere and with the point B on the positive z-axis, so that $B = (0,0,b)$. (See Figure 18.7.4(b).) By the symmetry argument used in Example 1 the x- and y-components of $\mathbf{E}(B)$ are 0. We need only find its z-component, which is $\mathbf{E}(B)\cdot\mathbf{k}$.

Let (x, y, z) be a point on the sphere S. Then

$$\mathbf{r} = (0\mathbf{i} + 0\mathbf{j} + b\mathbf{k}) - (x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) = -x\mathbf{i} - y\mathbf{j} + (b - z)\mathbf{k}. \tag{18.7.6}$$

Hence

$$\frac{\widehat{\mathbf{r}}}{r^2} = \frac{\mathbf{r}}{r^3} = \frac{-x\mathbf{i} - y\mathbf{j} + (b-z)\mathbf{k}}{(\sqrt{x^2 + y^2 + b^2 - 2bz + z^2})^3} = \frac{-x\mathbf{i} - y\mathbf{j} + (b-z)\mathbf{k}}{(a^2 + b^2 - 2bz)^{3/2}}.$$

Only its z-component is needed:

$$\frac{b-z}{(a^2 + b^2 - 2bz)^{3/2}}.$$

The $\mathbf{k}$-component of $\mathbf{E}(B)$ is therefore

$$\frac{\sigma}{4\pi\epsilon_0} \int\limits_{\mathcal{S}} \frac{b-z}{(a^2 + b^2 - 2bz)^{3/2}}\, dS. \tag{18.7.7}$$

To evaluate the surface integral in (18.7.7) introduce spherical coordinates in the standard position. In this context, $dS = a^2 \sin(\phi)\, d\phi\, d\theta$ and $z = a\cos(\phi)$. Thus,

$$\frac{\sigma}{4\pi\epsilon_0} \int\limits_{\mathcal{S}} \frac{b-z}{(a^2 + b^2 - 2bz)^{3/2}}\, dS = \int_0^{\pi}\int_0^{2\pi} \frac{(b - a\cos(\phi))a^2\sin(\phi)}{\left(a^2 + b^2 - 2ab\cos(\phi)\right)^{3/2}}\, d\theta\, d\phi \tag{18.7.8}$$

After the inner integration, with respect to θ, the iterated integral in (18.7.8) reduces to

$$2\pi a^2 \int_0^{\pi} \frac{(b - a\cos(\phi))\sin(\phi)}{(a^2 + b^2 - 2ab\cos(\phi))^{3/2}}\, d\phi. \tag{18.7.9}$$

The evaluation of the remaining integral, (18.7.9), can be completed with the use of two substitutions, paying special attention to the limits of integration:

$$\begin{aligned}
2\pi a^2 \int_0^{\pi} \frac{(b - a\cos(\phi))\sin(\phi)}{(a^2 + b^2 - 2ab\cos(\phi))^{3/2}}\, d\phi &= -2\pi a^2 \int_1^{-1} \frac{(b - au)}{(a^2 + b^2 - 2abu)^{3/2}}\, du && (\text{ substitution: } u = \cos(\phi)\text{ }) \\
&= 2\pi a^2 \int_{(b-a)^2}^{(b+a)^2} \frac{2b^2 + (v - a^2 - b^2)}{2bv^{3/2}}\, \frac{dv}{2ab} && (\text{ substitution: } v = a^2 + b^2 - 2abu\text{ }) \\
&= \frac{\pi a}{2b^2} \int_{(b-a)^2}^{(b+a)^2} \frac{v - a^2 + b^2}{v^{3/2}}\, dv && (\text{ simplification }).
\end{aligned} \tag{18.7.10}$$

Now, after expressing the integrand as the sum of $v^{-1/2}$ and $(b^2 - a^2)v^{-3/2}$, use the fundamental theorem of calculus (and the fact that $b > a$) to show that (18.7.10) equals $4\pi a^2/b^2$. Combining this with (18.7.7) and noting that $4\pi a^2 \sigma = Q$, completes the derivation that

$$\mathbf{E}(B) = \frac{\sigma}{4\pi\epsilon_0}\frac{4\pi a^2}{b^2}\mathbf{k} = \frac{Q}{4\pi\epsilon_0 b^2}\mathbf{k} \qquad \text{for } b > a.$$

Summary

This section showed how the central inverse-square field, $\hat{\mathbf{r}}/r^2$, arises naturally when looking at fields created by a charged particle. The divergence theorem is the key to developing an important property of electric fields produced by a charged particle: its flux across a closed surface is either 4π or 0, depending on whether the center of the central inverse-square field is inside or outside the region bounded by the surface.

Also, using Gauss's law and the symmetry of the sphere, we showed that a uniform charge on a sphere exerts a force outside as if all the charge were at the center of the sphere.

The electric field at a point P due to a point charge q at a point C is given by the formula

$$\mathbf{E}(P) = \frac{1}{4\pi\epsilon_0}\frac{q\hat{\mathbf{r}}}{r^2},$$

where $\mathbf{r} = \overrightarrow{CP}$. This field produces a force $q_0\mathbf{E}(P)$ on a charge q_0 located at P. The field due to a distribution of charge is obtained by an integration.

Lastly, Gauss's law for electricity, not to be confused with Gauss's theorem, states that the flux of the field $\mathbf{E}$ across a surface produced by the charge Q within that surface is Q/ϵ_0.

EXERCISES for Section 18.7

1. Describe Coloumb's law, that is, the force $\mathbf{F}$ exerted by one charged particle upon another.

2. State Gauss's law for electricity using no mathematical symbols.

3. (a) Define the electrostatic field $\mathbf{E}$ due to a charged particle.
(b) How is $\mathbf{E}$ related to the field $\mathbf{F}$ in Exercise 1?

4. A charge is distributed uniformly along an infinite straight wire. The charge on a section of length l is λl. Find the electric field $\mathbf{E}$ due to the charge.

(a) Use symmetry to say as much as you can about the direction and magnitude of $\mathbf{E}$.
(b) Find the magnitude of $\mathbf{E}$ by applying Gauss's law for electricity to the cylinder of radius r and height h shown in Figure 18.7.6
(c) Find the electric field directly by an integral over the line, as in Example 2.

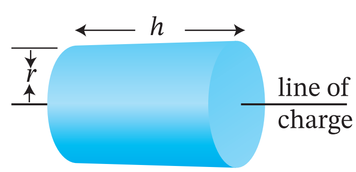


Figure 18.7.6

5. Exercise 4 concerned the electric field $\mathbf{E}$ due to a charge uniformly spread on an infinite line. If the charge density is λ, $\mathbf{E}$ at a point at a distance a from the line is $(\lambda/(2\pi a\epsilon_0))\,\mathbf{j}$.

Assume instead that the line occupies only the right half of the x-axis, $[0,\infty)$.

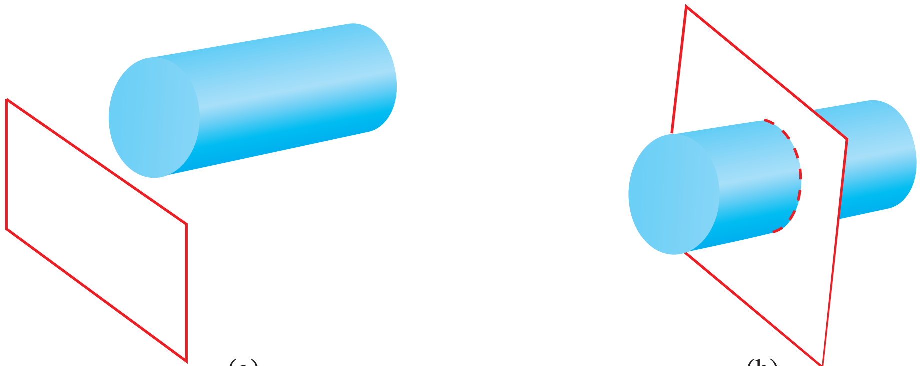
(a) Using the result in Exercise 4, show that the $\mathbf{j}$-component of $\mathbf{E}(0,a)$ is $\dfrac{\lambda}{4\pi a\epsilon_0}\mathbf{j}$.
(b) By integrating over $[0,\infty)$, show that the $\mathbf{i}$-component of $\mathbf{E}$ at $(0,a)$ is $-\dfrac{\lambda}{4\pi a\epsilon_0}\mathbf{i}$.
(c) What angle does $\mathbf{E}(0,a)$ make with the y-axis?
(d) Why is Gauss's law for electricity of no use in determining the $\mathbf{i}$-component of $\mathbf{E}$?

6. A charge of surface density σ is distributed uniformly over an infinite plane. Find the field $\mathbf{E}$ due to the charge at any point P not in the plane.

(a) Use symmetry to say as much as you can about $\mathbf{E}$. Be sure to discuss its direction.

(b) Show that the magnitude is constant by applying Gauss's law for electricity to a cylinder whose axis is perpendicular to the plane and which does not intersect the plane, as shown in Figure 18.7.7(a).

(c) Find the magnitude of $\mathbf{E}$ by applying Gauss's law for electricity to the cylinder in Figure 18.7.7(b) which intersects the plane of the charge. Denote by A the area of the circular cross section.



(a) (b)

Figure 18.7.7

7. Suppose that there is a uniform distribution of charge Q throughout a ball of radius a. Use Gauss's law for electricity to find the electrostatic field $\mathbf{E}$ produced by the charge at point
(a) at points outside the ball and (b) at points inside the ball.

8. Assume $f(r)$ denotes the magnitude of the field in Exercise 7 at a distance r from the center of the ball. Graph $f(r)$ for $r \geq 0$.

9. Show that a charge Q distributed in a solid region $\mathscr{R}$ outside a closed surface $\mathscr{S}$ induces zero flux across $\mathscr{S}$.

10. We showed that $\mathbf{E}(P) = \dfrac{\delta}{4\pi\epsilon_0} \displaystyle\int_{\mathscr{R}} \frac{\widehat{\mathbf{r}}}{r^2}\, dV$ if the charge density is constant. Find the integral for $\mathbf{E}(P)$ when the charge density varies. Do not attempt to evaluate the integral.

11. Find the field $\mathbf{E}$ in Exercise 6 by integrating over the entire xy-plane. Do not use Gauss's law for electricity.

12. A charge Q lies partly inside a closed surface $\mathscr{S}$ and partly outside. Let Q_1 be the amount inside and Q_2 the amount outside. What is the flux across $\mathscr{S}$ of the charge Q?

13. Find the field $\mathbf{E}$ of the charge in Example 1 at a point on the surface of the sphere. Why is Gauss's law for electricity not applicable here?

14. Find the field $\mathbf{E}$ of the charge in Example 1 at the center of the sphere.

15. Fill in the omitted details in the calculation in Example 2.

16. In Example 2, we used an integral to find the electrostatic field outside a uniformly charged sphere. Carry out similar calculations to find the field inside the sphere.

17. Show that $\displaystyle\int_{\mathscr{S}} \frac{\widehat{\mathbf{r}}}{r^2} \cdot \mathbf{n}\, d\mathscr{S} = 4\pi$ if the center of the field is within the solid that $\mathscr{S}$ bounds, and is 0 if the center is outside. Use the divergence theorem for one or two surfaces.

18. Solve Exercise 17 using solid angles.

18.8 Expressing Vector Functions in Other Coordinate Systems

Earlier in this chapter we defined gradient, divergence, and curl in rectangular coordinates:

$$\begin{aligned}
\mathbf{grad}\, f &= \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k}, \\
\operatorname{div}(P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}) &= \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}, \\
\mathbf{curl}(P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}) &= \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{pmatrix}.
\end{aligned}$$

This is convenient when $f = f(x, y, z)$ or $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ are given in rectangular coordinates. But what if they are given in polar, cylindrical, or spherical coordinates? Then it would be convenient to have expressions for **grad**, div, and **curl** in the same coordinate system. This section develops the corresponding expressions. The keys to doing this are the coordinate-free descriptions of three concepts developed earlier in this chapter:

grad f at P	div $\mathbf{F}$ at P	**curl** $\mathbf{F}$ at P
$\mathbf{grad}\, f(P)\cdot\mathbf{n} = D_{\mathbf{n}} f(P)$ for every unit vector $\mathbf{n}$.	$\operatorname{div}\mathbf{F}(P) = \lim_{\mathscr{R}\to P} \frac{1}{\text{Volume}(\mathscr{R})} \int_{\mathscr{S}} \mathbf{F}\cdot\mathbf{n}\, dS$, where $\mathscr{R}$ is a region that contains P and $\mathscr{S}$ is its surface.	$\mathbf{curl}\,\mathbf{F}(P)\cdot\mathbf{n} = \lim_{\mathscr{R}\to P} \frac{1}{\text{Area}(\mathscr{R})} \oint_C \mathbf{F}\cdot\mathbf{T}\, ds$ for every unit vector $\mathbf{n}$, where $\mathscr{R}$ is a region that lies in the plane through P that is perpendicular to $\mathbf{n}$ and contains P, and C is the boundary of $\mathscr{R}$ (oriented by the right-hand rule).

Table 18.8.1

Observation 18.8.1: *Practical Choices in the Definitions of* grad, div, *and* curl

1. While the definition of **grad** f says "every unit vector", it will be enough to check the defining property for 2 or 3 appropriately chosen unit vectors.
2. The regions $\mathscr{R}$ that appear in the definition of div $\mathbf{F}$ are often balls or cubes, but any solid bounded by a surface can be used.
3. Similarly, the regions $\mathscr{S}$ in the definition of **curl** $\mathbf{F}$ are usually disks or squares.
4. When appropriate, unit vectors, solid or plane regions determined by the coordinate systems are chosen to simplify the computations.

The first step towards expressing **grad**, div, and **curl** in a specific coordinate system is to identify the unit vectors in that coordinate system that play the same role that the vectores **i**, **j**, and **k** play in rectangular coordinates.

The Unit Vectors Associated with a Coordinate System

The unit vectors associated with a coordinate system indicate the direction one moves when one coordinate is increased while the others are kept fixed.

For example, in the xy-coordinate system, moving from (x, y) to $(x + \Delta x, y)$ is motion in the direction **i**. If, instead, one keeps x fixed but moves from (x, y) to $(x, y + \Delta y)$, the direction is given by **j**.

The corresponding cases for polar coordinates are shown in Figure 18.8.1.

In Figure 18.8.1(a) a small positive change Δr in the radius moves the point (r,θ) to a new point $(r + \Delta r,\theta)$ a distance Δr along the ray from the pole O. The direction of the change is recorded by the unit vector $\hat{\mathbf{r}}$ along that ray. In Figure 18.8.1(b) a small change $\Delta\theta$ in θ moves the point (r,θ) to a point $(r,\theta + \Delta\theta)$ along the circle of radius

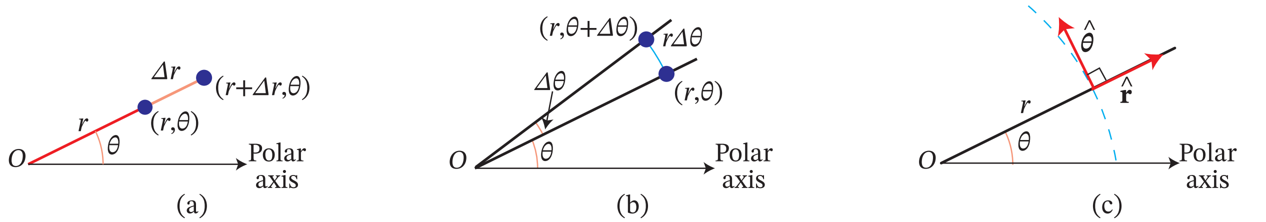


Figure 18.8.1

r centered at O. Changing the angle by $\Delta\theta$ moves the point a distance that is approximately $r\ \Delta\theta$ from the point with polar coordinates (r,θ). A tangent to a circle is perpendicular to the radius to the point of contact. As a result, the unit vector $\widehat{\boldsymbol{\theta}}$ is the same as the counter-clockwise vector $\mathbf{T}$ tangent to the circle of radius r with center O. Note that $\widehat{\mathbf{r}}$ is perpendicular to $\widehat{\boldsymbol{\theta}}$ because the vector from O to (r,θ) is perpendicular to the circle of radius r with center O. (See Figure 18.8.1(c).)

In the case of cylindrical coordinates the three unit vectors that record the directions a point moves with small changes in r, θ, and z are $\widehat{\mathbf{r}}$, $\widehat{\boldsymbol{\theta}}$, and $\mathbf{k}$. And, in spherical coordinates the three coordinate unit vectors are $\widehat{\boldsymbol{\rho}}$, $\widehat{\boldsymbol{\theta}}$, and $\widehat{\boldsymbol{\phi}}$. A sketch shows that $\widehat{\mathbf{r}}$, $\widehat{\boldsymbol{\theta}}$, and $\mathbf{k}$ are mutually perpendicular, as are $\widehat{\boldsymbol{\rho}}$, $\widehat{\boldsymbol{\theta}}$, and $\widehat{\boldsymbol{\phi}}$. (See Exercise 46.)

Directional Derivatives and Coordinate Systems

Let $f(r,\theta)$ be a scalar function given in polar coordinates. We will compute the two directional derivatives with respect to the unit vectors $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$: $D_{\widehat{\mathbf{r}}}f$ and $D_{\widehat{\boldsymbol{\theta}}}f$ at (r,θ).

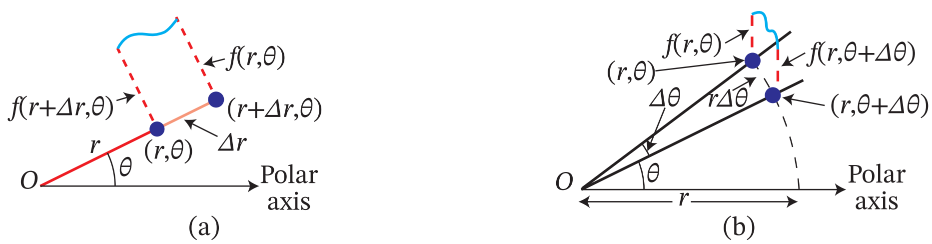


Figure 18.8.2

We begin with $D_{\widehat{\mathbf{r}}}f$ at (r,θ). As Figure 18.8.2(a) illustrates, $D_{\widehat{\mathbf{r}}}f$ is the limit as $\Delta r \to 0$ of the quotient $f(r+\Delta r,\theta)-f(r,\theta)$ divided by the distance from (r,θ) to $(r+\Delta r,\theta)$:

$$\lim_{\Delta r\to 0}\frac{f(r+\Delta r,\theta)-f(r,\theta)}{\Delta r}.$$

That limit is just the partial derivative $\partial f/\partial r$ at (r,θ). Figure 18.8.2(a) shows a trace of the graph of $z=f(r,\theta)$ with θ held fixed; its slope is $\partial f/\partial r$.

Next we find $D_{\widehat{\boldsymbol{\theta}}}f$ at (r,θ). (See Figure 18.8.2(b).) This is the limit as $\Delta\theta \to 0$ of $f(r,\theta+\Delta\theta)-f(r,\theta)$ divided by the distance from (r,θ) to $(r,\theta+\Delta\theta)$. When $\Delta\theta$ is small that distance is approximately $r\Delta\theta$. Thus

$$D_{\widehat{\boldsymbol{\theta}}}f = \lim_{\Delta\theta\to 0}\frac{f(r,\theta+\Delta\theta)-f(r,\theta)}{r\Delta\theta} = \frac{1}{r}\lim_{\Delta\theta\to 0}\frac{f(r,\theta+\Delta\theta)-f(r,\theta)}{\Delta\theta} = \frac{1}{r}\frac{\partial f}{\partial\theta}.$$

Thus $D_{\widehat{\boldsymbol{\theta}}}f$ is not $\partial f/\partial\theta$. It is $(\partial f/\partial\theta)/r$. The extra factor of r comes from the expression $r\Delta\theta$, which measures the distance a point moves when the angle changes by $\Delta\theta$. Figure 18.8.2(b) shows a trace of the graph of $z=f(r,\theta)$ with r held fixed. Its slope is $D_{\widehat{\boldsymbol{\theta}}}f=(\partial f/\partial\theta)/r$.

To summarize, in polar coordinates the derivatives of a scalar function f with respect to the unit vectors $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$ are

$$D_{\widehat{\mathbf{r}}}f=\frac{\partial f}{\partial r}, \quad \text{and} \quad D_{\widehat{\boldsymbol{\theta}}}f=\frac{1}{r}\frac{\partial f}{\partial\theta}.$$

A similar phenomenon occurs in other coordinate systems. For instance, in cylindrical coordinates:

$$D_{\widehat{\mathbf{r}}}f = \frac{\partial f}{\partial r}, \quad D_{\widehat{\boldsymbol{\theta}}}f = \frac{1}{r}\frac{\partial f}{\partial \theta}, \quad \text{and} \quad D_{\mathbf{k}}f = \frac{\partial f}{\partial z}.$$

And, in spherical coordinates:

$$D_{\widehat{\boldsymbol{\rho}}}f = \frac{\partial f}{\partial \rho}, \quad D_{\widehat{\boldsymbol{\theta}}}f = \frac{1}{\rho\sin(\phi)}\frac{\partial f}{\partial \theta}. \quad \text{and} \quad D_{\widehat{\boldsymbol{\phi}}}f = \frac{1}{\rho}\frac{\partial f}{\partial \phi},$$

The coefficient $1/\rho$ that appears in $D_\phi f$ and $1/(\rho\sin(\phi))$ that appears in $D_\theta f$ arise because a change $\Delta\phi$ in spherical coordinates moves a point about a distance $\rho\Delta\phi$, and a change $\Delta\theta$ moves a point about $\rho\sin(\phi)\Delta\theta$.

In rectangular coordinates, because the changes Δx, Δy, and Δz coincide with the distances a point is moved in each direction,

$$D_{\mathbf{i}}f = \frac{\partial f}{\partial x}, \quad D_{\mathbf{j}}f = \frac{\partial f}{\partial y}, \quad \text{and} \quad D_{\mathbf{k}}f = \frac{\partial f}{\partial z}.$$

No extra factors are needed.

Gradient in Polar Coordinates

Let $f(r,\theta)$ be a scalar function expressed in polar coordinates. Its gradient is a vector, and can be written in the form $A(r,\theta)\widehat{\mathbf{r}} + B(r,\theta)\widehat{\boldsymbol{\theta}}$ for some scalar functions $A(r,\theta)$ and $B(r,\theta)$, which we will call A and B for short.

We know that $D_{\widehat{\mathbf{r}}}f = \mathbf{grad}\, f\cdot\widehat{\mathbf{r}}$ and have just shown that $D_{\widehat{\mathbf{r}}}f = \partial f/\partial r$. When we recall that $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathbf{r}}$ are perpendicular, $\widehat{\boldsymbol{\theta}}\cdot\widehat{\mathbf{r}} = 0$, we learn that

$$\frac{\partial f}{\partial r} = \left(A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}\right)\cdot\widehat{\mathbf{r}} = A.$$

It then follows that $A = \dfrac{\partial f}{\partial r}$.

Next we find B by starting with

$$D_{\widehat{\boldsymbol{\theta}}}f = (\mathbf{grad}\, f)\cdot\widehat{\boldsymbol{\theta}}.$$

We have seen that $D_{\widehat{\boldsymbol{\theta}}}f = (\partial f/\partial\theta)/r$. Hence

$$\frac{1}{r}\frac{\partial f}{\partial \theta} = \left(A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}\right)\cdot\widehat{\boldsymbol{\theta}} = B.$$

By the same logic as before, $B = (\partial f/\partial\theta)/r$. Combined, these two results provide a general formula for the gradient of a scalar function in polar coordinates.

Formula 18.8.1: *Gradient in Polar Coordinates*

In polar coordinates, the gradient of $f(r,\theta)$ is given by

$$\mathbf{grad}\, f(r,\theta) = \frac{\partial f}{\partial r}\widehat{\mathbf{r}} + \frac{1}{r}\frac{\partial f}{\partial \theta}\widehat{\boldsymbol{\theta}}.$$

EXAMPLE 1. Find $\mathbf{grad}(r^2\theta^3)$.

SOLUTION $\mathbf{grad}(r^2\theta^3) = \dfrac{\partial}{\partial r}\left(r^2\theta^3\right)\widehat{\mathbf{r}} + \dfrac{1}{r}\dfrac{\partial}{\partial \theta}\left(r^2\theta^3\right)\widehat{\boldsymbol{\theta}} = 2r\theta^3\widehat{\mathbf{r}} + \dfrac{1}{r}\left(3r^2\theta^2\right)\widehat{\boldsymbol{\theta}} = 2r\theta^3\widehat{\mathbf{r}} + 3r\theta^2\widehat{\boldsymbol{\theta}}$. ◆

Divergence in Polar Coordinates

In rectangular coordinates, the divergence of $\mathbf{F}(x,y) = P(x,y)\mathbf{i} + Q(x,y)\mathbf{j}$ is $\dfrac{\partial P}{\partial x} + \dfrac{\partial Q}{\partial y}$. However, in polar coordinates, the divergence of $\mathbf{G}(r,\theta) = A(r,\theta)\widehat{\mathbf{r}} + B(r,\theta)\widehat{\boldsymbol{\theta}}$ is not the sum of $\dfrac{\partial A}{\partial r}$ and $\dfrac{\partial B}{\partial \theta}$.

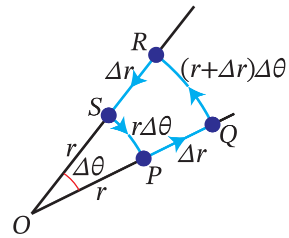


Figure 18.8.3

An informal derivation of $\operatorname{div}\mathbf{G}(r,\theta)$ at (r,θ) can be completed using the coordinate-free description of divergence in the plane:

$$\operatorname{div}\mathbf{G} = \lim_{\text{Length of } C \to 0} \frac{1}{\text{Area bounded by } C} \oint_C \mathbf{G}\cdot\mathbf{n}\, ds. \tag{18.8.1}$$

We are free to choose the small closed curve C to make it easy to estimate the quantities needed in (18.8.1): the flux across C, the length of C, and the area of the region bounded by C. The curve C shown in Figure 18.8.3 is formed by making small changes Δr and $\Delta\theta$ from the point P with polar coordinates (r,θ). It bounds a polar rectangle that will be convenient to use to find the divergence of $\mathbf{G} = A\hat{\mathbf{r}} + B\hat{\boldsymbol{\theta}}$ at P.

To estimate the flux across C, we estimate the flux across the four parts of the curve. Because they are short when Δr and $\Delta\theta$ are small, we may estimate the integral over each part by multiplying the value of the integrand at any point of the section by the length of the section. As usual, $\mathbf{n}$ denotes an exterior unit vector perpendicular to the corresponding section of C.

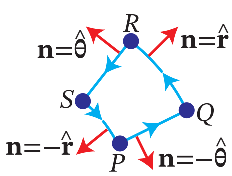


Figure 18.8.4

Since $\operatorname{div}(G) = \operatorname{div}(A\hat{\mathbf{r}}) + \operatorname{div}(B\hat{\boldsymbol{\theta}})$, we can look separately at the contributions from $A\hat{\mathbf{r}}$ and $B\hat{\boldsymbol{\theta}}$.

First, for $A\hat{\mathbf{r}}$. Because $\hat{\mathbf{r}}$ and $\mathbf{n}$ are perpendicular on PQ and RS, both $\int_{PQ} A\hat{\mathbf{r}}\cdot\mathbf{n}\, ds$ and $\int_{RS} A\hat{\mathbf{r}}\cdot\mathbf{n}\, ds$ are zero. (See Figure 18.8.4.)

On SP the radius is r and $\hat{\mathbf{r}}$ and $\mathbf{n}$ are unit vectors that point in directly opposite directions. Hence $\hat{\mathbf{r}}\cdot\mathbf{n}$ is -1. Since the length of SP is $r\Delta\theta$, the flux of $A\hat{\mathbf{r}}$ across SP is approximately

$$A(r,\theta)(\hat{\mathbf{r}}\cdot\hat{\mathbf{n}})r\Delta\theta = -rA(r,\theta)\Delta\theta.$$

On QR the radius is $r+\Delta r$ and $\mathbf{n} = \hat{\mathbf{r}}$, hence $\hat{\mathbf{r}}\cdot\mathbf{n}$ is 1. The contribution on QR, which has length $(r+\Delta r)\Delta\theta$, is approximately

$$A(r+\Delta r,\theta)\ (\hat{\mathbf{r}}\cdot\mathbf{n})\ (r+\Delta r)\ \Delta\theta = (r+\Delta r)\ A(r+\Delta r,\theta)\ \Delta\theta.$$

The approximate total contribution of $A\hat{\mathbf{r}}$ to the flux across C is

$$(r+\Delta r)A(r+\Delta r,\theta)\ \Delta\theta - rA(r,\theta)\ \Delta\theta. \tag{18.8.2}$$

Two geometric properties of the closed curve C are pivotal to evaluating the limit in (18.8.1). First, the area bounded by C is approximately $r\Delta r\Delta\theta$. Second, the length of C, $\Delta r + (r+\Delta r)\Delta\theta + \Delta r + r\Delta\theta$, approaches 0 when both Δr and $\Delta\theta$ approach 0.

With these facts in mind, and (18.8.2), we can evaluate the limit in (18.8.1) when $\mathbf{G} = A\hat{\mathbf{r}}$:

$$\begin{aligned}
\operatorname{div}(A\hat{\mathbf{r}})(P) &= \lim_{\text{Length of } C \to 0} \frac{1}{\text{Area bounded by } C} \oint_C A\hat{\mathbf{r}}\cdot\mathbf{n}\, ds \\
&= \lim_{\Delta r,\ \Delta\theta\to 0} \frac{(r+\Delta r)\ A(r+\Delta r,\theta)\ \Delta\theta - rA(r,\theta)\Delta\theta}{r\Delta r\Delta\theta} \\
&= \frac{1}{r}\lim_{\Delta r,\ \Delta\theta\to 0} \frac{(r+\Delta r)A(r+\Delta r,\theta) - rA(r,\theta)}{\Delta r} \\
&= \frac{1}{r}\frac{\partial(rA)}{\partial r}.
\end{aligned}$$

In addition to the factor $1/r$, the function being differentiated is rA.

We now turn our attention to the flux of $B\hat{\boldsymbol{\theta}}$ across the four segments of C. There is no contribution from QR or SP because $\hat{\boldsymbol{\theta}}\cdot\mathbf{n}$ is 0 on these sides. On RS, $\hat{\boldsymbol{\theta}}$ points in the same direction as $\mathbf{n}$. (See Figure 18.8.4.) Thus, on RS, $\hat{\boldsymbol{\theta}}\cdot\mathbf{n}$ is 1. However, on PQ, $\hat{\boldsymbol{\theta}}$ points in the opposite direction and $\hat{\boldsymbol{\theta}}\cdot\mathbf{n}$ is -1. Across PQ the flux is approximately

$$B(r,\theta)\ (\hat{\boldsymbol{\theta}}\cdot\mathbf{n})\ \Delta r = -B(r,\theta)\Delta r$$

and across RS the flux is approximately $B(r,\theta+\Delta\theta)\Delta r$. The approximate total contribution of $B\widehat{\boldsymbol{\theta}}$ to the flux across C is

$$B(r,\theta+\Delta\theta)\Delta r - B(r,\theta)\Delta r. \tag{18.8.3}$$

When (18.8.3), and the geometric properties of C, are inserted into (18.8.1) with $\mathbf{G} = B\widehat{\boldsymbol{\theta}}$, the limit is easily evaluated as before; its value is: In the same way, (18.8.3) leads to

$$\begin{aligned}
\operatorname{div}(B\widehat{\boldsymbol{\theta}})(P) &= \lim_{\text{Length of } C \to 0} \frac{1}{\text{Area bounded by } C} \oint_C B\widehat{\boldsymbol{\theta}} \cdot \mathbf{n}\, ds \\
&= \lim_{\Delta r,\ \Delta\theta \to 0} \frac{B(r,\theta+\Delta\theta)\Delta r - B(r,\theta)\Delta r}{r\Delta r\Delta\theta} \\
&= \frac{1}{r} \lim_{\Delta r,\ \Delta\theta \to 0} \frac{B(r,\theta+\Delta\theta) - B(r,\theta)}{\Delta\theta} \\
&= \frac{1}{r}\frac{\partial B}{\partial\theta}.
\end{aligned}$$

Here there is only the factor $1/r$; the function being differentiated is simply B.

The sum of these two limits is the divergence of $A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}$ in polar coordinates.

Formula 18.8.2: *Divergence in Polar Coordinates*

In polar coordinates, the divergence of $\mathbf{F}(r,\theta) = A(r,\theta)\widehat{\mathbf{r}} + B(r,\theta)\widehat{\boldsymbol{\theta}}$ is given by

$$\operatorname{div}\mathbf{F} = \operatorname{div}(A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}) = \frac{1}{r}\frac{\partial(rA)}{\partial r} + \frac{1}{r}\frac{\partial B}{\partial\theta}. \tag{18.8.4}$$

DISCLAIMER: We emphasize that the preceding argument is not a mathematically rigorous proof of the divergence in polar coordinates. Such proofs are often addressed in detail in an advanced calculus or multivariate analysis course. (See Exercise 43.) The basic ideas used here are generally useful and illuminating.

Our next objective is to obtain the curl in polar coordinates. Then, the corresponding results in cylindrical and spherical coordinates are stated, with the details of the derivations deferred to Exercises 39 to 42.

Curl in Polar Coordinates

The curl of $\mathbf{F}(x,y) = P(x,y)\mathbf{i} + Q(x,y)\mathbf{j} + 0\mathbf{k}$, a vector field in the plane, is given by

$$\mathbf{curl}\,\mathbf{F} = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k}.$$

What is the formula for the curl when the field is described in polar coordinates: $\mathbf{G}(r,\theta) = A(r,\theta)\widehat{\mathbf{r}} + B(r,\theta)\widehat{\boldsymbol{\theta}}$? To find out we will reason as we did with divergence. This time we use

$$(\mathbf{curl}\,\mathbf{G}) \cdot \mathbf{k} = \lim_{\text{Length of } C \to 0} \frac{1}{\text{Area bounded by } C} \oint_C \mathbf{G} \cdot \mathbf{T}\, ds. \tag{18.8.5}$$

where C is a closed curve around a point in the (r,θ) plane, and the limit is taken as the length of C approaches 0. We will use (18.8.5) to find the $\mathbf{k}$-component of the curl of $\mathbf{G}$ at point P with polar coordinates (r,θ).

We compute the circulation of $\mathbf{G} = A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}$ around the same curve used in the derivation of divergence in polar coordinates.

On SP and QR, $A\widehat{\mathbf{r}}$, being perpendicular to the curve, contributes nothing to the circulation of $\mathbf{G}$ around C. On PQ it contributes approximately

$$A(r,\theta)(\widehat{\mathbf{r}} \cdot \mathbf{T})\Delta r = A(r,\theta)\Delta r.$$

And, on RS, since there $\widehat{\mathbf{r}} \cdot \mathbf{T} = -1$, $A\widehat{\mathbf{r}}$ contributes to the total circulation

$$A(r,\theta+\Delta\theta)(\mathbf{r}\cdot\mathbf{T})\Delta r = -A(r,\theta+\Delta\theta)\Delta r.$$

A similar computation shows the contribution of $B\widehat{\boldsymbol{\theta}}$ towards the total circulation is approximately

$$(r+\Delta r)\ B(r+\Delta r,\theta)\ \Delta\theta - r\ B(r,\theta)\ \Delta\theta.$$

Therefore $(\mathbf{curl}\,\mathbf{G})\cdot\mathbf{k} = (\mathbf{curl}(A\widehat{\mathbf{r}}))\cdot\mathbf{k} + (\mathbf{curl}(B\widehat{\boldsymbol{\theta}}))\cdot\mathbf{k}$ where

$$\begin{aligned}(\mathbf{curl}(A\widehat{\mathbf{r}}))\cdot\mathbf{k} &= \lim_{\Delta r,\ \Delta\theta\to 0} \frac{A(r,\theta)\Delta r - A(r,\theta+\Delta\theta)\Delta r}{r\Delta r\Delta\theta} \\ &= -\frac{1}{r}\lim_{\Delta r,\ \Delta\theta\to 0}\frac{A(r,\theta+\Delta\theta)-A(r,\theta)}{\Delta\theta} \\ &= -\frac{1}{r}\frac{\partial A}{\partial\theta}\end{aligned}$$

and

$$\begin{aligned}(\mathbf{curl}(B\widehat{\boldsymbol{\theta}}))\cdot\mathbf{k} &= \lim_{\Delta r,\ \Delta\theta\to 0}\frac{(r+\Delta r)\ B(r+\Delta r,\theta)\ \Delta\theta - r\ B(r,\theta)\ \Delta\theta}{r\Delta r\Delta\theta} \\ &= \frac{1}{r}\lim_{\Delta r,\ \Delta\theta\to 0}\frac{(r+\Delta r)\ B(r+\Delta r,\theta) - r\ B(r,\theta)}{\Delta r} \\ &= \frac{1}{r}\frac{\partial(rB)}{\partial r}.\end{aligned}$$

Thus, when combined, we have the following formula for the curl in polar coordinates:

Formula 18.8.3: *Curl in Polar Coordinates*

In polar coordinates, the curl of the vector field $\mathbf{F}(r,\theta) = A(r,\theta)\widehat{\mathbf{r}} + B(r,\theta)\widehat{\boldsymbol{\theta}}$ is given by

$$\mathbf{curl}\,\mathbf{F} = \mathbf{curl}(A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}) = \left(-\frac{1}{r}\frac{\partial A}{\partial\theta} + \frac{1}{r}\frac{\partial(rB)}{\partial r}\right)\mathbf{k}. \tag{18.8.6}$$

EXAMPLE 2. Find the divergence and curl of $\mathbf{F} = r\theta^2\widehat{\mathbf{r}} + r^3\tan(\theta)\widehat{\boldsymbol{\theta}}$.

SOLUTION The calculations use (18.8.4) and (18.8.6). First, the divergence:

$$\begin{aligned}\operatorname{div}\mathbf{F} &= \frac{1}{r}\frac{\partial}{\partial r}\left(r\cdot r\theta^2\right) + \frac{1}{r}\frac{\partial}{\partial\theta}\left(r^3\tan(\theta)\right) \\ &= \frac{1}{r}\left(2r\theta^2\right) + \frac{1}{r}\left(r^3\sec^2(\theta)\right) \\ &= 2\theta^2 + r^2\sec^2(\theta).\end{aligned}$$

Then, the curl:

$$\begin{aligned}\mathbf{curl}\,\mathbf{F} &= \left(-\frac{1}{r}\frac{\partial}{\partial\theta}\left(r\theta^2\right) + \frac{1}{r}\frac{\partial}{\partial r}\left(r\cdot r^3\tan(\theta)\right)\right)\mathbf{k} \\ &= \left(-\frac{1}{r}(2r\theta) + \frac{1}{r}\left(4r^3\tan(\theta)\right)\right)\mathbf{k} \\ &= \left(-2\theta + 4r^2\tan(\theta)\right)\mathbf{k}.\end{aligned}$$

◆

Cylindrical Coordinates

To work in cylindrical coordinates let g be a scalar function of r, θ, and z and let $G(r,\theta,z) = G_r(r,\theta,z)\widehat{\mathbf{r}}+G_\theta(r,\theta,z)\widehat{\boldsymbol{\theta}}+G_z(r,\theta,z)\mathbf{k}$. The following formulas are obtained by reasoning as we did in polar coordinates.

Formula 18.8.4: *Gradient in Cylindrical Coordinates*

In cylindrical coordinates, the gradient of $g(r,\theta,z)$ is given by

$$\mathbf{grad}\, g = \frac{\partial g}{\partial r}\widehat{\mathbf{r}} + \frac{1}{r}\frac{\partial g}{\partial \theta}\widehat{\boldsymbol{\theta}} + \frac{\partial g}{\partial z}\mathbf{k} \tag{18.8.7}$$

Formula (18.8.7) differs from the case of polar coordinates only by the extra term $(\partial g/\partial z)\mathbf{k}$. Its derivation is similar to the one for the formula for the gradient in polar coordinates. (See Exercise 39.)

To obtain the formula for the divergence of $\mathbf{G}$ in cylindrical coordinates, use the relation between $\operatorname{div}\mathbf{G}$ and the flux across the small surface determined by small changes Δr, $\Delta\theta$, and Δz. (See Exercise 41.)

Formula 18.8.5: *Divergence in Cylindrical Coordinates*

In cylindrical coordinates, the divergence of the vector field $\mathbf{G}(r,\theta,z) = G_r\widehat{\mathbf{r}} + G_\theta\widehat{\boldsymbol{\theta}} + G_z\mathbf{k}$ is given by

$$\operatorname{div}\mathbf{G} = \frac{1}{r}\frac{\partial(rG_r)}{\partial r} + \frac{1}{r}\frac{\partial G_\theta}{\partial\theta} + \frac{\partial G_z}{\partial z}. \tag{18.8.8}$$

The formula for the curl of $\mathbf{G}$ in cylindrical coordinates is obtained by considering the circulation around three small closed curves lying in planes perpendicular to $\widehat{\mathbf{r}}$, $\widehat{\boldsymbol{\theta}}$, and $\mathbf{k}$. (See Exercise 40.)

Formula 18.8.6: *Curl in Cylindrical Coordinates*

In cylindrical coordinates, the curl of the vector field $\mathbf{G} = G_r\widehat{\mathbf{r}} + G_\theta\widehat{\boldsymbol{\theta}} + G_z\mathbf{k}$ is given by a determinant:

$$\begin{aligned}\mathbf{curl}\,\mathbf{G} &= \frac{1}{r}\det\begin{pmatrix} \widehat{\mathbf{r}} & r\widehat{\boldsymbol{\theta}} & \mathbf{k} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial\theta} & \frac{\partial}{\partial z} \\ G_r & rG_\theta & G_z \end{pmatrix} \\ &= \frac{1}{r}\left(\frac{\partial G_z}{\partial\theta} - \frac{\partial(rG_\theta)}{\partial z}\right)\widehat{\mathbf{r}} + \left(\frac{\partial G_r}{\partial z} - \frac{\partial G_z}{\partial r}\right)\widehat{\boldsymbol{\theta}} + \frac{1}{r}\left(\frac{\partial(rG_\theta)}{\partial r} - \frac{\partial G_r}{\partial\theta}\right)\mathbf{k}\end{aligned}$$

Spherical Coordinates

Warning: Notational Differences Between Mathematics and Physics/Engineering

In mathematics texts, spherical coordinates are typically denoted ρ, ϕ, θ. In physics and engineering a different notation is standard. There ρ is replaced by r, θ is the angle with z-axis, and ϕ plays the role of θ; that is, the roles of ϕ and θ are interchanged. The formulas we state are in the mathematicians' notation.

The three basic unit vectors for spherical coordinates are denoted $\widehat{\boldsymbol{\rho}}$, $\widehat{\boldsymbol{\theta}}$, and $\widehat{\boldsymbol{\phi}}$. (See Figure 18.8.5(a).) For instance, $\widehat{\boldsymbol{\rho}}$ points in the direction of increasing ρ. Note that at the point P, $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\phi}}$ are tangent to the sphere through P and centered at the origin, while $\widehat{\boldsymbol{\rho}}$ is perpendicular to that sphere. Also, any two of $\widehat{\boldsymbol{\rho}}$, $\widehat{\boldsymbol{\theta}}$, and $\widehat{\boldsymbol{\phi}}$ are perpendicular.

To obtain the formula for the three components of $\operatorname{div}\mathbf{G}$, use the solid region corresponding to small changes $\Delta\rho$, $\Delta\theta$, and $\Delta\phi$, shown in Figure 18.8.5(b) and (c). To obtain the formula for $\mathbf{curl}\,\mathbf{G}$, use three of the surfaces of that region. Those computations yield the following formulas:

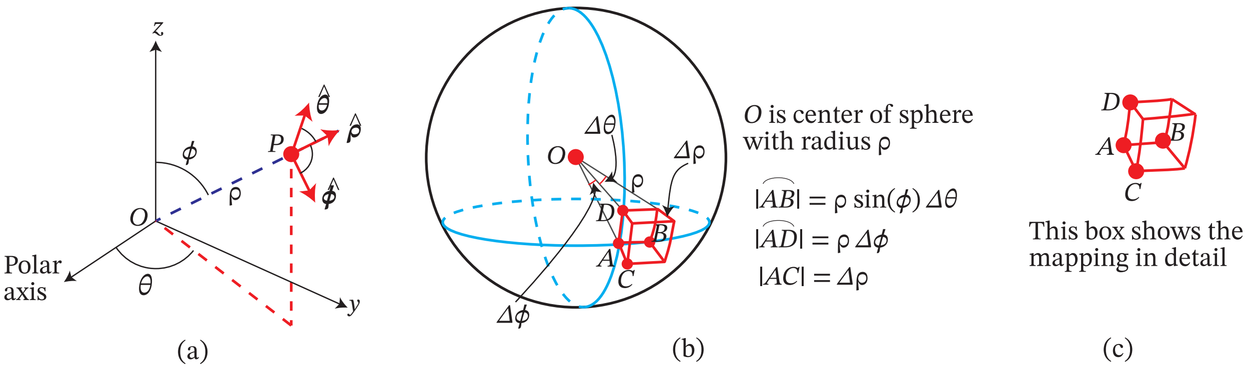


Figure 18.8.5

Formula 18.8.7: ***Gradient, Divergence, and Curl in Spherical Coordinates***

In spherical coordinates, the gradient of the scalar function $g(\rho,\theta,\phi)$ is given by

$$\mathbf{grad}\, g = \frac{\partial g}{\partial \rho}\widehat{\boldsymbol{\rho}} + \frac{1}{\rho}\frac{\partial g}{\partial \phi}\widehat{\boldsymbol{\phi}} + \frac{1}{\rho\sin(\phi)}\frac{\partial g}{\partial \theta}\widehat{\boldsymbol{\theta}}.$$

In spherical coordinates, the divergence and curl of the vector field $\mathbf{G}(\rho,\theta,\phi) = G_\rho\widehat{\boldsymbol{\rho}} + G_\phi\widehat{\boldsymbol{\phi}} + G_\theta\widehat{\boldsymbol{\theta}}$ are given by

$$\operatorname{div}\mathbf{G} = \frac{1}{\rho^2}\frac{\partial(\rho^2 G_\rho)}{\partial\rho} + \frac{1}{\rho\sin(\phi)}\frac{\partial G_\theta}{\partial\theta} + \frac{1}{\rho\sin(\phi)}\frac{\partial(\sin(\phi)G_\phi)}{\partial\phi}$$

and

$$\begin{aligned}\mathbf{curl}\,\mathbf{G} &= \frac{1}{\rho^2\sin(\phi)}\det\begin{pmatrix} \widehat{\boldsymbol{\rho}} & \rho\widehat{\boldsymbol{\phi}} & \rho\sin(\phi)\widehat{\boldsymbol{\theta}} \\ \dfrac{\partial}{\partial\rho} & \dfrac{\partial}{\partial\phi} & \dfrac{\partial}{\partial\theta} \\ G_\rho & \rho G_\phi & \rho\sin(\phi)G_\theta \end{pmatrix} \\ &= \frac{1}{\rho\sin(\phi)}\left(\frac{\partial}{\partial\phi}\left(\sin(\phi)G_\theta\right) - \frac{\partial G_\phi}{\partial\theta}\right)\widehat{\boldsymbol{\rho}} + \frac{1}{\rho}\left(\frac{1}{\sin(\phi)}\frac{\partial G_\rho}{\partial\theta} - \frac{\partial}{\partial\rho}\left(\rho G_\theta\right)\right)\widehat{\boldsymbol{\phi}} + \frac{1}{\rho}\left(\frac{\partial}{\partial\rho}\left(\rho G_\phi\right) - \frac{\partial G_\rho}{\partial\phi}\right)\widehat{\boldsymbol{\theta}}.\end{aligned}$$

Each of the above three formulas can be obtained by the method we used for polar coordinates. The change in ϕ or θ is not the same as the distance the point moves. However, a change in ρ is the same as the distance the point moves. When, in Section 17.5, we set up iterated integrals for spherical coordinates, we found the distance between (ρ,ϕ,θ) and $(\rho,\phi+\Delta\phi,\Delta\theta)$ is approximately $\rho\Delta\phi$ and the distance between (ρ,ϕ,θ) and $(\rho,\phi,\theta+\Delta\theta)$ is approximately $\rho\sin(\phi)\Delta\theta$. (See Exercise 33.) With these differences in mind, the derivations of the derivative operators in spherical coordinates are not difficult to complete. The easiest of the three is the gradient, which is obtained in Exercise 42.

An Application to Rotating Fluids

If a fluid is rotating in a cylinder, for instance, in a centrifuge, it rotates as a rigid body and its velocity at a distance r from the axis of rotation is $\mathbf{G}(r,\theta) = cr\widehat{\boldsymbol{\theta}}$, where c is a positive constant describing the angular speed.

Then

$$\mathbf{curl}\,\mathbf{G} = \frac{1}{r}\frac{\partial(cr^2)}{\partial r}\mathbf{k} = 2c\mathbf{k}.$$

The curl is independent of r. Now imagine that a paddle wheel is placed in the fluid with its axis parallel to the axis of the cylinder. The wheel is kept in the same place but is free to rotate about its axis. That the curl is constant says that the wheel will turn at the same rate no matter how far it is from the axis of the cylinder.

In the more general case with $\mathbf{G}(r,\theta) = cr^n\widehat{\boldsymbol{\theta}}$, with n an integer, we have

$$\mathbf{curl}\,\mathbf{G} = \frac{1}{r}\frac{\partial(cr^{n+1})}{\partial r}\mathbf{k} = c(n+1)r^{n-1}\mathbf{k}.$$

We just considered $n = 1$. For $n > 1$ the curl increases as r increases. The paddle wheel rotates faster if placed farther from the axis of rotation. The direction of rotation is the same as that of the fluid, counterclockwise.

When $n = -2$ the speed of the fluid decreases as r increases and

$$\mathbf{curl}\,\mathbf{G} = c(-2+1)r^{-2-1}\mathbf{k} = -cr^{-3}\mathbf{k}.$$

The minus sign on the right-hand side means the paddle wheel spins clockwise even though the fluid rotates counterclockwise. The farther the paddle wheel is from the axis, the slower it rotates.

When $n = -1$, we have the case that approximates water exiting a sink or bathtub. A paddlewheel placed in the whirling water will tend not to rotate.

Summary

We expressed three derivative operators, gradient, divergence, and curl, in polar, cylindrical, and spherical coordinate systems. Though the basic unit vectors in each system may change direction from point to point, they remain perpendicular to each other. That simplified the computation of directional derivatives, flux, and circulation. The formulas are more complicated than those in rectangular coordinates because the amount a parameter changes is not always the same as the distance the corresponding point moves.

EXERCISES for Section 18.8

In Exercises 1 through 4 find and draw the gradient of the function of (r,θ) at $\left(2, \frac{\pi}{4}\right)$.

1. r **2.** $r^2\theta$ **3.** $e^{-r}\theta$ **4.** $r^3\theta^2$

In Exercises 5 through 8 find the divergence of the function.

5. $5\widehat{\mathbf{r}} + r^2\theta\widehat{\boldsymbol{\theta}}$ **6.** $r^3\theta\widehat{\mathbf{r}} + 3r\theta\widehat{\boldsymbol{\theta}}$ **7.** $r\widehat{\mathbf{r}} + r^3\widehat{\boldsymbol{\theta}}$ **8.** $r\sin(\theta)\widehat{\mathbf{r}} + r^2\cos(\theta)\widehat{\boldsymbol{\theta}}$

In Exercises 9 through 12 compute the curl of the function.

9. $r\widehat{\boldsymbol{\theta}}$ **10.** $r^3\theta\widehat{\mathbf{r}} + e^r\widehat{\boldsymbol{\theta}}$ **11.** $r\cos(\theta)\widehat{\mathbf{r}} + r\theta\widehat{\boldsymbol{\theta}}$ **12.** $\frac{1}{r^3}\widehat{\boldsymbol{\theta}}$

13. Find the directional derivative of $r^2\theta^3$ in the direction (a) $\widehat{\mathbf{r}}$ and (b) $\widehat{\boldsymbol{\theta}}$.

14. What property of rectangular coordinates makes the formulas for gradient, divergence, and curl in those coordinates relatively simple?

15. Where did we use the fact that $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$ are perpendicular when developing the expression for divergence in polar coordinates?

16. Estimate the flux of $\mathbf{F}(r,\theta) = r^2\theta^3\widehat{\boldsymbol{\theta}}$ across the circle of radius 0.01 with center at $(r,\theta) = (2,\pi/6)$.

17. Estimate the circulation of the field in the preceding exercise around the same circle.

When translating between rectangular and polar coordinates, it may be necessary to express $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$ in terms of $\mathbf{i}$ and $\mathbf{j}$ and also $\mathbf{i}$ and $\mathbf{j}$ in terms of $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$. Exercises 18 and 19 concern this matter.

18. Let (r,θ) be a point that has rectangular coordinates (x,y).

(a) Show that $\widehat{\mathbf{r}} = \cos(\theta)\mathbf{i} + \sin(\theta)\mathbf{j} = \dfrac{x\mathbf{i} + y\mathbf{j}}{\sqrt{x^2+y^2}}$.

(b) Show that $\widehat{\boldsymbol{\theta}} = -\sin(\theta)\mathbf{i} + \cos(\theta)\mathbf{j} = \dfrac{-y\mathbf{i} + x\mathbf{j}}{\sqrt{x^2+y^2}}$.

(c) Draw a picture to accompany the calculations done in (a) and (b).

19. Show that if (x,y) has polar coordinates (r,θ), then $\mathbf{i} = \cos(\theta)\,\widehat{\mathbf{r}} - \sin(\theta)\,\widehat{\boldsymbol{\theta}}$ and $\mathbf{j} = \sin(\theta)\,\widehat{\mathbf{r}} + \cos(\theta)\,\widehat{\boldsymbol{\theta}}$.

In Exercises 20 through 23 (a) Find the gradient of the function, using the formula for gradient in rectangular coordinates and expressing the result in polar coordinates. (b) Find the gradient by first expressing the function in polar coordinates and using the formula for the gradient in polar coordinates. (c) Show that the gradients found in (a) and (b) agree.

20. x^2+y^2 **21.** $\sqrt{x^2+y^2}$ **22.** $3x+2y$ **23.** $\dfrac{x}{\sqrt{x^2+y^2}}$

In Exercises 24 through 27 (a) Find the gradient of the function, using its formula in polar coordinates. (b) Find the gradient by expressing the function in rectangular coordinates and then expressing the result in polar coordinates. (c) Show that the gradients in (a) and (b) agree.

24. r^2 **25.** $r^2\cos(\theta)$ **26.** $r\sin(\theta)$ **27.** e^r

---

In Exercises 28 and 29 (a) Find the divergence of the vector field in rectangular coordinates. (b) Find the divergence by expressing the function in polar coordinates. (c) Show that the divergences in (a) and (b) agree.

28. $x^2\mathbf{i} + y^2\mathbf{j}$ **29.** $xy\mathbf{i}$

In Exercises 30 and 31 (a) Find the curl of the vector field in rectangular coordinates, (b) Find the curl by expressing the function in polar coordinates. (c) Show that the curls found in (a) and (b) agree.

30. $xy\mathbf{i} + x^2y^2\mathbf{j}$ **31.** $\dfrac{x}{\sqrt{x^2+y^2}}\mathbf{i}$

Exercises 32 and 33 are useful in developing the formula for the gradient in cylindrical and spherical coordinates.

32. In cylindrical coordinates, find the distance from the point (r,θ,z) to each of the points (a) $(r+\Delta r,\theta,z)$, (b) $(r,\theta+\Delta\theta,z)$, and (c) $(r,\theta,z+\Delta z)$.

33. In spherical coordinates, approximate the distance from the point (ρ,θ,ϕ) to each of the points (a) $(\rho+\Delta\rho,\theta,\phi)$, (b) $(\rho,\theta+\Delta\theta,\phi)$, and (c) $(\rho,\theta,\phi+\Delta\phi)$.

---

34. Using the formula for the gradient of $g(\rho,\phi,\theta)$ in spherical coordinates, find the directional derivative of g in the direction (a) $\widehat{\boldsymbol{\rho}}$, (b) $\widehat{\boldsymbol{\theta}}$, and (c) $\widehat{\boldsymbol{\phi}}$.

35. Using the formula for the gradient of $g(r,\theta,z)$ in cylindrical coordinates, find the directional derivative of g in the direction (a) $\widehat{\mathbf{r}}$, (b) $\widehat{\boldsymbol{\theta}}$, and (c) $\mathbf{k}$.

36. Using as few mathematical symbols as you can, state the formula for the divergence of a vector field given in polar coordinates.

37. Using as few mathematical symbols as you can, state the formula for the curl of a vector field given in polar coordinates.

38. In the formula for the divergence of $A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}$, why do the terms rA and $\frac{1}{r}$ appear in $\frac{1}{r}\frac{\partial}{\partial r}(rA)$?

39. Obtain the formula for the gradient in cylindrical coordinates.

40. Obtain the formula for curl in cylindrical coordinates.

41. Obtain the formula for divergence in cylindrical coordinates.

42. Obtain the formula for the gradient in spherical coordinates.

In Exercise 43 you will obtain the formula for the gradient of $g(r,\theta)$ in polar coordinates by starting with the formula for the gradient of $f(x, y)$ in rectangular coordinates. It is not surprising that, during the calculations, some terms cancel and that $\cos^2(\theta) + \sin^2(\theta) = 1$ simplifies expressions.

43. Assume $g(r,\theta) = f(x, y)$, where $x = r\cos(\theta)$ and $y = r\sin(\theta)$. To express $\nabla f = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j}$ in terms of polar coordinates, it is necessary to express $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\mathbf{i}$, and $\mathbf{j}$ in terms of partial derivative of $g(r,\theta)$, $\widehat{\mathbf{r}}$, and $\widehat{\boldsymbol{\theta}}$.

(a) Show that $\frac{\partial r}{\partial x} = \cos(\theta)$, $\frac{\partial r}{\partial y} = \sin(\theta)$, $\frac{\partial \theta}{\partial x} = -\sin(\theta)/r$, and $\frac{\partial \theta}{\partial y} = \cos(\theta)/r$.

(b) Use the chain rule to express $\frac{\partial r}{\partial x}$ and $\frac{\partial r}{\partial y}$ in terms of partial derivatives of $g(r,\theta)$.

(c) Use the results of Exercise 19 to obtain the gradient of $g(r,\theta)$ in polar coordinates.

---

44. In Exercise 26 of Section 18.3 we found the divergence of $\mathbf{F} = r^n\widehat{\mathbf{r}}$ using rectangular coordinates. Find the divergence using polar coordinates. Compare and contrast the two approaches.

45. Use the formula for divergence in polar coordinates to show that an incompressible central field in the plane must have the form $\mathbf{F}(\mathbf{r}) = \frac{k}{r}\widehat{\mathbf{r}}$.

NOTE: Exercise 5 in Section 18.4 is the converse of Exercise 45 (and in rectangular coordinates).

46. (a) Use a diagram to show that $\widehat{\mathbf{r}}$, $\widehat{\boldsymbol{\theta}}$, and $\widehat{\mathbf{z}}$ are mutually perpendicular.

(b) Use a diagram to show that $\widehat{\boldsymbol{\rho}}$, $\widehat{\boldsymbol{\theta}}$, and $\widehat{\boldsymbol{\phi}}$ are mutually perpendicular.

Definition: *Orthogonal Coordinate Systems*

Coordinate systems whose basic unit vectors are mutually perpendicular are called **orthogonal**. Orthogonality is useful in developing the formulas for **grad**, div, and **curl** in such systems.

Exercise 46 shows that cylindrical and spherical coordinates are orthogonal coordinate systems.

47. In this problem you will derive the formula for the flux of $B\widehat{\boldsymbol{\theta}}$ across a curve C in polar coordinates.

(a) Show that the flux of $B\widehat{\boldsymbol{\theta}}$ across C (see Figure 18.8.3 is $\oint_C B\widehat{\boldsymbol{\theta}}\cdot\mathbf{n}\,ds = \int_r^{r+\Delta r} \left(B(u,\theta+\Delta\theta) - B(u,\theta)\right)\,du$.

(b) Use the mean value theorem to show that the second integrand in (a) equals $-\frac{\partial B}{\partial \theta}(u, d(u))\,\Delta\theta$ for some $d(u)$ between θ and $\theta + \Delta\theta$.

(c) Deduce that $\nabla \cdot (B\widehat{\boldsymbol{\theta}})(P) = \lim_{\Delta r,\, \Delta\theta \to 0} \dfrac{\dfrac{\partial B}{\partial \theta}(r,\theta)\Delta r \Delta\theta}{r\, \Delta r\, \Delta\theta} = \dfrac{1}{r}\dfrac{\partial B}{\partial \theta}(r,\theta)$.

(d) From (a) and the formula for $\nabla \cdot (B\widehat{\boldsymbol{\theta}})(P)$, obtain the formula for divergence in polar coordinates.

48. In this problem you will derive the formula for the curl of $A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}$ in polar coordinates.

(a) Show that $A(u,\theta+\Delta\theta) - A(u,\theta) = \dfrac{\partial A}{\partial u}(u, c(u))\, \Delta\theta$ where $c(u)$ is between θ and $\theta + \Delta\theta$.

(b) Deduce that $\mathbf{curl}(A\widehat{\mathbf{r}}) \cdot \mathbf{k} = \lim_{\Delta r, \Delta\theta \to 0} \dfrac{-\dfrac{\partial A}{\partial \theta}(r,\theta)\Delta r \Delta\theta}{r \Delta r \Delta\theta} = -\dfrac{1}{r}\dfrac{\partial A}{\partial \theta}(r,\theta)$.

(c) Show that $\mathbf{curl}(B\widehat{\boldsymbol{\theta}}) \cdot \mathbf{k} = \dfrac{1}{r}\dfrac{\partial (rB)}{\partial r}(r,\theta)$.

(d) From (b) and (c) deduce the formulae for curl in polar coordinates.

49. **SAM:** I can prove the formula for $\nabla \cdot (A\widehat{\mathbf{r}})$ in polar coordinates without using the mean value theorem.

JANE: How?

SAM: First I estimate the flux across SP by approximating A on that arc by its value at P. Then I estimate the flux across QR by approximating A by its value at Q. When I add those, and divide by the area, I get the limit found in the definition of $\dfrac{1}{r}\dfrac{dA}{dr}$.

JANE: Nice, but how do you justify those approximations?

SAM: Well, when Δr and $\Delta\theta$ are small, both arcs are short, so I can approximate A on each side by its value anywhere on the arc.

JANE: So what happens if you approximate A on SP and QR by its values at P and R instead of P and Q?

SAM: I haven't tried that, but I'm sure you get the same result.

(a) Fill in the details of Sam's argument and of Jane's suggested modification.

(b) Is Sam right?

(c) Is Jane's modification any better?

18.9 Maxwell's Equations

At any point in space there is an electric field **E** and a magnetic field **B**. The electric field is due to charges (electrons and protons), whether stationary or moving. The magnetic field is due to moving charges.

To assure yourself that the magnetic field **B** is everywhere, hold up a pocket compass. The Earth's magnetic field moves the needle so it points north.

All electrical phenomena and their applications can be described by four equations called **Maxwell's equations**. In this context the fields **B** and **E** are functions of both time and location. We state them here for the simpler case when **B** and **E** do not depend on time: $\partial \mathbf{B}/\partial t = \mathbf{0}$ and $\partial \mathbf{E}/\partial t = \mathbf{0}$. We met the first equation, Gauss's law of electricity, in Section 18.7. Here are all four equations:

Definition: *Maxwell's Equations – Global Form*

The following four equations are know as **Maxwell's equations** for an electric field **E** and magnetic field **B** that do not depend on time:

(I) **Gauss's law for electricity**: $\int_{\mathcal{S}} \mathbf{E} \cdot \mathbf{n}\, dS = Q/\epsilon_0$, where $\mathcal{S}$ is a surface bounding a spatial region and Q is the charge in that region.

(II) **Faraday's law of induction**: $\oint_C \mathbf{E} \cdot d\mathbf{r} = 0$ for a closed curve C.

(III) **Gauss's law for magnetism**: $\int_{\mathcal{S}} \mathbf{B} \cdot \mathbf{n}\, dS = 0$ for a surface $\mathcal{S}$ that bounds a spatial region.

(IV) **Ampere's law**: $\oint_C \mathbf{B} \cdot d\mathbf{r} = \mu_0 \int_{\mathcal{S}} \mathbf{J} \cdot \mathbf{n}\, dS$, where C bounds the surface $\mathcal{S}$ and **J** is the electric current flowing through $\mathcal{S}$.

The statements of these equations in terms of integrals can be translated into relationships between the fields **E** or **B** at each point. The following differential equations are the equivalent local forms:

Definition: *Maxwell's Equations — Local Form*

The following four differential equations are also known as Maxwell's equations.

(I') $\operatorname{div}\mathbf{E} = \rho/\epsilon_0$, where ρ is the charge density (**Coulomb's Law**)

(II') $\mathbf{curl}\,\mathbf{E} = \mathbf{0}$

(III') $\operatorname{div}\mathbf{B} = 0$

(IV') $\mathbf{curl}\,\mathbf{B} = \mu_0\mathbf{J}$

In either form, these equations involve two fundamental constants. The vacuum permittivity, ϵ_0, was introduced in Section 18.7. The other constant, μ_0 (pronounced "mew naught") is called the **vacuum permeability**,

$$\mu_0 = 4\pi \times 10^{-7}\ \text{Henry/m} \approx 1.25664 \times 10^{-6}\ \text{Newton/Ampere}^2.$$

It turns out that $1/(\mu_0\epsilon_0)$ equals the square of the speed of light, c^2. Maxwell's discovery of this fact is an astonishing story told in CIE 26, How Maxwell Did It, at the end of this chapter.

Observation 18.9.1: *Comparing Global and Local Forms*

The global, or nonlocal, form of Maxwell's equations involve surface and/or line integrals of **E** and **B**. In contrast, the local form of Maxwell's equations are differential equations involving the divergence or curl of **E** and **B** that applies pointwise.

The remainder of this section (Examples 1 to 3) and the first group of exercises (Exercises 1 to 5) show that each of the four equations in the global form is equivalent to the corresponding equation in local form. The time-dependent Maxwell's equations are introduced in Exercises 12 to 15.

Going Back and Forth Between the Local and Global Forms of Maxwell's Equations

Examples 1 and 2 show that Gauss's law for electricity is equivalent to Coulomb's law.

EXAMPLE 1. Obtain Gauss's law for electricity (I) from Coulomb's law (I').

SOLUTION Let $\mathcal{V}$ be any solid region whose boundary is the surface S. Then

$$\begin{aligned}
\int\limits_{\mathcal{S}} \mathbf{E}\cdot\mathbf{n}\,dS &= \int\limits_{\mathcal{V}} \nabla\cdot\mathbf{E}\,dV && (\text{ divergence theorem })\\
&= \int\limits_{\mathcal{V}} \frac{\rho}{\epsilon_0}\,dV && (\text{ Coulomb's law })\\
&= \frac{1}{\epsilon_0}\int\limits_{\mathcal{V}} \rho\,dV && (\text{ simplification })\\
&= \frac{Q}{\epsilon_0} && (\text{ charge is } Q = \int\limits_{\mathcal{V}} \rho\,dV\text{ }).
\end{aligned}$$

◆

EXAMPLE 2. Obtain Coulomb's law (I') from Gauss's law for electricity (I).

SOLUTION Let $\mathcal{V}$ be any region in space bounded by a surface $\mathcal{S}$. Let Q be the total charge in $\mathcal{V}$. Then

$$\begin{aligned}
\frac{Q}{\epsilon_0} &= \int\limits_{\mathcal{S}} \mathbf{E}\cdot\mathbf{n}\,dS && (\text{ Gauss's law })\\
&= \int\limits_{\mathcal{V}} \nabla\cdot\mathbf{E}\,dV && (\text{ divergence theorem }).
\end{aligned}$$

On the other hand, $Q = \int_{\mathcal{V}} \rho \, dV$, where ρ is the charge density. Thus

$$\int_{\mathcal{V}} \frac{\rho}{\epsilon_0} \, dV = \int_{\mathcal{V}} \nabla \cdot \mathbf{E} \, dV \qquad \text{or} \qquad \int_{\mathcal{V}} \left(\frac{\rho}{\epsilon_0} - \nabla \cdot \mathbf{E} \right) dV = 0,$$

for all spatial regions. Since the integrand is assumed to be continuous, the zero-integral principle in space (Theorem 18.5.4) implies the integrand is identically 0 throughout $\mathcal{V}$. That is,

$$\frac{\rho}{\epsilon_0} - \nabla \cdot \mathbf{E} = 0,$$

which gives us Coulomb's law. ♦

EXAMPLE 3. Show that (II) implies (II'). That is, $\oint\limits_C \mathbf{E} \cdot d\mathbf{r} = 0$ for any closed curve C implies $\mathbf{curl}\,\mathbf{E} = \mathbf{0}$.

SOLUTION We know that $\oint_C \mathbf{E} \cdot d\mathbf{r} = 0$ for any closed curve C. Therefore, by Stokes' theorem, for any orientable surface $\mathcal{S}$ bounded by a closed curve,

$$\int_{\mathcal{S}} (\mathbf{curl}\,\mathbf{E}) \cdot \mathbf{n} \, dS = 0.$$

The zero-integral principle in space (Theorem 18.5.4) implies that $(\mathbf{curl}\,\mathbf{E}) \cdot \mathbf{n} = 0$ at each point on the surface. Choosing $\mathcal{S}$ such that $\mathbf{n}$ is parallel to $\mathbf{curl}\,\mathbf{E}$ (if $\mathbf{curl}\,\mathbf{E}$ is not $\mathbf{0}$), implies that the magnitude of $\mathbf{curl}\,\mathbf{E}$ is 0, hence $\mathbf{curl}\,\mathbf{E}$ is $\mathbf{0}$. ♦

The exercises present the analogy of the equations in integral form for the general case where **B** and **E** vary with time. In this generality they are known as **Maxwell's equations**, in honor of James Clerk Maxwell (1831-1879), who announced his findings in their final form in 1865.

Historical Note: Mathematics and Electricity

Benjamin Franklin, in his book *Experiments and Observations Made in Philadelphia*, published in 1751, made electricity into a science. For his accomplishments, he was elected a Foreign Associate of the French Academy of Sciences, an honor bestowed on no other American for over a century. In 1865, Maxwell completed the theory that Franklin had begun.

At the time that Newton published his *Principia* on gravity (1687) electricity and magnetism were subjects of little scientific study. The experiments of Franklin, Oersted, Henry, Ampère, Faraday, Volta, and others in the late eighteenth and early nineteenth centuries gradually built up a mass of information leading to interesting mathematical analysis. All the phenomena could be summarized in four equations, which in their final form appeared in Maxwell's *Treatise on Electricity and Magnetism*, published in 1873.

Summary

We stated the four Maxwell equations that describe electrostatic and magnetic fields that do not vary with time in both their global form and their local form. Then we showed how to use the divergence theorem or Stokes' theorem to translate between their global and local forms. The exercises include the four Maxwell equations in their general form, where **E** and **B** vary with time.

EXERCISES for Section 18.9

Exercises 1 to 5 complete the proof that the global and local forms of Maxwell's equations are equivalent.

See also Examples 1 to 3.

1. Obtain (II) from (II').

2. Obtain (III') from (III).

3. Obtain (III) from (III').

4. Obtain (IV') from (IV).

5. Obtain (IV) from (IV').

In Exercises 6 to 9 use terms such as circulation, flux, current, and charge density to express the equation in words.

6. (I)

7. (II)

8. (III)

9. (IV)

10. Which of Maxwell's laws imply that there is a relationship between an electric current and a magnetic field?

11. Which of Maxwell's laws imply that there is a relationship between an electric field and charge?

We have assumed that the fields **E** and **B** do not vary in time, that is, $\frac{\partial \mathbf{E}}{\partial t} = \mathbf{0}$ and $\frac{\partial \mathbf{B}}{\partial t} = \mathbf{0}$. The general case, in empty space, where **E** and **B** depend on time, is also described by four equations, which we call (i), (ii), (iii), and (iv).

Definition: ***Time-Dependent Maxwell's Equations***

(i) $\operatorname{div} \mathbf{E} = \frac{\rho}{\epsilon_0}$, where ρ is the charge density

(ii) $\mathbf{curl}\, \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$

(iii) $\operatorname{div} \mathbf{B} = 0$

(iv) $\mathbf{curl}\, \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}$, where **J** is the current density.

NOTE: Equations (i) and (iii) do not involve time; these two equations are identical to (I') and (III').

Exercises 12 to 15 are about these equations.

12. Which equation provides a relationship between a changing magnetic field and an electric field?

13. Which equation provides a relationship between a changing electric field and a magnetic field?

14. Show that (ii) is equivalent to $\oint_C \mathbf{E} \cdot d\mathbf{r} = -\frac{\partial}{\partial t} \int_{\mathscr{S}} \mathbf{B} \cdot \mathbf{n}\, dS$. Here, C bounds $\mathscr{S}$.

15. Show that (iv) is equivalent to $\oint_C \mathbf{B} \cdot d\mathbf{r} = \mu_0 \int_{\mathscr{S}} \mathbf{J} \cdot \mathbf{n}\, dS + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \int_{\mathscr{S}} \mathbf{E} \cdot \mathbf{n}\, dS$.

NOTE: The circulation of **B** is related to the total current through the surface $\mathscr{S}$ that C bounds and to the rate at which the flux of **E** through $\mathscr{S}$ changes.

18.S Chapter Summary

The first six sections developed three theorems: Green's theorem, Gauss's theorem (also called the divergence theorem), and Stokes' theorem. The other sections applied them to nonrectangular coordinate systems and to physics.

Green's theorem can be viewed as the planar version of both the divergence theorem and Stokes' theorem.

Name	Mathematical Expression	Description
Green's theorem	$\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \int_{\mathscr{R}} \operatorname{div}\mathbf{F}\,dA$	flux of **F** across C
	$\oint_C (-Q\,dx + P\,dy) = \int_{\mathscr{R}} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) dA$	differential form for flux across C
Green's theorem	$\oint_C \mathbf{F}\cdot\mathbf{T}\,ds = \oint_C \mathbf{F}\cdot d\mathbf{r} = \int_{\mathscr{R}} (\mathbf{curl}\,\mathbf{F})\cdot\mathbf{k}\,dA$	circulation of **F** around C
	$\oint_C (P\,dx + Q\,dy) = \int_{\mathscr{R}} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA$	differential form for circulation around C
Gauss' theorem (divergence theorem)	$\int_{\mathscr{S}} \mathbf{F}\cdot\mathbf{n}\,dS = \int_{\mathscr{R}} \operatorname{div}\mathbf{F}\,dV$	flux across a closed surface equals the integral of the divergence over the spatial region it bounds
Stokes' theorem	$\oint_C \mathbf{F}\cdot\mathbf{T}\,ds = \int_{\mathscr{S}} (\mathbf{curl}\,\mathbf{F})\cdot\mathbf{n}\,dS$ (surface $\mathscr{S}$ has boundary curve C with **n** compatible with orientation of C)	circulation along a curve equals the surface integral over the surface it bounds of the component of the curl perpendicular to the surface

Table 18.S.1

Though $\operatorname{div}\mathbf{F}$ and $\mathbf{curl}\,\mathbf{F}$ were defined in rectangular coordinates, they have a meaning that is independent of coordinates. For instance, if **F** is a vector field in space, the divergence of **F** at a point multiplied by the volume of a small region containing the point approximates the flux of **F** across the surface of the small region. More precisely,

$$\operatorname{div}\mathbf{F}(P) = \lim_{\mathscr{R}\to P} \frac{1}{\text{Volume of } \mathscr{R}} \int_{\mathscr{S}} \mathbf{F}\cdot\mathbf{n}\,dS$$

where $\mathscr{S}$ is the boundary of $\mathscr{R}$. A field is called incompressible (or divergence-free) when its divergence is 0.

The description of the curl of **F** at P is more complicated. For each unit vector **n**,

$$\mathbf{curl}\,\mathbf{F}(P)\cdot\mathbf{n} = \lim_{\mathscr{S}\to P} \frac{1}{\text{Area of } \mathscr{S}} \oint_C \mathbf{F}\cdot d\mathbf{r}$$

where $\mathscr{S}$ is a region in the plane that contains P and is perpendicular to **n** and C is the boundary of $\mathscr{S}$ oriented according to the right-hand rule. A field whose curl is **0** is called irrotational.

Of particular interest are conservative fields. A field **F** is conservative if its circulation along a curve depends only on its endpoints. If the domain of **F** is simply connected, **F** is conservative if and only if $\mathbf{curl}\,\mathbf{F} = \mathbf{0}$. A conservative field is expressible as the gradient of a scalar function.

Among the conservative fields are the central fields. If, in addition, they are divergence-free, they take a special form that depends on the dimension, as shown in the table below.

In the table below, $\mathbf{R}$ stands for the set of real numbers, $\mathbf{R}^2$ for the set of pairs (x, y), $\mathbf{R}^3$ for the set of triplets (x, y, z), and $\mathbf{R}^n$ for the set of n-tuplets $(x_1, x_2, \ldots, x_n)$.

Geometry	General Form of Divergence-Free Central Fields	Description
$\mathbf{R}^2$ (plane)	$c\frac{\widehat{\mathbf{r}}}{r}$	inverse first power
$\mathbf{R}^3$ (space)	$c\frac{\widehat{\mathbf{r}}}{r^2}$	inverse square power
$\mathbf{R}^n$	$c\frac{\widehat{\mathbf{r}}}{r^{n-1}}$	inverse $n-1$ power

Table 18.S.2

When $\mathbf{F}$ is irrotational ($\mathbf{curl}\,\mathbf{F} = \mathbf{0}$) we can replace an integral $\int_{AB} \mathbf{F} \cdot d\mathbf{r}$ by an integral over another curve joining A and B, which can be helpful if the new line integral is easier to evaluate than the original one. Similarly, in a region where $\nabla \cdot \mathbf{F} = 0$ we can replace an integral $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$ with a more convenient integral over a different surface.

In applications in space the most important field is the inverse-square central field, $\mathbf{F} = \widehat{\mathbf{r}}/r^2$. Its flux over a closed surface that does not enclose the origin is 0, but its flux over a surface that encloses the origin is 4π. This is established with the aid of the divergence theorem. If one thinks in terms of steradians, it is clear why the second integral is 4π: the flux of $\widehat{\mathbf{r}}/r^2$ also measures the solid angle subtended by a surface. Also, the first case becomes clear in terms of solid angles when the parts of the surface where $\mathbf{n} \cdot \mathbf{r}$ is positive and where it is negative are treated separately.

EXERCISES for Section 18.S

1. Explain, in words, the difference between $f(\mathbf{r})\widehat{\mathbf{r}}$ and $f(r)\widehat{\mathbf{r}}$.

2. Use Green's theorem to evaluate $\oint_C (xy\, dx + e^x\, dy)$, where C is the curve that goes from $(0,0)$ to $(2,0)$ on the x-axis and returns from $(2,0)$ to $(0,0)$ on the parabola $y = 2x - x^2$.

3. Let $\mathbf{F}(x, y, z) = 3\mathbf{i}$.

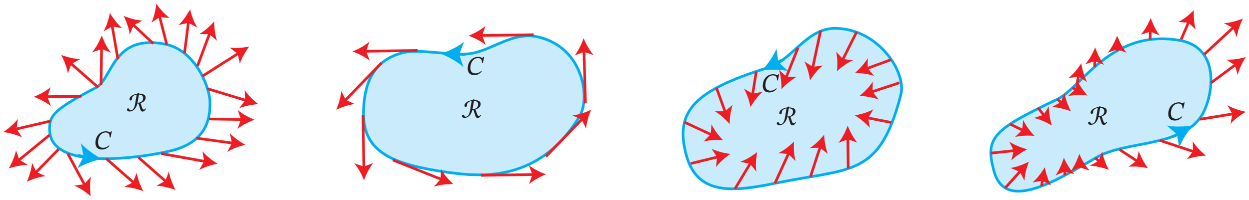
(a) Find the flux of $\mathbf{F}$ across each face of the tetrahedron with vertices $(1,0,0)$, $(0,1,0)$, $(0,0,1)$, and $(0,0,0)$.
(b) Find the total flux of $\mathbf{F}$ across the surface of the tetrahedron.
(c) Verify the divergence theorem for $\mathbf{F}$.

4. Let $\mathcal{V}$ be the cylinder bounded by the surfaces $x^2 + y^2 = 1$, $z = a$, and $z = b$ and let $\mathcal{S}$ be its bounding surface. Verify the divergence theorem for $\mathbf{F} = x\mathbf{i} + y\mathbf{j}$ in this case.

5. Let $\mathbf{F}$ be the uniform field $\mathbf{F}(x, y, z) = 2\mathbf{i} + 3\mathbf{j} + 0\mathbf{k}$. Carry out the three parts of Exercise 3 for this field.

6. Suppose you placed the point at which $\mathbf{E}$ is evaluated at $(a, 0, 0)$ instead of at $(0, 0, a)$. (a) What integral in spherical coordinates arises? (b) Would you like to evaluate it?

See Exercise 13 in Section 18.7



(a) (b) (c) (d)

Figure 18.S.1

In Exercises 7 to 10 the vector field **F** is defined on the whole plane but indicated only at points on a curve C bounding a region $\mathscr{R}$. What can be said about $\int_{\mathscr{R}} \nabla \cdot \mathbf{F}\, dA$?

7. See Figure 18.S.1(a). **8.** See Figure 18.S.1(b). **9.** See Figure 18.S.1(c). **10.** See Figure 18.S.1(d).

In Exercises 11 to 14, the vector field **F** is defined on the whole plane but indicated only at points on a curve C bounding a region $\mathscr{R}$. What can be said about $\int_{\mathscr{R}} (\nabla \times \mathbf{F}) \cdot \mathbf{k}\, dA$ in each case?

11. See Figure 18.S.1(a). **12.** See Figure 18.S.1(b). **13.** See Figure 18.S.1(c). **14.** See Figure 18.S.1(d).

15. (a) How many steradians does one face of a cube subtend at one of the four vertices not on that face?
(b) How many steradians does one face of a cube subtend at the center of the cube?

16. A tetrahedron is made of four congruent equilateral triangles. Find the steradians subtended by one face at the centroid of the tetrahedron. No integration is necessary.

17. Let C be the circle of radius 1 with center $(0,0)$.
(a) What does Green's theorem say about the line integral $\oint_C \left((x^2 - y^3)\, dx + (y^2 + x^3)\, dy\right)$?
(b) Use Green's theorem to evaluate the integral.
(c) Evaluate the line integral in (a) directly.

18. Which of the sets are connected? simply connected?
(a) A circle $(x^2 + y^2 = 1)$ in the xy-plane
(b) A disk $(x^2 + y^2 \leq 1)$ in the xy-plane
(c) The xy-plane from which a circle is removed
(d) The xy-plane from which a disk is removed
(e) The xy-plane from which one point is removed
(f) xyz-space from which one point is removed
(g) xyz-space from which a sphere is removed
(h) xyz-space from which a ball is removed
(i) A solid torus
(j) xyz-space from which a solid torus is removed
(k) A coffee cup with one handle

19. In Example 5, Section 18.1, we computed $\oint_C \mathbf{F}\cdot d\mathbf{r}$ where $\mathbf{F} = \dfrac{-y\mathbf{i}+x\mathbf{j}}{x^2+y^2}$ and C is the unit circle centered at $(0,0)$.

(a) Without doing any new computations, evaluate $\oint_C \mathbf{F}\cdot d\mathbf{r}$ where C is the square path with corners $(1,0)$, $(2,0)$, $(2,1)$, and $(1,1)$.

(b) Evaluate the integral in (a) by breaking the integral into four integrals, one over each edge.

20. Let $\mathbf{F}(x,y) = (x+y)\mathbf{i} + x^2\mathbf{j}$ and let C be the counterclockwise path around the triangle whose vertices are $(0,0)$, $(1,1)$, and $(-1,1)$.

(a) Use the planar divergence theorem to evaluate $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$, where $\mathbf{n}$ is the outward unit normal.

(b) Evaluate the line integral in (a) directly.

21. Let b and c be positive and $\mathscr{S}$ the infinite rectangle parallel to the xy-plane, consisting of the points (x,y,c) such that $0 \le x \le b$ and $y \ge 0$.

(a) If b were replaced by ∞, what is the solid angle that $\mathscr{S}$ subtends at the origin?

(b) Find the solid angle subtended by $\mathscr{S}$ when b is finite.

(c) Is the limit of your answer in (b) as $b \to \infty$ the same as your answer in (a)?

22. Show that $\mathbf{curl}(\mathbf{F}\times\mathbf{G}) = (\mathbf{G}\cdot\nabla)\mathbf{F} - (\mathbf{F}\cdot\nabla)\mathbf{G} + \mathbf{F}(\nabla\cdot\mathbf{G}) - \mathbf{G}(\nabla\cdot\mathbf{F})$.

NOTATION: If $\mathbf{F} = F_1\mathbf{i}+F_2\mathbf{j}+F_3\mathbf{k}$ and $\mathbf{G} = G_1\mathbf{i}+G_2\mathbf{j}+G_3\mathbf{k}$, then $(\mathbf{G}\cdot\nabla)\mathbf{F} = G_1\dfrac{\partial\mathbf{F}}{\partial x} + G_2\dfrac{\partial\mathbf{F}}{\partial y} + G_3\dfrac{\partial\mathbf{F}}{\partial z}$.

23. Look back at the fundamental theorem of calculus (Section 6.4), Green's theorem (Section 18.2), the divergence theorem (Section 18.5), and Stokes' theorem (Section 18.6). What single theme runs through all of them?

24. Any vector field $\mathbf{G}$ that is the curl of a vector field has the property that the integral of $\mathbf{G}\cdot\mathbf{n}$ over a surface depends only on the curve bounding the surface. So two surfaces with the same boundary provide the same value of the integral. Is this property shared by all vector fields, not just those that are the curl of another field?

In Exercises 25 to 29 a fluid rotates in a right circular cylinder with the velocity $\mathbf{F}(r,\theta) = r^n\widehat{\boldsymbol{\theta}}$ relative to polar coordinates whose pole is at the center of a circular cross section.

25. Sketch enough vectors to indicate $\mathbf{F}$ when (a) $n=2$ and (b) $n=-2$.

26. A wheel with blades is placed in the fluid with its axis perpendicular to the surface of the fluid. In which direction will it turn when (a) $n=2$ and (b) $n=-2$. See Exercise 25.

27. Compute $\mathbf{curl}\,\mathbf{F} = \mathbf{curl}(r^n\widehat{\boldsymbol{\theta}})$ in these steps:

(a) Express r^n and $\widehat{\boldsymbol{\theta}}$ in terms of x, y, $\mathbf{i}$, and $\mathbf{j}$.

(b) Using (a), express $\mathbf{F}$ in rectangular coordinates.

(c) Using (b), compute $\mathbf{curl}\,\mathbf{F}$ in rectangular coordinates.

(d) Using (c), show that $\mathbf{curl}\,\mathbf{F} = (n+1)r^{n-1}\mathbf{k}$.

28. Using (d) in the preceding Exercise, answer the questions in Exercise 26.

29. What does the immersed wheel do when $n=-1$?

Calculus is Everywhere # 25

Heating and Cooling

Engineers who design a car radiator or a home air conditioner are interested in the distribution of temperature in a fin. We present one of the mathematical tools they use as an example showing how Green's theorem is applied.

Consider a sheet of metal that occupies a plane region $\mathscr{A}$. Heating and cooling devices hold the temperature along the border constant, independent of time. Assume that the temperature in $\mathscr{A}$ eventually stabilizes. The steady-state temperature at point P in $\mathscr{A}$ is denoted $T(P)$. What does that imply about the function T?

Heat tends to flow from points with high temperatures to points with low temperatures, that is, heat flows in the direction of $-\nabla T$. According to Fourier's law, flow is proportional to the conductivity of the material k (a positive constant) and the magnitude of the gradient ∇T. Thus, for any closed curve C in $\mathscr{A}$,

$$\oint_C (-k\nabla T)\cdot \mathbf{n}\, ds$$

measures the rate of heat loss across C.

Since the temperature in the metal is at a steady state, the heat in the region bounded by C remains constant. Thus, the rate of heat loss across C is zero. That is, for any curve C in $\mathscr{A}$,

$$\oint_C (-k\nabla T)\cdot \mathbf{n}\, ds = 0.$$

Green's theorem can be used to translate this line integral over C to an equivalent double integral over the region $\mathscr{R}$ bounded by C to conclude that

$$\int_{\mathscr{R}} \nabla\cdot(-k\nabla T)\; dA = 0$$

for any region $\mathscr{R}$ bounded by a curve in the metal plate. Since the constant k is not zero, we conclude that

$$\int_{\mathscr{R}} (\nabla\cdot\nabla T)\, dA = 0 \tag{C.25.1}$$

for every region $\mathscr{R}$ contained in $\mathscr{A}$. Then, by the zero-integral theorem, the integrand must be 0 throughout $\mathscr{A}$, so $\nabla\cdot\nabla T = 0$ at every point (x, y) in $\mathscr{A}$.

The left-hand side of this equation, $\nabla\cdot\nabla T$, is frequently called the **Laplacian** of T:

$$\Delta T = \nabla^2 T = \nabla\cdot\nabla T.$$

When the specific combination of differential operators in the Laplacian are expanded in Cartesian coordinates in the xy-plane, we find:

$$\begin{aligned}\nabla^2 T = \nabla\cdot\nabla T &= \left(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j}\right)\cdot\left(\frac{\partial T}{\partial x}\mathbf{i} + \frac{\partial T}{\partial y}\mathbf{j}\right)\\ &= \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}.\end{aligned}$$

This reduces the study of the temperature distribution to solving a second-order partial differential equation known as **Laplace's equation**:

$$\nabla\cdot\nabla T = 0. = \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0. \tag{C.25.2}$$

In Cartesian coordinates, Laplace's equation takes the form:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0. \tag{C.25.3}$$

A solution $T(x,y)$ of (C.25.3) is called a harmonic function. More generally, a function of n variables, $f(x_1, x_2, \ldots, x_n)$ is said to be **harmonic** if

$$\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2} = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2} = 0.$$

EXERCISES for CIE C.25

In Exercises 1–4, show that the given function is harmonic.

1. $x^2 - y^2$

2. $e^x \sin(y)$

3. $\ln(x^2 + y^2)$

4. $\frac{1}{r}$, where $r = \sqrt{x^2 + y^2 + z^2}$

---

5. (a) Show that if $r = \sqrt{x^2 + y^2}$, then $\frac{\partial r}{\partial x} = \frac{x}{r}$. (b) Show that if $r = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$, then $\frac{\partial r}{\partial x_i} = \frac{x_i}{r}$.

6. For which values of k is $f(x,y) = r^k$ harmonic?

7. For which values of k is $f(x,y,z) = r^k$ harmonic?

8. For which values of k is $f(x_1, x_2, x_3, x_4) = r^k$ harmonic?

9. If $f(x,y)$ and $g(x,y)$ are harmonic, (a) is $f+g$ harmonic? (b) is fg harmonic?

For any positive integer n, the set of all "points" $(x_1, x_2, \ldots, x_n)$ is called n-**dimensional Euclidean space**, denoted $\mathbb{R}_n$. Exercise 10 uses an induction argument to show that if r_n is the distance in $\mathbb{R}_n$ from the origin to $(x_1, x_2, \ldots, x_n)$, then $r_n^2 = x_1^2 + x_2^2 + \cdots + x_n^2$.

10. (a) Show that $r_2^2 = x_1^2 + x_2^2$.

(b) Show that $r_2^2 + x_3^2 = r_3^2$.

(c) Show that $r_3^2 = x_1^2 + x_2^2 + x_3^2$.

(d) Show that, in $\mathbb{R}_n$, the origin, $(x_1, x_2, \ldots, x_{n-1}, 0)$, and $(x_1, x_2, \ldots, x_{n-1}, x_n)$ form a right triangle.

(e) Deduce by an induction using (a) and (d) that, for $n \geq 2$, $r_n^2 = x_1^2 + x_2^2 + \cdots + x_n^2$.

(f) Dispose of the dull case when $n = 1$: $r_1^2 = x_1^2$.

Calculus is Everywhere # 26

How Maxwell Did It

In a letter to his cousin Charles Cay, dated January 5, 1865, James Maxwell wrote:

> "I have also a paper afloat containing an electromagnetic theory of light, which, till I am convinced to the contrary, I hold to be great guns."
>
> REFERENCE: Everitt, F., James Clerk Maxwell: a force for physics, *Physics World*, December 2006, http://physicsworld.com/cws/article/print/26527

It indeed was "great guns", for with dazzling insight Maxwell predicted that light is an electrical phenomenon. In this section we will see how that prediction came out of the equations the time-dependent Maxwell equations introduced in Exercises 12 to 15 in Section 18.9 (on page 1096):

(i) $\operatorname{div}\mathbf{E} = \dfrac{\rho}{\epsilon_0}$ (ii) $\mathbf{curl}\,\mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t}$ (iii) $\operatorname{div}\mathbf{B} = 0$ (iv) $\mathbf{curl}\,\mathbf{B} = \mu_0\mathbf{J} + \mu_0\epsilon_0\dfrac{\partial \mathbf{E}}{\partial t}$

where **E** and **B** are the electric and magnetic fields, respectively, **J** is the current density, and q is the charge density.

The following sequence of computations show how equations (i), (ii), (iii), and (iv) imply that **E** satisfies a wave equation with wave speed $1/\sqrt{\mu_0\epsilon_0}$:

$$\frac{\partial^2 E}{\partial x^2} = \mu_0\epsilon_0\frac{\partial^2 E}{\partial t^2}.$$

Maxwell's reasoning begins with equations (i) and (iv) with the assumptions that there is no charge ($\rho = 0$), no current $\mathbf{J} = 0$, and the fields **B** and **E** vary with time.

Differentiating (iv) with respect to time t yields

$$\frac{\partial}{\partial t}(\mathbf{curl}\,\mathbf{B}) = \mu_0\epsilon_0\frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{C.26.1}$$

The operator $\partial/\partial t$ can be moved past the **curl** to operate directly on **B**. Thus (C.26.1) becomes

$$\mathbf{curl}\,\frac{\partial \mathbf{B}}{\partial t} = \mu_0\epsilon_0\frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{C.26.2}$$

The negative of (ii) is

$$-\,\mathbf{curl}\,\mathbf{E} = \frac{\partial \mathbf{B}}{\partial t}. \tag{C.26.3}$$

Taking the **curl** of both sides of (C.26.3) we get

$$\mathbf{curl}\,(-\,\mathbf{curl}\,\mathbf{E}) = \mathbf{curl}\,\frac{\partial \mathbf{B}}{\partial t}. \tag{C.26.4}$$

Combining (C.26.2) and (C.26.4) gives an equation that involves **E** alone:

$$\mathbf{curl}\,(-\,\mathbf{curl}\,\mathbf{E}) = \mu_0\epsilon_0\frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{C.26.5}$$

Equation (i), under the assumption that there is no charge ($\rho = 0$), reduces to $\operatorname{div}\mathbf{E} = 0$. Thus, the vector identity obtained in Exercise 2 $\mathbf{curl}\,(\mathbf{curl}\,\mathbf{E}) = \mathbf{grad}\,(\operatorname{div}\mathbf{E}) - \left(\operatorname{div}\mathbf{grad}\right)\mathbf{E}$ simplifies to

$$\mathbf{curl}\,(\mathbf{curl}\,\mathbf{E}) = -\left(\operatorname{div}\mathbf{grad}\right)\mathbf{E}. \tag{C.26.6}$$

Combining (C.26.5) and (C.26.6) leads to

$$\left(\operatorname{div}\mathbf{grad}\right)\mathbf{E} = \mu_0\epsilon_0\frac{\partial^2\mathbf{E}}{\partial t^2} \qquad \text{or} \qquad \frac{\partial^2 E}{\partial t^2} - \frac{1}{\mu_o\epsilon_0}\operatorname{div}\mathbf{grad}\,\mathbf{E} = \mathbf{0} \tag{C.26.7}$$

Recall, from CIE 25, that the differential operator div **grad** is the Laplacian: div **grad** $= \nabla\cdot\nabla = \nabla^2 = \Delta$. And, in Cartesian coordinates,

$$\Delta = \nabla^2 = \nabla\cdot\nabla = \left(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}\right)\cdot\left(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}\right) = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{C.26.8}$$

The Laplacian of a vector field $\mathbf{E}$ is obtained by applying the Laplacian operator to each component of $\mathbf{E}$. Thus $\Delta\mathbf{E}$ and $\partial^2\mathbf{E}/\partial t^2$ are both vectors, and (C.26.7) makes sense.

For simplicity, suppose E has only a $\mathbf{j}$-component, which depends on x and t. Thus $\mathbf{E}(x,y,z,t) = E(x,t)\mathbf{j}$, where E is a scalar function. We have

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)E(x,t) = \frac{\partial^2 E}{\partial x^2} + 0 + 0 = \frac{\partial^2 E}{\partial x^2}.$$

Then (C.26.7) becomes

$$\frac{\partial^2}{\partial t^2}E(x,t)\mathbf{j} - \frac{1}{\mu_0\epsilon_0}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)E(x,t)\mathbf{j} = \mathbf{0},$$

from which it follows that

$$\frac{\partial^2}{\partial t^2}E(x,t) - \frac{1}{\mu_0\epsilon_0}\frac{\partial^2 E}{\partial x^2} = 0. \tag{C.26.9}$$

Multiply (C.26.9) by $-\mu_0\epsilon_0$ to obtain

$$\frac{\partial^2 E}{\partial x^2} - \mu_0\epsilon_0\frac{\partial^2 E}{\partial t^2} = 0. \tag{C.26.10}$$

This is an instance of the wave equation ((16.3.10) in Section 16.3). The solutions are waves traveling with speed $1/\sqrt{\mu_0\epsilon_0}$.

The wave equation was also discussed in CIE 23 at the end of Chapter 16 and CIE 24 in Chapter 17.

To see that, in fact, $1/\sqrt{\mu_0\epsilon_0}$ has the units of speed, look at the dimensions of the ϵ_0 and μ_0.

The vacuum permittivity (ϵ_0) first appears in the description of the force between two static electrically charged particles (see (18.7.3)). This force has magnitude

$$|\mathbf{F}| = \frac{qq_0}{4\pi\epsilon_0 r^2} \tag{C.26.11}$$

and dimensions of mass times acceleration

$$\text{mass}\cdot\frac{\text{length}}{\text{time}^2},$$

or, in symbols,

$$m\frac{L}{T^2}.$$

The number 4π is a pure number, without any physical dimensions. The quantity qq_0 has the dimensions of charge squared, q^2, and r^2 has dimensions L^2.

To find the dimensions of ϵ_0, solve (C.26.11) for ϵ_0:

$$\epsilon_0 = \frac{qq_0}{4\pi|\mathbf{F}|r^2}.$$

Thus, the dimensions of ϵ_0 are

$$\frac{T^2}{mL}\cdot\frac{q^2}{L^2} = \frac{T^2q^2}{mL^3}.$$

To find the dimensions of the vacuum permeability (μ_0), we call upon the fact (from physics) that μ_0 is involved in the calculation of the force between two wires of length L each carrying currents I_1 and I_2 in the same direction and separated by a distance R. Each generates a magnetic field that draws the other towards it. The magnitudes of these forces are

$$|\mathbf{F}| = \mu_0 \frac{I_1 I_2 L}{2\pi R}.$$

It follows that

$$\mu_0 = \frac{2\pi R|\mathbf{F}|}{I_1 I_2 L}.$$

Since R has the dimensions of length L and $|\mathbf{F}|$ has dimensions mL/T^2, the numerator has dimensions mL^2/T^2. The currents I_1 and I_2 have units of "charge q per second," so $I_1 I_2$ has dimensions q^2/T^2. The dimensions of the denominator are, therefore,

$$\frac{q^2 L}{T^2}.$$

Hence μ_0 has the dimensions

$$\frac{mL^2}{T^2} \cdot \frac{T^2}{q^2 L} = \frac{mL}{q^2}.$$

This shows the dimensions of the product $\mu_0\epsilon_0$ are

$$\frac{mL}{q^2} \cdot \frac{T^2 q^2}{mL^3} = \frac{T^2}{L^2}.$$

Thus the reciprocal $1/\mu_0\epsilon_0$ has the same dimensions as the square of speed. In short, $1/\sqrt{\mu_0\epsilon_0}$ has the dimensions of speed, namely length divided by time.

Maxwell then compares $1/\sqrt{\mu_0\epsilon_0}$, which equals in his terminology "the ratio of electrical units," with the velocity of light:

"In the following table, the principal results of direct observation of the velocity of light, are compared with the principal results of the calculation of electrical units.

It is significant that the velocity of light and the ratio of the units are quantities of the same order of magnitude. Neither of them can be said to be determined as yet with such a degree of accuracy as to enable us to assert that the one is greater or less than the other. It is to be hoped that, by further experiment, the relation between the magnitude of the two quantities may be more accurately determined.

In the meantime our theory, which asserts that these two quantities are equal, and assigns a physical reason for this equality, is certainly not contradicted by the comparison of these results such as they are."

REFERENCE: James Clerk Maxwell, Treatise on Electricity and Magnetism, Vol. 2, third edition, Oxford University Press, (1904), first edition 1873, p. 436

Velocity of light (meters per second)		Ratio of electrical units	
Fizeau	314,000,000	Weber	310,740,000
Sun's Parallax	308,000,000	Maxwell	288,000,000
Foucault	298,360,000	Thomson	282,000,000

Table C.26.1

On this basis Maxwell concluded that light is an electromagnetic disturbance, uniting the fields of optics and electromagnetism. Earlier, in 1848, Gustav Kirchoff had noticed that the expression involving the two electrical units was near the velocity of light, but viewed it as a coincidence.

By 1890 experiments gave the velocity of light as 299,766,000 meters per second and $\sqrt{1/\mu_0\epsilon_0}$ as 299,550,000 meters per second, overwhelming evidence for Maxwell's conjecture.

Newton, in his *Principia* of 1687, related gravity on Earth to gravity in the heavens. Benjamin Franklin with his kite experiment showed that lightning was an electric phenomenon. From then through the early nineteenth century experimenters showed that electricity and magnetism were inseparable. It is still not known whether there is any connection between gravity and electromagnetism.

EXERCISES for CIE C.26

1. Justify the assertion that $\dfrac{\partial}{\partial t}\,\mathbf{curl}\,\mathbf{B} = \mathbf{curl}\left(\dfrac{\partial \mathbf{B}}{\partial t}\right)$.

2. Verify the vector identity $\mathbf{curl}\,(\mathbf{curl}\,\mathbf{E}) = \mathbf{grad}\,(\operatorname{div}\mathbf{E}) - \left(\operatorname{div}\mathbf{grad}\right)\mathbf{E}$.

Appendix A

Tables of Derivatives and Antiderivatives

Derivatives

1. $\dfrac{d}{dx}\left(x^n\right) = nx^{n-1}$
2. $\dfrac{d}{dx}(\ln|x|) = \dfrac{1}{x}$
3. $\dfrac{d}{dx}(\sin(x)) = \cos(x)$
4. $\dfrac{d}{dx}(\cos(x)) = -\sin(x)$
5. $\dfrac{d}{dx}(\tan(x)) = \sec^2(x)$
6. $\dfrac{d}{dx}(\sec(x)) = \sec(x)\tan(x)$
7. $\dfrac{d}{dx}(\cot(x)) = -\csc^2(x)$
8. $\dfrac{d}{dx}(\csc(x)) = -\csc(x)\cot(x)$
9. $\dfrac{d}{dx}(\arcsin(x)) = \dfrac{1}{\sqrt{1-x^2}}$
10. $\dfrac{d}{dx}(\arctan(x)) = \dfrac{1}{1+x^2}$
11. $\dfrac{d}{dx}(\operatorname{arcsec}(x)) = \dfrac{1}{|x|\sqrt{x^2-1}}$
12. $\dfrac{d}{dx}\left(e^x\right) = e^x$
13. $\dfrac{d}{dx}\left(a^x\right) = a^x(\ln(a))$
14. $\dfrac{d}{dx}(\sinh(x)) = \cosh(x)$
15. $\dfrac{d}{dx}(\cosh(x)) = \sinh(x)$

Basic Antiderivatives

16. $\int x^n \, dx = \frac{1}{n+1} x^{n+1} \quad n \neq -1$
17. $\int \frac{dx}{x} = \ln(x),\ x > 0 \quad \text{or} \quad \ln|x|,\ x \neq 0$
18. $\int e^x \, dx = e^x$
19. $\int \sin(x) \, dx = -\cos(x)$
20. $\int \cos(x) \, dx = \sin(x)$
21. $\int \tan(x) \, dx = \ln|\sec(x)| = -\ln|\cos(x)|$
22. $\int \cot(x) \, dx = \ln|\sin(x)| = -\ln|\csc(x)|$
23. $\int \sec(x) \, dx = \ln|\sec(x) + \tan(x)| = \ln\left|\tan\left(\frac{x}{2} + \frac{\pi}{4}\right)\right|$
24. $\int \csc(x) \, dx = \ln|\csc(x) - \cot(x)| = \ln\left|\tan\left(\frac{x}{2}\right)\right|$
25. $\int \frac{dx}{x^2 + a^2} = \frac{1}{a} \arctan\left(\frac{x}{a}\right)$
26. $\int \frac{dx}{\sqrt{a^2 - x^2}} = \arcsin\left(\frac{x}{a}\right),\ a > 0$
27. $\int \frac{dx}{|x|\sqrt{x^2 - a^2}} = \frac{1}{a} \operatorname{arcsec}\left(\frac{x}{a}\right)$

Expressions Containing $ax+b$

28. $\int (ax+b)^n\,dx = \frac{1}{a(n+1)}(ax+b)^{n+1}$
29. $\int \frac{dx}{ax+b} = \frac{1}{a}\ln|ax+b|$
30. $\int \frac{dx}{(ax+b)^2} = \frac{-1}{a(ax+b)}$
31. $\int \frac{x\,dx}{(ax+b)^2} = \frac{b}{a^2(ax+b)} + \frac{1}{a^2}\ln|ax+b|$
32. $\int \frac{dx}{x(ax+b)} = \frac{1}{b}\ln\left|\frac{x}{ax+b}\right|$
33. $\int \frac{dx}{x^2(ax+b)} = \frac{-1}{bx} + \frac{a}{b^2}\ln\left|\frac{ax+b}{x}\right|$
34. $\int \sqrt{ax+b}\,dx = \frac{2}{3a}\sqrt{(ax+b)^3}$
35. $\int x\sqrt{ax+b}\,dx = \frac{2(3ax-2b)}{15a^2}\sqrt{(ax+b)^3}$
36. $\int \frac{dx}{\sqrt{ax+b}} = \frac{2}{a}\sqrt{ax+b}$
37. $\int \frac{\sqrt{ax+b}}{x}\,dx = 2\sqrt{ax+b} + b\int \frac{dx}{x\sqrt{ax+b}}$
38. $\int \frac{dx}{x\sqrt{ax+b}} = \frac{1}{\sqrt{b}}\ln\left|\frac{\sqrt{ax+b}-\sqrt{b}}{\sqrt{ax+b}+\sqrt{b}}\right|,\ b>0$
39. $\int \frac{dx}{x\sqrt{ax+b}} = \frac{2}{\sqrt{-b}}\arctan\sqrt{\frac{ax+b}{-b}},\ b<0$
40. $\int \frac{dx}{x^2\sqrt{ax+b}} = \frac{-\sqrt{ax+b}}{bx} - \frac{a}{2b}\int \frac{dx}{x\sqrt{ax+b}}$
41. $\int \sqrt{\frac{cx+d}{ax+b}}\,dx = \frac{\sqrt{ax+b}\sqrt{cx+d}}{a} + \frac{ad-bc}{2a}\int \frac{dx}{\sqrt{ax+b}\sqrt{cx+d}}$

Expressions Containing ax^2+c, $x^2 \pm p^2$, and p^2-x^2 $(p>0)$

42. $\displaystyle\int \frac{dx}{p^2-x^2} = \frac{1}{2p}\ln\left|\frac{p+x}{p-x}\right|$

43. $\displaystyle\int \frac{dx}{ax^2+c} = \begin{cases} \frac{1}{\sqrt{ac}}\arctan\left(x\sqrt{\frac{a}{c}}\right) & a>0,\ c>0 \\ \frac{1}{2\sqrt{-ac}}\ln\left|\frac{x\sqrt{a}-\sqrt{-c}}{x\sqrt{a}+\sqrt{-c}}\right| & a>0,\ c<0 \\ \frac{1}{2\sqrt{-ac}}\ln\left|\frac{\sqrt{c}+x\sqrt{-a}}{\sqrt{c}-x\sqrt{-a}}\right| & a<0,\ c>0 \end{cases}$

44. $\displaystyle\int \frac{dx}{(ax^2+c)^n} = \frac{1}{2(n-1)c}\frac{x}{(ax^2+c)^{n-1}} + \frac{2n-3}{2(n-1)c}\int \frac{dx}{(ax^2+c)^{n-1}} \qquad n>1$

45. $\displaystyle\int x(ax^2+c)^n\,dx = \frac{1}{2a}\frac{(ax^2+c)^{n+1}}{n+1} \qquad n\neq -1$

46. $\displaystyle\int \frac{x}{ax^2+c}\,dx = \frac{1}{2a}\ln\left|ax^2+c\right|$

47. $\displaystyle\int \sqrt{x^2\pm p^2}\,dx = \frac{1}{2}\left(x\sqrt{x^2\pm p^2} \pm p^2\ln\left|x+\sqrt{x^2\pm p^2}\right|\right)$

48. $\displaystyle\int \sqrt{p^2-x^2}\,dx = \frac{1}{2}\left(x\sqrt{p^2-x^2} + p^2\arcsin\left(\frac{x}{p}\right)\right)$

49. $\displaystyle\int \frac{dx}{\sqrt{x^2\pm p^2}} = \ln\left|x+\sqrt{x^2\pm p^2}\right|$

50. $\displaystyle\int (p^2-x^2)^{3/2}\,dx = \frac{x}{4}\left(p^2-x^2\right)^{3/2} + \frac{3p^2x}{8}\sqrt{p^2-x^2} + \frac{3p^4}{8}\arcsin\left(\frac{x}{p}\right)$

Expressions Containing ax^2+bx+c

51. $\displaystyle\int \frac{dx}{ax^2+bx+c} = \begin{cases} \frac{1}{\sqrt{b^2-4ac}}\ln\left|\frac{2ax+b-\sqrt{b^2-4ac}}{2ax+b+\sqrt{b^2-4ac}}\right| & b^2>4ac \\ \frac{2}{\sqrt{4ac-b^2}}\arctan\left(\frac{2ax+b}{\sqrt{4ac-b^2}}\right) & b^2<4ac \\ \frac{-2}{2ax+b} & b^2=4ac \end{cases}$

52. $\displaystyle\int \frac{dx}{(ax^2+bx+c)^{n+1}} = \frac{2ax+b}{n(4ac-b^2)(ax^2+bx+c)^n} + \frac{2(2n-1)a}{n(4ac-b^2)}\int \frac{dx}{(ax^2+bx+c)^n}$

53. $\displaystyle\int \frac{x\,dx}{ax^2+bx+c} = \frac{1}{2a}\ln\left|ax^2+bx+c\right| - \frac{b}{2a}\int \frac{dx}{ax^2+bx+c}$

54. $\displaystyle\int \frac{dx}{\sqrt{ax^2+bx+c}} = \begin{cases} \frac{1}{\sqrt{a}}\ln\left|2ax+b+2\sqrt{a}\sqrt{ax^2+bx+c}\right| & a>0 \\ \frac{1}{\sqrt{-a}}\arcsin\left(\frac{-2ax-b}{\sqrt{b^2-4ac}}\right) & a<0 \end{cases}$

55. $\displaystyle\int \frac{x\,dx}{\sqrt{ax^2+bx+c}} = \frac{\sqrt{ax^2+bx+c}}{a} - \frac{b}{2a}\int \frac{dx}{\sqrt{ax^2+bx+c}}$

56. $\displaystyle\int \sqrt{ax^2+bx+c}\,dx = \frac{2ax+b}{4a}\sqrt{ax^2+bx+c} + \frac{4ac-b^2}{8a}\int \frac{dx}{\sqrt{ax^2+bx+c}}$

Expression Containing Powers of Trigonometric Functions

57. $\int \sin^2(ax)\,dx = \frac{x}{2} - \frac{\sin(2ax)}{4a}$
58. $\int \sin^3(ax)\,dx = \frac{-1}{a}\cos(ax) + \frac{1}{3a}\cos^3(ax)$
59. $\int \sin^n(ax)\,dx = -\frac{\sin^{n-1}(ax)\cos(ax)}{na} + \frac{n-1}{n}\int \sin^{n-2}(ax)\,dx$, $n \geq 2$ positive integer
60. $\int \cos^2(ax)\,dx = \frac{x}{2} + \frac{\sin(2ax)}{4a}$
61. $\int \cos^3(ax)\,dx = \frac{1}{a}\sin(ax) - \frac{1}{3a}\sin^3(ax)$
62. $\int \cos^n(ax)\,dx = \frac{\cos^{n-1}(ax)\sin(ax)}{na} + \frac{n-1}{n}\int \cos^{n-2}(ax)\,dx$, $n \geq$ positive integer
63. $\int \tan^2(ax)\,dx = \frac{1}{a}\tan(ax) - x$
64. $\int \tan^3(ax)\,dx = \frac{1}{2a}\tan^2(ax) + \frac{1}{a}\ln|\cos(ax)|$
65. $\int \tan^n(ax)\,dx = \frac{\tan^{n-1}(ax)}{a(n-1)} - \int \tan^{n-2}(ax)\,dx$, $n \neq 1$
66. $\int \sec^2(ax)\,dx = \frac{1}{a}\tan(ax)$
67. $\int \sec^3(ax)\,dx = \frac{1}{2a}\sec(ax)\tan(ax) + \frac{1}{2a}\ln|\sec(ax) + \tan(ax)|$
68. $\int \sec^n(ax)\,dx = \frac{\sec^{n-2}(ax)\tan(ax)}{a(n-1)} + \frac{n-2}{n-1}\int \sec^{n-2}(ax)\,dx$, $n \neq 1$
69. $\int \frac{dx}{1 \pm \sin(ax)} = \mp\frac{1}{a}\tan\left(\frac{\pi}{4} \mp \frac{ax}{2}\right)$

Expressions Containing Algebraic and Trigonometric Functions

70. $\int \sin(ax)\cos(bx)\,dx = \frac{-\cos((a-b)x)}{2(a-b)} - \frac{\cos((a+b)x)}{2(a+b))} \qquad a^2 \neq b^2$
71. $\int x\sin(ax)\,dx = \frac{1}{a^2}\sin(ax) - \frac{x}{a}\cos(ax)$
72. $\int x\cos(ax)\,dx = \frac{1}{a^2}\cos(ax) + \frac{x}{a}\sin(ax)$
73. $\int x^n\sin(ax)\,dx = \frac{-1}{a}x^n\cos(ax) + \frac{n}{a}\int x^{n-1}\cos(ax)\,dx \qquad n$ positive
74. $\int x^n\cos(ax)\,dx = \frac{1}{a}x^n\sin(ax) - \frac{n}{a}\int x^{n-1}\sin(ax)\,dx \qquad n$ positive

Expressions Containing Exponential and Logarithmic Functions

75. $\displaystyle\int e^{ax}\,dx = \frac{1}{a}e^{ax}$

76. $\displaystyle\int b^{ax}\,dx = \frac{b^{ax}}{a\ln(b)}$

77. $\displaystyle\int xe^{ax}\,dx = \frac{1}{a^2}e^{ax}(ax-1)$

78. $\displaystyle\int xb^{ax}\,dx = \frac{1}{a^2}\frac{b^{ax}}{(\ln(b))^2}(a\ln(b)x-1)$

79. $\displaystyle\int x^n e^{ax}\,dx = \frac{1}{a}x^n e^{ax} - \frac{n}{a}\int x^{n-1}e^{ax}\,dx$

80. $\displaystyle\int e^{ax}\sin(bx)\,dx = \frac{e^{ax}}{a^2+b^2}(a\sin(bx)-b\cos(bx))$

81. $\displaystyle\int e^{ax}\cos(bx)\,dx = \frac{e^{ax}}{a^2+b^2}(a\cos(bx)+b\sin(bx))$

82. $\displaystyle\int \ln(ax)\,dx = x(\ln(ax)-1)$

83. $\displaystyle\int x^n\ln(ax)\,dx = x^{n+1}\left(\frac{\ln(ax)}{n+1}-\frac{1}{(n+1)^2}\right) \qquad n=0,1,2,\ldots$

84. $\displaystyle\int (\ln(ax))^2\,dx = x\left((\ln(ax))^2-2\ln(ax)+2\right)$

85. $\displaystyle\int \frac{\ln(ax)}{x}\,dx = \frac{a}{2}(\ln(ax))^2$

Expressions Containing Inverse Trigonometric Functions

86. $\displaystyle\int \arcsin(ax)\,dx = x\arcsin(ax)+\frac{1}{a}\sqrt{1-a^2x^2}$

87. $\displaystyle\int \arccos(ax)\,dx = x\arccos(ax)-\frac{1}{a}\sqrt{1-a^2x^2}$

88. $\displaystyle\int \operatorname{arcsec}(ax)\,dx = x\operatorname{arcsec}(ax)-\frac{1}{a}\ln\left|ax+\sqrt{a^2x^2-1}\right|$

89. $\displaystyle\int \operatorname{arccsc}(ax)\,dx = x\operatorname{arccsc}(ax)+\frac{1}{a}\ln\left|ax+\sqrt{a^2x^2-1}\right|$

90. $\displaystyle\int \arctan(ax)\,dx = x\arctan(ax)-\frac{1}{2a}\ln\left(1+a^2x^2\right)$

91. $\displaystyle\int \operatorname{arccot}(ax)\,dx = x\operatorname{arccot}(ax)+\frac{1}{2a}\ln\left(1+a^2x^2\right)$

Some Special Definite Integrals

92. $\displaystyle\int_0^{\pi/2}\sin^n(x)\,dx = \int_0^{\pi/2}\cos^n(x)\,dx = \begin{cases} \frac{1\cdot3\cdot5\cdot7\cdots(n-1)}{2\cdot4\cdot6\cdot8\cdots(n)}\frac{\pi}{2} & n \text{ even} \\ \frac{2\cdot4\cdot6\cdot8\cdots(n-1)}{1\cdot3\cdot5\cdot7\cdots(n)} & n \text{ odd} \end{cases}$

93. $\displaystyle\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$

Appendix B

Collected Hints

Hint for Exercise 34 (§1.1)

Reflect B across the line L.

Hint for Exercise 36(b) (§1.1)

This will be a piecewise-defined function.

Hint for Exercise 51 (§1.2)

Assume that it is rational, that is, equal to m/n for some integers m and n, and obtain a contradiction.

Hint for Exercise 6 (§1.5)

Experiment and form a conjecture.

Hint for Exercise 17 (§1.5)

How small can the income be? How large?

Hint for Exercise 19 (§1.5)

Start with the equation $5^x = 3^7$.

Hint for Exercise 21(a) (§1.5)

Start with $b^{\log_b(x)} = x$.

Hint for Exercise 23(b) (§1.5)

Think of the final multiplication.

Hint for Exercise 23(c) (§1.5)

n is a power of 2 when $n = 2^k$, k a positive integer.

Hint for Exercise 8(b) (§1.S)

Take logarithms to the base 2 of both sides of the equation $3^{\log_3(17)} = 17$.

Hint for Exercise 9(a) (§1.S)

See 10(e).

Hint for Exercise 10(b) (§1.S)

Start with $a^{\log_a(b)} = b$ and take logarithms to the base b.

Hint for Exercise 17 (§1.S)

Consider drawing an appropriately labeled right triangle for each part.

Hint for Exercise 27 (§1.S)

Display x and $\sin(x)$ using a unit circle, for two different values of x, a and b.

Hint for Exercise 28 (§1.S)

The function $f(x) = x - \sin(x)$ is increasing for all numbers x. (See Exercise 27.)

Hint for Exercise 29 (§1.S)

Either exhibit positive a, b, and c for which the equation does not hold or else prove it always holds.

Hint for Exercise 7 (§1.2)

Write $\tan(x) = \sin(x)/\cos(x)$.

Hint for Exercise 9 (§2.2)

How does $b^3 - a^3$ factor?

Hint for Exercise 22 (§2.2)

Let $u = 3t$.

Hint for Exercise 22 (§2.2)

Resist the temptation to cancel the sin's. Instead, do a little algebra.

Hint for Exercise 35 (§2.2)

Multiply $\sqrt{x^2+x} - x$ by $\frac{\sqrt{x^2+x}+x}{\sqrt{x^2+x}+x}$, an operation that removes square roots from the numerator.

Hint for Exercise 45 (§2.2)

Write b as $e^{\ln(b)}$ and use $\lim_{x\to 0} \frac{e^x-1}{x} = 1$.

Hint for Exercise 34 (§2.3)

$\sin(a+b) = \sin(a)\cos(b) + \cos(a)\sin(b)$.

Hint for Exercise 35 (§2.3)

$\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$.

Hint for Exercise 21 (§2.5)

Use the permanence property.

Hint for Exercise 24 (§2.S)

Use (a).

Hint for Exercise 30 (§2.S)

Your solution will involve the sum of 100 numbers in a geometric progression.

Hint for Exercise 31 (§2.S)

The technique in Exercises 29 and 30 can help you decide. Assume that $\lim_{n\to\infty} 1/2^n = 0$.

Hint for Exercise 35 (§2.S)

Replace h in the denominator by $(hk)/k$.

Hint for Exercise 36 (§2.S)

What happens to P as $r \to \infty$?

Hint for Exercise 16 (§3.1)

Be careful when sketching the graph; the tangent line crosses the curve at $(0,0)$.

Hint for Exercise 22 (§3.1)

To find $(2+h)^3$, multiply out the product $(2+h)(2+h)(2+h)$.

Hint for Exercise 49 (§3.1)

Choose particular values of x and use your calculator to create a table of the results.

Hint for Exercise 50 (§3.1)

Use your calculator.

Hint for Exercise 1 (§3.2)

$\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$.

Hint for Exercise 1 (§3.2)

In Exercise 4, use a calculator to estimate the coefficient that appears.

Hint for Exercise 16 (§3.2)

In Exercise 20, use the identity $\tan(A-B) = \frac{\tan(A)-\tan(B)}{1+\tan(A)\tan(B)}$.

Hint for Exercise 8 (§1.3)

Write $(\tan(x))^2$ as $\tan(x)\tan(x)$.

Hint for Exercise 19 (§3.3)

See Exercise 18.

Hint for Exercise 22 (§3.3)

See Exercise 21.

Hint for Exercise 23 (§3.3)

See Exercise 18.

Hint for Exercise 41 (§3.3)

multiply it out first.

Hint for Exercise 45 (§3.3)

First write fgh as $(f)(gh)$. Then use the product rule twice.

Hint for Exercise 51 (§3.3)

$x^2 = x \cdot x$.

Hint for Exercise 54 (§3.3)

Review how we obtained the formula for the derivative of a product.

Hint for Exercise 15 (§3.5)

Express $\log_{10}$ in terms of $\ln$.

Hint for Exercise 16 (§3.5)

Express $\log_x$ in terms of ln.

Hint for Exercise 49 (§3.5)

To simplify the calculation, first use the property $\ln(p/q) = \ln(p) - \ln(q)$.

Hint for Exercise 59 (§3.5)

Assume $b^2 > 4ac$ and use properties of ln before differentiating.

Hint for Exercise 24 (§3.7)

Let $\theta = f(t)$ be the angle of deviation from the vertical at time t.

Hint for Exercise 39 (§3.7)

Assume each acceleration is constant.

Hint for Exercise 14 (§3.8)

See Exercise 13.

Hint for Exercise 17 (§3.8)

$|\sin(x)| \le 1$ for all x.

Hint for Exercise 33 (§3.8)

See Example 5.

Hint for Exercise 34 (§3.8)

$f(x)^2 - A^2 = (f(x) - A)(f(x) + A)$, and control the size of each factor.

Hint for Exercise 35 (§3.8)

To use the two limits, write $f(x)$ as $A + (f(x) - A)$ and $g(x)$ as $B + (g(x) - B)$.

Hint for Exercise 4 (§3.9)

Draw a suitable vertical band for the given value of ϵ.

Hint for Exercise 17 (§3.9)

Draw a picture.

Hint for Exercise 27 (§3.9)

Factor $x^2 - 9$.

Hint for Exercise 29 (§3.9)

Factor $x^2 + 5x - 24$.

Hint for Exercise 630.95 (§3.S)

Think of y as $y(x)$ and remember to use the chain rule when differentiating y^n with respect to x.

Hint for Exercise 70 (§3.S)

The identity $\tan(a+b) = \frac{\tan(a)+\tan(b)}{1-\tan(a)\tan(b)}$ will help.

Hint for Exercise 71 (§3.S)

Use the two inequalities that squeezed $\sin(x)/x$ toward 1.

Hint for Exercise 72 (§3.S)

Use the x and h notation.

Hint for Exercise 75 (§3.S)

Recognize the limit as the derivative of a function at a certain input. Keep in mind that x is constant in this limit.

Hint for Exercise 2 (§C.4)

Draw a right triangle with acute angles C and $\frac{\pi}{2} - C$.

Hint for Exercise 50 (§4.1)

Consider the cases with n even and n odd separately.

Hint for Exercise 60 (§4.2)

This function is difficult to graph in one picture. Instead, create separate sketches for $x > 0$ and for $x < 0$.

Hint for Exercise 60 (§4.2)

Watch out for points where f is not differentiable.

Hint for Exercise 30 (§4.3)

Do not attempt to find exact locations of the intercepts or critical points.

Hint for Exercise 17 (§4.4)

Why do you want to show that if a and x are in the interval, then $f(x) > f(a) + f'(a)(x-a)$? Treat the cases $x < a$ and $x > a$ separately.

Hint for Exercise 54 (§4.S)

Part (c) may come in handy.

Hint for Exercise 55 (§4.S)

Show the graph is symmetric with respect to the inflection point. Why can one assume it is enough to show this for the special case with $a = 1$ and $d = 0$, that is, for $x^3 + bx^2 + cx$?

Hint for Exercise 56 (§4.S)

Show that $f'(x) > 0$.

Hint for Exercise 64 (§4.S)

Review the mean value theorem.

Hint for Exercise 15 (§5.1)

Square the area to avoid square roots.

Hint for Exercise 21 (§5.1)

Think of the area as a function of x and solve.

Hint for Exercise 57 (§5.1)

It is more convenient to maximize $\tan(\theta)$ than θ itself. Use $\tan(A-B) = (\tan(A)-\tan(B))\,/\,(1+\tan(A)\tan(B))$.

Hint for Exercise 63 (§5.1)

No calculus is needed for this.

Hint for Exercise 10 (§5.2)

Do not find the radius and height of the largest can. (Keep in mind that the surface area is constant.)

Hint for Exercise 11 (§5.2)

Start with $x = \tan(y)$.

Hint for Exercise 12 (§5.2)

Start with $x = \sin(y)$.

Hint for Exercise 25 (§5.2)

What do you know about dy/dx at the highest and lowest points on the graph of a function?

Hint for Exercise 23 (§5.3)

In (a) you found $\dot{h}$ when $h = 3$, 4, and 5 feet.

Hint for Exercise 26 (§5.4)

Let $y = e^x$ and compare with Exercise 25.

Hint for Exercise 28 (§5.4)

See Exercise 27.

Hint for Exercise 29 (§5.4)

If $x \to 0^+$, then $t \to \infty$.

Hint for Exercise 31 (§5.4)

Use the result obtained in Exercises 27 and 28.

Hint for Exercise 40 (§5.5)

Use the definition of the derivative.

Hint for Exercise 41 (§5.5)

See also Exercise 17 in Section 4.4.

Hint for Exercise 74 (§5.6)

See Exercise 73.

Hint for Exercise 79 (§5.6)

See Exercise 77.

Hint for Exercise 80 (§5.6)

See Exercise 78.

Hint for Exercise 70 (§5.S)

Introduce an imaginary house C such that the midpoint of B and C is on the road and the segment BC is perpendicular to the road. That is, reflect B across the road to become C.

Hint for Exercise 80 (§5.S)

Sketch graphs, then explain.

Hint for Exercise 86 (§5.S)

The angle θ produces the shortest ladder to reach the building and stay above the fence.

Hint for Exercise 88 (§5.S)

Assume $e < 3$.

Hint for Exercise 89 (§5.S)

Assume π is not algebraic.

Hint for Exercise 92 (§5.S)

If $P(x)$ has degree n, what are the degrees of $xP(x)$ and $x^2P(x)$?

Hint for Exercise 103 (§5.S)

There is a short way and a long way to answer this question.

Hint for Exercise 104 (§5.S)

Sketch a diagram of the circles and the chord.

Hint for Exercise 28 (§6.2)

First show that a cross section by a plane perpendicular to the axis of the cone and a distance x from the vertex is a disk of radius ax/h. See Exercise 30.

Hint for Exercise 32 (§6.2)

Review Exercise 12 and Figure 6.2.7.

Hint for Exercise 33 (§6.2)

Is this a telescoping sum?

Hint for Exercise 41 (§6.2)

Because $r = b^{1/n}$ you could use the x^y key on a calculator.

Hint for Exercise 41 (§6.2)

Start by taking ln of both sides of the equation $(r(n))^n = b$.

Hint for Exercise 24 (§6.3)

Read the equation out loud — without using "integral".

Hint for Exercise 40 (§6.4)

For which one value of x is $f(x)$ easy to compute?

Hint for Exercise 47 (§6.4)

Use the chain rule.

Hint for Exercise 49 (§6.4)

Rewrite the integral as $\int_{2x}^{0} t\tan(t)\,dt + \int_{0}^{3x} t\tan(t)\,dt$ and manipulate the first integral so 0 is its lower limit.

Hint for Exercise 57 (§6.4)

Not all of the information provided is needed.

Hint for Exercise 63 (§6.4)

Either give two interpretations of “it” or explain why “it” has only one meaning.

Hint for Exercise 71 (§6.4)

Write $\frac{x}{e^x}$ as xe^{-x}.

Hint for Exercise 72 (§6.4)

Write the integral as the sum of two integrals.

Hint for Exercise 15 (§6.5)

Consider the simplest case, $\int_0^1 x^2\,dx$.

Hint for Exercise 16 (§6.5)

Consider the simplest case, $\int_0^1 x^4\,dx$.

Hint for Exercise 34 (§6.5)

Draw the tangent line at the point $\big((a+b)/2, f((a+b)/2)\big)$.

Hint for Exercise 1 (§6.S)

FTC I refers to $F(b)-F(a)$.

Hint for Exercise 2 (§6.S)

FTC II refers to the derivative of $\int_a^x f(t)\,dt$.

Hint for Exercise 64 (§6.S)

It will include a definite integral with integrand $\sqrt{x^2-1}$.

Hint for Exercise 69 (§6.S)

Use trial and error.

Hint for Exercise 72 (§6.S)

Differentiate! And, be careful.

Hint for Exercise 77 (§6.S)

Draw a tangent to the curve at each of the midpoints.

Hint for Exercise 34 (§7.1)

You may have to solve an equation to find an endpoint of the interval of integration.

Hint for Exercise 35 (§7.1)

You may have to solve an equation to find an endpoint of the interval of integration.

Hint for Exercise 38 (§7.1)

Make sure that $f'(0) > 0$.

Hint for Exercise 0 (§7.2)

A jar lid or soda can works fine for drawing circles and circular arcs. Credit cards and ID badges make good straightedges.

Hint for Exercise 5 (§7.3)

Start by drawing a good picture of the local approximation.

Hint for Exercise 7 (§7.3)

Use square cross sections.

Hint for Exercise 14 (§7.3)

Assume that dx is small.

Hint for Exercise 24 (§7.3)

Include a clear picture on which your local approximation is based.

Hint for Exercise 25 (§7.3)

Include a clear picture on which your local approximation is based.

Hint for Exercise 28 (§7.3)

Estimate the area of the narrow band shown in Figure 7.3.10.

Hint for Exercise 0 (§7.5)

When evaluating the definite integral, use the book's Table of Integrals (in Appendix A).

Hint for Exercise 21 (§7.5)

Use FTC I.

Hint for Exercise 22 (§7.5)

Use FTC I.

Hint for Exercise 11 (§7.7)

The integral involves only a, b, and $A(x)$.

Hint for Exercise 12 (§7.7)

Express the integral in terms of a, b, $c(x)$, and h.

Hint for Exercise 21 (§5.8)

This integrand is undefined at both endpoints.

Hint for Exercise 29 (§7.8)

Don't forget that you can use the book's Table of Integrals (in Appendix A), in particular, (80).

Hint for Exercise 37 (§7.8)

Show that $S_n \geq n$ for all positive integers n.

Hint for Exercise 9 (§7.S)

Write $A(t) = \int_t^{t+1} f(x)\, dx = \int_0^{t+1} f(x)\, dx - \int_0^t f(x)\, dx$, then use FTC II.

Hint for Exercise 2 (§C.9)

What information is needed to complete this calculation?

Hint for Exercise 3 (§C.9)

What information is needed to complete this calculation?

Hint for Exercise 15 (§8.1)

$\frac{a+b}{c} = \frac{a}{c} + \frac{b}{c}$

Hint for Exercise 17 (§8.1)

First, expand the integrand.

Hint for Exercise 35 (§8.1)

Use Formula 55 first, followed by Formula 54.

Hint for Exercise 57 (§8.2)

Let $u = 3x + 1$.

Hint for Exercise 59 (§8.2)

Let $u = x + 1$.

Hint for Exercise 64 (§8.2)

Show that $\int_{-a}^{0} f(x)\, dx = -\int_{0}^{a} f(x)\, dx$ by using the substitution $u = -x$. Do not refer to areas.

Hint for Exercise 65 (§8.2)

Show that $\int_{-a}^{0} f(x)\, dx = \int_{0}^{a} f(x)\, dx$ by using the substitution $u = -x$. Do not refer to areas.

Hint for Exercise 40 (§8.3)

Use the substitution $u = \frac{\pi}{2} - \theta$.

Hint for Exercise 46 (§8.4)

You might find it helpful to know $u^4 + 1 = (u^2 + \sqrt{2}u + 1)(u^2 - \sqrt{2}u + 1)$.

Hint for Exercise 54 (§8.4)

First, multiply both sides of the partial fraction representation by $x - c$.

Hint for Exercise 28 (§8.5)

Let $u = \sqrt[6]{x}$.

Hint for Exercise 60 (§8.5)

Confirm that $\theta = 2\arctan(u)$, then differentiate.

Hint for Exercise 10 (§8.S)

Exercise 10 may be helpful.

Hint for Exercise 20 (§8.S)

Does the book's Table of Integrals (in Appendix A) have any suggestions for evaluating this integral?

Hint for Exercise 56 (§8.S)

Use integration by parts and $d(F(x)-1) = f(x)\,dx$.

Hint for Exercise 57 (§8.S)

Use integration by parts with $u = x$ and $dv = f(x)\,dx$.

Hint for Exercise 59 (§8.S)

Start with integration by parts.

Hint for Exercise 62 (§8.S)

Use the Binomial Theorem, as discussed before Exercise 26 in Section 5.5.

Hint for Exercise 66 (§8.S)

The graph of g is the graph of f shifted to the right by a.

Hint for Exercise 74 (§8.S)

Make the substitution $t = x - \mu$.

Hint for Exercise 82 (§8.S)

First show that there is a constant, B, such that $|f(x)| \leq B$ for all x in $[0, \infty)$.

Hint for Exercise 83 (§8.S)

See Exercise 82.

Hint for Exercise 86 (§8.S)

To get started, use $\cos^2(\theta) = 1 - \sin^2(\theta)$.

Hint for Exercise 87 (§8.S)

Use the fact that $\int e^{x^2}\, dx$ is not elementary.

Hint for Exercise 95 (§8.S)

Avoid labor.

Hint for Exercise 101 (§8.S)

First, substitute $u = x^2$ in both. Then, in the second integral, let $u = v + \pi$.

Hint for Exercise 102 (§8.S)

Express $\cos(\theta)$ in terms of the constants A and B.

Hint for Exercise 4 (§C.11)

In one of these two cases the Cauchy-Schwarz inequality will help.

Hint for Exercise 33 (§9.1)

Convex is defined in Section 2.5. Find the points on the curve farthest to the left and compare them to the point on the curve corresponding to $\theta = \pi$.

Hint for Exercise 230.55 (§9.2)

You may need to approximate a limit of integration.

Hint for Exercise 310.55 (§9.2)

Consider the cases with n even or odd separately.

Hint for Exercise 4 (§9.3)

For (d), use the identity $\cos^2(t) + \sin^2(t) = 1$.

Hint for Exercise 13 (§9.3)

A good start is to introduce x and y.

Hint for Exercise 14 (§9.2)

A good start is to introduce x and y.

Hint for Exercise 240.30 (§9.3)

Eliminate t.

Hint for Exercise 26 (§9.3)

Work with the horizontal distance traveled, x, not the distance along the hill.

Hint for Exercise 37 (§9.3)

Make a sketch of the curve near $(0,0)$ and show on it the geometric meaning of the quotients $f(t)/g(t)$ and $f'(t)/g'(t)$.

Hint for Exercise 38 (§9.3)

If $(x(t), y(t))$ is a point on a curve parameterized in the way described for this problem, then $y(t) = tx(t)$.

Hint for Exercise 24 (§9.4)

Express all quantities appearing in this average in terms of θ.

Hint for Exercise 32 (§9.4)

It is known that $\int x^p(1+x)^q \, dx$ is elementary for rational numbers p and q only when at least one of p, q, and $p+q$ is an integer.

Hint for Exercise 38 (§9.4)

Assume that one of the dogs starts on the positive x-axis.

Hint for Exercise 39 (§9.4)

L'Hôpital's Rule does not help. For simplicity, assume $a = 0 = f(0)$.

Hint for Exercise 21 (§9.5)

The identity $1 + \cos(\theta) = 2\cos^2(\theta/2)$ may be helpful.

Hint for Exercise 25 (§9.5)

Let b approach a from the left.

Hint for Exercise 31 (§9.5)

See Exercise 29.

Hint for Exercise 12 (§9.6)

Since the curve is part of a circle of radius a, shouldn't the answer be a?

Hint for Exercise 13 (§9.6)

How does one find the minimum of a function?

Hint for Exercise 32 (§9.6)

For (b) and (c) draw the little triangle whose hypotenuse is like a short piece of arc length ds on the curve and whose legs are parallel to the axes.

Hint for Exercise 32 (§9.6)

Use (b) and (c) to find an equation that involves x, y, and κ. Also, think about antiderivatives.

Hint for Exercise 7 (§9.S)

Using polar coordinates may help.

Hint for Exercise 12 (§10.1)

A limit in Section 2.2 will help.

Hint for Exercise 13 (§10.1)

A limit in Section 2.2 will help.

Hint for Exercise 17 (§10.1)

Write $f(n)^{g(n)}$ as $e^{g(n)\ln(f(n))}$.

Hint for Exercise 39 (§10.2)

This answer depends on the choice of a_0.

Hint for Exercise 5 (§10.3)

Use, for example, in Exercise 6, $f(x) = x^2 - 15$ with $a_0 = 3$ and $b_0 = 4$.

Hint for Exercise 5 (§10.4)

The sequence $\{x_n\}$ and the sequence $\left\{\frac{1}{2}\left(x_n + \frac{a}{x_n}\right)\right\}$ have the same limit.

Hint for Exercise 5 (§10.4)

Review Finding the Limit of a Recursive Sequence in Section 10.2.

Hint for Exercise 17 (§10.S)

Set $R_n = \dfrac{1}{C_n}$.

Hint for Exercise 19 (§10.S)

Disregard higher powers of z_k.

Hint for Exercise 22 (§10.S)

Recall from Section 3.4 (page 133) that $\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e$.

Hint for Exercise 37 (§11.1)

Use $\arctan(1) = \frac{\pi}{4}$

Hint for Exercise 24 (§11.2)

Do not mimic the treatment for $a = 1$.

Hint for Exercise 30 (§11.2)

Show that $1 + x \le e^x$ for $x > 0$.

Hint for Exercise 31 (§11.2)

Assume the series can be multiplied term-by-term.

Hint for Exercise 33 (§11.3)

Use the first three terms and control the tail end by comparing it to the sum of a geometric series.

Hint for Exercise 46 (§11.3)

Use the Cauchy-Schwarz inequality, which appears in CIE 11 at end of Chapter 9 and also in Exercise 29 in Section 16.7.

Hint for Exercise 47 (§11.3)

First show that $|x_k| \leq x_k^2 + 1$.

Hint for Exercise 21 (§11.5)

The only difference is that the k^{th} term is $(-1)^{k+1} p_k$ instead of $(-1)^{k+1}/k$.

Hint for Exercise 17 (§11.6)

First show that there is a number s, $s > 1$, such that for some integer N, $|a_{k+1}| > s|a_k|$ for $k \geq N$.

Hint for Exercise 23 (§11.6)

Begin by using the first few positive summands until their sum exceeds $\sqrt{2}$. Then use the first few negative summands until the total sum so far is less than $\sqrt{2}$. Continue in this manner.

Hint for Exercise 36 (§11.S)

Rewrite the inequality as $a_{k+1}/b_{k+1} \le a_k/b_k$.

Hint for Exercise 15 (§12.1)

Examine $R_n(x;1)$. Consider the cases $x < 1$ and $x > 1$ separately.

Hint for Exercise 11 (§12.2)

Write $4 - x$ as $4(1 - x/4)$ and factor 4 out of the radical.

Hint for Exercise 17 (§12.2)

Use the fact that $e^{-5x^2} < e^{-5x}$ for $x > 1$.

Hint for Exercise 17 (§12.2)

Use the Maclaurin series for e^{-5x^2}.

Hint for Exercise 18 (§12.2)

Use the fact that $|\cos(x)| \le 1$.

Hint for Exercise 18 (§12.2)

Use the Maclaurin series for $\cos(x)$.

Hint for Exercise 27 (§12.2)

See Formula 92 in the Table of Integrals (in Appendix A).

Hint for Exercise 18 (§12.4)

Keep the first five terms of e^x.

Hint for Exercise 25 (§12.4)

To control the error, compare the tail of the series to a geometric series.

Hint for Exercise 28 (§12.4)

This gives $\ln(1/2)$, which is $-\ln(2)$.

Hint for Exercise 28 (§12.4)

Use Newton's method.

Hint for Exercise 29 (§12.4)

Separate the left-hand side into two integrals. Then take the limit as $n \to \infty$.

Hint for Exercise 34 (§12.4)

To get started, and to help see the pattern, write the first four terms. For the general case, use the equation obtained in Exercise 33.

Hint for Exercise 25 (§12.5)

Draw $8 + 8\sqrt{3}i$.

Hint for Exercise 27 (§12.5)

See Exercise 26.

Hint for Exercise 43 (§12.5)

Draw some pictures.

Hint for Exercise 54 (§12.5)

See Exercise 53.

Hint for Exercise 55 (§12.5)

See Exercise 53.

Hint for Exercise 58 (§12.5)

Find a polar equation for this curve.

Hint for Exercise 30 (§12.6)

Use the fact that $1+\cos(\theta)+\cos(2\theta)+\cdots+\cos((n-1)\theta)=\sum_{m=0}^{n-1}\cos(m\theta)$ is the real part of $1+e^{\theta i}+e^{2\theta i}+\cdots+e^{(n-1)\theta i}$.

Hint for Exercise 31 (§12.6)

Use the fact that $\sin(\theta)+\sin(2\theta)+\cdots+\sin((n-1)\theta)=\sum_{m=1}^{n-1}\sin(m\theta)$ is the imaginary part of $1+e^{\theta i}+e^{2\theta i}+\cdots+e^{(n-1)\theta i}$.

Hint for Exercise 37 (§12.6)

Use $|a+bi|\leq\sqrt{a^2+b^2}$.

Hint for Exercise 4 (§12.7)

Use the fact that $\frac{2}{a^2-1} = \frac{1}{a-1} - \frac{1}{a+1}$.

Hint for Exercise 12 (§12.7)

Note that this is a closed interval. What happens at the endpoints?

Hint for Exercise 2 (§12.S)

Show that its natural logarithm approaches a.

Hint for Exercise 7 (§12.S)

Use Exercise 6.

Hint for Exercise 23 (§12.S)

Use Euler's formula.

Hint for Exercise 48 (§13.1)

Separate and use partial fractions.

Hint for Exercise 54 (§13.1)

Differentiate the differential equation with respect to t and remember that $P = P(t)$.

Hint for Exercise 30 (§13.4)

The general solution will be a superposition of four independent solutions.

Hint for Exercise 4 (§13.S)

Factor the right-hand side.

Hint for Exercise 18 (§13.S)

Use partial fractions.

Hint for Exercise 26 (§14.1)

Draw a picture.

Hint for Exercise 27 (§14.1)

Draw a picture.

Hint for Exercise 31 (§14.1)

What angles do they make with the x-axis?

Hint for Exercise 32 (§14.1)

First, draw them.

Hint for Exercise 32 (§14.1)

Draw $\mathbf{i}$, $\mathbf{u}_1$, and $\mathbf{u}_2$.

Hint for Exercise 36 (§14.1)

Use the Pythagorean Theorem twice.

Hint for Exercise 30 (§14.2)

Look at the proof in the text for vectors in the plane.

Hint for Exercise 45 (§14.2)

Use the dot product.

Hint for Exercise 48 (§14.2)

Show that its four faces are equilateral triangles.

Hint for Exercise 48 (§14.2)

θ is the angle between $\overrightarrow{DA}$ and $\overrightarrow{DB}$.

Hint for Exercise 24 (§14.3)

Think of $\mathbf{A} \times \mathbf{B}$ as a single vector, $\mathbf{E}$.

Hint for Exercise 26 (§14.3)

Think of $\mathbf{B} + \mathbf{C}$ as a single vector $\mathbf{E}$.

Hint for Exercise 39 (§14.3)

Draw a diagram.

Hint for Exercise 40 (§14.3)

Draw a large, clear picture.

Hint for Exercise 44 (§14.4)

Draw a picture and think in terms of vectors.

Hint for Exercise 61 (§14.4)

Use parametric equations but give the parameters of the lines different names, such as t and s.

Hint for Exercise 6 (§14.S)

It can be helpful to start by drawing the diagram for yourself. Be sure it is large enough to be labeled.

Hint for Exercise 11 (§14.S)

Recall that normal means perpendicular.

Hint for Exercise 30 (§14.S)

Set up coordinate systems in the plane of the circle and in the plane of its shadow, which we can take to be the xy-plane. Choose the axes for the coordinate systems to be as convenient as possible. Then express the equation of the shadow in terms of x and y by utilizing the equation of the circle.

Hint for Exercise 31 (§14.S)

Use the Pythagorean Theorem 3 times.

Hint for Exercise 32 (§14.S)

See Exercises 15 and 16.

Hint for Exercise 34 (§14.S)

Think about direction angles and direction cosines. See also Exercise 65.

Hint for Exercise 32 (§15.1)

At time $t = 0$, the rock is at $(0,0)$ and the x-axis is horizontal. Time is in seconds and distance is in feet.

Hint for Exercise 33 (§15.1)

See Section 9.4.

Hint for Exercise 44 (§15.1)

$\mathbf{v}(t)$ is approximated by $\Delta\mathbf{r}/\Delta t$.

Hint for Exercise 32 (§15.2)

Start with the definition, $\kappa = |d\mathbf{T}/ds|$.

Hint for Exercise 34 (§15.2)

See Exercise 27 and start with $\mathbf{r} \cdot \mathbf{r} = a^2$.

Hint for Exercise 21 (§15.4)

Is the net outward flow positive or negative?

Hint for Exercise 21 (§15.4)

Answer on the basis of your diagram in (a).

Hint for Exercise 33 (§15.4)

Recall that $\mathbf{r} = \overrightarrow{OP}$ where P is a point on the closed curve C.

Hint for Exercise 33 (§15.4)

Think about when $\mathbf{r} \cdot \mathbf{n}$ is negative and when it is positive.

Hint for Exercise 10 (§15.S)

Let $\mathbf{s}$ is a vector function. To differentiate $|\mathbf{s}|$, start with $\mathbf{s} \cdot \mathbf{s} = |\mathbf{s}|^2$.

Hint for Exercise 4 (§C.20)

No calculus is required.

Hint for Exercise 1 (§C.21)

It always lies within the sun.

Hint for Exercise 1 (§C.22)

Draw a picture.

Hint for Exercise 10 (§C.22)

What is the vector identity for $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$?

Hint for Exercise 12 (§C.22)

Use the vector identity $(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C} = \mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$.

Hint for Exercise 12 (§C.22)

Use the same vector identity as in (a).

Hint for Exercise 15 (§16.1)

Compute $f(1,1)$.

Hint for Exercise 30 (§16.2)

Should these two mixed second-order derivatives be equal?

Hint for Exercise 53 (§16.2)

Start with the definition of $g'(y)$ where $g(y) = \int_a^b f(x,y)\,dx$.

Hint for Exercise 19 (§16.3)

Use the relation between rectangular coordinates (x, y) and polar coordinates (r,θ).

Hint for Exercise 44 (§16.4)

To simplify the notation, write g as $A\cos(\theta) + B\sin(\theta)$.

Hint for Exercise 12 (§16.5)

The proof is a slight modification of the proof of Theorem 16.5.1.

Hint for Exercise 17 (§16.5)

Let P_0 and P_1 be two points on $\mathscr{S}$. Take a curve with parameterization $(x(t), y(t), z(t))$ and show that on it $|\mathbf{r}|$ is constant.

Hint for Exercise 36 (§16.6)

Express f in terms of θ.

Hint for Exercise 47 (§16.6)

Show it is perpendicular to each curve on the surface that passes through P.

Hint for Exercise 29 (§16.7)

Let $x_i = \frac{a_i}{(\sum_{i=1}^n a_i^2)^{1/2}}$ and $y_i = \frac{b_i}{(\sum_{i=1}^n b_i^2)^{1/2}}$. See also the Average Speed CIE (CIE 11) at the end of Chapter 7.

Hint for Exercise 8 (§16.8)

Duplicate the steps leading to (16.8.9).

Hint for Exercise 13 (§16.9)

The computations are simpler than expected because four of the eight summands for the composite function cancel.

Hint for Exercise 16 (§16.9)

Use the definition of magnification.

Hint for Exercise 17 (§16.9)

First review how magnification of a mapping from uv-space to xy-space was defined and then expressed as a determinant. Instead of a cross product the formula in Section 14.3 for the volume of a parallelepiped spanned by three vectors is needed. The 3×3 matrix is also called a Jacobian.

Hint for Exercise 21 (§16.9)

Use the fact that $e^{x+iy} = e^x \left(\cos(y) + i \sin(y)\right)$.

Hint for Exercise 27 (§16.S)

Consider a slight change in x_i and x_j, with the other x_k's held fixed.

Hint for Exercise 21 (§17.3)

Choose the pole wisely.

Hint for Exercise 29 (§17.3)

Use polar coordinates.

Hint for Exercise 33 (§17.5)

The answer depends on r and h.

Hint for Exercise 14 (§17.6)

What is the equation of the circle in polar coordinates when the polar axis is along the positive x-axis?

Hint for Exercise 38 (§17.6)

Do not use an iterated integral; use symmetry.

Hint for Exercise 40 (§17.6)

For some ρ, $\sqrt{H^2+\rho^2 A-2\rho H}$ equals $H-\rho$ and for some it equals $\rho-H$.

Hint for Exercise 45 (§17.6)

Examine the integral over the region between concentric spheres of radii a and t, and let $t \to 0^+$.

Hint for Exercise 23 (§17.7)

Note where $\hat{\mathbf{r}} \cdot \mathbf{n}$ is positive and where it is negative.

Hint for Exercise 34 (§17.7)

See Section 15.4 for fluid flow in a plane region.

Hint for Exercise 35 (§17.7)

Differentiate the first equation.

Hint for Exercise 36 (§17.7)

Exercise 35 may be useful.

Hint for Exercise 37 (§17.7)

Break $\mathscr{S}$ into three parts. In one part $\cos(\gamma)$ is positive; in another $\cos(\gamma)$ is negative; and in the other it is negative.

Hint for Exercise 31 (§17.8)

Use a table of integrals or techniques from Chapter 8.

Hint for Exercise 12 (§17.9)

See Exercise 7 in Section 16.9.

Hint for Exercise 15 (§17.9)

Sketch the images of the edges of $\mathscr{R}$.

Hint for Exercise 21 (§17.S)

Review the proof of the Cauchy-Schwarz inequality presented in the CIE 11 (Average Speed and Class Size) at the end of Chapter 8

Hint for Exercise 22 (§17.S)

See Exercise 21.

Hint for Exercise 36 (§17.S)

Show that the x'-axis and the y'-axis are balancing lines.

Hint for Exercise 37 (§17.S)

Use repeated integrals in cylindrical coordinates.

Hint for Exercise 38 (§17.S)

Why is the iterated integral less than or equal to 0?

Hint for Exercise 8 (§18.1)

See Section 16.2 for the definitions of open and closed sets.

Hint for Exercise 15 (§18.1)

No computations are needed.

Hint for Exercise 7 (§18.2)

When the fingers of your right hand copy the direction of the flow, your thumb points in the direction of the curl, up or down.

Hint for Exercise 11 (§18.2)

Is the domain of **F** is simply connected? Is $\nabla \times \mathbf{F} = \mathbf{0}$?

Hint for Exercise 43 (§18.2)

Assume C is given parametrically as $x = r(\theta)\cos(\theta)$, $y = r(\theta)\sin(\theta)$, for $\alpha \le \theta \le \beta$.

Hint for Exercise 14 (§18.3)

Write **F** with a common denominator.

Hint for Exercise 24 (§18.3)

Recall that there are several equivalent line integrals for finding the area of the enclosed region.

Hint for Exercise 26 (§18.3)

First express **F** in rectangular coordinates.

Hint for Exercise 28 (§18.3)

Break $\mathscr{R}$ into two regions that have no holes, as in Exercise 27.

Hint for Exercise 32 (§18.3)

At each point P on a stream line, $\mathbf{F}(P)$ is tangent to the streamline.

Hint for Exercise 5 (§18.4)

First write $\mathbf{F}(x, y)$ in terms of x and y.

Hint for Exercise 26 (§18.4)

The integration is much easier in polar coordinates.

Hint for Exercise 11 (§18.5)

$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$.

Hint for Exercise 30 (§18.5)

Choose your coordinate system carefully.

Hint for Exercise 4 (§18.6)

Recall that $\mathbf{r} = r\hat{\mathbf{r}}$.

Hint for Exercise 31 (§18.6)

Break the integral into integrals over $\mathscr{S}_1$ and $\mathscr{S}_2$.

Hint for Exercise 13 (§18.7)

Let the point be $(0, 0, a)$. Steradians might help.

Hint for Exercise 14 (§18.7)

Use symmetry, you should not have a need to evaluate any integrals.

Hint for Exercise 16 (§18.7)

Is the square root of $(b-a)^2$ still $b-a$?

Hint for Exercise 19 (§18.8)

Solve the equations for $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$ found in Exercise 18 for $\mathbf{i}$ and $\mathbf{j}$.

Hint for Exercise 14 (§18.9)

Assume that $\frac{\partial}{\partial t}\int_{\mathscr{S}} \mathbf{B}\cdot\mathbf{n}\,dS$ equals $\int_{\mathscr{S}} \frac{\partial B}{\partial t}\cdot\mathbf{n}\,dS$.

Hint for Exercise 15 (§18.S)

No integration is necessary.

Hint for Exercise 21 (§18.S)

No integration is needed.

Hint for Exercise 21 (§18.S)

See Exercise 86 in Section 8.S.

Hint for Exercise 27 (§18.S)

Make a sketch with (x, y) in the first quadrant.

Hint for Exercise 6 (§C.25)

Consider k positive, negative, and zero.

Hint for Exercise 7 (§C.25)

Consider k positive, negative, and zero.

Hint for Exercise 8 (§C.25)

Consider k positive, negative, and zero.

Hint for Exercise 10 (§C.25)

Draw a picture.

Hint for Exercise 10 (§C.25)

Draw a picture.

Index