The Newton-Raphson fractal: what is it *really* all about?

James M. Christian and George S. Jensen,
Joule Physics Laboratory, University of Salford, U.K.
(email: j.christian@salford.ac.uk)

The Newton-Raphson (NR) method is a black-box device for finding roots to equations like $f(x) = 0$. Insert guess at input, crank handle, re-insert output as better guess. Repeat until happy with answer. NR is elegant in its simplicity and easy to implement, beloved by mathematicians, physicists, and computer scientists everywhere. Why? Not just because of its speed and efficiency, but because of its innate capacity to create strikingly intricate patterns when the inputs are complex numbers instead of real numbers.

In what follows, we will look for the cube roots of -1 so that $f(z) \equiv z^3 + 1 = 0$. Here, $z = x + iy$ is a complex number. The real numbers x and y are its real and imaginary parts, respectively, while $i \equiv \sqrt{-1}$ is the imaginary unit. All the usual arithmetic for adding, subtracting, multiplying, and dividing real numbers also holds for complex numbers, but now we have the extra rule that $i^2 = -1$.

The ordered pair (x, y) represents a point in the complex plane, just like in coordinate geometry, and the three points we are looking for are the solutions $z = -1 + i0$, $(1 + i\sqrt{3})/2$, and $(1 - i\sqrt{3})/2$. They mark the corners a_1, a_2, and a_3 of an equilateral triangle, centred on the origin O, and sitting inside a circle with radius 1. That we already know the answers for z is not important; we are not interested in root-finding per se. Rather, we want to unpack *how* NR creates its famous pattern and what some of it means.

On the complex plane, the NR machinery for our cube-roots problem is

$$z_{n+1} = F(z_n), \quad \text{where} \quad F(\chi) \equiv \chi - \frac{\chi^3 + 1}{3\chi^2} \tag{1}$$

and χ denotes the input to black box F. The discrete index $n = 0, 1, 2, 3, ...$ is the iterate number, often interpreted as labelling points in time that are separated by some arbitrary fixed interval. For any starting point z_0, repeated application of F generates a sequence of numbers $z_1 = F(z_0)$, $z_2 = F(z_1) = F(F(z_0))$, $z_3 = F(z_2) = F(F(F(z_0)))$ and so on. At each step, the output is fed back into F as a new input. The number z_n in the sequence $z_1, z_2, z_3, ..., z_{n-1}, z_n, z_{n+1}, ...$ is therefore computed through n applications of F when the initial condition is z_0 (see Fig. 0.1).

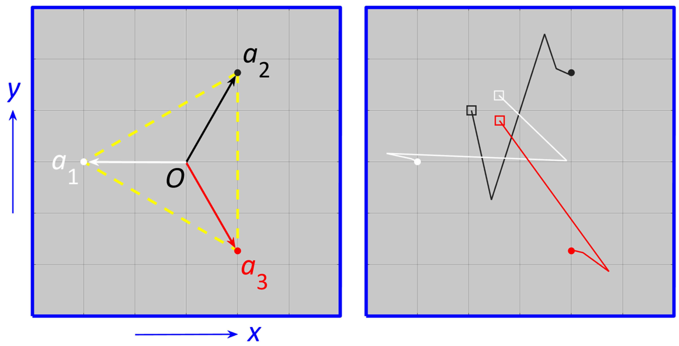

Figure 0.1: The winning attractor is evidently not always the root closest to the initial condition (squares).

Attractors and their basins

Suppose there exists a number a with the property $a = F(a)$ so that feeding a into F returns the number a. Given $z_0 = a$, under iteration map F must generate a very simple sequence $a, a, a, a, a, \ldots$ so that input and output are identical to each other. Terminology-wise, we say that a is a fixed point of F. For our chosen F, there exist three such fixed points—they are the roots of $a^3 + 1 = 0$ which wrote down earlier. It is not a coincidence that these particular values of a have appeared. By design, the NR scheme naturally has the roots we want as the fixed points of F.

A sequence produced by F for a given initial condition z_0 is called a trajectory (or, sometimes, an orbit). For almost any z_0, the trajectory typically bounces around in the complex plane, gradually zeroing-in on a_1, a_2, or a_3. It is as though the trajectory is being pulled by invisible forces, attracted simultaneously towards all three fixed points until one of them ultimately 'wins'. With that conceptual picture in mind, the three fixed points are known as attractors. But which one wins? And how does the winner vary with the choice of z_0?

There is an appealing graphical way to start addressing those questions on computer. We consider a region of the complex plane, typically dividing it up into a square grid of points. Taking each point on that grid in turn, the trajectory is computed and the winning attractor identified. The attractors are colour-coded (white for a_1, red for a_2 and black for a_3) and the collection of winners then overlaid back on top of the complex plane. The result is an abstract kind of map where the colour at each grid point indicates where a trajectory starting at that point will end up in the long term (see Fig. 0.2).

The white area, comprising all the points whose trajectories converge on a_1, is the basin of attraction for a_1. In the same way, the black and red areas

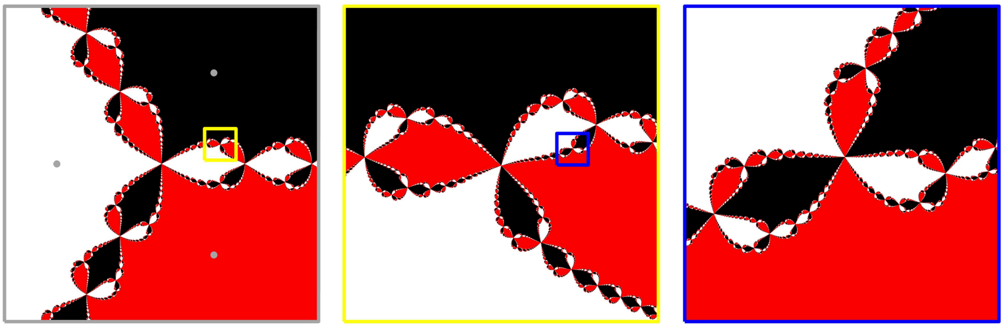

Figure 0.2: Magnifying regions of the basin boundaries by a factor of 10, then by another factor of 10. The three grey dots in the left-hand pane are the fixed point of map F.

are the basins of attraction for a_2 and a_3, respectively. The pattern along the arms where the three colours are intertwined is incredibly complicated and detailed. Zoom-in on any small region and we see the same sting-of-pearls motif repeated only at a shorter length-scale. This string-of-pearls is the Newton-Raphson fractal for the cube-roots problem, an absolute classic example of a scale-free—or self-similar—pattern. It survives under arbitrary magnification and it *never* goes away.

At a visual level, we can infer the presence of boundaries between the colours just from inspection. Whenever the colour changes, we have implicitly crossed a boundary. But there is so much more to it than that. What, exactly, *is* the boundary? How is it defined? What properties does it have? What role does it play? We will now attempt to answer some of those questions from a practical perspective.

Period-2 orbits

Suppose we can find two numbers, say b and $c \neq b$, such that, upon iteration, map F produces a sequence $b, c, b, c, b, c, b, c, \ldots$. The output of F now alternates, jumping back-and-forth between $z = b$ and $z = c$ according to the rules $c = F(b)$ and $b = F(c)$. The numbers b and c are called period-2 points and, taken together, they prescribe a period-2 orbit.

Eliminating c, it follows that $b = F(F(b)) \equiv F^{(2)}(b)$ and where the symbol $F^{(2)}$ denotes the second iterate of F. Symbolically, $F^{(2)}(b)$ means that the number $F(b)$ is fed back into the input of F to generate a new output, $F(F(b))$—so whenever we start with $z = b$, two iterations later we are guaranteed to return to $z = b$. Moreover, since the labelling of b and c is arbitrary, we must also have $c = F^{(2)}(c)$. The important point to take

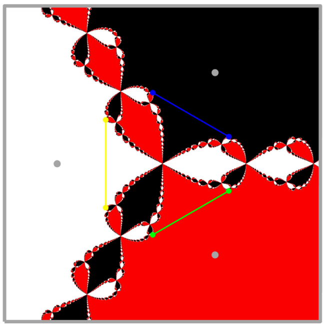

Figure 0.3: Period-2 points (dots) and their orbits (lines).

away is that b and c are two distinct values of z satisfying the equation $z = F^{(2)}(z)$.

These very general concepts can now be applied to the cube-roots problem. A period-2 point can be identified by the condition $z_{n+2} = z_n$, whereby the value of z is repeated every two iterations. Since $z_{n+1} = F(z_n)$, replacing index n with $n+1$ throughout gives $z_{n+2} = F(z_{n+1}) = F(F(z_n)) = F^{(2)}(z_n) = z_n$. We can thus find all the period-2 points of F by solving

$$z = F(F(z)) = z - \frac{z^3+1}{3z^2} - \frac{\left(z - \dfrac{z^3+1}{3z^2}\right)^3 + 1}{3\left(z - \dfrac{z^3+1}{3z^2}\right)^2}, \tag{2a}$$

which is a decidedly daunting task.

With a bit of algebra and bit of patience, we can show that its roots are also the roots of a simpler equation that can be written as the product of two polynomials:

$$(z^3+1) \times (20z^6 - 5z^3 + 2) = 0. \tag{2b}$$

The possibility that $z^3 + 1 = 0$ must be discarded because its solutions are the three fixed points of F which we already know. Recall that fixed points correspond to sequences like $a, a, a, a, ...$, clearly repeating after every two iterations but not period-2 in the $b, c, b, c, ...$ 'flipping' sense. They are reappearing here after being unintentionally snagged by the condition $z_{n+2} = z_n$. The remaining true period-2 points must instead be the six solutions to $20z^6 - 5z^3 + 2 = 0$, which eagle-eyed observers will recognize as a quadratic in z^3. Solve to find $z^3 = (5 \pm \sqrt{-135})/40 = (5 \pm i3\sqrt{15})/40$, then take the cube roots (see Fig. 0.3).

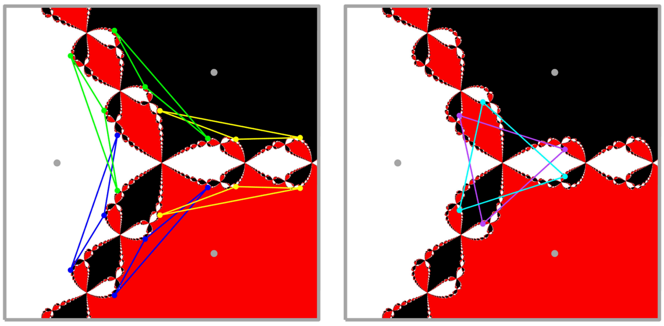

Figure 0.4: Period-3 orbits break up into two distinct groups with different shapes.

Period-3 orbits

We can do the whole thing again, only this time looking for a triplet of different numbers d, e, and f such that F produces the sequence $d, e, f, d, e, f, d, e, f, \ldots$. The output now repeats after every three iterations according to the rules $e = F(d)$, $f = F(e)$, and $d = F(f)$. Using exactly the same type of elimination as before, it is easy to show that d, e, and f must each be a unique value of z satisfying $z = F^{(3)}(z)$, where $F^{(3)} \equiv F(F(F))$ denotes the third iterate of F. Those numbers are referred to as period-3 points and together they prescribe a period-3 orbit.

The period-3 condition is $z_{n+3} = z_n$, and by shifting the index according to $n \to n+2$ we end up with $z_{n+3} = F(z_{n+2}) = F(F(z_{n+1})) = F(F(F(z_n))) \equiv F^{(3)}(z_n)$. Hence, period-3 points can be found by identifying the solutions of $z = F^{(3)}(z)$ which is an equation far too big to write out here! The salient points are that it gives rise to a degree-27 polynomial, but where three of the roots are the fixed points of F and so can be factorized away (just as they were in the period-2 case).

After a lot more algebra and a lot more patience, we are obliged to solve

$$\begin{aligned} &19,456z^{24} - 98,368z^{21} + 195,820z^{18} - 140,530z^{15} \\ &\quad + 60,493z^{12} - 14,413z^{9} + 2,149z^{6} - 196z^{3} + 16 = 0. \end{aligned} \tag{3}$$

Each period-3 orbit comprises three points and the set of $24/3 = 8$ such orbits naturally separates into two groups (see Fig. 0.4). The first contains six scalene triangle-shaped orbits distributed about the arms of the NR fractal. The second contains two equilateral triangle-shaped orbits centred on the origin that are slightly rotated versions of each other. The organization of sets of periodic orbits into distinct families is absolutely essential, and it follows from symmetries present in map F (not by accident, the three reflections and three rotations of an equilateral triangle).

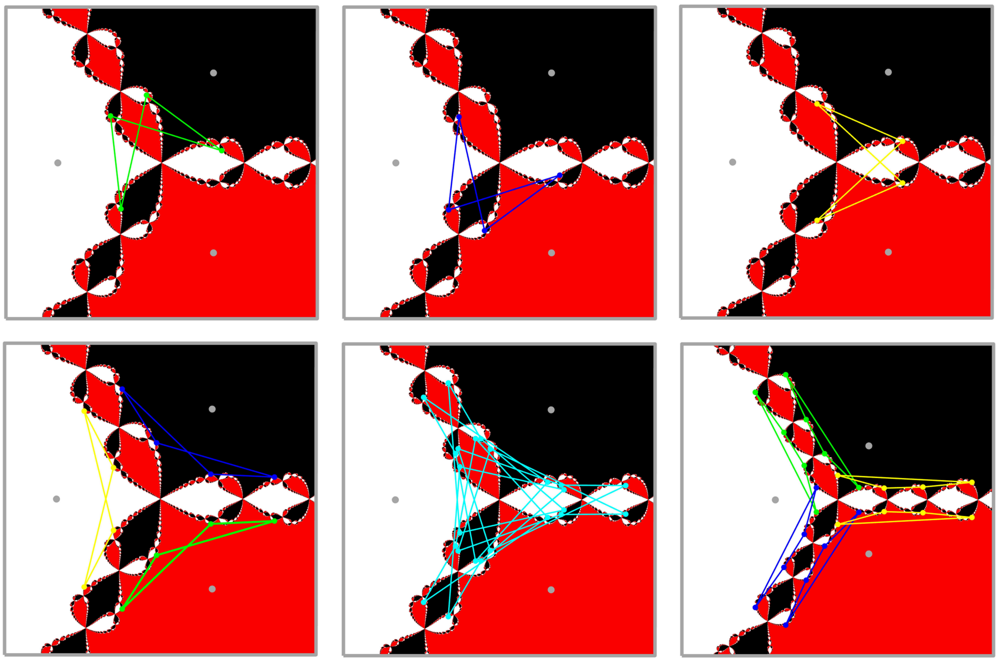

Figure 0.5: The 72 period-4 points comprise two groups of bat orbits (top row and bottom left), a central cobweb, and stretched quadrilaterals along the arms.

Beyond period-3 orbits

The barriers to finding period-N orbits with $N = 4, 5, 6, \ldots$ are fairly stark. Prior to discarding any unwanted solutions, the equation $z = F^{(N)}(z)$ leads to a degree-3^N polynomial and so the size of the task increases exponentially with N. For example, the case of $N = 4$ gives degree-81. After removing the three fixed points and six period-2 points, that 81 is reduced to 72. For Chalkdust readers' pleasure, we have persevered and solved the period-4 problem. The $72/4 = 18$ orbits are plotted in Fig. 0.5.

The situation is even worse for $N = 5$, where degree-243 reduces only slightly to 240 and there exist $240/5 = 48$ orbits to find and classify. One can reasonably expect to find all kinds of exotic solutions, but we are not brave enough to attempt that here.

As suggested by the terminology, the three fixed-point attractors are all stable in the sense that a trajectory following an $a, a, a, a, \ldots$ sequence is robust. Any small disturbance will knock the trajectory off course a little bit, but it will always tend back toward $a, a, a, a, \ldots$ as $n \to \infty$. Crucially, the same is not true of the periodic solutions—all of which are unstable. Reconsider our description of a period-3 orbit. For an initial condition such as $z_0 = d$, map F would lead to a sequence $d, e, f, d, e, f, \ldots$ that repeats indefinitely. But what about when $z_0 = d+\epsilon$, where ϵ is an arbitrarily small disturbance? In that case, F will produce a sequence that is *approximately* $d, e, f, d, e, f, ...$ in the short term but 'approximately' is not good enough. Tiny differences due to the ϵ seed grow rapidly under repeated iteration and the trajectory eventually wanders off in the complex plane, never to return.

The periodic orbits in F are said to be repellors since perturbed trajectories tend to move away from them. Such is their precariousness that they do not tend to survive for very long in numerical calculations, a fact due to the finite size (and, thus, precision) of computers. Hardcore mathematicians will have already clocked, for example, that periodic points tend to involve irrational numbers and so they cannot be reproduced as a finite decimal with perfect accuracy.

The Julia set of F

The final conceptual leap ties everything together. Take all of the periodic points of F, starting with the six period-2 points then onto the 24 period-3 points, 72 period-4 points, 240 period-5 points, ..., $515, 377, 520, 732, 011, 331, 036, 460, 411, 867, 633, 580, 846, 032, 026{,}400$ period-100 points, and so on. That uncountable collection of repelling periodic points is the Julia set of F, denoted by $J(F)$. Why is it so important? Because $J(F)$ *is* the boundary between the three basins of attraction (see Fig. 0.6).

Julia sets are some of the most fascinating objects in modern mathematics. They are nearly always fractal and have many mind-boggling properties but we mention only two. Take any point in $J(F)$ and imagine drawing a circle of radius ϵ around it. No matter how arbitrarily small we make ϵ, there are always uncountably-many other points of $J(F)$ within that circle. Also inside the same circle are points lying in all three basins of attraction. Any point in J is therefore on the boundary between *all three colours*. Put another way, no two colours can ever form a solid interface because the third always manages to squeeze inbetween. That is the essence of the Wada property, and it is by no means an intuitive concept to grasp.

Concluding remarks

We have hopefully left readers with a sense that fractals are endlessly fascinating (literally!) and they are enormous fun to play with, just for the pleasure of it. More seriously, they often appear at the border between many research problems in applied mathematics and theoretical physics. The key to unlocking the mystery of a Newton-Raphson fractal turns out to be its Julia set. But that is really just the start; so far as this article goes, we have

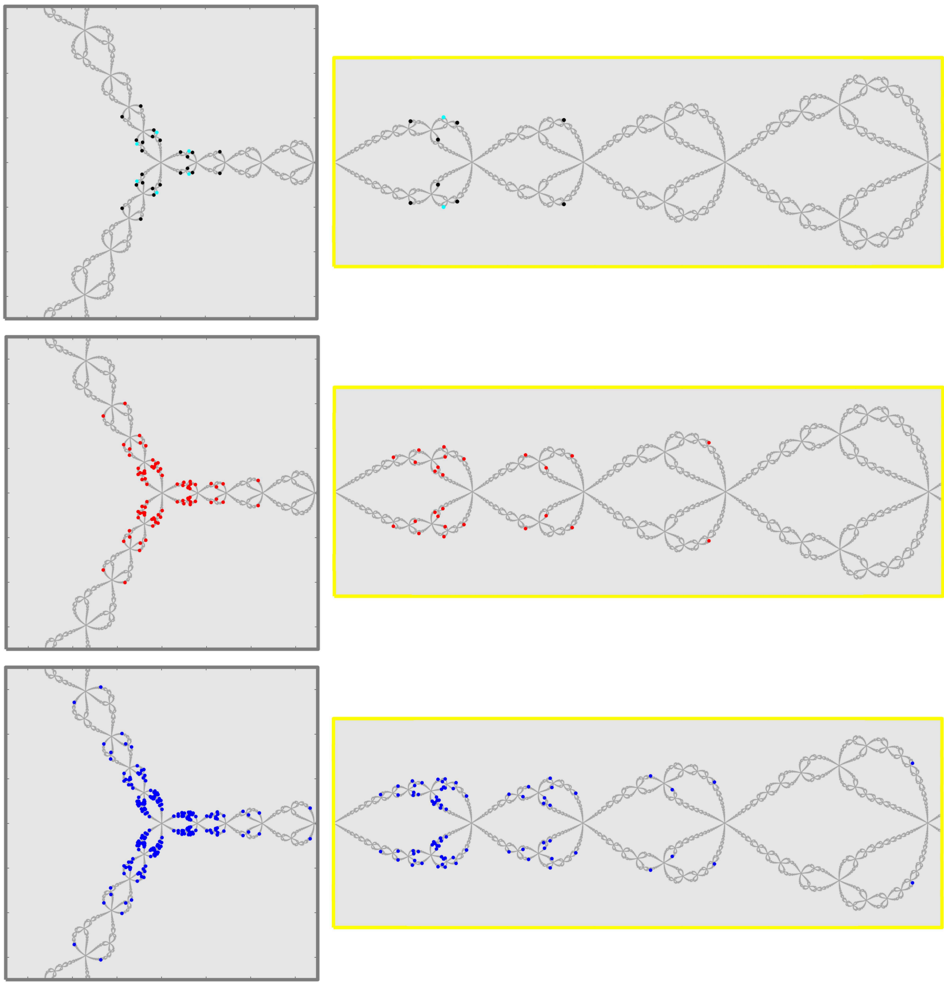

Figure 0.6: The Julia set (grey) showing the period-2 and period-3 points (cyan and black, respectively; top), period-4 points (middle), and period-5 points (bottom). The distribution of points is necessarily mirror-symmetric about a line down the centre of each arm.

probably raised more questions than we could possibly answer. Why are the fixed points attracting (or, more correctly, super-attracting)? Why are the periodic orbits repelling (some more so than others)? What happens when a trajectory lands on the origin (and which trajectories end up there in the first place)? What are the implications of fractal basin boundaries for predictability (and for physics more generally)? Why is all this stuff so utterly addictive? The list is almost as endless as the Julia set itself...