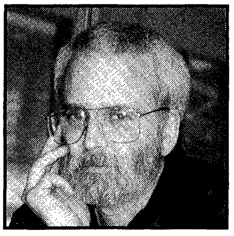
*Can We See the Mandelbrot Set?**

John Ewing



John Ewing is a professor of mathematics at Indiana University, where he currently serves as chair of the department. Editor of the *American Mathematical Monthly*, he also served as its associate editor from 1981–1991. From 1980 to 1986 he was editor-in-chief of the *Mathematical Intelligencer,* and he now serves on editorial boards for several Springer-Verlag book series. His research has been in algebraic topology and related areas, including work in number theory and topological dynamics. In 1976 he shared the Lester R. Ford Award for expository writing, and in 1991 he was named the first George Pólya Lecturer.

Almost every reader will have seen beautiful pictures of the Mandelbrot set, pictures that become more and more intricate as we zoom in at finer and finer scales. They make terrific tee shirts and lovely posters, and they have captured the imagination of mathematicians and non-mathematicians alike. While we admire the beauty of such pictures, we ought to be equally impressed by the mathematical feat of producing them. If the entire Mandelbrot set were placed on an ordinary sheet of paper, the tiny sections of boundary we examine would not fill the width of a hydrogen atom. Physicists *think about* such tiny objects; only mathematicians have microscopes fine enough to actually observe them.

But are these really pictures of the Mandelbrot set? When we use the computer to study a small intricate pattern along the boundary, are we viewing the Mandelbrot set with a powerful microscope, or are we seeing merely a distorted shadow of something real? Can we be *sure* that those pictures are accurate? In fact, how much do we *know* about the Mandelbrot set?

Drawing Pictures

The Mandelbrot set is a complicated object with a simple definition. The definition is easy to give using complex numbers, $c = a + bi$, and we use nothing more sophisticated than the fact that complex numbers can be identified with points in the plane.

For each complex number c we can define the "c-process." Start with 0, square and add c, square *that* and add c, and keep on repeating:

$$0 \to c \to c^2 + c \to (c^2 + c)^2 + c \to \cdots .$$

It is instructive to try this out for a few values of c, and readers with access to a computer algebra program or a calculator that works with complex numbers can experiment. For example, when $c = 0.1 + 0.2i$ the process soon reaches $0.070 + 0.223i$ and gets stuck; that is, there is a fixed point. When $c = -0.9 + 0.1i$ the process settles down quickly to oscillate between just two numbers, $-0.907 + 0.122i$ and $-0.904 - 0.123i$. When $c = -0.5 + 0.6i$ the process wanders hopelessly, re-

*A written version of the Mary P. Dolciani Lecture delivered at Hunter College in April 1993.

turning close to various numbers, but at irregular intervals. And when $c = -0.4 + 0.8i$ the process seems to wander for a few repetitions and then gets large quickly; after 14 repetitions both real and imaginary parts of the number exceed 10^{113}.

Precisely what happens for each value of c is complicated and a bit mysterious. One aspect of the behavior is simple, however: For each c, either the c-process stays bounded or it does not.

Definition. The Mandelbrot set M is the set of all c for which the c-process stays bounded.

How do we draw pictures? That's simple too. Each complex number $c = a + bi$ can be identified with a point (a, b) in the plane. Subdivide a section of the plane into tiny squares (pixels) and use the center of each pixel as your c; color that pixel black if it is in the Mandelbrot set—otherwise, leave it white. Figure 1 shows the first crude picture of M drawn in this way.

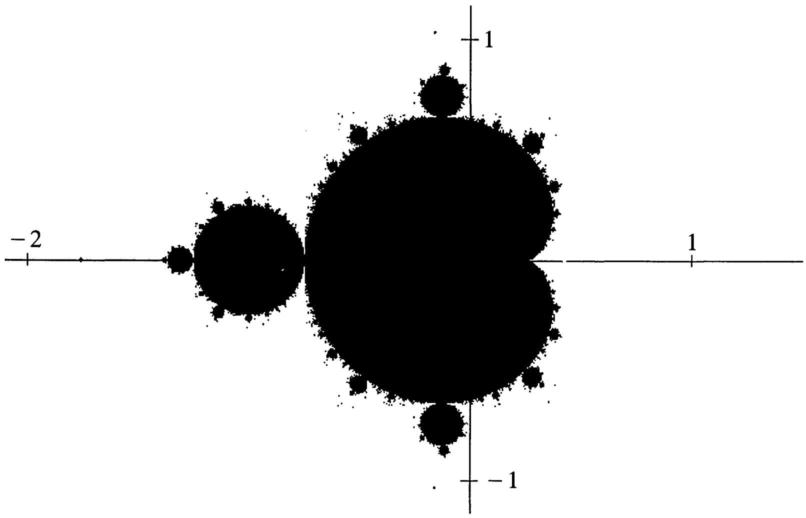


Figure 1

Of course, the astute reader may have noticed a loose end in this discussion. How do we know when the process stays bounded? The answer is: For many points, we don't, not for sure. On the other hand, often we can be sure that the process is *unbounded*. The idea is clear: The square of a *large* complex number is much larger still, even when we add c. So when the process reaches a large number, the next number is even larger, and so on. Here is the precise result.

Lemma. *If the c-process ever reaches a number beyond radius* 2, *then it is unbounded.*

The proof uses only elementary ideas—essentially the triangle inequality—but we omit the details of the proof.

This small lemma not only tells us how to recognize values of c that are *not* in the Mandelbrot set; it also gives us a way to draw more informative (and more attractive) pictures. Again, we subdivide a section of the plane into pixels and perform the c-process using the center of each. This time, however, when the process goes beyond radius 2, we stop (since we know it is unbounded) and color the pixel a color that reflects how many repetitions it required to get there. Those are all points *outside* the Mandelbrot set, but the colors reflect (in some sense) how far outside they are.

Figure 2 shows the Mandelbrot set drawn according to this scheme, but we have replaced the colors with different shades of grey. The image in Figure 3 is a tiny portion of the boundary of the image in Figure 2. By using the computer to zoom in on that image, we are able to draw a portion of the boundary inside a box that is about 10^{-10} cm on a side—no larger than the size of a hydrogen atom—and we can draw it in exquisite detail. The computer is in this case a powerful microscope indeed.

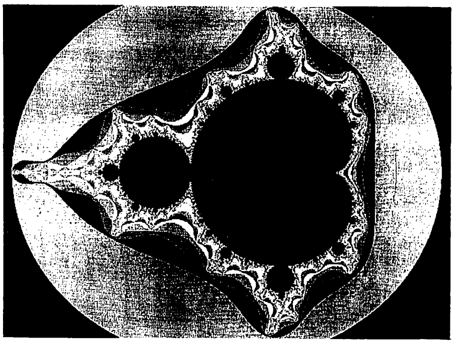


Figure 2
The Mandelbrot set, outside shadings.

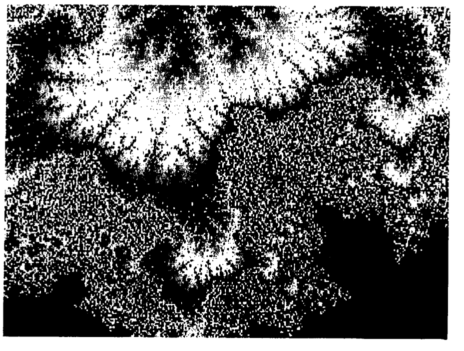


Figure 3
Part of the boundary, size of hydrogen atom.

Our new scheme still does not tell us when a point is in the Mandelbrot set. If we perform the c-process for a thousand repetitions, we may begin to suspect that the process is bounded—but we are not certain. After another thousand repetitions, we might give up trying. In order to draw pictures, therefore, we simply

Facts about the Mandelbrot Set

- We know from the lemma that the entire set M is contained in the disk of radius 2.
- We know from the definition that M is closed (since its complement is open—if the c-process is unbounded, then it is unbounded for nearby values of c).
- Therefore, we know that M is compact (Heine-Borel).
- The main part, which looks like a cardioid, is in fact a cardioid, and consists of those values of c for which the c-process approaches a single value (a fixed point). It is not hard to write down the explicit formula.
- To the left of that cardioid, there is a disk, which consists of those values for which the c-process oscillates closer and closer to a pair of points (a 2-cycle). Again, we can find a formula.

We can describe other specific features, but with more and more difficulty as we view finer and finer detail on the boundary. See [2].

agree to establish a threshold T (the give-up threshold or GUT) and declare values of c to be in the Mandelbrot set when the c-process has not gone beyond radius 2 after T repetitions. Of course, the picture we draw may depend on our GUTs.

As indicated in the "Facts" box, we *can* prove that *some* points are in the Mandelbrot set. These are interior points. Whether all interior points can be described in a similar way—as values of c for which the c-process approaches a cyclic orbit of points—is not known.

What's the Problem?

Why should we worry about the accuracy of these pictures? Two reasons: Computers make mistakes, and so do people.

Computers must approximate numbers in almost every computation, and even when a computer exactly represents a number, it may not exactly represent its square. In fact, the *only* numbers most modern computers can handle with complete precision are numbers with finite binary expansion. The mistakes (round-off errors) are quite small, of course, but for *some* values the c-process is sensitive to even incredibly tiny errors. After many repetitions, we may be far away from the true value.

The mistakes people make are those GUTs we referred to before. For any c we perform the c-process until either we get beyond radius 2 (and then we know the c is not in the Mandelbrot set) or we give up (and then we declare it in the Mandelbrot set due to exhaustion). We may be giving up too early, and changing the give-up threshold changes the picture we draw.

Here's a good analogy. Suppose we try to draw a graph of a reasonably complicated function, say $y = \sin(50x)$ on the interval $[-1, 1]$. Secretly, we know that the graph oscillates up and down about 16 times between -1 and 1, and looks like the familiar sine curve. But we "draw" the graph in the same way that we "draw" the Mandelbrot set. We subdivide the square into pixels, and test the center of each to see whether it is on the graph; that is, for the point (x, y) we check to see whether $y = \sin(50x)$. The graph appears in Figure 4.

Of course, we should not be surprised that the graph is virtually invisible: Few centers of the pixels will be *precisely* on the graph. We need to test whether they are nearby, not exactly on, the graph. We therefore introduce a *tolerance* (analogous to the give-up threshold). When y is within the tolerance of $\sin(50x)$, we color the pixel black. Figures 5, 6, and 7 show pictures of the graph with tolerances

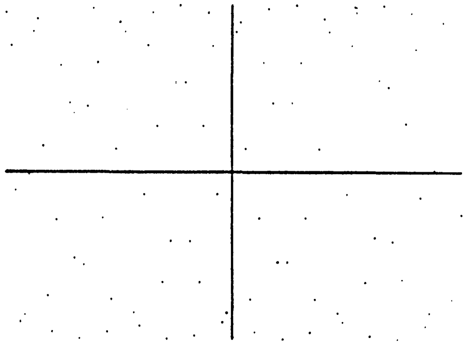


Figure 4
The graph of $y = \sin(50x)$.

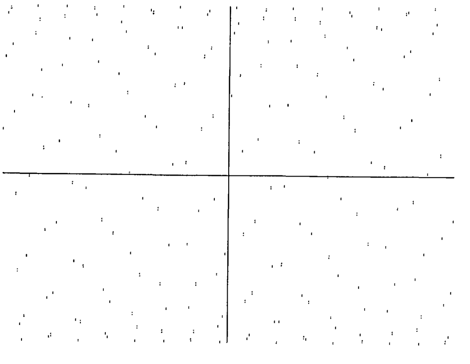


Figure 5
Tolerance = 0.01.

of 0.01, 0.1, and 1. Notice that the picture with what appears to be a foolish tolerance of 1 gives the most accurate (although crude) picture of the graph. By carefully adjusting the tolerance (and the number of pixels) we can produce a good picture of the graph. But in this case, we *know* what the real picture should look like.

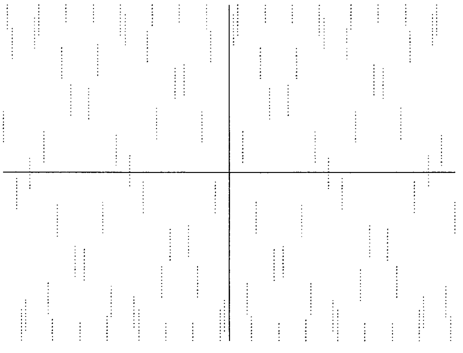


Figure 6
Tolerance = 0.1.

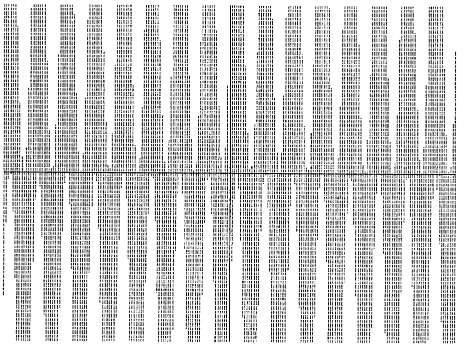


Figure 7
Tolerance = 1.

What about the Mandelbrot set? Figures 8 and 9 show the same small piece of the boundary, one with give-up threshold of 250 and the other with threshold 400. The difference is striking. At first, we might be tempted to say that the *larger* threshold provides the more accurate picture; after all, we should be less likely to make a mistake when we work harder. But in the $\sin(50x)$ example above, the *crudest* tolerance gave the most accurate picture. Could the same be true for the Mandelbrot set—could a smaller threshold give a better likeness?

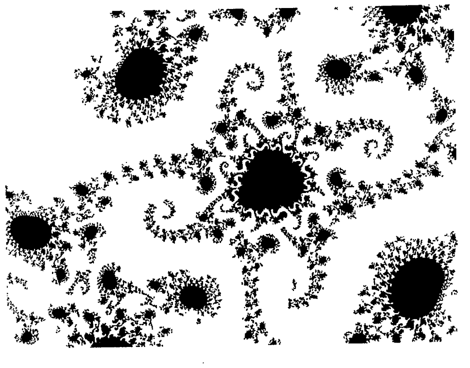


Figure 8
Threshold = 250.

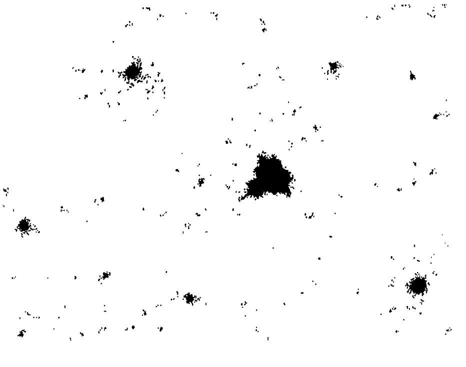


Figure 9
Threshold = 400.

Do Pictures Mislead?

The Mandelbrot set is a complicated object, and like all pictures of complicated mathematical objects (for example, the Cantor set) whatever we draw on paper cannot show all detail. We want pictures to represent basic properties, not the fine detail. Do pictures represent the *basic* features of the Mandelbrot set?

Not always. In fact, almost all pictures of tiny sections of the boundary show disconnected pieces of M, often resembling miniature copies of the whole set, floating nearby. One might easily conclude from those pictures that M is not connected—that it consists of a main body with an infinite number of islands nearby (and an infinite number of islands near each of these, and so on).

The pictures are misleading.

Theorem (1982, Douady and Hubbard). *The Mandelbrot set is connected.*

The proof they presented [**3**] does not depend on pictures at all. The key is *analysis*, not pictures. Here are the ideas. First, we can think of the Mandelbrot set as a subset of the Riemann sphere Σ using stereographic projection. (Rest a sphere on the plane with its south pole at the origin. Straight lines from each point in the plane to the north pole intersect the sphere at one point, and this identifies the plane with the sphere—with the north pole mapping to the point at infinity.) The cover illustration for this issue shows the projection of M on Σ. Now the Mandelbrot set is *connected* precisely when its complement (in the sphere) is *simply connected*, that is, when its complement has no holes.

Simply connected open subsets of the sphere play a central role in complex analysis. In fact, the famous Riemann mapping theorem says that such a set which is not the entire plane is *equivalent* to the unit disk Δ. The word "equivalent" means that there is a *homeomorphism* (a one-to-one, continuous, onto map) that is also *analytic* (i.e., has a power series expansion).

Hubbard and Douady turned the Riemann mapping theorem around; they showed that there must exist such a map from the disk Δ to the complement of M (in the sphere), and consequently the complement of M is simply connected, showing that M itself is connected. Their map $\psi\colon \Delta \to \Sigma - M$ must have the form

$$\psi(z) = \frac{1}{z} + b_0 + b_1 z + b_2 z^2 + b_3 z^3 + \cdots$$

and, of course, if we knew all the coefficients b_k then we would know exactly how to map the disk to the complement of M. We would *parametrize* the complement of M.

Open Problem

The map ψ is a homeomorphism from the open disk to the (open) complement of M. It does *not* extend to a homeomorphism on the boundaries. Does it even extend to a continuous map? This is still unknown. From classical results we know it extends if and only if the boundary of M is *locally connected*. For this reason, a great deal of effort has been spent trying to prove what appears to be an arcane topological result about the boundary.

In a lovely paper [**5**] Jungreis gave an algorithm for calculating those coefficients. Knowing all coefficients b_k for $k < n$, Jungreis showed how to compute b_n. The process was complicated, however, and it was hard to know how much error arose in the computation. Because *some* of those coefficients could be calculated explicitly [1], [6], one could check the Jungreis algorithm to see how large the error might be. All this was done in [4], where Glenn Schober and I computed the first 500,000 coefficients and showed that the error was likely to be within 10^{-18} by comparing the calculated values for certain coefficients with the known values.

Why should we be so interested in the map to the complement of M? Because this gives a *new* way to draw pictures of M, different from the pixel method. Think of the mapping

$$z \to \psi(z) = \frac{1}{z} + b_0 + b_1 z + b_2 z^2 + \cdots$$

in two steps. The reciprocal $z \to 1/z$ maps the interior of the unit disk Δ onto the exterior, reversing orientation and sending the origin to ∞. The power series terms $b_0 + b_1 z + b_2 z^2 + \cdots$ add a small distortion to make the image of ψ the exterior of M rather than the exterior of Δ. The image under ψ of a circle $|z| = r$, $r < 1$, is a simple closed curve bounding a region M_r which contains M. As $r \to 1$ these regions "shrink-wrap" M, as indicated in Figure 10.

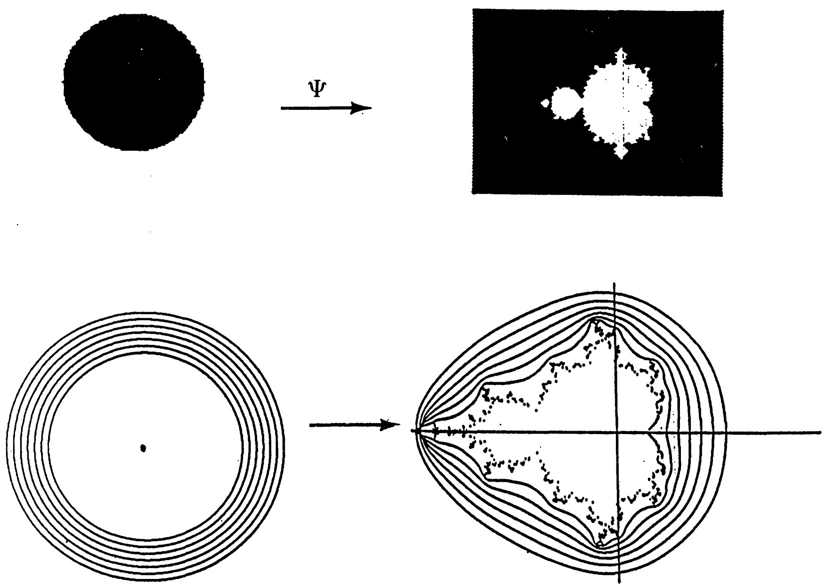


Figure 10
Mapping the disk to the outside of M.

We therefore get two pictures of the Mandelbrot set, shown in Figure 11—one using the pixel method described at the beginning, and one using our approximation to the map ψ. Which picture is correct? Are they really different?

The Area of *M*

This last question is the important one. Since we have already agreed that the Mandelbrot set is complicated, so that *any* picture will not represent all the detail, we should ask whether the two pictures are all that different. And perhaps the best

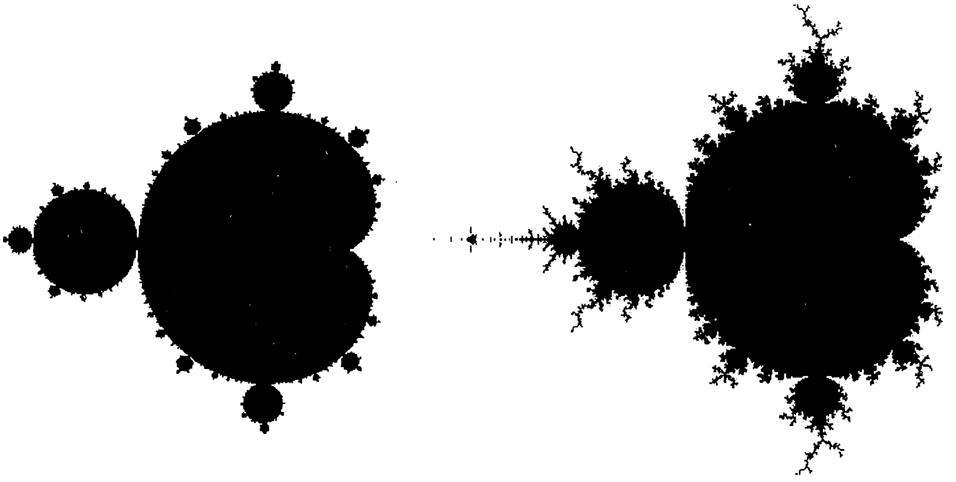


Figure 11
Two views of the Mandelbrot set.

way to measure the difference of two sets in the plane is to calculate the area of each. Our aim therefore is to use each method of drawing M—the pixel method and the ψ method—to calculate the area. We hope the answers are close, and hence the two pictures are not too far apart.

Calculating the area using the pixel method is relatively straightforward, similar to the way you would calculate the area of a large, strangely shaped room by counting floor tiles. We subdivide a region of the plane containing M into pixels and calculate the fraction that are inside M. The area is that fraction of the region's area. The trick is, of course, to know when a pixel is *inside* (or partly inside) the Mandelbrot set. We have already seen that this depends in part on the give-up threshold, and the answer we get depends both on that threshold and on the size of pixels we choose. By varying both the threshold and the pixel size, however, we can watch how the answer varies and extrapolate to a single number:

The area of the Mandelbrot set found by the pixel method is 1.52.

How do we use the map ψ to calculate the area? There is a beautiful trick (well known to people who work in complex variables) to do this. It is an elegant application of one of the basic theorems of calculus, Green's theorem, which allows one to compute the area of a (nice) subset of the plane by integrating around the boundary. Integrating around the boundary of the Mandelbrot set is not so easy, of course, but the map $w = \psi(z)$ provides a change of variable that converts the problem of integrating around the Mandelbrot set to the problem of integrating around a circle—and that certainly is much easier.

Here are some of the details.

Green's theorem (symmetric form). *For nice subsets S in the (u, v)-plane,*

$$\text{Area } S = \int\!\!\int_S du\,dv = \frac{1}{2}\int_{\partial S} u\,dv - v\,du.$$

If we introduce a complex variable $w = u + iv$, then an elementary calculation shows another compact form of Green's theorem:

$$\text{Area } S = \frac{1}{2i}\int_{\partial S} \bar{w}\, dw.$$

(*Careful*: You need to use the fact that $u\,du$ and $v\,dv$ are "exact," which means their integrals around the boundary are zero.)

We can't apply Green's theorem directly to M, but *can* apply it to the sets M_r, which are the sets containing M and bounded by the image of $|z| = r$ under ψ. We'll find the area of M by *squeezing* in as $r \to 1$. Applying Green's theorem and making a change of variable $w = \psi(z)$, we see that

$$\text{Area } (M_r) = \frac{1}{2i}\int_{\partial M_r} \bar{w}\, dw = \frac{1}{2i}\int_{|z|=r} \overline{\psi(z)}\psi'(z)\, dz.$$

We can use the power series expression for ψ to compute this last integral. Since $z\bar{z} = r^2$ we have

$$\overline{\psi(z)} = \frac{z}{r^2} + \bar{b}_0 + \frac{\bar{b}_1 r^2}{z} + \frac{\bar{b}_2 r^4}{z^2} + \frac{\bar{b}_3 r^6}{z^3} + \cdots.$$

By differentiating term-by-term, we have

$$\psi'(z) = \frac{-1}{z^2} + b_1 + 2b_2 z + 3b_3 z^2 + \cdots.$$

Multiplying these together and integrating around $|z| = r$ gives the result. This is easier than it looks because *most* of the terms in the product disappear when we integrate:

$$\int_{|z|=r} z^k\, dz = \begin{cases} 2\pi i & \text{if } k = -1 \\ 0 & \text{otherwise.} \end{cases}$$

Collecting the coefficients of $1/z$ in the product and integrating, we obtain a formula for the area of M_r. (*Careful*: Because ψ reverses orientation, we have to integrate *clockwise*!)

$$\begin{aligned} \text{Area } M_r &= \frac{1}{2i}\int_{|z|=r} \overline{\psi(z)}\psi'(z)\, dz \\ &= \frac{1}{2i}\left(\frac{-1}{r^2} + |b_1|^2 r^2 + 2|b_2|^2 r^4 + 3|b_3|^2 r^6 + \cdots\right)(-2\pi i) \\ &= \pi\left(\frac{1}{r^2} - \sum_{n\geq 1} n|b_n|^2 r^{2n}\right). \end{aligned}$$

Finally, taking the limit as $r \to 1$ gives an elegant formula for the area of M in terms of the coefficients for ψ:

$$\text{Area } M = \pi\left(1 - \sum_{n\geq 0} n|b_n|^2\right).$$

We can use the first 500,000 coefficients to get an approximation (actually, an upper bound) for the area:

The area of the Mandelbrot set found by the ψ method is 1.72.

Conclusion

Which is it? Is the area 1.52? Is it 1.72? The answer is—we don't know, not for sure.

Does that surprise you? We don't know the area of the Mandelbrot set to within 10% accuracy. How then can we use the computer to zoom in on small pieces of the boundary that are no larger than a hydrogen atom? Are those really pictures of the Mandelbrot set, or are they intricate shadows produced mainly by round-off error in our computer and a badly chosen threshold? Again, the answer is: We don't know—not for sure.

What is the best guess for the area? The series for ψ converges slowly—very slowly—and it is likely that one needs not half a million terms but many orders of magnitude more terms to produce a reasonable upper bound.

Pixel counting, however, has its own drawbacks. By using more sophisticated algorithms (see [7]), we can approximate the distance of a point (outside) to the set M; this gives a better way to measure whether a pixel should be counted inside M. But the algorithm uses iteration, and there has not been a satisfactory analysis of the error. Although the final answer is quite likely much closer to 1.52 than 1.72, providing a convincing argument is not easy.

There is one strange possibility that has not been ruled out. Pixel counting measures the area *inside* the Mandelbrot set, that is, the area of its interior. The ψ-method gives an upper bound for the area of the entire set M, including its boundary. Could the two answers be so different because the area of the boundary is positive?

We don't know—not for sure.

So... the next time you see one of those gorgeous pictures of the Mandelbrot set, with swirls and dots and dainty patterns, that claims to represent the fine detail of an amazingly complicated set, I hope you will admire the artistry... and question the mathematics. I hope you will be a skeptic.

I am.

References

1. B. Bielefeld, Y. Fisher, and F. v. Haeseler, Computing the Laurent series of the map ψ, *Max Planck Institut Preprint* 46 (1988) 1–16.
2. Bodil Branner, The Mandelbrot set, in R. Devaney and Linda Keen, eds., *Chaos and Fractals* (Proceedings of Symposia in Applied Mathematics, vol. 39), American Mathematical Society, Providence, 1988, 75–105.
3. A. Douady and J. Hubbard, Itération des polynômes quadratiques complexes, *Comptes Rendus, Académie des Sciences, Paris* 294 (1982) 123–126.
4. J. Ewing and G. Schober, The area of the Mandelbrot set, *Numerische Mathematik* 61 (1992) 59–72.
5. I. Jungreis, The uniformization of the complement of the Mandelbrot set, *Duke Mathematics Journal* 52 (1985) 935–938.
6. G. M. Levin, On the arithmetic properties of a certain sequence of polynomials, *Russian Mathematical Surveys* 43 (1988) 245–246.
7. J. Milnor, Self-similarity and hairiness in the Mandelbrot set, in M. Tangora, ed., *Computers in Geometry and Topology* (Lecture Notes in Pure and Applied Mathematics, vol. 114), Dekker, New York, 1989, 211–257.