You Could Learn a Lot from a Quadratic: Newton Squares the Circle

Author: *Henry G. Baker, http://home.pipeline.com/˜hbaker1/home.html; hbaker1@pipeline.com*

"Newton's Method" for the iterative approximation of square roots—not just on the real line, but also in the complex plane—provides one of the most beautiful and elegant applications of 'linear fractional transformations,' also known as 'Möbius transformations.'

Briefly, Newton's Method for the square root of the real number N consists of guessing an initial approximation z_0, and then successively computing the members of the sequence $z_1 = (z_0 + N/z_0)/2$, $z_2 = (z_1 + N/z_1)/2$, ..., $z_{n+1} = (z_n + N/z_n)/2$. As we have discussed in previous columns, this iteration converges rapidly to $\pm\sqrt{N}$ so long as $z_0 \neq 0$, for positive N, real z_0.

But the real beauty and understanding of Newton's method becomes evident by looking at its operation in the complex plane. In other words, we will consider the action of the Newton iteration for *complex* numbers $z_n = x_n + iy_n$. We will see that Newton's square root method consists of a Möbius transformation of the initial guess to a new "coordinate system" for the complex plane, followed by a number of simple *squarings* of the transformed initial guess, followed by the inverse Möbius transformation back to the original coordinate system.

We must first review a few properties of complex numbers. A complex number z may be expressed in either *rectangular* form as $z = x + iy$, or in *polar* form as $z = |z|e^{i\theta} = |z|(\cos\theta + i\sin\theta)$, where $|z|$ is the *absolute value* of z, i.e., $|z| = |x + iy| = \sqrt{x^2 + y^2}$, and θ is the *arctangent* of y/x, i.e., $\theta = \text{atan}\,(y/x) = \text{atan2}(y, x)$.

If we square a complex number, we get $z^2 = (x + iy)^2 = (x^2 - y^2) + i(2xy)$ in rectangular form, and $z^2 = (|z|e^{i\theta})^2 = |z|^2 e^{i2\theta}$ in polar form. Thus, when squaring a complex number in polar form, we note that its (real) absolute value is *squared*, and its angle θ is *doubled*.

Consider, for example, those complex numbers whose absolute value is exactly *one*, i.e., numbers z such that $|z| = 1$. These numbers form the *unit circle* in the complex plane. In this case, $z = |z|e^{i\theta} = e^{i\theta}$, so z is exactly representable as a pure imaginary exponential. For these complex numbers, $z = e^{i\theta} = \cos\theta + i\sin\theta$, so that

$$\begin{aligned} z^2 &= e^{i2\theta} \\ &= \cos 2\theta + i\sin 2\theta \\ &= (\cos\theta + i\sin\theta)^2 \\ &= (\cos^2\theta - \sin^2\theta) + i(2\cos\theta\sin\theta) \end{aligned}$$

This equation trivially shows the classic "double angle" formulas so laboriously memorized by high school trigonometry students.

Squaring a complex number z has the following property—if the number $|z| > 1$, then $|z^2| > |z| > 1$, and additional squarings will grow the absolute value towards infinity—i.e., $|z^{2^n}|$ grows very fast. Similarly, if $|z| < 1$, then $|z^2| < |z| < 1$, and additional squarings will shrink the absolute value towards zero—i.e., $|z^{2^n}|$ shrinks very fast. Finally, if $|z|$ is *exactly* one, then squaring z will not change this, and z^{2^n} will remain on the unit circle for all $n > 1$. Newton's square root method makes essential use of this growing/shrinking property of squaring to converge quickly to the correct value.

We are particularly interested in sets of complex numbers whose absolute values are all the same—these form concentric *circles* about the origin. The squaring function maps one such circle into a second such circle, with the mapping "covering" the image circle twice. If we square again, we map the original circle into a final circle *four* times, and if we square the original circle n times, we map the original circle into a final circle 2^n times.

If we *invert* a general complex number—i.e., perform the operation $z \to 1/z$, then in polar form this operation becomes $|z|e^{i\theta} \to (1/|z|)e^{-i\theta}$. For the case where $|z| = 1$, $1/z = (1/|z|)e^{-i\theta} = e^{-i\theta}$, so that $1/z$ is identical to the *conjugate* $\bar{z} = x - iy$ of $z = x + iy$.

The next complex operation we will need is that of 'linear fractional transformation' or 'Möbius transformation,' for short. A Möbius transformation transforms $z \to (Az+B)/(Cz+D)$, where A, B, C, and D are complex numbers, and we will need the additional condition that $AD - BC \neq 0$. These transformations are elegant and important because they are one-to-one and onto functions of the complex plane extended with the additional point $\infty = 1/0$. These transformations have inverses—e.g., the inverse transformation to $z \to (Az+B)/(Cz+D)$ is $z \to (Dz-B)/(-Cz+A)$:

$$\begin{aligned}&\frac{A(Dz-B)/(-Cz+A)+B}{C(Dz-B)/(-Cz+A)+D}\\ &=\frac{A(Dz-B)+B(-Cz+A)}{C(Dz-B)+D(-Cz+A)}\\ &=\frac{ADz-AB-BCz+AB}{CDz-BC-CDz+AD}\\ &=\frac{(AD-BC)z}{AD-BC}\\ &=z\end{aligned}$$

Working with Möbius transforms as ratios works, but we can see the properties much more clearly if we write these transforms as 2×2 matrices:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} z \\ 1 \end{pmatrix} = \begin{pmatrix} Az+B \\ Cz+D \end{pmatrix}$$

Once we have this bidirectional mapping between Möbius transformations and matrices, the other properties of Möbius transformations become obvious. Thus, the requirement that $AD-BC \neq 0$ is simply the requirement that the *determinant* of the 2×2 matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ be non-zero, so that the matrix has an inverse. Thus, the inverse of this matrix is

$$\frac{1}{AD-BC}\begin{pmatrix} D & -B \\ -C & A \end{pmatrix}$$

However, since we intend to form the *ratio* of the two elements of the final column matrix, we can safely ignore any "scale factors"—so long as any such scale factor (in this case $AD-BC$) is non-zero.

One of the truly remarkable properties of Möbius transforms is that they *preserve circles*—i.e., given *any* circle in the complex plane, a Möbius transform will transform that circle into another circle. For this property to work, however, we must define "circles" as including the limiting case of *straight lines*, which are circles having an "infinite radius."

For example, the transform $z \to (z-1)/(z+1)$ transforms the imaginary (or "y") axis into the unit circle—i.e., 0 transforms into -1, $\pm i$ transforms into $\pm i$, and ∞ transforms into 1.

Proving that Möbius transforms preserve circles is a bit tricky, since Möbius transforms do *not* preserve the *centers* of the circles. Thus, the center of a circle does *not* usually map into the the center of the Möbius-transformed circle. As a result, one cannot simply transform the origin to the center of the circle and show that a transformed rotating clock hand sweeps out the transformed circle.

There is another characterization of a circle, however, that was known to the ancient Greeks, but not to most modern high school students. This characterization describes a circle as the locus of points the ratio of whose distance to two given fixed points is a constant. Thus, we start with two arbitrary points P and Q in the plane, and then find those points Z such that $d(Z,P)/d(Z,Q)=k$, where k is a real (positive) constant and $d(Z,P)$ means the (Euclidean) distance between the point Z and the point P. The locus of these points Z will be a circle or a straight line.

We will now prove this fact twice—once in traditional x,y coordinates, and once in complex numbers. Hopefully, you will agree that the complex number formulation is much shorter and sweeter.

Recall that the equation of a circle in the x,y plane is in the form $(x-O_x)^2+(y-O_y)^2=r^2$, where (O_x,O_y) is the center of the circle, and r is the radius of the circle. If we now form the ratio of the distances from (x,y) to $P=(P_x,P_y)$ and $Q=(Q_x,Q_y)$, respectively, we get by brute algebraic force:

$$\frac{\sqrt{(x-P_x)^2+(y-P_y)^2}}{\sqrt{(x-Q_x)^2+(x-Q_y)^2}}=k$$

Squaring both sides, we get

$$\frac{(x-P_x)^2+(y-P_y)^2}{(x-Q_x)^2+(y-Q_y)^2}=k^2$$

Clearing fractions, we get

$$(x-P_x)^2+(y-P_y)^2=k^2(x-Q_x)^2+k^2(y-Q_y)^2.$$

Without even finishing this derivation, we notice that the coefficients of both x^2 and y^2 are the same $(1-k^2)$, so that if the radius squared of the resulting equation is positive, we will have a "real" circle.

Let us now consider the same derivation in complex numbers. We know that the "length" of a complex number $z=x+iy$ is the distance from z to the origin, or $|z|=\sqrt{x^2+y^2}=\sqrt{z\bar{z}}=\sqrt{(x+iy)(x-iy)}$. So the equation of a circle in complex coordinates is $|z-O|=r$, where O is the center and $r > 0$ is the radius. This equation is more usually written as

$$\begin{aligned}|z-O|^2 &= (z-O)(\bar{z}-\bar{O})\\ &= [(x-O_x)+i(y-O_y)]\\ &\quad [(x-O_x)-i(y-O_y)]\\ &= (x-O_x)^2+(y-O_y)^2\\ &= r^2\end{aligned}$$

So the complex formulation of a circle is identical to the Cartesian formulation. We can now go back to our *ratio* formulation for a circle:

$$\frac{|z-P|^2}{|z-Q|^2} = k^2$$

Clearing fractions, we have

$$\begin{aligned}|z-P|^2 &= k^2|z-Q|^2\\ \left|z-\frac{P-k^2Q}{1-k^2}\right|^2 &= \frac{|P-k^2Q|^2}{(1-k^2)^2} - \frac{|P|^2-k^2|Q|^2}{1-k^2}\\ \left|z-\frac{P-k^2Q}{1-k^2}\right|^2 &= \frac{k^2|P-Q|^2}{(1-k^2)^2}\\ &= \left(\frac{|k||P-Q|}{1-k^2}\right)^2\\ &= r^2\end{aligned}$$

Thus, in the general case, we have a circle of positive radius $|P-Q||k|/(1-k^2)$ centered at $(P-k^2Q)/(1-k^2)$. If $k=0$, then we have a circle of zero radius centered at P, while if $k=\pm 1$, we have a circle of infinite radius—i.e., a straight line perpendicularly bisecting the line PQ. Note that the sign of k is irrelevant, and we can therefore take it to be positive.

Now we can use this ratio characterization of circles to easily prove that Möbius transforms preserve circles. We simply plug in the result of the transformation and see that it is still the ratio of distances, as before.

Consider the circle defined by the two points P, Q and the ratio k. Then $|z-P|^2 = k^2|z-Q|^2$ is a circle. Consider transforming $z \to (Az+B)/(Cz+D)$. Its inverse is $z \to (Dz-B)/(-Cz+A)$, as we found above. Plugging this into the equation for the circle, we have:

$$\begin{aligned}\left|\tfrac{Dz-B}{-Cz+A}-P\right|^2 &= k^2\left|\tfrac{Dz-B}{-Cz+A}-Q\right|^2\\ |Dz-B+PCz-PA|^2 &= k^2|Dz-B+QCz-QA|^2\\ |(CP+D)z-(AP+B)|^2 &= k^2|(CQ+D)z-(AQ+B)|^2\\ \left|z-\tfrac{AP+B}{CP+D}\right|^2 &= \left|k\tfrac{CQ+D}{CP+D}\right|^2\left|z-\tfrac{AQ+B}{CQ+D}\right|^2\\ |z-P'|^2 &= k'^2|z-Q'|^2\end{aligned}$$

But we now recognize this as the equation of another circle defined by the points P', Q' and the ratio k', where P' is the Möbius-transformed P, Q' is the Möbius-transformed Q, and $k' = k|CQ+D|/|CP+D|$. Thus, we have proved that *Möbius transforms preserve circles* (suitably defined).

We are finally ready for the *pièce de résistance*—we show how Möbius transforms factor the Newton square root iteration into its essential parts.

Consider a circle defined by the two points $P=\sqrt{N}$, $Q=-\sqrt{N}$ and a ratio k. We would like to 1) map this circle onto a circle centered on the origin; 2) *square* the resulting quantity; and then 3) apply the mapping inverse to that in #1. In functional form, this sequence consists of a mapping $z \to f(z)$, following by a squaring, followed by the inverse mapping $z \to f^{-1}(z)$, or more succinctly, $z \to f^{-1}(f(z)^2)$. If we now apply this mapping twice—i.e., we *compose* the mapping with itself—we get the mapping $f^{-1}(f(f^{-1}(f(z)^2))^2) = f^{-1}(f(z)^4)$. If we continue iterating, then the n-th iterate will produce the mapping $z \to f^{-1}(f(z)^{2^n})$.

Consider now the function $f(z) = (z-\sqrt{N})/(z+\sqrt{N})$. $f(z)$ is a Möbius mapping which maps

$$\begin{aligned}0 &\to -1\\ \infty &\to 1\\ i\sqrt{N} &\to i\\ -i\sqrt{N} &\to -i\\ \sqrt{N} &\to 0\\ -\sqrt{N} &\to \infty\end{aligned}$$

In short, $f(z)$ maps $\sqrt{N} \to 0$, $-\sqrt{N} \to \infty$, and the y-axis onto the unit circle. If we now square the result of this mapping, the unit circle will map to the unit circle, while circles smaller than the unit circle will get smaller, and circles larger than the unit circle will get larger.

But in the "transformed coordinate system" produced by the mapping $f(z)$, in which the origin stands for $\sqrt{N}$ and ∞ stands for $-\sqrt{N}$, we see that the squaring achieves the effect of mapping circles which were originally close to $\sqrt{N}$ even closer to $\sqrt{N}$, and circles which were originally close to $-\sqrt{N}$ even closer to $-\sqrt{N}$.

Let us now prove that $f^{-1}(f(z)^2) = (z + N/z)/2$:

$$f^{-1}(f(z)^2) = f^{-1}\left(\left(\frac{z-\sqrt{N}}{z+\sqrt{N}}\right)^2\right)$$
$$= \sqrt{N}\frac{\left(\frac{z-\sqrt{N}}{z+\sqrt{N}}\right)^2 + 1}{-\left(\frac{z-\sqrt{N}}{z+\sqrt{N}}\right)^2 + 1}$$
$$= \sqrt{N}\frac{(z-\sqrt{N})^2 + (z+\sqrt{N})^2}{-(z-\sqrt{N})^2 + (z+\sqrt{N})^2}$$
$$= \sqrt{N}\frac{2z^2 + 2N}{4\sqrt{N}z}$$
$$= \frac{z^2 + N}{2z}$$
$$= \frac{z + N/z}{2}$$

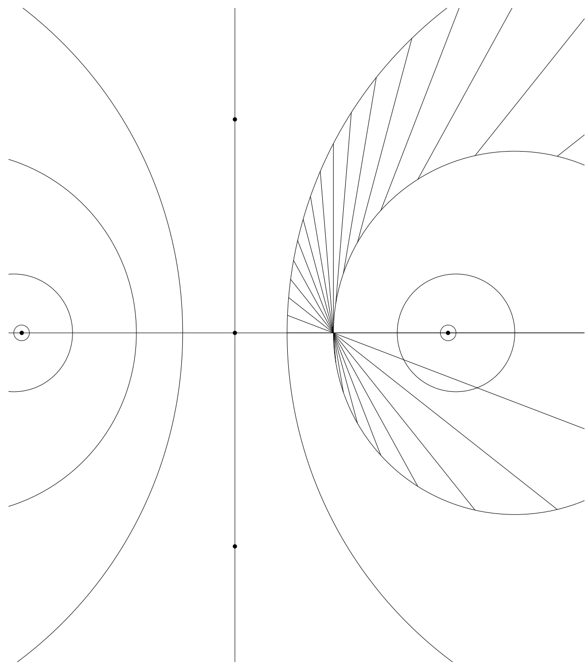
Instead of transforming to the new coordinate system for each iteration, we can conceptually think of performing a number of iterations in the transformed coordinate system before transforming back. In fact, we can perform sufficient squarings in the transformed coordinate system so that when we transform back, we will be sufficiently close to either $\sqrt{N}$ or $-\sqrt{N}$. These insights have thus given us a *closed form* solution for the n-th Newton iteration:

$$z \rightarrow \sqrt{N}\frac{(z+\sqrt{N})^{2^n} + (z-\sqrt{N})^{2^n}}{(z+\sqrt{N})^{2^n} - (z-\sqrt{N})^{2^n}}$$

Our explorations have shown us that instead of focussing upon the sequence of *points* z_0, z_1, ..., z_n, we should instead focus upon the sequence of *circles* C_0, C_1, ..., C_n on which these points lie, since all of the points of the circles will be treated essentially alike by the Newton iterations. In fact, unless z_0 lies exactly *on* the y-axis, then these circles will very quickly converge to zero-radius circles about either $\sqrt{N}$ or $-\sqrt{N}$.

Given an initial guess z_0, how do we find the appropriate circle on which it lies? The answer should now be obvious: these circles are defined by the two points $\sqrt{N}$, $-\sqrt{N}$, and the ratio $k = |z-\sqrt{N}|/|z+\sqrt{N}|$, and the n-th converging circle is defined by the same two points $\sqrt{N}$, $-\sqrt{N}$, and the ratio k^{2^n}.

If we plot a circle of points having the same ratio of distances to $\sqrt{N}$, $-\sqrt{N}$, respectively, and then plot the images of those points under the Newton square root iteration, we find that a) these image points form a circle; and b) that the intersections I_1, I_2 of this image circle with the x-axis have some interesting properties: 1) the intersection point I_2 further from the origin is the *center* of the first circle; and 2) the intersection point I_1 lies on every line which connects a point on the first circle with its image point under the Newton square root mapping.



In the figure, we have plotted a sequence of circles generated by successive Newton iterations, and we see that these circles converge very quickly to $\sqrt{N}$. We have also plotted a number of points on the initial circle and drawn lines to their images under a single Newton iteration. These points are 10 degrees apart on the circle in the coordinate system in which the squarings take place, and we readily see that they are not equally spaced in the original coordinate system.

References

[Beeler72] Beeler, M, Gosper, R.W., and Schroeppel, R. *HAKMEM*. MIT AI Memo No. 239, Feb. 29, 1972, items #3, #126. http://home.pipeline.com/~hbaker1/hakmem/hakmem.html

[Cayley79] Cayley, A. "Application of the Newton-Fourier Method to an Imaginary Root of an Equation." *Quart. J. Pure & Appl. Math.*, xvi (1879), 179-185. http://home.pipeline.com/~hbaker1/quaternion/cayley/Newton.ps.gz

[Corless98] Corless, Robert M. "Variations on a Theme of Newton." *Math. Mag.*, 71, 1 (Feb. 1998), 34-41.

[Curry83] Curry, J., Garnett, L., and Sullivan, D. "On the iteration of rational funtions: Computer experiments with Newton's method." *Commun. Math. Phys. 91* (1983), 267-277.

[Friedman86] Friedman, J. "On Newton's Method for Polynomials." *IEEE 27th Symp. on FOCS*, Oct. 1986, 153-161.

[Klein56] Klein, Felix. *Lectures on the Icosahedron and the Solution of Equations of the Fifth Degree*. Morrice, G.G., transl. from 1888, 1913 manuscript. Dover Publs. Inc., New York, 1956.

[Schwerdtfeger79] Schwerdtfeger, H. *Geometry of Complex Numbers*. Dover Publ., New York, 1979.

Henry Baker has been diddling bits for 35 years, with time off for good behavior at MIT and Symbolics. In his spare time, he collects garbage and tilts at windbags. *This column appeared in ACM Sigplan Notices 33,6 (Jun 1998), 27-31.*