Visualization of Homotopy's and their Properties

Paul Vrbik : 250389673 : AM 9505 : Spring 2009

May 25, 2009

1 Preliminaries

To begin we define the notation and objects used throughout. Let $k[x_1, \ldots, x_n]$ denote the polynomials in variables $x_1, \ldots, x_n$ with coefficients in any arbitrary ring k. Interpret k^n as the set obtained by applying the cartesian product n times, namely $k^n = k \underbrace{\times \cdots \times}_{n \text{ times}} k$. Lastly, bold symbols, like $\boldsymbol{F}$ and $\boldsymbol{x}$, are used to indicate that an object is a vector.

2 Introduction

One of the nicest (in this author's opinion) mathematical algorithms is:

Definition 1 (n-dimensional Newton's Method)**.** Let $\boldsymbol{F} \in \mathbb{Q}[x_1, \ldots, x_n]^n$, $\boldsymbol{x}_0 \in \mathbb{C}^n$, and $\varepsilon > 0$ be given. If we iterate as

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} - \left[\mathrm{Jac}(\boldsymbol{F})|_{\boldsymbol{x}_{i-1}}\right]^{-1} F(\boldsymbol{x}_{i-1})$$

then we will eventually produce $\boldsymbol{x}_N$, $N \neq \infty$, such that

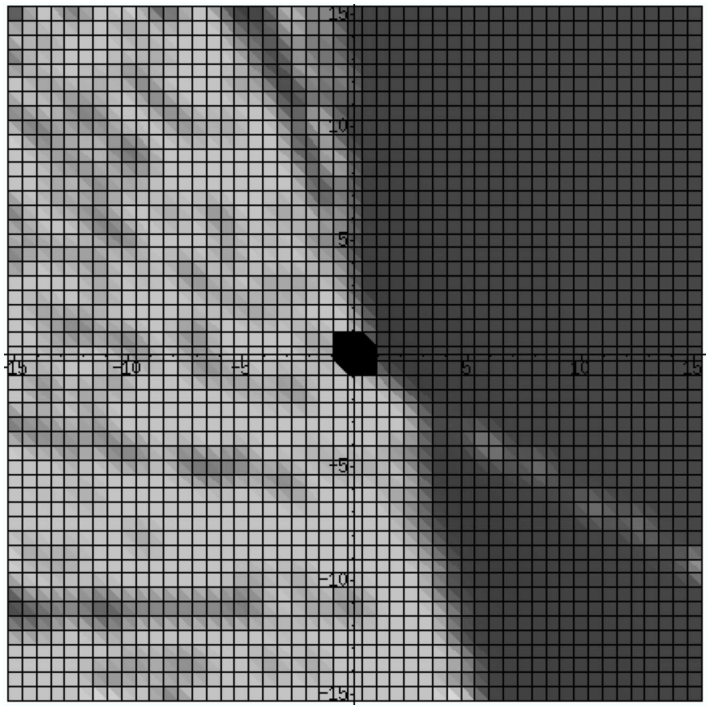
$$|\boldsymbol{F}(\boldsymbol{x}_N) - \boldsymbol{F}(\mathrm{RootOf}(F))| < \varepsilon.$$

This definition is far from being complete. For instance, we are assuming that the Jacobian is non-singular. Looking more deeply we see that in order to guarantee the existence of x_N with the desired property we must also assume we have infinite precision, which is fine theoretically but absurd from a computational standpoint.

However, despite these problems Newton's method is still very powerful. For one, at least superficially, it is a simple recursive formula which is easy to implement. If this wasn't already a good enough reason to use it, we also know that the method converges quadratically. This means for every iteration we expect to double the number of correct digits of the root we are building (a rare case of high pay off for little work).

But this isn't to say that Newton's method is without fault. The method is highly dependent on a good selection of the initial point $\boldsymbol{x}_0$. That is, we can observe highly varying behavior of Newton's method depending on the input (a generally undesirable feature). It is hard to see this in the one variable case as there is much less opportunity for the tangent to throw the algorithm off (in contrast to the multivariate case). This can only happen at singularities which are easily detectable for univariate polynomials. For the multivariate case we have the following example.

Example 1. Density plots show the values of a function of two variables at a regular array of points. Shading indicats the "height" of the surface. Below is a density plot measuring the speed of convergence for Newton's method given the system $\boldsymbol{F} = \left\langle x^2 - \frac{1}{2}y^2, 2x^2 + xy - 3x - 1 \right\rangle$.



In this case lighter regions correspond to those initial values that yield fast convergence. The solutions of $\boldsymbol{F}$ are (approximately) $\{(5.0, -6.7), (1.1, 1.6)\}$. Notice how easily one could pick an initial point in the "bad" darker region. The black region at the origin are points that seeded runs that exceeded fifty iterations (and as a result were cut off).

The natural question to ask now is : can we devise a method to generate better $\boldsymbol{x}_0$'s (initial guesses)? The answer to this question is of course "yes", accomplished by doing something called homotopy continuation. The development of this method claims a large portion of this paper. It is the author's hope that by highlighting the theory with detailed illustrations and examples that this method will become clear. The later part of this paper will be an exposé of the deficiencies of this method (after all no method is perfect), also exemplified through illustrations.

3 Homotopy Continuation

Homotopy continuation, like Newton's method, is an iterative approach for finding the isolated complex roots of a polynomial (or system of polynomials). But unlike Newton's method this process is much less sensitive to its input (because as we see this input will be controlled). First, to be somewhat rigorous, let $\boldsymbol{p}(\boldsymbol{z}), \boldsymbol{q}(\boldsymbol{z}) \in \mathbb{Q}[x_1, \ldots, x_n]^n$, with $\boldsymbol{z} \in \mathbb{C}^n$ (so $\boldsymbol{p}, \boldsymbol{q} : \mathbb{C}^n \to \mathbb{C}^n$). The underlying idea is to connect the solutions of an *exactly solvable* system $\boldsymbol{q}(\boldsymbol{z}) = \boldsymbol{0}$ with the solutions of the desired (target) system $\boldsymbol{p}(\boldsymbol{z}) = \boldsymbol{0}$ via a homotopy map:

$$\boldsymbol{H}(\boldsymbol{z}, t) = t\boldsymbol{q}(\boldsymbol{z}) + (1 - t)\boldsymbol{p}(\boldsymbol{z}),$$

where $t \in [0,1]$. By homotopy we mean that it is a continuous transformation from one function to another. We use the term 11homotopy continuation" in reference to how we use this homotopy to "track" (or follow) the roots of $\boldsymbol{H}(z,1) = \boldsymbol{q}(\boldsymbol{z})$ to the desired roots of $\boldsymbol{H}(\boldsymbol{z},0) = \boldsymbol{p}(\boldsymbol{z})$.

To make our goal explicit, we are searching for $\boldsymbol{z}_N$ such that $\boldsymbol{H}(\boldsymbol{z}_N, 0) = \boldsymbol{q}(\boldsymbol{z}) = 0$ starting with the initial data $\boldsymbol{H}(\boldsymbol{z}_0, 1) = \boldsymbol{p}(\boldsymbol{z}) = 0$. For now assume that $\boldsymbol{p}(\boldsymbol{z})$ is exactly solvable (for those impatient readers $\boldsymbol{p}(\boldsymbol{z})$ can easily be something like $\boldsymbol{p}(\boldsymbol{z}) = \boldsymbol{z}^n - 1$, but this is not ideal). Consider:

Definition 2. The first order taylor expansion of $\boldsymbol{H} = \langle H_1, \ldots, H_n \rangle$ is

$$\boldsymbol{H}(\boldsymbol{z} + \Delta\boldsymbol{z}, t + \Delta t) = \boldsymbol{H}(\boldsymbol{z}, t) + \boldsymbol{H}_{\boldsymbol{z}}(\boldsymbol{z}, t)\Delta\boldsymbol{z} + \boldsymbol{H}_t(\boldsymbol{z}, t)\Delta t \tag{1}$$

where $\boldsymbol{H}_z = \partial H/\partial \boldsymbol{z}$ is the $n \times n$ Jacobian and $\boldsymbol{H}_t = \partial H/\partial t$ is the size $n \times 1$ gradient, (note that Δt is negative as we are going from $1 \to 0$).

From this we see that we can get a point $(\boldsymbol{z}_1, t_1) = (\boldsymbol{z}_0 + \Delta z, 1 + \Delta t)$ such that $\boldsymbol{H}(\boldsymbol{z}_1, t_1) \approx 0$. Remember, $\boldsymbol{H}(z_0, 1) = 0$ by our design and an appropriate value of Δt is chosen by us (more on this later). Solving for $\Delta\boldsymbol{z}$ in (1) we get

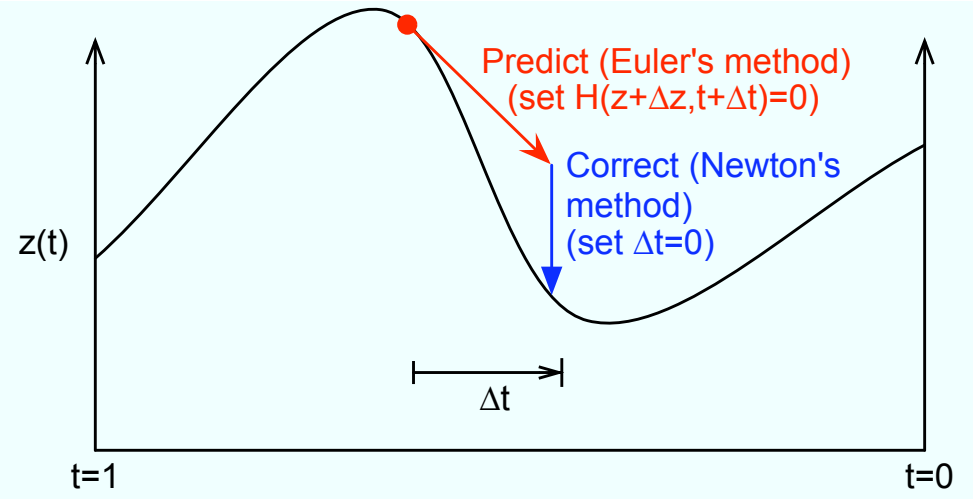
$$\Delta\boldsymbol{z} = -\boldsymbol{H}_{\boldsymbol{z}}^{-1}(\boldsymbol{z}_1, t_1)\boldsymbol{H}_t(\boldsymbol{z}_1, t_1)\Delta t,$$

which we recognize to be Euler's method. We call this the "prediction" step.

The next step, called "correction", is necessary because Euler's method is not that great. That is, it will likely be the case that $\boldsymbol{H}(\boldsymbol{z}_1, t_1)$ is not as close to zero as we would like. We can fix this by using Newton's method to refine it. Fortunately we have a good starting point as $(\boldsymbol{z}_1, t_1)$ is reasonably close to the curve. This means Newton's method is likely to converge quickly (in fact we will classify anything that takes longer than three iteration as a failure). We may see this as setting $\Delta t = 0$ in (1) to get:

$$\boldsymbol{z}_1 = -\boldsymbol{H}_{\boldsymbol{z}}^{-1}(\boldsymbol{z}_0 + \Delta\boldsymbol{z}, t_1)\boldsymbol{H}(\boldsymbol{z}_0, t_1).$$

Example 2. Below we illustrate how given $(\boldsymbol{z}_1, t_1)$ such that $H(\boldsymbol{z}_1, t_1) \approx 0$ one can predict a new approximate solution at $t_1 + \Delta t$.



We now have $(\boldsymbol{z}_1, t_1)$ such that $\boldsymbol{H}(\boldsymbol{z}_1, t_1) \approx 0$. Repeating what we did above we can get another point $(\boldsymbol{z}_2, t_2)$ with the same property. As $1 = t_0 > t_1 > t_2 > \cdots > 0$ what this process eventually leaves us with is $(\boldsymbol{z}_{N-1}, t_{n-1})$ with $t_N \approx 0$. As a final step we use Newton's method, seeded by $(\boldsymbol{z}_{N-1}, 0)$ to calculate $\boldsymbol{z}_N$ such that $\boldsymbol{H}(\boldsymbol{z}_N, 0) = \boldsymbol{q}(\boldsymbol{z}_N) \approx 0$. This is one of the desired solutions to our original system.

To conclude this section we note an issue surrounding the implementation of this idea. Initially you may choose Δt to be some small value like 1/100. Generally as Δt become larger the new points generated by Euler's method will become much worse approximations. This may mislead one to choose a Δt as small as possible but recall that we are trying to reach $t = 0$. The smaller our step length the longer it will take for our algorithm to run. To strike a balance between these two ideas a good strategy is to cut Δt in half if Newton's method fails to converge in three iterations and double Δt after m successful corrections. A choice of m in the range of two to five works well.

4 Path Tracking In Action

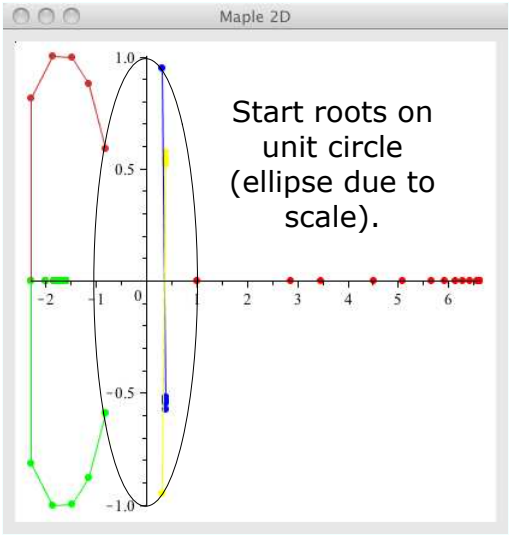
We have so far only discussed what it means to follow a single root from $\boldsymbol{p}(z)$ to a root of $\boldsymbol{q}(z)$. In fact what we would like to do is follow enough paths to recover all isolated roots of $\boldsymbol{q}(z)$. We do this by to chosing $\boldsymbol{p}(z)$ so that it has at least as many roots as $\boldsymbol{q}(z)$ (to prove that each root of $\boldsymbol{q}(z)$ is associated with at least one path is non-trivial so I refer to Sommese's and Wampler's book). The simplest start system we can construct is $\boldsymbol{p}(z) = \boldsymbol{z}^d - 1$ where d is the product of the largest degree for each variable over all polynomials in the target system. That is, for a target system $\langle f_1, \ldots, f_n \rangle$ with $f_i \in \mathbb{Q}[x_1, \ldots, x_n]$

$$d = \prod_{i=1}^{n} \max(\deg_{x_i}(f_1), \ldots, \deg_{x_i}(f_n)).$$

Example 3. For $F = \left\langle x^2 - \frac{1}{2}y^2, 2x^2 + xy - 3x - 1 \right\rangle$, $d = 2 \times 2$ and we would choose the start system $\boldsymbol{p}(\boldsymbol{z}) = \boldsymbol{z}^4 - 1$.

Now that we have multiple paths we may now begin to see things go wrong. It is possible that a tracker will actually jump paths and converge to the wrong root.

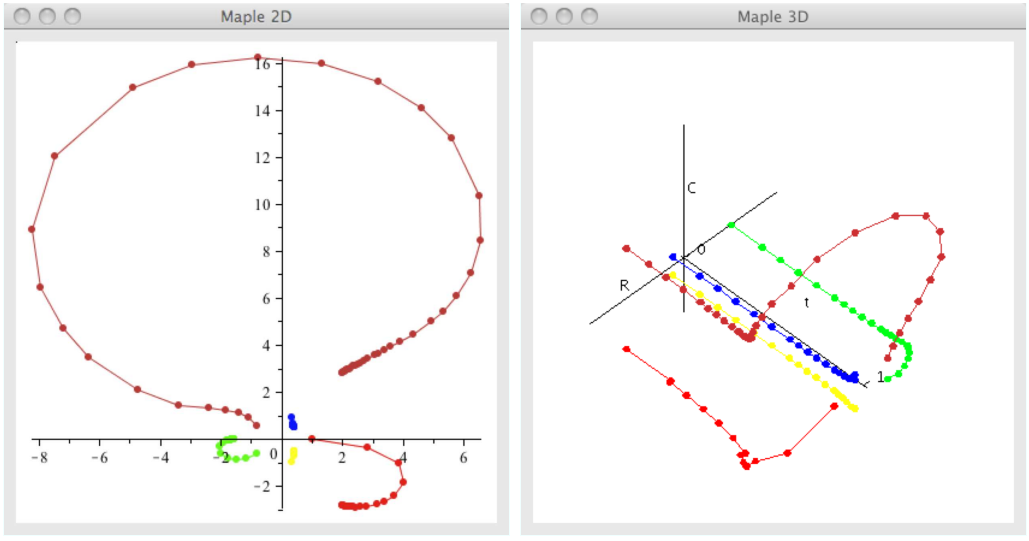
Example 4. An example where tracking fails. Below are the paths for solving the system $-7x^5 + 22x^4 - 55x^3 - 94x^2 + 87x - 56$. The brown and green paths converge to the real root -1.6 whereas the blue and yellow paths converge to the complex root 0.4-0.5i. The red root is escaping to infinity and is (quickly) flagged as a failed path. We do not find all roots.



To fix this we introduce the random co-ordinates $\theta, \phi \in [-\pi, \pi]$ and modify our original homotopy to use $\boldsymbol{q}(\boldsymbol{z}) = te^{i\theta}(\boldsymbol{z}^d - e^{i\phi})$. To explain why this helps consider walking from the middle of a circle to its boundary. If many people are walking in this circle it is likely that some paths will intersect with one another. The variable θ can be thought of as the cardinal direction one must start walking in, this will avoid some intersections but not all. The addition of ϕ allows us to place our roots along separate great circles of the sphere S^2 given by the co-ordinates (θ, ϕ). The paths now travel through the interior of the sphere and can only collide within a set of measure 0 (i.e. with probability one).

Now we can observe the paths are (more or less) well behaved.

Example 5. In the diagrams below the paths start from the unit circle. The diagram on the left can be interpreted as looking at the diagram on the right (which has the added dimension t) head on down the t axis.



Using $\boldsymbol{q}(\boldsymbol{z}) = te^{i\theta}(\boldsymbol{z}^d - e^{i\phi})$ as we defined above is what we call a total degree homotopy. It is not optimal because it usually the case that a system equations has far fewer roots than it's

theoretical maximum (given by d). This means a lot of paths will either converge to the same root or just diverge. In both cases we are wasting computation. A smarter heuristic is to estimate the number of roots of the target system using mixed volume computations. The mixed volume is a much more strict bound then the total degree. For information about this and to find some smarter start systems I refer to the Thesis of T-Y Li.

5 Basins of Attraction

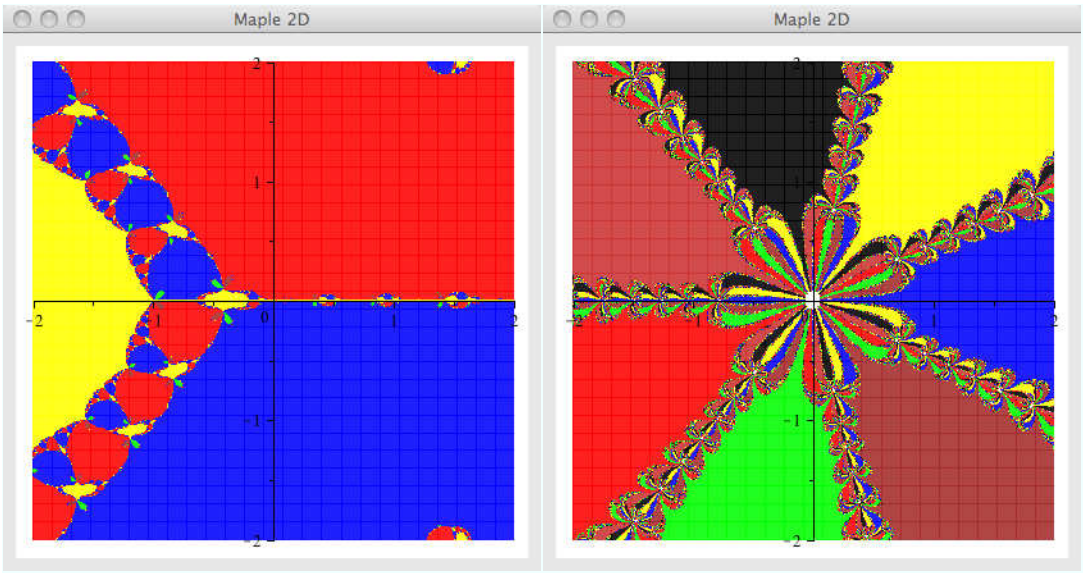
Another unfortunate deficiency of Newton's method is that Newton's method will converge to different roots depending on what initial value it is seeded. Thereby at the correction step, Newton's method can throw the path off course by converging to the wrong root. Unfortunately it is hard (nearly impossible) to predict if this is going to happen. We begin by defining:

Definition 3 (Basin of Attraction)**.** For some polynomial $f(x) \in \mathbb{Q}[x]$ suppose that we use Newton's method to find the roots $\zeta_1, \ldots, \zeta_n \in \mathbb{C}$ of f. We say

$$\text{Basin of attraction of } \zeta_i = \{\text{initial points yielding the root}\}.$$

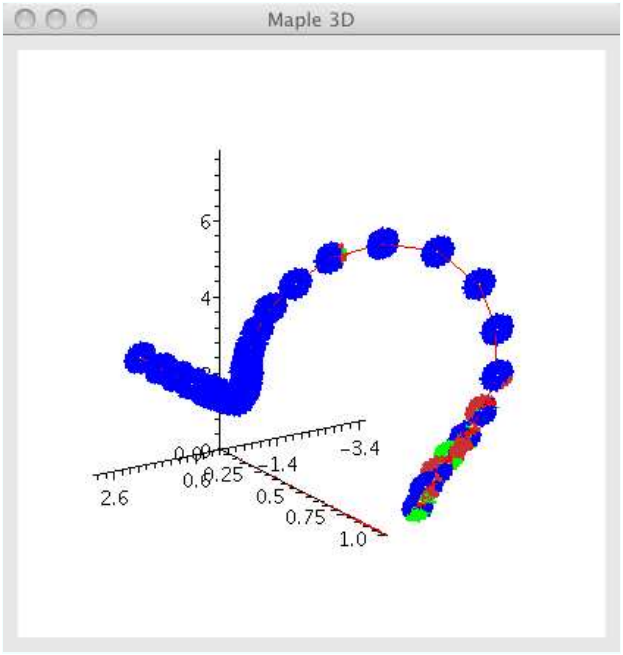
We can visualize basins by coloring points in the complex plane by which basis they lie in.

Example 6. The basin of attraction for $-7x^5 + 22x^4 - 55x^3 - 94x^2 + 87x - 56$ and $x^7 - 1$. In both cases we see that the boundaries are fractal in nature, and therefore hard to study.



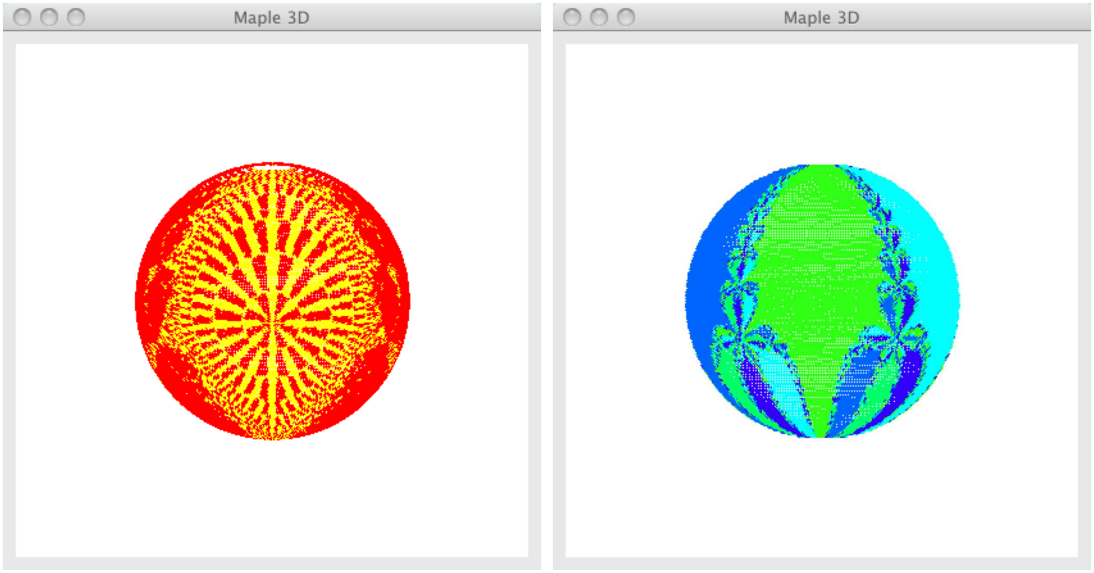
However our interest is in seeing how these basins effect our algorithm at each time step t_i. For this we have what we call a "tube of attraction". Which we illustrate in the following example.

Example 7. We take the brown path from Example 5 and plot a small disc basin for each time step. Notice how the path stabilized when the basin is dominated by one root. This is interesting as it is in conflict with the literature which says that it is the "end game" or final stage of the algorithm which is the most unstable. However, it is clear from the picture below that our path must deal with being in basins which are highly irregular at the *beginning*.



To visualize the basin of attraction for the *entire* complex plane we can use a stereographic projection to plot on a Riemann sphere. This gives us some global insight of how the basins act.

Example 8. From left to right is the basin of attraction for $x^3 - 1$ (from the bottom of the sphere) and $x^5 + 1$ (from the side).



Fractals are very interesting as they are quite beautiful to look at, but very problematic to study mathematically. Some future work in this area include:

1. characterizing the boundaries of these basins.
2. finding the largest basins.
3. boundary avoidance algorithms.

6 Diverging Paths

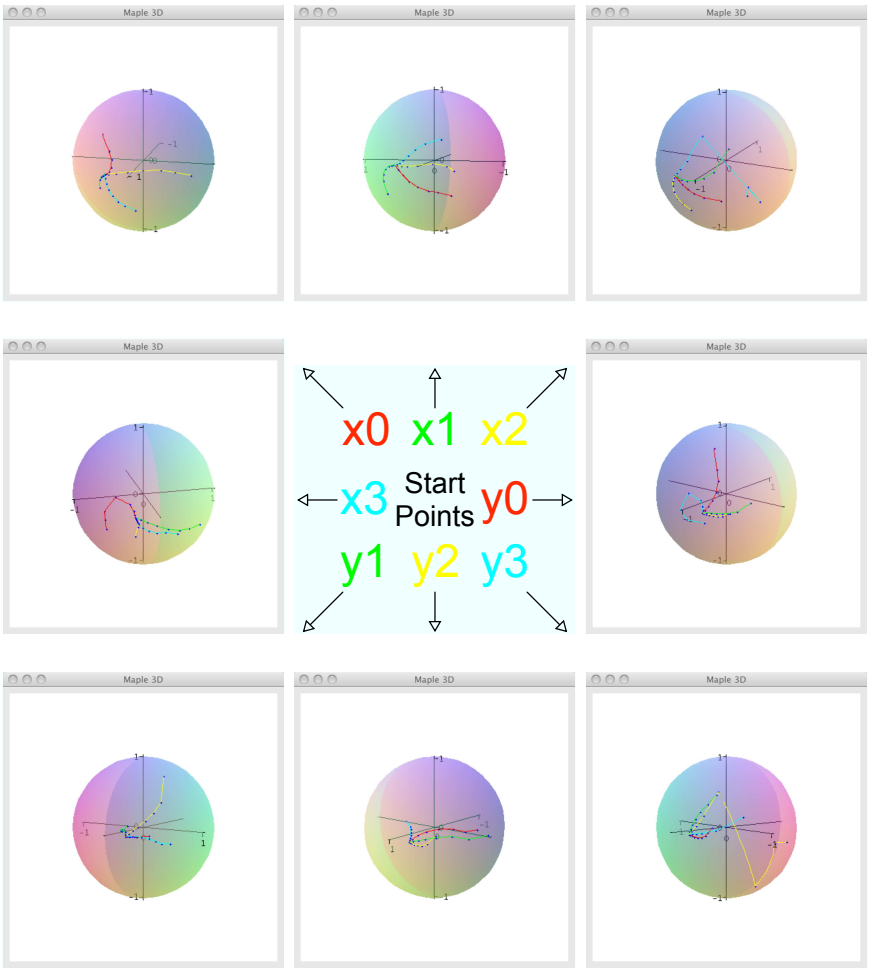
A phenomena which we do not observe in the univariate case is diverging paths. These paths are interesting because we can never be completely sure as to when a path is diverging. It is always possible that a path heading towards infinity can turn around and converge to a root. However, as we can not afford to hold on to explosively large values, or run an algorithm for an infinite amount of time, such paths must be cut off. This is not ideal, but it is the nature of the field to have to make such compromises.

Consider the target system:

$$p(z) = \begin{bmatrix} x^3y + xy^2 + 1 \\ x^4 + xy^2 + 1 \end{bmatrix}$$

Now $\frac{\partial H}{\partial z}$ (the Jacobian) can become singular and our predictor will point to infinity. To visualize this we again use stereographic projection to plot the paths on a Riemann sphere so we may see paths converge to infinity (the north pole).

The target system $p(z)$ may have up to 16 roots so we must track 16 paths. All possible pairs (x_i, y_j) with $0 \leq i, j \leq 3$ constitute our 16 start points. Each sphere represents a path that a single component of the solution (x or y) takes. We will observe that half the paths diverge.



7 Conclusion

Homotopy continuation is an effective way to determine the solutions to systems of polynomials equations. It often works in many situations where there are no other numerical methods that will work. It has made its mark, through software packages like Bertini, as being one of the fastest and most reliable methods, and is a big contender to the triangular decomposition and Gröbner basis methods.

Yet there is still much to be gained by studying the nature of this method through visualization. Illustrations can often lead to insights that are obfuscated by mathematical expressions. In addition to this visualization is sometimes the most effective way to explain a concept.