Solving Equations, an Elegant Legacy

Jerry L. Kazdan

Solving equations—one of the primary themes in mathematics. In discussing this my secondary theme is that classical and modern mathematics are tightly intertwined, that contemporary mathematics contributes real insight and techniques to understand traditional problems.

In the long second section I discuss some procedures that help to solve equations. There the discussion of symmetry is extensive because it is treated so inadequately as a fundamental thread throughout mathematics courses. The third section gives two different techniques to prove that equations have solutions. Most of the sections are independent; thus you can skip to examples that are more appealing.

One ingredient in solving equations that I have not emphasized adequately is the basic role of inequalities. They are lurking here and there: the Euclidean algorithm and the application of the Brouwer fixed point theorem, to name two less obvious instances. To give inequalities their due would have changed the character of this article, which is drawn from the longer version [**10**].

1. INTRODUCTION. Although by 1535 mathematicians had a formula to solve the cubic $p(x) := x^3 + bx^2 + cx + d = 0$, even without the formula it is easy to show that if the roots are x_1, x_2, x_3, then expanding $p(x) = (x - x_1)(x - x_2)(x - x_3)$ we get for instance

$$x_1 + x_2 + x_3 = -b. \tag{1}$$

An immediate consequence is that if the coefficients in a cubic polynomial are rational and if two of the roots are rational, then so is the third root. This result introduces one thread in our story: even without a formula for the solution of an equation it may be possible to obtain useful information.

Shortly after the cubic, the general quartic polynomial was solved. The next challenge was the quintic. If the coefficients of

$$p(x) := x^5 + bx^4 + cx^3 + dx^2 + ex + f, \tag{2}$$

are real, for all large positive x we have $p(x) > 0$, while for all large negative x we have $p(x) < 0$. Thus if you graph the polynomial $y = p(x)$, it is geometrically evident that it crosses the x-axis at least once and hence there is at least one real root x_1 of $p(x) = 0$. The polynomial $q(x) := p(x)/(x - x_1)$ is then a quartic polynomial for whose four roots there are formulas. Thus it was known that every quintic polynomial has five (some possibly repeated or complex) roots. But Abel showed that despite knowing these five roots *exist*, there cannot be a general *formula* for them involving only the usual arithmetic operations along with taking roots. Formulas similar to (1) were essential in Abel's reasoning.

Mathematicians found themselves in the fascinating dilemma of having proved that these roots exist but also having proved that there can never be an algebraic formula for them. The existence proof is what we now call the *Fundamental Theorem of Algebra*, while understanding the obstructions to finding formulas for

the roots is *Galois theory*. Both were vital pillars in the future development of mathematics. As a twist of fate, except for their fundamental historic role, the formulas for the solutions of the cubic and quartic have become museum pieces, rarely used because they are so complicated.

The proof that the quintic (2) always has at least one real root was one of the first "pure" existence proofs. Although this proof was regarded as obvious, in the nineteenth century mathematicians became concerned because this proof presumes that the real number line has no "holes." What would happen if there were a hole in the number line exactly where the root should have been? How can one precisely define this "no holes" property? This became the *completeness* property. It is essential for problems involving limiting processes.

2. STEPS TOWARD SOLVING EQUATIONS. In solving equations, the most primitive question is to decide if there are any solutions at all. From our understanding of polynomial equations, we have learned to separate this from the important problem of explicitly finding solutions. Moreover, in the many cases where we know there is a solution but there is no "formula," you need qualitative properties of the solution.

2.1 Find a formula for a solution. Usually there is no formula of any sort. A power or Fourier series solution may be difficult to understand. A numerical solution may be a jumble of numbers that you need to decipher to learn anything useful. Hamming's assertion: "The purpose of computing is insight, not numbers," applies to most scientific computations.

There are elementary problems where there are no formulas for the solution, but there is an algorithm for finding a solution. Even in such cases occasionally you may prefer a non-constructive proof that a solution exists.

An example is solving $ax \equiv b \pmod{m}$, where a and m are relatively prime. Since the solution is $x \equiv a^{-1}b \pmod{m}$, we need to find $a^{-1} \pmod{m}$. One approach is to observe that the numbers $a, 2a, \ldots, (m-1)a$ are all distinct $\pmod{m}$ so one of them must be $1 \pmod{m}$. This proof that a^{-1} exists gives no hint of how to find it except by trial and error. This is a non-constructive existence proof for the solution of $ax \equiv 1 \pmod{m}$. One constructive proof considers the equivalent problem of solving $ax - my = 1$ for integers x, y. The Euclidean algorithm solves this explicitly (see [**4**, Sec. I.8]). Since at the kth step in this algorithm the absolute value of the remainder can be chosen to be at most half the value of the previous remainder, this new remainder is at most $a/2^k$ so you need at most $\log a/\log 2$ steps (this is one of the few places that we consider the important issue of the efficiency of an algorithm).

2.2 Find an equivalent problem that is simpler. Making a *change of variable* is perhaps the most familiar technique to simplify a problem. A small example of this is the cubic polynomial $p(x) = ax^3 + bx^2 + cx + d$. View this as a Taylor series. Since the second derivative is zero at the point where $6ax + 2b = 0$, the change of variables $z = 6ax + 2b$ (or just the translation $z = x + b/3a$) yields a simpler polynomial $q(z) = \alpha z^3 + \gamma z + \delta$ without a quadratic term. If the coefficients of the original equation were rational, then so are those of the new equation and the rational roots of the new equation correspond to those of the original equation. This is a generalization of the procedure of "completing the square." Similarly, by a translation one can eliminate the coefficient a_{n-1} in $p(x) = x^n + a_{n-1}x^{n-1} +$ lower order terms.

We can use this to show that *every double root of a cubic polynomial with rational coefficients is rational.* Using our change of variable, it is enough to show this for $q(z) = \alpha z^3 + \gamma z + \delta$. Thus, we must show that if $q(r) = 0$ and $q'(r) = 0$, then r is rational. But $0 = q'(r) = 3\alpha r^2 + \gamma$ implies that $\alpha r^3 = -(\gamma/3)r$. Thus $0 = q(r) = -(\gamma/3)r + \gamma r + \delta$, that is, $r = -3\delta/2\gamma$, which is rational. From (1), since $x_1 = x_2 = r$, the third root of q (and hence of p) is also rational.

For cubic polynomials with rational coefficients and having a double root r (necessarily rational) you can now find all *rational points* (x, y), that is, both x and y are rational, on the "elliptic curve" $y^2 = p(x)$. They are the points where straight lines through $(r, 0)$ and having rational slope intersect the curve. A related exercise is to show that the rational points on the circle $x^2 + y^2 = 1$ are where the straight lines through $(1, 0)$ with rational slope intersect the circle. An easy consequence is a formula for all the "Pythagorean triples": the integers a, b, c with $a^2 + b^2 = c^2$.

Another instance of finding an equivalent problem that is simpler is the change of variable (that is, a change of basis) in a matrix equation to diagonalize the matrix (if possible). We can use the same idea for a system of differential equations

$$Lu := u' + Au = f, \tag{3}$$

where $u(t)$ and $f(t)$ are vectors and $A(t)$ is a square matrix. We seek a change of variables $u = Sv$, where $S(t)$ is an invertible matrix, to transform this to a simpler equation. In some applications this is called a *gauge transformation.* To find a useful S we compute

$$f = Lu = u' + Au = (Sv)' + A(Sv) = Sv' + (S' + AS)v. \tag{4}$$

The right side of this is simplest if S is a solution of the matrix equation

$$LS = S' + AS = 0, \quad \text{say with} \quad S(0) = I; \tag{5}$$

we use $S(0) = I$ to ensure that S is invertible. Then solving (4) is just integrating $v' = g$, where $g = S^{-1}f$.

With this choice of S and writing $D := d/dt$ it is instructive to rewrite (4) as $f = Lu = SDv = SDS^{-1}u$. In particular, $L = SDS^{-1}$. One sees that *every* linear ordinary differential operator is "conjugate" or "gauge equivalent" to D. We thus come to the possibly surprising conclusion that *any* first order linear differential operator L is equivalent to the simple operator D; this makes studying linear ordinary differential operators far easier than partial differential operators. We also have formally $L^{-1} = SD^{-1}S^{-1}$. Since D^{-1} is just integration (and adding a constant), an immediate consequence is that the general solution of the inhomogeneous equation $Lu = f$ is

$$u(t) = L^{-1}f = S(t)C + S(t)\int_0^t S^{-1}(\tau)f(\tau)\,d\tau, \tag{6}$$

where $C = u(0)$. The matrix S defined by (5) is the usual *fundamental matrix solution* one meets for ordinary differential equations. Unfortunately it is presented frequently as a trick to solve the inhomogeneous equation rather than as a straightforward approach to reduce the study of L to the simpler differential operator D. It is sometimes useful to introduce *Green's function* $G(t, \tau) := S(t)S^{-1}(\tau)$ and rewrite (6) as

$$u(t) = u(0) + \int_0^t G(t, \tau)f(\tau)\,d\tau. \tag{7}$$

We then think of the integral operator with kernel $G(t, \tau)$ as L^{-1}. This integral can be interpreted physically and gives another (equivalent) approach to solving (3).

Usually S cannot be found explicitly. However in special cases such as a single equation or a 2×2 system with constant coefficients, you can carry out the computations and obtain the classical formulas quickly.

What we call a "change of variable" is part of a fundamental procedure known to everyone since childhood. As an illustration, say you have a problem $\mathcal{P}$ that is stated in another language, perhaps Latin. To solve it, first translate (T) it into your language, solve the translated version $\mathcal{Q}$, and then translate it back (T^{-1}). Symbolically, reading from right to left, $\mathcal{P} = T^{-1}\mathcal{Q}T$; see Figure 1. The goal is to choose the new setting and T so the new problem $\mathcal{Q}$ is easier than $\mathcal{P}$. Diagonalizing a matrix and using a Laplace transform are two familiar mathematical examples.

$$\begin{tikzcd}
\textit{old} \arrow[r, "\mathcal{P}", "\text{(original version)}"'] \arrow[d, "T"'] & \textit{old} \\
\textit{new} \arrow[r, "\mathcal{Q}", "\text{(new version)}"'] & \textit{new} \arrow[u, "T^{-1}"']
\end{tikzcd}$$

Figure 1

The same idea—but with a different twist—is also useful in discussing symmetry (Section 2.6). There we see that finding a T so that the new version is the *same* as the old, $\mathcal{P} = T^{-1}\mathcal{P}T$, is how one identifies symmetries of the problem $\mathcal{P}$. As a silly linguistic illustration, one observes the phrase "madam I'm Adam" reads the same backwards. Here T is the operation of reading backward.

The *calculus of variations* offers a radical way to reformulate some problems. In $\mathbb{R}^n$, if A is a self-adjoint matrix, then solving $Ax = b$ is equivalent to finding a critical point of the scalar-valued function $J(x) = \frac{1}{2}\langle Ax, x\rangle - \langle x, b\rangle$, where we use the standard inner product in $\mathbb{R}^n$. To see this, if x is a critical point of $J(x)$ and if we set $\varphi(\epsilon) = J(x + \epsilon v)$, where v is a vector and ϵ a scalar, then by the definition of critical point, $\varphi'(0) = 0$. But $\varphi'(0) = \langle Ax - b, v\rangle$. Thus $\langle Ax - b, v\rangle = 0$ for *any* vector v. Since v is arbitrary this implies $Ax - b = 0$, as asserted.

Similarly, we claim that solving the Laplace equation $\Delta u = 0$ in a region Ω, with u satisfying the boundary conditions $u(x, y) = f(x, y)$ for (x, y) on the boundary of Ω, is equivalent to finding a critical point of the functional

$$J(u) := \tfrac{1}{2}\iint_{\Omega}\left(u_x^2 + u_y^2\right) dx\, dy \tag{8}$$

among all suitable smooth functions satisfying the same boundary conditions.

To show this, for simplicity, say u minimizes J. Let $h(x, y)$ be any smooth function in Ω that is zero in a neighborhood of the boundary. Then $u + th$ has the

same boundary values as u and the function $\varphi(t) = J(u + th)$ has a minimum at $t = 0$, so

$$0 = \left.\frac{dJ(u + th)}{dt}\right|_{t=0} = \iint_\Omega (u_x h_x + u_y h_y)\, dx\, dy.$$

Now we integrate by parts (the divergence theorem) and observe there are no boundary terms because h is zero on the boundary:

$$0 = -\iint_\Omega (u_{xx} + u_{yy}) h\, dx\, dy = -\iint_\Omega (\Delta u) h\, dx\, dy. \tag{9}$$

This implies that u satisfies the Laplace equation $\Delta u = 0$ since if, say, $\Delta u > 0$ in some small disk in Ω, then we could choose h to be positive in this disk and zero outside. But then the right side of (9) would be positive, a contradiction. To summarize, we see that the minima of J are harmonic functions, $\Delta u = 0$. In fact, all we really used was that u is a critical point of $J(u)$. The equation $\Delta u = 0$ is called the *Euler-Lagrange equation* of the variational problem of finding critical points of $J(u)$.

One virtue of introducing a variational problem is that some properties, such as the existence of a solution or a "conservation law," may be more accessible. We meet this again in Sections 2.6d and 3.1 (see also the books **[6]**, **[7]**, **[9]**).

2.3 Duality: Find a related problem that is useful. To me, duality is the most vague and mysterious item in this article. My impression is that duality appeared first in projective geometry where one interchanges the roles of points and lines. Lagrange introduced the adjoint of a differential operator in the eighteenth century, while the adjoint of a matrix seems to have been used significantly only in the nineteenth century. Lagrangian and Hamiltonian mechanics are dual: Lagrangian living on the tangent bundle, Hamiltonian on the cotangent bundle. There are dual problems in the calculus of variations—including linear programming. Cohomology is the dual of homology. Duality is even a standard device in rhetoric: "Do unto others as you would want others do unto you," and J. F. Kennedy's "... ask not what your country can do for you, ask what you can do for your country". I do not know how to make the concept of duality precise enough to fit all known mathematical instances and ease introduction of new dual objects.

In Section 2.5 we mention the use of duality in linear algebra. Here we follow Lagrange and define the formal adjoint L^* of a linear differential operator L. Use the inner product for real-valued functions: $\langle \varphi, \psi \rangle = \int \varphi\psi\, dx$. Then L^* is defined by the usual rule $\langle u, L^* v\rangle = \langle Lu, v\rangle$ for all smooth functions u and v that are zero outside a compact set; we choose functions that are zero outside a compact set to avoid having boundary terms when we integrate by parts. We use the word "formal" since the strict adjoint requires a (complete) Hilbert space and the consideration of boundary conditions.

If $L := d/dt$, then an integration by parts reveals that

$$\langle Lu, v\rangle = \int u' v\, dt = -\int u v'\, dt = \langle u, L^* v\rangle.$$

Thus, the formal adjoint of $L := d/dt$ is $L^* = -d/dt$. Similarly, if $A(t)$ is a matrix and $u(t)$ is a vector, then the formal adjoint of $Lu := u' + A(t)u$ is $L^* v = -v' + A^*(t)v$. Two integrations by parts show that the formal adjoint of the second order system $Mu := u'' + A(t)u$ is $M^* v = v'' + A^*(t)v$. In particular, if A is symmetric then M is formally self-adjoint, a fact that is basic in quantum

mechanics where, with a complex inner product, the self-adjoint operator id/dt appears in the Schrödinger equation.

If u is a solution of the homogeneous system $Lu := u' + A(t)u = 0$ and v is a solution of the adjoint system, $L^*v = -v' + A^*(t)v = 0$, then their pointwise inner product $v \cdot u$ is a constant. Indeed,

$$\frac{d}{dt}(v \cdot u) = v' \cdot u + v \cdot u' = A^* v \cdot u - v \cdot Au = 0. \tag{10}$$

Observing that $v \cdot u$ is the matrix product $v^* u$, a similar computation shows that if $S(t)$ and $T(t)$ are (not necessarily square) matrix solutions of $LS = 0$ and $L^*T = 0$, respectively, then

$$T^*(t)S(t) = \text{constant}. \tag{11}$$

In particular, if A, S, and T are square matrices with $S(0) = T(0) = I$ (as in (5), S and T are then *fundamental matrix solutions*), we have

$$T^*(t)S(t) = I \quad \text{that is,} \quad T(t) = S^{-1*}(t). \tag{12}$$

If formula (11) appears boring, the disguise is perfect. It is a wide-sweeping generalization both of $e^t e^{-t} = 1$, which is the special case of $Lu := u' + u$, so $L^* v = -v' + v$, as well as $\cos^2 t + \sin^2 t = 1$. In a physical context it may express some conservation law. To prove $\cos^2 t + \sin^2 t = 1$, consider the second order system $Mw := w'' + Cw = 0$, where $C(t)$ is an $n \times n$ matrix, with corresponding adjoint system $M^*z = z'' + C^*z$. Let $\varphi(t)$ and $\psi(t)$ be (vector or matrix) solutions of $M\varphi = 0$ and $M^*\psi = 0$, respectively. Then $\cos^2 t + \sin^2 t = 1$ is the special case of the identity $\psi^{*\prime}(t)\varphi(t) - \psi^*(t)\varphi'(t) = \text{constant}$ when C is the 1×1 identity matrix, $\varphi(t) = \cos t$, and $\psi(t) = \sin t$. This identity is a routine consequence of the basic identity (11) and requires no additional insight to discover; merely rewrite $w'' + C(t)w = 0$ as a first order system by the usual procedure of letting $u_1 := w$ and $u_2 := w'$.

2.4 Understand the family of all solutions. How many solutions are there? Is uniqueness desirable? If so, what conditions would ensure uniqueness of the solution? If you slightly modify some parameters in the problem, do the solutions change only slightly? This continuous dependence on parameters is basic in real-life problems where the data are known only approximately. It is also important for problems solved using a computer that introduces both round-off errors (computers use only a finite number of decimal places) and truncation errors (computers approximate limiting processes such as integration by finite discrete operations).

For instance, by Rouche's theorem in complex analysis the roots of a polynomial $p(z)$ depend continuously on the coefficients, that is, if $p(z)$ has k roots in the small disk $|z - c| < r$ and if we perturb the coefficients of p only slightly, then this perturbed polynomial also has exactly k roots in this disk. A corollary is that the eigenvalues of a matrix depend continuously on the elements of the matrix. The example $x^2 = \pm\epsilon$ shows that these assertions may be *false* if one considers only real roots.

This example $x^2 = \epsilon$ for ϵ near zero also shows that, even allowing complex roots, the solution may *not* be a differentiable function of the parameter. By contrast, the implicit function theorem shows easily that simple roots do depend smoothly on parameters. Here is a proof. Say we have a polynomial $p(x,c)$ depending smoothly on a parameter c and at $c = c_0$ we have a root x_0, so $p(x_0, c_0) = 0$. Since x_0 is a simple root, $\partial p(x, c_0)/\partial x|_{x=x_0} \neq 0$. Thus, by the

implicit function theorem, for all c near c_0, there is a unique solution $x = x(c)$ near x_0 of $p(x, c) = 0$. This solution depends smoothly on c. This proof does not require p to be a polynomial. A consequence is that simple eigenvalues of a matrix depend smoothly on the elements of the matrix.

The study of what happens when one cannot apply the implicit function theorem is carried out in *bifurcation theory* and in the study of *singularities of maps*; these are the same subjects, although they arose from different origins with different viewpoints; see [**1**] and [**8**]. The key new phenomenon is that several solutions can branch—or solutions can disappear—as occurs for the real solutions of $x^2 = \epsilon$; for $\epsilon < 0$ there are no real solutions while for $\epsilon > 0$ there are two solutions. An early notable appearance of bifurcation theory was Euler's classical study of the buckling of a slender column under compression [**16**, pp. 167–169].

Our next example sheds light on the family of all solution of certain polynomial equations. In high school we solve quadratic equations in one variable and systems of linear equations. These are the first steps in understanding the solutions of a system of k polynomial equations $f_1(z) = 0, f_2(z) = 0, \ldots, f_k(z) = 0$ in n unknowns $z = (z_1, \ldots, z_n)$. From experience we know that it is simplest to allow complex numbers as solutions. If there are more equations than unknowns ($k > n$), then usually there will be no solutions, that is, no common zeroes [CHALLENGE: restate this precisely and then prove it, say for any smooth functions f_j], while if there are more unknowns than equations there are usually infinitely many solutions. If there are the same number of equations as unknowns, then usually there are only finitely many solutions. While plausible, this is not obvious and is false for smooth functions: $\sin x = 0$, which is one equation in one unknown, has infinitely many solutions—hardly a surprise if one views its Taylor series as a polynomial of infinite degree.

Bezout made this precise for two polynomial equations in two variables:

$$f(x,y) = 0, \qquad g(x,y) = 0. \tag{13}$$

If f has degree k and g degree l, he proved that there are exactly kl solutions, possibly complex, unless f and g have a common (non-constant) polynomial factor [**15**].

In geometric language think of the two equations (13) as defining two curves C_1 and C_2. The common solutions of (13) are the points where the curves intersect.

We examine (13) in the special case where f and g are both cubic polynomials. Say they intersect at the nine points $p_1 = (x_1, y_1), \ldots, p_9$. So far this is quite general. But now assume that six of these, $p_1, \ldots, p_6$, happen to lie on a conic Γ, so they are also roots of the quadratic polynomial $q(x,y) = 0$ that defines Γ. By Bezout, we know that C_1 and Γ intersect in six points, so it is quite special if C_2 and Γ intersect in the *same* six points. For simplicity also assume that the conic Γ is irreducible, that is, it is not the product of two non-constant polynomials of lower degree (this is the case if Γ is the product of two linear polynomials and hence is just two straight lines). We claim that the remaining three points p_7, p_8, p_9 lie on a straight line.

Here is an algebraic proof. For any linear combination $h(x, y) := \alpha f(x, y) + \beta g(x, y)$, notice that the cubic curve C defined by $h = 0$ automatically contains the points where C_1 and C_2 intersect. Pick another point v on the conic Γ and choose α and β so that v is also a zero of h. Then the cubic curve C also intersects the conic Γ at the *seven* points $v, p_1, \ldots, p_6$. But by Bezout's theorem C and Γ have $3 \cdot 2 = 6$ points of intersection unless h and q have a common factor.

Thus there must be a common factor. Because q is irreducible, the factor must be q itself, so $h(x, y) = q(x, y)r(x, y)$ where, by matching degrees, $r(x, y)$ is a linear polynomial. Thus p_7, p_8, p_9, which are zeroes of $h = 0$ but not $g = 0$, are roots of the linear polynomial $r = 0$ and thus lie on a straight line.

One can reinterpret this to obtain a classical theorem of Pascal. Connect any six points $p_1, \ldots, p_6$ on a conic to obtain a "hexagon", probably with self-intersections. Some terminology for hexagons: a pair of sides separated by two sides is called *opposite* (as $\overline{p_1 p_2}$ and $\overline{p_4 p_5}$) while the points of intersection of opposite sides are called *diagonal points*. Thus a hexagon has three diagonal points (circled in Figure 2). Pascal's theorem asserts that these three points always lie on a straight line.

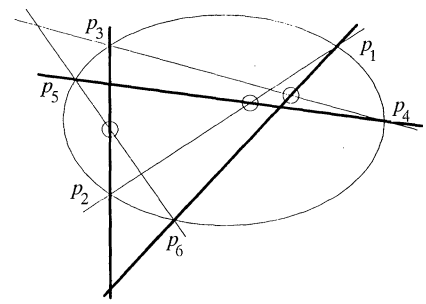

Figure 2

To prove it, take the alternate edges of the hexagon, $\overline{p_1 p_2}$, $\overline{p_3 p_4}$, $\overline{p_5 p_6}$, and $\overline{p_2 p_3}$, $\overline{p_4 p_5}$, $\overline{p_6 p_1}$, to obtain two triangles whose sides contain these edges. To each triangle we associate a cubic polynomial by taking the product of the three linear polynomials determined by the edges of the triangle. Note that here a triangle is the union of the three entire lines, not just the segments joining vertices. Then the points $p_1, \ldots, p_6$ plus the three diagonal points are the nine points of intersection of these two triangles. Now apply the preceding algebraic result. To include the possibility that some pairs of opposite sides might be parallel—so the corresponding points of intersection are at infinity—it is better if one works in the projective plane.

The algebraic reasoning generalizes immediately: *Let $f(x, y) = 0$, $g(x, y) = 0$ be polynomials of degree n that intersect at n^2 points. If kn of these points lie on an irreducible curve defined by a polynomial of degree k, then the remaining $n(n - k)$ points lie on a curve defined by a polynomial of degree $n - k$.* This generalization illustrates the power of the algebraic approach.

2.5 If a solution does not always exist, find the obstructions. If an equation does not have a solution, it is important to understand the reason. If you are trying to fit a straight line $p = at + b$ to the k data points that were found experimentally, then it is unlikely that there will be a choice of the coefficients a and b that fits the data exactly. In this situation one seeks an "optimal" approximate solution. A typical approach to solving $F(x) = y$ approximately is to find a solution x_0 that minimizes the error: $E(x) = \|F(x) - y\|$. An important human decision is choosing a norm (or some other metric) for measuring the error. One often uses a norm arising from an inner product; the procedure is then called the *Method of Least Squares*.

Now, say you want to solve an equation that you believe *should* have an exact solution under suitable conditions. You thus need to determine and understand these conditions.

The simplest example is a system of linear algebraic equations $Ax = y$. A basic result in linear algebra uses the adjoint equation (duality) and says that for a given y there is at least one solution if and only if y is orthogonal to all the solutions z of the homogeneous adjoint equation, $A^*z = 0$. The same assertion also holds for linear elliptic boundary value problems—where this is known as the *Fredholm Alternative*. See **[10]** for a discussion along with a proof for the instructive ordinary differential equations case.

The next example gives the flavor of the issues for a simple nonlinear differential equation. Recall that the curvature $k(x)$ of a smooth curve $y = y(x)$ is given by

$$k(x) = \frac{y''}{\left(1 + y'^2\right)^{3/2}} = \left(\frac{y'}{\sqrt{1 + y'^2}}\right)'. \tag{14}$$

The inverse problem is, given a smooth function $k(x)$, $0 < x < 1$, to find a smooth curve $y = y(x)$ having this function as its curvature.

A circle of radius R has curvature $1/R$. Thus if $k(x) \equiv 2$, then a semi-circle of radius $1/2$ solves our problem. However, if $k(x) \equiv 4$, then the circle of radius $1/4$ supplies a solution for only half the desired interval $0 < x < 1$. This leads us to suspect that if there is a solution, then the curvature can't be too large for too much of the interval.

To find an obstruction integrate both sides of (14):

$$\int_0^x k(t)\,dt = \frac{y'(x)}{\sqrt{1 + y'(x)^2}} - \frac{y'(0)}{\sqrt{1 + y'(0)^2}}. \tag{15}$$

Let $\gamma = y'(0)/\sqrt{1 + y'(0)^2}$ so $|\gamma| \le 1$. Then (15) implies the obstruction

$$\int_0^x k(t)\,dt \le 1 - \gamma \le 2, \qquad 0 \le x \le 1.$$

This inequality embodies our suspicion that "the curvature can't be too large for too much of the interval." In the case of constant curvature $k(x) \equiv c > 0$, for $x = 1$ this condition is $c \le 2$, which is sharp. For non-constant k a necessary and sufficient condition is that there is a constant $\gamma \in [-1, 1]$ such that $\left|\int_0^x k(t)\,dt + \gamma\right| < 1$ for all $0 < x < 1$. If we assume the curve is convex, that is, $k(x) > 0$, then we may choose $\gamma = -1$ and find that a necessary and sufficient condition is simply $\int_0^1 k(t)\,dt \le 2$. The necessity is immediate from (15), while the sufficiency follows by solving (15) for $y'(x)$ and integrating. Implicitly we have not permitted vertical tangents ($y'(x) = \pm\infty$) inside the interval but do allow them at the boundary points—as in the case of a semicircle of radius $1/2$.

A standard version of this problem is to impose boundary conditions such as $y(0) = y(1) = 0$. I leave you the pleasure of discovering necessary and sufficient conditions for solving this boundary value problem in the special case of a convex curve. Assuming existence, is the solution of this boundary value problem unique?

The difficulties here are because this problem is *global* for the whole interval $0 < x < 1$. If we are satisfied with a *local* solution, defined only in some neighborhood of $x = 0$ then a solution always exists.

Our understanding of obstructions to the existence of a solution of a nonlinear differential equation is very incomplete; many of the known obstructions use Noether's theorem mentioned in Section 2.6d.

2.6 Exploit symmetry. a) *Simple symmetry* One familiar example of symmetry in algebra occurs for a polynomial $p(z) = a_n z^n + \cdots + a_0$ with real coefficients. Then the coefficients are invariant under complex conjugation, $\bar{a}_j = a_j$, so for any complex number z we have $\overline{p(z)} = \overline{\Sigma a_k z^k} = \Sigma a_k \bar{z}^k = p(\bar{z})$. Thus if z is a complex root, then so is $\bar{z}$. The nature of complex conjugation as a symmetry is clearer if one uses different notation for the complex conjugation operator: write $T(z) = \bar{z}$. Thus $T^2 =$ Identity and $(Tp)(z) = T(p(z))$. For a polynomial with real coefficients $\overline{p(z)} = p(\bar{z})$ means $Tp = pT$, that is, T and p commute; it may be clearer if we write this as $TpT^{-1} = p$, so p is fixed under the automorphism T. Galois' deep contribution to the theory of solving polynomial equations was to show how to exploit related symmetries.

A variant of this reasoning is also useful to solve the equation $F(x) = c$. Assume that F commutes with some map T, so $TF = FT$, and that c is invariant under T: $T(c) = c$. If x_0 is a solution of $F(x) = c$, then x_0 is not necessarily invariant, but $T(x_0)$ is also a solution. If you also know that the solution of $F(x) = c$ is *unique*, then $T(x_0) = x_0$, that is, this solution x_0 is invariant under T. Here are three similar instances.

i) Let f be a homeomorphism of the sphere $S^2 \in \mathbb{R}^3$, and let $\varphi:(x, y, z) \mapsto (x, y, -z)$ be a reflection across the equator. Assume that $c \in S^2$ is fixed by φ, that is, $\varphi(c) = c$ so c is on the equator $z = 0$, and assume that f and φ commute, $f \circ \varphi = \varphi \circ f$. If $f(p_0) = c$, then $p_0 = (x_0, y_0, z_0)$ is also invariant under φ and hence p_0 is also on the equator. Thus f maps the equator onto itself.

ii) For this example $u(x, t)$ is the solution of the wave equation $u_{xx} - u_{tt} = 0$ on the interval $-1 \le x \le 1$. Assume u satisfies the boundary conditions $u(-1, t) = u(1, t) = 0$. If the initial position $u(x, 0)$ and the initial velocity $u_t(x, 0)$ are both even functions, that is, invariant under the map T: $x \mapsto -x$, then so is the solution $u(x, t)$. This follows from the uniqueness of the solution of the wave equation with given initial conditions.

The uniqueness is easy to prove. If u and v are both solutions with the same initial position and velocity, then $w := u - v$ is also a solution of the wave equation, but with $w(x, 0) = w_t(x, 0) = 0$. Let $E(t) = \frac{1}{2}\int_{-1}^{1}(w_t^2 + w_x^2)\,dx$ be the "energy" at time t. Then by a computation involving an integration by parts, $dE/dt = 0$ so $E(t) = E(0)$ (*conservation of energy*). But $E(0) = 0$, so $E(t) \equiv 0$ for all t. Hence $w(x, t)$ is constant. Since $w(x, 0) = 0$, then $w(x, t) \equiv 0$ so $u(x, t) \equiv v(x, t)$.

Using the linearity of this problem, even without uniqueness we could have obtained an invariant solution. Let $\varphi(x, t)$ be any solution. Since $T^2 = I$, then the average $u := \frac{1}{2}(\varphi + T\varphi)$ is an invariant solution. One generalizes this construction by the important procedure of averaging over the group of symmetries. One application of this in electrostatics is the *method of images*.

iii) A Markov chain example. In an experiment you are placed in a five room "house"; see Figure 3. Every hour the doors are opened and you must move from your current room to one of the adjacent rooms. Assuming the rooms are all equally attractive, what percentage of the time will you spend in each room?

To solve this problem one introduces the 5×5 *transition matrix* $M = (m_{ij})$ of this *Markov chain*: if you are currently in room j, then m_{ij} is the probability you

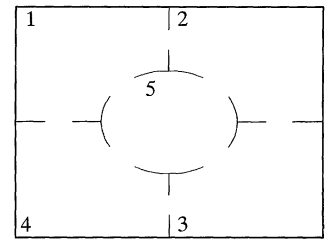

Figure 3

will next be in room i. One checks that

$$M = \begin{pmatrix} 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{4} \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}.$$

The elements of M are non-negative and the sum of every column is 1: no matter where you are now, at the next step you will certainly be in one of the rooms.

It is useful to introduce column *probability vectors* $P = (p_1, \ldots, p_5)$ with the property that p_j gives the probability of being in the jth room at a given time. Then $0 \le p_j \le 1$ and $\Sigma p_j = 1$. If P_{now} describes the probabilities of your current location, then $P_{\text{next}} = MP_{\text{now}}$, gives the probabilities of your location at the next time interval. If one begins in Room 1, then $P_0 = (1, 0, 0, 0, 0)$, and after the first hour $P_1 = (0, \frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3}) = MP_0$. In the same way, at the end of the second hour $P_2 := MP_1 = M^2P_0$, and $P_k := MP_{k-1} = M^kP_0$.

In seeking the long-term probabilities, we are asking if the probability vectors $P_k = M^kP_0$, $k = 1, 2, \ldots$ converge to some "equilibrium" vector P *independent* of the initial probability vector P_0. If so, then in particular $P = \lim M^{k+1}P_0 = \lim MM^kP_0 = MP$, that is, $P = MP$ so P is an eigenvector of M with eigenvalue 1.

Although $\lambda = 1$ is always an eigenvalue of M (since it is an eigenvalue of M^* with eigenvector $(1, \ldots, 1)$), the limit M^kP_0 does not always exist. For example, it does not exist for the transition matrix $M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ for a two room "house." If $M = I$, then the limit of M^kP_0 exists but is not independent of P_0. However the limit M^kP_0 does exist and is independent of the initial probability vector P_0 if all of the elements of M—or some power of M—are positive. The simplest proof I know for the convergence that does not assume M is diagonalizable is in [**2**, p. 257]. In our case all the elements of M^2 are positive. It remains to find this limiting probability distribution P by solving $P = MP$.

Here is where we can use symmetry. Since the four corner rooms are identical, M must commute with the matrices T_{ij} that interchange the probabilities of being in the corner rooms, p_i and p_j for $1 \le i, j \le 4$. Since $M(T_{ij}P) = T_{ij}MP = T_{ij}P$, we see that $T_{ij}P$ is also a probability eigenvector with eigenvalue $\lambda = 1$. Thus, by uniqueness of this probability eigenvector, $T_{ij}P = P$ so "by symmetry" P has the special form $P = (x, x, x, x, y)$ with $1 = \Sigma p_i = 4x + y$. The system of equations $P = MP$ now involves only two unknowns x, y. Its first equation is $x = \frac{1}{3}x + \frac{1}{3}x + \frac{1}{4}y$, that is $4x = 3y$. Combined with $4x + y = 1$ one finds $x = \frac{3}{16}$, $y = \frac{1}{4}$. Therefore 25% of the time is spent in the center room and 18.75% in each

of the corner rooms. Symmetry turned a potentially messy computation into a simple one.

Figure 1 gives added insight. To exploit symmetry one seeks changes of variable T so that the old problem $\mathcal{P}$ and new problem $\mathcal{Q}$ are *identical*: $\mathcal{P} = T^{-1}\mathcal{P}T$.

b) *Translation invariance* If there are families of symmetries, one can obtain more information. We first discuss this for a linear differential equation with constant coefficients, $Lu = au'' + bu' + cu$. Here L commutes with *all* the *translation operators* T_α defined by $(T_\alpha u)(x) := u(x + \alpha)$, so $T_\alpha L = LT_\alpha$ for all complex α. These translations T_α are a continuous group of symmetries: $T_\alpha T_\beta = T_{\alpha+\beta}$. The eigenfunctions of translations are just exponentials: $T_\alpha e^{cx} = \mu e^{cx}$, where $\mu = e^{c\alpha}$. We claim that these exponentials are also eigenfunctions of L. While this is simple to show directly, we prove more generally that this is true for any linear map L that commutes with all translations; some other instances are constant coefficient linear difference and linear partial differential equations.

Write $q(x; \lambda) := Le^{\lambda x}$. Since $T_\alpha e^{\lambda x} = e^{\lambda\alpha}e^{\lambda x}$, we have

$$T_\alpha Le^{\lambda x} = T_\alpha(q(x;\lambda)) = q(x+\alpha;\lambda) \quad \text{and} \quad LT_\alpha(e^{\lambda x}) = e^{\lambda\alpha}Le^{\lambda x} = e^{\lambda\alpha}q(x;\lambda).$$

Comparing these at $x = 0$, we see that if the linear map L commutes with translations, then $q(\alpha; \lambda) = q(0; \lambda)e^{\lambda\alpha}$ for any α,; equivalently, $q(x; \lambda) = q(0; \lambda)e^{\lambda x}$. Writing $Q(\lambda) := q(0; \lambda)$, we conclude

$$Le^{\lambda x} = Q(\lambda)e^{\lambda x}. \tag{16}$$

Thus $e^{\lambda x}$ is an eigenfunction of L for any λ, and the corresponding eigenvalue is $Q(\lambda)$.

Working formally, we apply (16) to find some solution of $Lu = f$. Write $f(x) = \Sigma f_\lambda e^{\lambda x}$ and seek a solution u as $u(x) = \Sigma u_\lambda e^{\lambda x}$. Then by (16) $Lu = \Sigma u_\lambda Q(\lambda)e^{\lambda x}$. To solve the homogeneous equation $Lu = 0$ pick λ to be any root of $Q(\lambda)$, while for the inhomogeneous equation $Lu = f$ match coefficients to conclude that $u_\lambda = f_\lambda/Q(\lambda)$. Thus a solution is $u(x) = \Sigma[f_\lambda/Q(\lambda)]e^{\lambda x}$. One recognizes these formulas as the standard Fourier series/integrals and Laplace transform methods. This is why Fourier series and Fourier and Laplace transforms are so useful for constant coefficient differential equations. The value of $Q(\lambda)$ is determined separately for each problem. Since $Q(\lambda)$ appears in the denominator of the solution, its zeros play an important role. The point is that just by using translation invariance we know how to proceed.

As a quick application, return to the special case $Lu = au'' + bu' + cu$, where a, b, and c are constants. Then $Le^{\lambda x} = (a\lambda^2 + b\lambda + c)e^{\lambda x}$, so $Q(\lambda) = a\lambda^2 + b\lambda + c$. In particular, if $Q(r) = 0$, then $u = e^{rx}$ solves the homogeneous equation $Lu = 0$, while if $Q(r) \neq 0$, then $u(x) = e^{rx}/Q(r)$ is a particular solution of the inhomogeneous equation $Lu = e^{rx}$; if $Q(r) = 0$ but $Q'(r) \neq 0$, then one can take the derivative of (16) with respect to λ and evaluate at $\lambda = r$ to solve $Lu = e^{rx}$. Similarly, if r is a double root of $Q(\lambda) = 0$ then also $Q'(r) = 0$; here taking the derivative of equation (16) with respect to λ and evaluating at $\lambda = r$ reveals that $u(x) = xe^{\lambda x}$ is also a solution of the homogeneous equation, a fact that often is bewildering in elementary courses in differential equations.

There is an interesting cultural difference between the way mathematicians and physicists usually write the general solution of $u'' + u = 0$. Mathematicians write $u(x) = A\cos x + B\sin x$, which emphasizes the linearity of the space of solutions, while physicists write $u(x) = C\cos(x + \alpha)$, which emphasizes the translation invariance.

As an exercise, apply translation invariance to develop the theory of second order linear difference equations with constant coefficients, $au_{n+2} + bu_{n+1} + cu_n = f(n)$. The Fibonacci sequence $u_{n+2} = u_{n+1} + u_n$, with initial conditions $u_0 = 0$, $u_1 = 1$, is a special case.

Invariance under multiplication $x \mapsto cx$ is related closely to translation invariance: if we let $x = e^z$, then translating z multiplies x by a constant. With this hint, one can treat the Euler differential operator $Lu = \alpha x^2 u'' + \beta x u' + \gamma u$, which commutes with the stretching operator $x \mapsto cx$. Here the analog of the Fourier transform is called the Mellin transform.

The Laplace operator in Euclidean space is invariant under translations and orthogonal transformations; on a Riemannian manifold the Laplacian is invariant under all isometries. The wave equation is invariant under Lorentz transformations (see the end of this Section). The basic point is that invariance under some group automatically implies fundamental formulas.

c) *More complicated group invariance* In more complicated problems, there may be some symmetry, but it may not be obvious to find or use. Sophus Lie created the theory of what we now call *Lie groups* to exploit symmetries to solve differential equations. His vision was to generalize Galois theory to differential equations. The resulting theory has been extraordinarily significant throughout mathematics. As our first example, observe that the differential equation

$$\frac{dy}{dx} = \frac{ax^2 + by^2}{cx^2 + dy^2} \qquad a, b, c, d \text{ constants},$$

is invariant if one makes the change of variable (a stretching) $x \mapsto \lambda x$, $y \mapsto \lambda y$ for any value of $\lambda > 0$. In other words, if $y = \varphi(x)$ is a solution, then so is $\lambda y = \varphi(\lambda x)$, that is $y = \varphi(\lambda x)/\lambda$. This motivates us to introduce a new variable that is invariant under this stretching: $w = y/x$. Then w satisfies $xw' = (a + bw^2)/(c + dw^2) - w$, which can be solved by separation of variables. The equation $dy/dx = (ax + by + p)/(cx + dy + q)$ has the symmetry of stretching from the point of intersection of the lines $ax + by + p = 0$ and $cx + dy + q = 0$. Lie showed that many complicated formulas one has for solving differential equations are but special instances of invariance under a group of symmetries. His work showed that a daunting bag of tricks that demoralize undergraduates were all instances of exploiting symmetries. The next example is not as simple, so we'll be a bit more systematic.

Nonlinear equations of the form $\Delta u = f(x, u)$ arise frequently in applications. For instance the special cases where $f(x, u)$ has the forms $|x|^a u^b$ and $|x|^c e^u$ arise in astrophysics (Emden-Fowler equation), complex analysis, and conformal Riemannian geometry. We briefly discuss

$$\Delta u = |x|^c e^u \tag{17}$$

in $\mathbb{R}^n$ from the view of symmetry. While there are systematic approaches to seek symmetry, in practice one usually tries to guess; the method is of no help if finding symmetries is as difficult as solving the original problem.

The right side of (17) suggests we seek a symmetry group in the form $G{:}(x, u) \mapsto (\alpha x, u + \lambda)$, that is, we try the change of variables $\tilde{x} = \alpha x$, $\tilde{u} = u + \lambda$, where $\alpha > 0$, λ are constants. Let $\tilde{\Delta} = \partial^2/\partial \tilde{x}_1^2 + \cdots = \alpha^{-2}\Delta$ be the Laplacian in these new variables. Then $\tilde{u}(\tilde{x})$ is a solution of $\tilde{\Delta}\tilde{u} = [\alpha^{c+2} e^{\lambda}]^{-1} |\tilde{x}|^c e^{\tilde{u}}$. Thus if we pick $\alpha^{c+2} e^{\lambda} = 1$, so $\lambda = -(c + 2)\ln \alpha$, then $\tilde{u}(\tilde{x})$ is a solution of (17) for any

value of α. In other words, if $u = \varphi(x)$ is a solution then so is $u(x) - (c+2)\ln\alpha = \varphi(\alpha x)$ that is, $u(x) = \varphi(\alpha x) + (c+2)\ln\alpha$ for any $\alpha > 0$. The symmetry group is $G_\alpha\colon (x, u) \mapsto (\alpha x, u - (c+2)\ln\alpha)$. This is the identity map at $\alpha = 1$.

To go further, recall that the Laplacian is invariant under the orthogonal group: if $u(x)$ is a solution, so is $u(Rx)$ for any orthogonal transformation R. It thus is reasonable to seek special solutions $u = u(r)$, where $r = |x|$, that are also invariant under the orthogonal group. Writing the Laplacian in spherical coordinates leads us to consider

$$u'' + \frac{n-1}{r}u' = r^c e^u,$$

where $u' = du/dr$. We know this equation is invariant under the change of variables

$$\tilde{r} = \alpha r, \qquad \tilde{u} = u - (c+2)\ln\alpha. \tag{18}$$

For fixed r and u, as we vary α, (18) defines a curve in the $\tilde{r}, \tilde{u}$ plane. It is natural to define new coordinates in which these curves are straight lines, say parallel to the vertical axis. We want one function $s = s(\tilde{r}(r, u, \alpha), \tilde{u}(r, u, \alpha)) = s(\alpha r, u - (c+2)\ln\alpha)$ that is constant on each of these curves; this function is used to select which of these curves one is on. The other function $v = v(\tilde{r}(r, u, \alpha), \tilde{u}(r, u, \alpha)) = v(\alpha r, u - (c+2)\ln\alpha)$ is used as a normalized parameter along these curves, chosen so that the directional derivative of v along these curves is one; see Figure 4. Thus, the conditions are

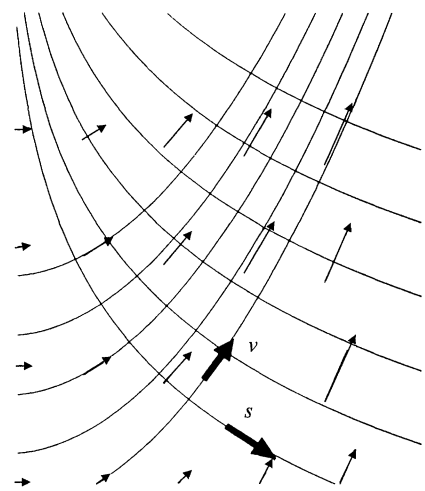

Figure 4

$$\left.\frac{\partial s}{\partial \alpha}\right|_{\alpha=1} = 0 \quad \text{and} \quad \left.\frac{\partial v}{\partial \alpha}\right|_{\alpha=1} = 1. \tag{19}$$

By the chain rule these can be rewritten as

$$rs_r - (c+2)s_u = 0 \quad \text{and} \quad rv_r - (c+2)v_u = 1, \tag{20}$$

where s_r etc. are the partial derivatives. Using the tangent vector field V to our curves,

$$V := \left.\frac{\partial \tilde{r}}{\partial \alpha}\right|_{\alpha=1}\frac{\partial}{\partial r} + \left.\frac{\partial \tilde{u}}{\partial \alpha}\right|_{\alpha=1}\frac{\partial}{\partial u} = r\frac{\partial}{\partial r} - (c+2)\frac{\partial}{\partial u},$$

we can rewrite (20) as

$$Vs = 0 \qquad \text{and} \qquad Vv = 1;$$

V is called the *infinitesimal generator* of the symmetry. In these new coordinates, by integrating (19) the invariance (18) becomes

$$\tilde{s} = s \qquad \text{and} \qquad \tilde{v} = v + \alpha. \tag{21}$$

An obvious particular solution of the second equation in (20) is $v = \ln r$; an equally obvious solution is $v = -u/(c+2)$, which would also work.

The first equation in (20) is straightforward to solve. The standard approach to solve $a(x,y)\psi_x + b(x,y)\psi_y = 0$ for $\psi(x,y)$ is to solve the ordinary differential equation $dy/dx = b/a$ and write its solution in the form $\psi(x,y) = C$, where C is the constant of integration. This $\psi(x,y)$ is a solution of the partial differential equation, as is any function of it. In our application the solution of $du/dr = -(c+2)/r$ is $u = -(c+2)\ln r + C$ so $s := \psi(r,u) = u + (c+2)\ln r$. [Alternative approach to obtain $s(r,u)$: eliminate α from the formulas (18) and find that $\tilde{u} + (c+2)\ln \tilde{r} = u + (c+2)\ln r$. Thus the function $s = (c+2)\ln r + u$ is constant along each of these curves]. Since any function of s has the same property one can use this flexibility to choose a "simple" s. In these new coordinates, $s = u + (c+2)\ln r$, $v = \ln r$. After a computation that is not painless one finds that $v(s)$ satisfies

$$\ddot{v} = (n-2)[1 - (c+2)\dot{v}]\dot{v}^2 - e^s\dot{v}^3,$$

where $\dot{v} = dv/ds$ and $\ddot{v} = d^2v/ds^2$. Since this does not involve v itself, the substitution $w = \dot{v}$ gives a *first* order equation for $w(s)$, which simplifies significantly if $n = 2$, exactly the case of interest in applications.

It is a useful exercise to repeat this analysis for $\Delta u = |x|^a u^b$ in $\mathbb{R}^n$ and notice that the resulting equation simplifies dramatically when $(a+2)/(b-1) = (n-2)/4$, again exactly the situation of applications to physics and geometry. By using symmetry one can solve some problems that are otherwise impenetrable.

One impressive application of symmetry was G. I. Taylor's computation of the energy in the first atomic explosion just by exploiting symmetry and taking measurements from publicly available photographs; see [**3**, Chapter 1] for an exposition. The monographs [**3**] and [**14**] show how to apply and exploit symmetry for ordinary and partial differential equations.

d) *Noether's Theorem* Most "natural" differential equations arise as Euler-Lagrange equations in the calculus of variations. Indeed, many believe one should always formulate fundamental equations using variational principles. E. Noether's theorem shows how symmetry invariance of a variational problem implies basic identities, including conservation laws. While shorter direct proofs of these conservation laws might be found after one knows what to prove, there is a view that the symmetry is considerably deeper and more basic. Moreover, symmetry gives a way of finding new conservation laws. See [**6**], [**7**], [**3**], and [**14**].

e) *Using symmetry for Pell's equation* Here is another way to use symmetry. We want all the integer solutions of

$$x^2 - 2y^2 = 1. \tag{22}$$

By experimentation you quickly find the solution $x = 3$, $y = 2$. Are there any others? Can you find *all* the solutions? They are the integer lattice points on the hyperbola (22).

Writing $X := (x, y)$ and $Q(X) := x^2 - 2y^2$, seek a symmetry of the hyperbola $Q(X) = 1$ as a linear change of variables R: $(x, y) \mapsto (ax + by, cx + dy)$ defined by the matrix $R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We want R to have the property $Q(RX) = Q(X)$; in more formal language, we want the group of automorphisms R of the quadratic form Q. If we can find R, and if we have one solution $X_1 = (x_1, y_1)$ of $Q(X) = 1$, then $X_2 := RX_1 = (ax_1 + by_1, cx_1 + dy_1)$ is another solution since $Q(X_2) = Q(RX_1) = Q(X_1) = 1$. Thus, knowing R enables us to construct new solutions from old ones.

These automorphisms R embody the symmetries of the polynomial $Q(X)$ much as the rotations T (orthogonal transformations) embody the symmetries of the more familiar polynomial $P(X) := x^2 + y^2$ since $P(TX) = P(X)$. If X_1 is a point on a circle centered at the origin, then $X_2 := TX_1$ is another point on the same circle.

For our quadratic polynomial the obvious symmetries are $x \to \pm x$ and $y \to \pm y$. We want more. Since

$$\begin{aligned} Q(RX) &= (ax + by)^2 - 2(cx + dy)^2 \\ &= (a^2 - 2c^2)x^2 + 2(ab - 2cd)xy + (b^2 - 2d^2)y^2, \end{aligned}$$

the condition $Q(RX) = Q(X)$ means $a^2 - 2c^2 = 1$, $ab - 2cd = 0$, and $(b^2 - 2d^2) = -2$. If we pick a and c to satisfy the first of these, which is just the original equation (22), then the other two conditions imply $d = \pm a$ and $b = \pm 2c$. This yields all the symmetries R of our quadratic polynomial.

For our purposes it is enough to use the solution $(3, 2)$ we found of (22) so $a = 3$, $c = 2$, $b = 4$, $d = 3$, and $R = \begin{pmatrix} 3 & 4 \\ 2 & 3 \end{pmatrix}$. We began with the solution $X_1 := (x_1, y_1) = (3, 2)$. Using this we find the solutions $X_2 = RX_1 = (17, 12)$, $X_3 = RX_2 = (99, 70)$, etc. of (22). Since $\det R = 1$ and the elements of R are integers, both the symmetry R and its inverse R^{-1} take integer lattice points to integer lattice points.

The mapping R has two basic geometric properties. To describe them take two points $V_1 := (x_1, y_1)$ and $V_2 := (x_2, y_2)$, both on the right $(x > 0)$ branch of the hyperbola $x^2 - 2y^2 = 1$. We call this right branch Γ, and say that V_1 is *below* V_2 (and write $V_1 \prec V_2$) if $y_1 < y_2$. The geometric properties are:

- *R preserves the branch*: if a point V is on Γ, then so is RV.
- *R preserves the order on* Γ: If $V_1 \prec V_2$ then $RV_1 \prec RV_2$.

Note that R^{-1} also has these properties. Since R is a continuous map from the hyperbola to itself, by connectedness, it maps the right branch, Γ, either to itself or to the left branch. Checking the image of one point, say $(1, 0)$ we see that the image is in Γ. Moreover, since R is invertible as a map of the whole plane, its restriction to Γ is invertible. Therefore it is either monotonic increasing or decreasing as a function of the y coordinate on Γ. Again checking the image of $(1, 0)$, we conclude that the restriction of R to Γ is an increasing function of the y coordinate. This implies that R preserves the order on Γ.

Our particular solution $X_1 := (3, 2)$ is the positive integral solution with the *smallest* possible positive value for y_1. Writing $X_0 = (1, 0)$, this means $X_0 \prec X_1$ and there is no other integral solution between X_0 and X_1. Since $Q(RX_1) = Q(X_1) = 1$ we see that $X_2 := RX_1 = (17, 12)$ is also a solution of (22). Similarly $X_k := (x_k, y_k) = RX_{k-1} = R^k X_0$ are all positive integer solutions for any positive

integer k. These solutions are distinct since their y coordinates are increasing, so $X_k \prec X_{k+1}$.

Moreover, these are *all* the positive integral solutions. If there were another, Z, then for some k we have $X_k \prec Z \prec X_{k+1}$. Therefore $R^{-1}Z$ is yet another solution and because R preserves the order of the points on the hyperbola,

$$X_{k-1} = R^{-1}X_k \prec R^{-1}Z \prec R^{-1}X_{k+1} = X_k.$$

Continuing, we obtain a solution $R^{-k}Z$ between X_0 and X_1 since

$$X_0 = R^{-k}X_k \prec R^{-k}Z \prec R^{-k}X_{k+1} = X_1.$$

This contradicts the fact that $X_1 = (x_1, y_1) = (3, 2)$ was the positive solution whose second coordinate was as small as possible. We conclude that $X_k = R^k X_0$, that is, the *orbit* of X_0 after repeated action by R, are *all* of the integer solutions.

The matrix R^k can be computed explicitly by first diagonalizing it. This gives $R^k = S\Lambda^k S^{-1}$, where Λ is the diagonal matrix of eigenvalues $3 \pm 2\sqrt{2}$ of R and S is the matrix whose columns are the corresponding eigenvectors $(\pm\sqrt{2}, 1)$; these vectors also determine the asymptotes of the hyperbola. Thus $X_k = R^k X_0$ has the explicit formula

$$X_k = \left(\frac{(3+2\sqrt{2})^k + (3-2\sqrt{2})^k}{2}, \frac{(3+2\sqrt{2})^k - (3-2\sqrt{2})^k}{2\sqrt{2}} \right), \tag{23}$$

which shows that such formulas may be more complicated—and possibly less desirable—than you might expect. Perhaps of greater value, this formula leads us to define R^t, $-\infty < t < \infty$, by the rule $R^t = S\Lambda^t S^{-1}$, so $R^{s+t} = R^s R^t$.

Using a similar computation, it is straightforward to find all linear changes of variable $x' = \alpha x + \beta t$, $t' = \gamma x + \delta t$ that preserve the wave operator $\partial^2/\partial t^2 - c^2\partial^2/\partial x^2$, where c is a constant (the speed of sound or light). By the chain rule,

$$u_{tt} - c^2 u_{xx} = (\delta^2 - c^2\gamma^2)u_{t't'} + 2(\beta\delta - c^2\alpha\gamma)u_{x't'} + (\beta^2 - c^2\alpha^2)u_{x'x'}.$$

Thus we want $\delta^2 - c^2\gamma^2 = 1$, $\beta\delta - c^2\alpha\gamma = 0$, and $\beta^2 - c^2\alpha^2 = -c^2$. First pick γ and δ so that $\delta^2 - c^2\gamma^2 = 1$, and then let $\beta = \neq c^2\gamma$, $\alpha = \pm\delta$. To preserve orientation we use the $+$ signs. Since $c^2\alpha^2 - \beta^2 = c^2$ and $\cosh^2\sigma - \sinh^2\sigma = 1$, it is traditional to write $\alpha = \cosh\sigma$, $\beta = c\sinh\sigma$. For any real σ a transformation that preserves the wave operator is

$$\begin{aligned} x' &= (\cosh\sigma)x + (c\sinh\sigma)t \\ t' &= \left(\frac{1}{c}\sinh\sigma\right)x + (\cosh\sigma)t. \end{aligned} \tag{24}$$

This is called a *Lorentz transformation.* Lorentz transformations also preserve arc length $ds^2 := dx'^2 - c^2\,dt'^2 = dx^2 - c^2\,dt^2$ in space-time and are fundamental in the study of the wave operator and special relativity.

In special relativity it is enlightening to replace the parameter σ in (24) by one that is physically more meaningful. If the x axis moves with constant velocity V relative to the x' axis, for an observer on the x' axis, $x'/t' = V$ is the constant velocity of the origin $x = 0$ of the x axis. But from (24) with $x = 0$

$$V = \frac{x'}{t'} = c\tanh\sigma,$$

so $\sinh\sigma = (V/c)/\sqrt{1-(V/c)^2}$ and $\cosh\sigma = 1/\sqrt{1-(V/c)^2}$. We can use

this to rewrite the Lorentz transformation (24) in terms of the velocity V as

$$x' = \frac{x + Vt}{\sqrt{1 - (V/c)^2}} \qquad t' = \frac{(V/c^2)x + t}{\sqrt{1 - (V/c)^2}}.$$

3. SOME PROCEDURES TO PROVE EXISTENCE. Existence of a solution of an equation may be approached in different ways. One should first try to find a "simple" expression for the solution, perhaps using some of the procedures already discussed. The following discussion assumes this has been used as much as possible.

There are two types of existence procedures: those that construct a specific solution, and those that merely prove a solution exists. As examples, I present two purely existential approaches to proving the existence of a solution. In addition to these I assume the reader has seen the standard constructive method of successive approximations involving fixed points of contracting mappings. It is presented in many texts and should be tried before these. Recall Hermann Weyl's: "Whenever you can settle a question by explicit construction, be not satisfied with purely existential arguments." In the light of this dictum it is useful to reflect on the constructive and non-constructive approaches discussed in Section 2.1 for solving $ax \equiv b \pmod{m}$.

3.1 Variational methods. An example illustrates the issues vividly. Say we want to solve the system of equations

$$\begin{aligned} x^3 + 2xy - 3y\cos x e^{\sin x} &= -7 \\ y^5 + x^2 - 3e^{\sin x} &= 5 \end{aligned}$$

Is there a solution? Without further insight it may not be obvious. But these two equations state that the gradient of the function

$$u(x, y) := \tfrac{1}{4}x^4 + \tfrac{1}{6}y^6 + x^2y - 3ye^{\sin x} + 7x - 5y$$

is zero. Thus, the solutions of our equation correspond to the critical points of $u(x,y)$. It is obvious that as one goes far from the origin then u becomes large. Thus, there is some point (x_0, y_0) where u takes on its minimum value. This minimum gives us one solution of our equations. To determine if there are others would require a more detailed investigation.

This approach is a useful technique for proving that certain differential equations always have at least one solution. The method is called the "direct method in the calculus of variations." In Section 2.2 we saw that the solutions of the Laplace Equation $\Delta u = 0$ in a region Ω with $u = f$ on the boundary of Ω are critical points of the functional

$$J(u) = \tfrac{1}{2}\iint_\Omega \left(u_x^2 + u_y^2\right) dx\,dy. \tag{25}$$

Following the example at the beginning of this section, to find a solution of $\Delta u = 0$, we instead can seek a minimum u of J among all functions with $u = f$ on the boundary. Since the functional J is non-negative, this leads one to assert that it attains its minimum at some function u, and proves the existence of a solution of $\Delta u = 0$ with the prescribed boundary values. This assertion is called *Dirichlet's Principle*.

After Riemann dramatically applied this reasoning in his work on complex analysis, Weierstrass pointed out this "principle" is not obvious since for some similar problems the corresponding functional J only has an infimum and does not attain a minimum value in the class of admissible functions. Nonetheless, everyone

—including Weierstrass—believed that Riemann's results were essentially correct. This issue was a catalyst in the clarification of the foundations of analysis done at the end of the nineteenth century. The gap remained until Hilbert's work in 1901 and 1909. In this context, it is interesting to note Nietzsche's remark: "Great men's errors are to venerated as more fruitful than little men's truths".

The calculus of variations is a powerful method to prove that certain equations do have solutions.

3.2 Fixed point methods. Another example. Say you want to solve the system of equations

$$3x - 5y = \frac{2x + ye^{2-\sin xy}}{7 + x^2 + y^4} - 13$$

$$2x + 71y = 9 - \cos(xy + 19e^{x-5y})$$

Is there at least one solution? Again, to most people this is not immediately obvious. You look at the equations... The equations look at you.

Eventually you may be led to write this in the form $LX = F(X)$, where $X = (x, y)$, L is the 2×2 matrix on the left side, and $F(X)$ is the nonlinear right side. The key observation is that the vector function $F(X)$ is bounded independent of X. In fact $\|F(X)\| \leq 100$ (the size of the bound is unimportant for us). Moreover, the matrix L is invertible, so we can rewrite our equations in the symbolic form $X = T(X)$ where $T(X) = L^{-1}F(X)$. If we view $T(X)$ as a map from the plane $\mathbb{R}^2$ to itself, then the equation $X = T(X)$ means that the solution X we seek is a *fixed point* of the map T. Since $\|F(X)\| \leq 100$, we know that $\|T(X)\| \leq R$ for some constant R that is independent of X (we can let $R = 10{,}000$, but that is irrelevant for our immediate concerns). Thus we have found the *a priori* inequality: if a solution of our equation exists, it must lie in the closed disk $B = \{\|X\| \leq R\}$. Since T maps any point X into B, in particular it maps B into B.

Now we can invoke the Brouwer fixed point theorem, a result customarily proved in topology courses (see [**5**, p. 75] for a slick proof using Stokes' theorem). It asserts that any continuous map of a closed disk to itself must have at least one fixed point. This fixed point is the solution we seek.

The Schauder fixed point theorem generalizes the Brouwer theorem to infinite dimensional spaces. This generalization requires an additional compactness assumption. If B is a Banach space and $S \subset B$, then a continuous map $T : S \mapsto B$ is *compact* if for any bounded set $Q \subset S$ the closed set $\overline{f(Q)}$ is compact. For example, consider the Banach spaces $C(S^1)$ and $C^1(S^1)$ of 2π-periodic continuous functions and periodic continuously differentiable functions on the circle S^1 with the usual norms

$$\|u\|_{C(S^1)} = \max_{0 \leq x \leq 2\pi} |u(x)| \quad \text{and} \quad \|u\|_{C^1(S^1)} = \max_{0 \leq x \leq 2\pi} |u(x)| + \max_{0 \leq x \leq 2\pi} |u'(x)|.$$

We should probably write C_{periodic} to emphasize the periodicity. The Arzelá-Ascoli theorem implies that the identity map $id : C^1(S^1) \hookrightarrow C(S^1)$ is compact. The Schauder fixed point theorem says that if $S \subset B$ is a closed, convex, bounded set and if $T : S \to S$ is a compact map, then T has a fixed point [**13**, p. 32]. Schauder devised it specifically for partial differential operators. As an application we prove the existence of at least one periodic solution $u(x)$ with period 2π of

$$u' + u = F(x, u),$$

assuming only that $F(x, s)$ is a smooth function, periodic with period 2π in x and uniformly bounded, $|F(x, s)| \leq k$, where the constant k is independent of x and s.

A key observation is that the linear equation $Lu = u' + u = f(x)$ has a unique 2π periodic solution for any smooth periodic function $f(x)$. A direct computation gives

$$u(x) = \frac{1}{e^{2\pi} - 1} \int_0^{2\pi} e^{t-x} f(t)\, dt + \int_0^x e^{t-x} f(t)\, dt$$

(solve for $u(x)$ as usual and then pick $u(0)$ to force the periodicity: $u(2\pi) = u(0)$). This formula also yields the inequality $\|u\|_{C(S^1)} \leq \|f\|_{C(S^1)} = \|Lu\|_{C(S^1)}$. However $|u'| = |Lu - u| \leq |Lu| + |u|$ so we obtain the estimate

$$\|u\|_{C^1(S^1)} \leq 3\|Lu\|_{C(S^1)}. \tag{26}$$

This asserts that $L^{-1} : C(S^1) \to C^1(S^1)$ is a continuous map. Rewrite our problem as $u = L^{-1}F(x, u)$. Thus we seek a fixed point of the map $T(u) := L^{-1}F(x, u)$. Since we defined T as the composition

$$C(S^1) \xrightarrow{F} C(S^1) \xrightarrow{L^{-1}} C^1(S^1) \overset{id}{\hookrightarrow} C(S^1),$$

it is a compact map. Because $F(x, s)$ is bounded, then for some constant K,

$$\|T(u)\|_{C(S^1)} \leq K \quad \text{for all} \quad u \in C(S^1).$$

This proves *a priori* that any solution u of our problem must satisfy $\|u\|_{C(S^1)} = \|T(u)\|_{C(S^1)} \leq K$. Thus let S be the ball

$$S := \{u \in C(S^1) : \|u\|_{C(S^1)} \leq K\}.$$

The Schauder theorem shows there is at least one periodic solution $u \in S$ and $u \in C^1(S^1)$. Using a bootstrap argument, if $F(x, s)$ is smooth, then so is this solution u.

There is a similar result for $Lu := -\Delta u + cu = F(x, u)$ with various boundary conditions, assuming L is invertible and F is bounded. However one must use more complicated function spaces, such as Sobolev spaces, to prove an analogue of the fundamental inequality (26).

4. AN OPEN QUESTION. One is not surprised to see a seemingly elementary unsolved problem in number theory. It is less well-known that there are many interesting and simple-looking nonlinear partial differential equations about which little is known. Let $f(x, y)$ be a smooth function. Is there always at least one solution $u(x, y)$ of the Monge-Ampère equation

$$u_{xx}u_{yy} - u_{xy}^2 = f(x, y)? \tag{27}$$

This is a modest question. We seek some solution in a possibly small neighborhood of the origin; no additional conditions such as initial or boundary conditions are imposed. Yet we still do not know the answer. Many cases have been treated. If $f(x, y)$ has a power series expansion, we can invoke the Cauchy-Kowalewskaya theorem to get a power series solution. If $f(0, 0) > 0$, we can use the theory of elliptic partial differential equations to prove that a solution exists, while if $f(0, 0) < 0$ we appeal to the theory of hyperbolic equations. The difficult case is when $f(0, 0) = 0$. This case has also been treated if either $f(x, y) \geq 0$ near the origin, or if $\nabla f(0, 0) \neq 0$, [**11**], [**12**]. Nothing more is known. Perhaps there are smooth functions with $f(0, 0) = 0$ for which no solutions exist.

A similar differential equation arises in geometry. Locally, an abstract two dimensional surface with a Riemannian metric is a neighborhood of the origin in

the u, v plane where one specifies the element of arc length

$$ds^2 = E(u,v)\,du^2 + 2F(u,v)\,du\,dv + G(u,v)\,dv^2 \tag{28}$$

of curves in that neighborhood. You always get an arc length of this form if you consider the curves $u(t), v(t)$ on a two-dimensional surface with local coordinates u, v in $\mathbb{R}^n$. Does this give all possible abstract Riemannian metrics for the special case of surfaces in $\mathbb{R}^3$? In other words, given any arc length ds^2 of the form (28), locally can one always find a surface $x = x(u,v), y = y(u,v), z = z(u,v)$ in $\mathbb{R}^3$ having this as its arc length? More briefly, can every abstract two-dimensional Riemannian manifold be locally isometrically embedded in $\mathbb{R}^3$? One can show that there is a surface in $\mathbb{R}^4$ having this arc length, but the more interesting $\mathbb{R}^3$ case is still open. In one approach, the partial differential equation to be solved is essentially (27). Here the Gauss curvature $K(x, y)$ plays the role of the function $f(x, y)$, so we know there is a local embedding if $K(0,0) \neq 0$. The difficult case remaining is when $K(0,0) = 0$.

Problems such as this are challenges for the future.

REFERENCES

1. V. I. Arnol'd, *The Theory of Singularities and its Applications*, Lezioni Fermiane, Accademia Nazionale del Lincei Scuola Normale Superiore, Pisa, published by the Press of the University Cambridge, England 1991.
2. R. Bellman, *Introduction to Matrix Analysis*, McGraw-Hill, 1960, reprinted (1995) by SIAM in their series "Classics in Applied Mathematics."
3. G. W. Bluman, and S. Kumei, *Symmetries and Differential Equations*, Springer-Verlag, 1989, New York. Applied Mathematical Sciences **81**.
4. H. Davenport, *The Higher Arithmetic*, Harper and Row, New York, 1952 (6th edition published by Cambridge Univ. Press, 1992).
5. M. P. do Carmo, *Differential forms and Applications*, Universitext, Springer-Verlag, New York, 1994.
6. I. M. Gelfand, and S. V. Fomin, *Calculus of Variations*, Prentice Hall, 1963 (translated from the Russian).
7. M. Giaquinta, and S. Hildebrandt, *Calculus of Variations*, Grundlehren Vols. 310–311, Springer-Verlag, Berlin, 1996.
8. M. Golubitsky, and D. G. Schaeffer, *Singularities and groups in bifurcation theory*, Applied mathematical sciences; 51, 69, Springer-Verlag, 1985, New York.
9. S. Hildebrandt, and A. Tromba, *The Parsimonious Universe*, Springer-Verlag, 1996, New York (an earlier version: *Mathematics and Optimal Form* was published in the Scientific American Library, 1985).
10. J. L. Kazdan, `http://www.math.upenn.edu/~kazdan/solving.html`.
11. C-S Lin, "The local isometric embedding in $\mathbf{R}^3$ of 2-dimensional Riemannian manifolds with nonnegative curvature," *J. Diff. Geom.* **21** (1985), 213–230.
12. C-S Lin, "The local isometric embedding in $\mathbf{R}^3$ of two-dimensional Riemannian manifolds with Gaussian curvature changing sign cleanly," *Comm. Pure Appl. Math.*, **39** (1986), 867–887.
13. L. Nirenberg, *Topics in nonlinear functional analysis*, New York University Lecture Notes, 1974.
14. P. J. Olver, *Applications of Lie groups to differential equations*, 2nd ed., in the series Graduate texts in mathematics, **107**, Springer-Verlag, 1993, New York.
15. R. J. Walker, *Algebraic Curves*, Princeton Univ. Press, 1950, reprinted by Dover, 1962.
16. C. R. Wilie, and L. C. Barrett, *Advanced Engineering Mathematics*, 6th ed, McGraw-Hill, New York, 1995.

JERRY L. KAZDAN was an undergraduate at Rensselaer Polytechnic Institute and a graduate student at the Courant Institute of Mathematical Sciences (New York University). His main interests are partial differential equations and differential geometry. Since 1966 he has been at the University of Pennsylvania, and has taught at Harvard, Berkeley, UC Davis, and Tokyo as well as an invited guest at many universities throughout the world.

University of Pennsylvania, Philadelphia, PA 19104-6395

kazdan@math.upenn.edu