On the Convergence of Newton's Method

Joel Friedman
University of California, Berkeley

0. Abstract

Let P_d be the set of polynomials over the complex numbers of degree d with all its roots in the unit ball. For $f \in P_d$, let Γ_f be the set of points for which Newton's method converges to a root, and let $A_f \equiv |\Gamma_f \cap B_2(0)|/|B_2(0)|$, i.e. the density of Γ_f in the ball of radius 2 (where $|\ |$ denotes Lebesgue measure on $\mathbf{C}$ viewed as $\mathbf{R}^2$). For each d we consider A_d, the worst-case density (i.e. infimum) of A_f for $f \in P_d$. In [S], S. Smale conjectured that $A_d > 0$ for all $d \geq 3$ (it was well-known that $A_1 = A_2 = 1$). In this paper we prove that

$$\left(\frac{1}{d}\right)^{cd^2 \log d} \leq A_d$$

for some constant c. In particular, $A_d > 0$ for all d.

Remark: Our definition of A_d differs slightly from that in [S], but the conclusions hold for A_d as defined in [S] as well.

1. Introduction

Newton's method is a method for finding the zeros of a function, f. One starts with an initial guess, z_0, of a zero of f, and then generates a sequence of successive guesses according to the rule

$$z_{i+1} \quad \longleftarrow \quad z_i - \frac{f(z_i)}{f'(z_i)}. \tag{1.1}$$

Intuitively, using the value of f and f', we locate the unique zero of the tangent line to f at z_i; this is the next guess, z_{i+1}.

It is known that for any differentiable function, f, and any root, r, of f, the sequence $\{z_i\}$ converges to r if our first guess, z_0, is close enough to r. We define the *basin of* r to be the set of points in $\mathbf{C}$ for which Newton's method converges to r.

There are various forms of Newton's method. We will assume that f is a polynomial

$$f(z) = a_0 z^n + \cdots + a_n, \qquad a_i \in \mathbf{C}, \tag{1.2}$$

so that f and f' are easy to compute. We will view equation (1.1) as a map

$$T_f(z) = z - \frac{f(z)}{f'(z)},$$

from $\mathbf{C} \cup \{\infty\}$ to itself. In this paper we will study the set Γ_f, defined to be the set of points, z_0, for which Newton's method works, i.e. $\{z_i\}$ converges to a root of f.

In this paper we will use a geometric interpretation of Newton's method, involving a relation that goes back at least 100 years–

Theorem 1.1 (Lucas, 1874)**:** *Let f(z) be a polynomial with coefficients in* $\mathbf{C}$*. Then the zeros of* f' *lie in the convex hull of the roots of f.*

Proof: For a set $S \subset \mathbf{C}$, we denote its closed convex hull by $\langle S \rangle$. It is easy to check that

$$f'(z) = f(z) \sum_{i=1}^{d} \frac{1}{z - r_i}, \tag{1.3}$$

where $r_1, \ldots, r_n$ are the roots of f. Assume that f' has a root, z, outside $\langle r_j \rangle$, the convex hull of the roots. Then the vectors from z to the r_i's all lie in one side of a half-plane through z. Then the vectors $z - r_i$ all lie to one side of a half-plane. Hence the $\frac{1}{z-r_i}$'s lie to one side of a half-plane (not the same half-plane, rather the one you get by reflecting through the x-axis). Hence the sum of the $\frac{1}{z-r_i}$'s cannot vanish, which is a contradiction.

∎

Given a polynomial, f, the Newton map for f,

$$T_f(z) = z - \frac{f(z)}{f'(z)}, \tag{1.4}$$

has a geometric interpretation in terms of f's roots. For a polynomial $f(z) = (z-r_1)\ldots(z-r_d)$, we can write T_f as

$$z + \frac{1}{\sum_{i=1}^{d} \frac{1}{r_i - z}}. \tag{1.5}$$

For example, if $z = 0$,

$$T_f(0) = \frac{1}{\sum_{i=1}^{d} \frac{1}{r_i}}. \tag{1.6}$$

This is sometimes called the *harmonic sum* of the r_i's. Looking at equation (1.5) we see that $T_f(z)$ looks like $T_f(0)$ with z as the origin. From equation (1.5) we can also see that if we change z and the roots of f, by a translation, rotation, and/or dialation, (i.e. a linear map $az + b$, with $a, b \in \mathbf{C}$), then the new $T_f(z)$ is just the old $T_f(z)$ transformed in the same way.

We will give an example of the geometric point of view. By a *wedge*, W, we mean a subset of $\mathbf{C}$ of the form

$$W = \{z : \theta_1 \leq \arg(z - r) \leq \theta_2\}, \tag{1.7}$$

where $\arg(z)$ is the angle z makes with the positive x-axis; r is called the *vertex* of the wedge, and $\theta_2 - \theta_1$ its *angle*. A wedge is convex iff its angle is $\leq \pi$.

Theorem 1.2: *If W is a convex wedge at z containing the $\{r_i\}$'s, then $T_f(z)$ is contained in W.*

Proof: The $r_i - z$'s lie in a wedge $V = W - z$ about the origin; by $W - z$ we mean the points $\{w - z : w \in W\}$. Thus the $\frac{1}{r_i - z}$'s lie in the wedge V', the reflection of V through the x-axis. Since V' is a convex wedge, then it also contains their sum $\sum \frac{1}{r_i - z}$. Its reciprocal is contained in the original wedge, V, and hence T_f is contained in $W = V + z$ $(= \{v + z : v \in V\})$.

∎

Returning to the study of Γ_f, we would like to know how likely a randomly chosen point in $\mathbf{C}$ will lie in Γ_f.

Let P_d denote the polynomials of degree d with roots in the unit ball. For a function f, let A_f denote the density of Γ_f in the ball or radius 2, i.e.

$$A_f \equiv \frac{|\Gamma_f \cap B_2(0)|}{|B_2(0)|}, \tag{1.8}$$

where $|C|$ denotes the area of C. For a positive integer d, let

$$A_d \equiv \inf\{A_f : f \in P_d\}. \tag{1.9}$$

In this paper we will prove that $A_d > 0$ for all d.

There are several reasons why we study A_d:

(1) *It is easy to ensure $f \in P_d$.*

Fact: *If $f(z) = a_0 z^d + \cdots$ and $|a_0| > \sum_{i=1}^{d} |a_i|$, then $f \in P_d$.*

Proof: If $|z| > 1$, then $a_0 z^d$ is larger in absolute value than the sum of all the lower order terms of f. Hence if $|z| > 1$, then z cannot be a root of f.

(2) *By rescaling we can assume that (1) holds.* Since

$$f(zc) = (a_0 c^d) z^d + (a_1 c^{d-1}) z^{d-1} + \cdots, \tag{1.10}$$

the $a_0 c^d$ will dominate if we take c large enough.

(3) *We need some restriction on f to prove a density theorem.* It is well known that for any $d \geq 3$ there is a polynomial f for which Newton's method does not converge on some open set in $\mathbf{C}$ (see, for example, [S]). It follows that for any bounded set we can find a polynomial of degree d for which Newton's method does not converge anywhere on this bounded set, simply by taking an appropriate translation and dilation of f.

The main theorem of the paper is

Theorem 1.4: *$A_d > 0$ for $d \geq 3$. More precisely, there is a constant c_1 such that*

$$\left(\frac{1}{d}\right)^{c_1 d^2 \log d} \leq A_d. \tag{1.11}$$

We caution the reader to note that A_f is not continuous in f's coefficients or its roots. For example, for $f(z) = z^3$ Newton's method works for any initial guess. Yet one can show that there exists a constant $c < 1$ such that there exist polynomials g, arbitrary close to f, with $A_g \leq c$.

We will introduce some useful notations for the rest of the paper.

Let E be a subset of $\mathbf{C}$. By sE, $s \in \mathbf{R}$, we mean the set E dilated by s, i.e.

$$sE \equiv \{s \cdot e : e \in E\}.$$

By sE_z we mean sE with z viewed as the origin, i.e.

$$sE_z \equiv \{z + s(e - z) : e \in E\}.$$

In particular $sE_0 = sE$.

Given a convex polygon, P, and a vertex, v, of the polygon, the *interior angle* at v is the angle determined by the two line segments of the boundary of P meeting v; the *exterior angle* at v it the angle opposite the interior angle (see figure 1).

Figure 1

The *exterior wedge* at v is the wedge bounded by the exterior angle at v. If P is a degenerate polygon, that is to say a line segment, then the exterior wedge is the ray from v to ∞ which is colinear and opposite the line segment.

In this paper we use many different constants. Rather that give each one a different name, we will denote them all by c (unless some confusion will occur).

In §2 we describe some regions in which Newton's method converges. In §3 we estimate the area of one of these regions, yielding the lower bound on A_d. Some of the calculations used in these sections are postponed until appendix A.

We remark that as we send this paper for publication, A. Manning proved that Γ_f has density $1/(d^2 \log^2 d)$ in the ball of radius d, see [M]; one can use this to improve our density bound to $1/d^{cd \log d}$. Also, an earlier version of this paper, [F], appeared in a conference.

The author would like to thank Steve Smale for many helpful discussions.

2. Invariant Curves

For background, let us begin with the question of how fast does Newton's method converge. One can do both local and assymptotic analysis as follows—

Let f be a polynomial of degree d, and let r be a root of f of multiplicity k. An easy calculation shows

$$T_f'(z) = \frac{f(z)f''(z)}{(f'(z))^2}. \tag{2.1}$$

It follows that at r we have

$$T_f'(r) = 1 - \frac{1}{k}. \tag{2.2}$$

Hence if r is a simple root of f we get $T_f'(r) = 0$ so that

$$T_f(r+\epsilon) = r + a\epsilon^2 + \cdots. \tag{2.3}$$

In other words, Newton's method is *quadratically convergent.* For a root of multiplicity > 1, Newton's method is *linearly convergent.* That is

$$T_f(r+\epsilon) = r + \left(1 - \frac{1}{k}\right)\epsilon + \ldots, \tag{2.4}$$

and so

$$T^n(r+\epsilon) = r + \left(1 - \frac{1}{k}\right)^n \epsilon + \cdots. \tag{2.5}$$

Assymptotically, for $|z| \gg r$'s, we have

$$\begin{aligned} T_f(z) &= z + \frac{1}{\sum \frac{1}{r_i - z}} \\ &\approx z + \frac{1}{\sum \frac{1}{-z}} \\ &= z + \frac{1}{-d/z} = z\left(1 - \frac{1}{d}\right). \end{aligned} \tag{2.6}$$

Notice the similarity in the right-hand side of (2.4) with that of (2.6); geometrically a d-tuple root and $|z| \gg r$'s look similar.

We would like to know what Newton's method looks like, not only very near the root or very far away, but also in between.

Theorem 2.1: *Let l be a line separating 0 from the roots, $\{r_i\}$, of f. Then $T_f(0)$ lies on the side opposite from 0 of the line $l' = \frac{1}{d}l$, i.e the line parallel to l and d times closer to 0.*

Figure 2 *For theorem 2.1*

Proof: (See figure 2). We recall that the map $f : \mathbf{C} \to \mathbf{C}$ given by $f(z) = \frac{1}{z}$ maps circles and lines to circles and lines (not respectively). If the roots, $\{r_i\}$, lie in the half-plane H, then $\{\frac{1}{r_i}\}$'s lie in a ball, B, through 0. Then, since B is convex, the sum of the $\{\frac{1}{r_i}\}$'s will lie in the ball dB; and its reciprocal hence lies in the half-place $\frac{1}{d}H$, whose boundary is $\frac{1}{d}l$.

■

We will give another instance where the geometry of Newton's method makes things a bit simpler. Before we saw that for a root, r, of the polynomial f,

$$|T_f(r)| = r, \qquad |T_f'(r)| < 1, \tag{2.7}$$

and concluded that near r, Newton's method converges to r; i.e. $B_\delta(r) \subset \{\text{basin of r}\}$ for some $\delta > 0$. We now want an estimate for δ.

Theorem 2.2: *Let $\eta = \min_{r_i \neq r} |r_i - r|$, and $\delta = \frac{\eta}{2d}$. Then $B_\delta(r) \subset \{\text{basin of r}\}$. Furthermore, $z \in B_\delta(r)$ implies $|T_f(z) - r| \leq (1-\epsilon)|z - r|$ for some $\epsilon > 0$ depending only on d.*

Proof: An easy calculation— see appendix A.

By an *invariant curve* we mean a curve $\phi\colon \mathbf{R} \to \mathbf{C}$ such that

$$T_f(\phi(t)) = \phi(t+1) \tag{2.8}$$

for all $t \in \mathbf{R}$. Note that if $\phi(t)$ lies in the basin of r, for some root r, for $t \in [0,1]$, then so would $\phi(t)$ for any $t < 0$ and any $t > 0$. We will prove that for each vertex, r, of $\langle r_j \rangle$, there will be an invariant curve from r to ∞ in the exterior wedge of r (and an open set about the curve) which lies in the basin of r (see figure 3).

Figure 3 *Some parts of the basins*

Lemma 2.3: *If z lies on the bisector of the exterior angle at r, then $T_f(z)$ lies in the exterior wedge at r.*

Proof: In appendix A.

Let r be a vertex of $\langle r_j \rangle$, and let z be a point within $B_\delta(r)$, δ as in theorem 2.2, and lying on the bisector of the exterior angle at r. Then $T_f(z)$ lies in $B_\delta(r)$ and in the exterior wedge at r. We construct an invariant curve $\phi: \mathbf{R} \to \mathbf{C}$ by defining $\phi(0) = z$, $\phi(1) = T_f(z)$, and

$$\text{for } 0 \le t \le 1 \qquad \phi(t) \equiv (1-t)z + (t)T_f(z). \tag{2.9}$$

Since $B_\delta(r)$ and the exterior wedge at r are convex, it follows that $\phi(t) \in B_\delta(r)$ and lies in the exterior wedge for all $t \in [0, 1]$.

(The reader may notice that the definition in equation (2.9) is rather arbitrary— all we need is that $\phi(t)$, for $t \in [0, 1]$, lies in the exterior wedge and $B_\delta(r)$.)

One can use equation (2.8) to extend the curve ϕ to all of $[0, \infty]$, but the important step is to extend ϕ backwards to $-\infty$.

We can always extend the curve backwards for a short time (uniquely) as long as $T_f^{-1} \ne \infty$ (or 0). By Lucas' theorem we have f, f', and f'' have all their roots in $\langle r_j \rangle$. Since

$$T_f'(z) = \frac{f(z)f''(z)}{(f'(z))^2}, \tag{2.10}$$

it follows that we can extend ϕ backwards at least as long as ϕ remains outside $\langle r_j \rangle$.

But on the other hand, when we extend ϕ backwards, we will never leave the exterior wedge at r— if ϕ left the wedge it would do so at some initial time, T_0. But it is easy to see that any point on the boudary of the exterior wedge is mapped to a point outside the exterior wedge (assuming the wedge is non-degenerate; see figure 4).

Hence we can extend ϕ backwards to all of $\mathbf{R}$.

Remark: In the degenerate case of all roots lying on one line segment, ϕ is just the continuation of the line segment.

Figure 4

We want to prove that as $t \to -\infty$, $\phi(t) \to \infty$. For this we need the following theorem—

Theorem 2.4: *Let the angle at r be α. Then*

$$\frac{|\phi(t-n)-r|}{|\phi(t)-r|} \geq \left(\frac{d}{d-1}\right)^n \left\{ \begin{array}{rl} \sin\alpha, & \text{if } \alpha > \pi/2 \\ 1, & \text{if } \alpha \leq \pi/2 \end{array} \right\}.$$

Figure 5

Proof: Let t be fixed and let $z = \phi(t-n)$. Let l and l' be the lines depicted in figure 5, through z and parallel to the interior angle boundaries. By theorem (2.1), $T_f(z) = \phi(t-n+1)$ and z must lie on opposite sides of both $l_r(1-\frac{1}{d})$ and $l'_r(1-\frac{1}{d})$. Applying T recursively it follows that $\phi(t)$ is contained in the diamond pictured above, bounded between $(1-\frac{1}{d})^n$ times l_r and l'_r and the exterior angle at r. It remains to prove the following—

Lemma 2.5: *Let $ABCD$ be a parallelogram. Let $\angle C = \alpha$, and let E be any point in*

ABCD. Then

$$|AC| \geq |EC| \begin{Bmatrix} 1 & \text{if} \quad \alpha \leq \dfrac{\pi}{2} \\ \sin\alpha & \text{if} \quad \alpha > \dfrac{\pi}{2} \end{Bmatrix}.$$

Proof: See appendix A.

∎

Corollary 2.6: *$\phi(t) \to \infty$ as $t \to -\infty$.*

At this point let us stop for a few remarks—

1. To estimate A_d we only need a sequence of points which are successive backward iterates of T_f and which are reasonably well-behaved. The invariant curves are a technical conveniece which gives us a lot of such sequences.
2. At this point we start to see some of the picture. We claim that the invariant curve is surrounded by an open region of points in the basin of r; more generally, Γ_f is an open set. The reason is that if z lies in the basin of r, then $T_f^n(z)$ lies in the interior of $B_\delta(r)$ for some n. Then $T_f^{-n} \equiv (T_f^{-1})^n$, defined locally at z, gives us a neighborhood about z which lies in the basin of r.

3. Estimating the Area of the Basin Near the Invariant Curves

For each invariant curve constructed in §2, there is an open set containing it and lying in the basin of a root. We want to estimate the area of these open sets to get a lower bound on the density of Γ_f. To do this, we will fix a vertex, r, of $\langle r_j \rangle$ whoose interior angle is $\leq \pi(1-\frac{2}{d})$. Such an r always exists since the average angle of an m-gon is $\pi(1-\frac{2}{m})$. Let ϕ be the invariant curve described in §2 determined by having $\phi(0)$ being the point on the bisector of the exterior angle satisfying $|\phi(0)-r| = \delta/2$. For a fixed r and ϕ we define for $z \in$ Image ϕ,

$$\rho(z) \equiv \sup\{\rho : B_\rho(z) \subset (\text{ basin } r)\}.$$

It is convenient to define

$$\theta(z) \equiv \frac{\rho(z)}{|z-r|}. \tag{3.1}$$

In this section we will prove—

Theorem 3.1:

$$\theta(z) > \left(\frac{1}{d}\right)^{cd^2 \log d}.$$

Corollary 3.2:

$$A_f > \pi\left(\frac{1}{2}\left(\frac{1}{d}\right)^{cd^2 \log d}\right)^2 \Big/ |B_2(0)| = \left(\frac{1}{d}\right)^{c'd^2 \log d}.$$

First we will give some intuition for why theorem 3.1 should be true. Far away, when $|z| \gg |r|$, $T_f(z) = z(1-\frac{1}{d}) +$ lower order terms. It follows that

$$\frac{|T_f(z)-r|}{|z-r|}, \quad \frac{\rho(T_f(z))}{\rho(z)} \text{ are both } 1-\frac{1}{d}+O\left(\frac{1}{|z|}\right),$$

so that

$$\frac{\theta(T_f(z))}{\theta(z)} = 1+O\left(\frac{1}{|z|}\right).$$

Hence

$$\begin{aligned}\frac{\theta(\phi(t-n))}{\theta(\phi(t))} &= \frac{\theta(\phi(t-n))}{\theta(\phi(t-n+1))}\cdots\frac{\theta(\phi(t-1))}{\theta(\phi(t))}\\ &= 1+O(\frac{1}{|\phi(t)|}+\frac{1}{|\phi(t-1)|}+\cdots)\\ &= 1+O(\frac{1}{|\phi(t)|}),\end{aligned}$$

because the sequence $|\phi(t)|$, $|\phi(t-1)|$,... behaves roughly like a geometric series, according to theorem 2.4.

So if z is far enough away from all the roots, there is little difference between $\theta(z)$ and $\theta(T_f^{-n}(z))$ for any positive integer n. There may be other ranges of z for which θ does not change much. When we cannot prove that this is the case we will use

Lemma 3.3: *There exist constants μ and c, independent of d, such that if $s \le 0$ and $\theta(\phi(s)) < \mu/d$ then*

$$\frac{\theta(\phi(s))}{\theta(\phi(s+1))} \ge \frac{c}{d^6}. \tag{3.2}$$

Proof: See appendix A.

It is in this lemma that we get the main contribution for the lower bound. It turns out that we will apply the lemma 3.3 about $d^2 \log d$ times to get a bound on θ, namely for $t \in [0,1]$ and integer n we have

$$\begin{aligned}\theta(\phi(t-n)) &= \left[\frac{\theta(\phi(t-n))}{\theta(\phi(t-n+1))}\cdots\right]\theta(\phi(t))\\ &\ge \left(\frac{c}{d^6}\right)^{O(d^2\log d)}\end{aligned}$$

The rest of the time θ will not change much.

Now we give more ranges of ϕ where the ratio of the θ's doesn't change much.

Lemma 3.4: *There exist μ and c independent of d such that the following is true. Let m and M be positive integers with $m < M$. Assume that for each $n \in [m, M]$, n integer, we have*

$$\frac{|r_i - r|}{|\phi(t-n) - r|} \left\{ \begin{array}{ll} < \dfrac{c}{d^2} & \text{for } i = 1, \ldots, k \\ > \dfrac{d^2}{c} & \text{for } i = k+1, \ldots, d \end{array} \right\}$$

for some $k \geq 2$. Then

$$\frac{|\phi(t-n) - r|}{|\phi(t-n+1) - r|} \leq 1 - \frac{1}{2d}$$

for $n \in [m, M]$. If, in addition, for $m < n \leq M$ we have

$$\theta(\phi(t-n)) < \frac{\mu}{d},$$

then

$$\theta(\phi(t-M)) \geq \frac{1}{2}\theta(\phi(t-m)).$$

Proof: See appendix A.

Lemma 3.5: *For $t \in [0, 1]$ we have*

$$\theta(\phi(t)) \geq 1.$$

Proof: Since $\phi(0)$ and $\phi(1)$ lie in $B_{\delta/2}(r)$, so does $\phi(t)$ for $t \in [0, 1]$, and hence $\rho(\phi(t)) \geq \delta/2$ so that $\theta(\phi(t)) \geq 1$.

■

We are now ready for the lower bound

Theorem 3.6: *For any $s \leq 1$ we have $\theta(\phi(s)) \geq (1/d)^{cd^2 \log d}$.*

Proof: Let $s \leq 1$. Let $t \in [0, 1]$ and n be a positive integer such that $s = t - n$. By lemma 3.5 we have

$$\theta(\phi(t)) \geq 1 \geq \frac{\mu}{d}, \tag{3.3}$$

if μ is sufficiently small. Let n_0 be the largest integer $\leq n$ such that $\theta(\phi(t - n_0)) \geq \mu/d$, where μ is sufficiently small to make equation (3.3) and lemmas 3.3 and 3.4 hold. We write

$$\begin{aligned} &\theta(\phi(t-n)) \\ &= \left[\frac{\theta(\phi(t-n))}{\theta(\phi(t-n+1))} \cdots \frac{\theta(\phi(t-n_0+1))}{\theta(\phi(t-n_0))}\right]\theta(\phi(t-n_0)). \end{aligned}$$

Let I denote the subset of positive integers m such that

$$\frac{c}{d^2} \leq \frac{|r_i - r|}{|\phi(t-m) - r|} \leq \frac{d^2}{c}, \text{ for some } i, \tag{3.4}$$

with c as in lemma 3.4, and let J denote the set of positive integers not in I. By theorem 2.4 we see that the size of I is at most $cd^2\log(d)$— for each of $d-1$ possible r_i's, each one can satisfy (3.4) for at most $cd\log d$ values of m. Hence, using (3.2),

$$\prod_{m\in I, n_0\leq m<n} \frac{\theta(\phi(t-m+1))}{\theta(\phi(t-m))} \geq \left(\frac{1}{d}\right)^{cd^2\log d}.$$

If $m \geq n_0$ and $m \in J$, then $\phi(t-m)$ must lie outside of $B_{\delta/2}$ (or else $\theta(\phi(t-m)) \geq 1$), and thus there are at least two roots r_i with

$$\frac{|r_i - r|}{|\phi(t-m) - r|} \leq \frac{c}{d^2}.$$

It follows that $J \cap [n_0, n-1]$ consists of a union of at most $d-1$ sequences of consecutive integers, each satisfying the conditions for lemma 3.4. Applying lemma 3.4 we have

$$\prod_{m\in J, n_0\leq m<n} \frac{\theta(\phi(t-m+1))}{\theta(\phi(t-m))} \geq \left(\frac{1}{2}\right)^{d-1}.$$

Hence

$$\begin{aligned}
&\theta(\phi(t-n)) = \\
&\left[\prod_{m\in I, n_0\leq m<n} \frac{\theta(\phi(t-m+1))}{\theta(\phi(t-m))}\right]\left[\prod_{m\in J, n_0\leq m<n} \frac{\theta(\phi(t-m+1))}{\theta(\phi(t-m))}\right]\theta(\phi(t-n_0)) \\
&\geq \left(\frac{1}{d}\right)^{cd^2\log d}.
\end{aligned}$$

■

Appendix A. Some Calculations

Lemma 2.5: *Let ABCD be a parallelogram. Let $\angle C = \alpha$, and let E be any point in ABCD. Then*

$$|AC| \geq |EC| \left\{ \begin{array}{ccc} 1 & \text{if} & \alpha \leq \frac{\pi}{2} \\ \sin\alpha & \text{if} & \alpha > \frac{\pi}{2} \end{array} \right\}.$$

Proof: Clearly we only need show the above for $E = B$ or D, and by symmetry only for $E = B$. Consider triangle ABC. Then $\angle B = \pi - \alpha$ and $\angle A = \beta$, $\angle C = \gamma$ with $\alpha = \beta + \gamma$. Then, by law of sines,

$$\frac{|AC|}{|BC|} = \frac{\sin\alpha}{\sin\beta}.$$

Since β satisfies $0 \le \beta \le \alpha$, we have

$$\frac{|AC|}{|BC|} \ge \min_{\beta \in [0,\alpha]} \frac{\sin\alpha}{\sin\beta} = \left\{ \begin{array}{ccc} 1 & \text{if} & \alpha \le \frac{\pi}{2} \\ \sin\alpha & \text{if} & \alpha > \frac{\pi}{2} \end{array} \right\}$$

■

Lemma A.1: *Let $r = r_1 = \cdots = r_k$, $k \ge 1$, and let $\eta = \min_{r_i \ne r} |r_i - r|$. Then $|T_f(z) - r| \le (1-\epsilon)|z - r|$ if $|z - r| \le \frac{\eta}{2d}$ for some ϵ depending only on d.*

Proof: After a linear map (a translation, rotation, and dilation) we may assume $z = 0$ and $r = -1/2$. Assuming $\frac{\eta}{2d} \ge |z - r| = \frac{1}{2}$, we have $\eta > d$. Thus the r_i's not equal to r lie outside of $B_{d-\frac{1}{2}}(0)$, and so

$$\left| \sum_{r_i \ne r} \frac{1}{r_i} \right| \le (d-1)\frac{1}{d - \frac{1}{2}} = \frac{2d-2}{2d-1}.$$

Furthermore, since $1/r = -2$ we have that

$$\sum \frac{1}{r_i} \in B_{\frac{2d-2}{2d-1}}(-2k)$$

which lies in the interior of

$$\{y : \Re(y) < -1\},$$

since $1 \le k \le d$ and $\frac{2d-2}{2d-1} < 1$. Since the map $y \to 1/y$ maps $\{y : \Re(y) \le -1\}$ to $B_{1/2}(-1/2)$, we have

$$\left\{ y : \frac{1}{y} \in B_{\frac{2d-1}{2d-d}}(-2k) \text{ for some } k = 1, \ldots, d \right\} \subset B_{\frac{1-\epsilon}{2}}(-1/2)$$

for some $\epsilon > 0$. Reversing the linear map sending z, r to $0, -1/2$ yields the lemma.

■

Theorem 2.2: *Let $\eta = \min_{r_i \ne r} |r_i - r|$, and $\delta = \frac{\eta}{2d}$. Then $B_\delta(r) \subset \{\text{basin of } r\}$. Furthermore, $z \in B_\delta(r)$ implies $|T_f(z) - r| \le (1-\epsilon)|z - r|$ for some $\epsilon > 0$ depending only on d.*

Proof: By lemma A.1, $z \in B_\delta(r)$ implies $|T_f(z) - r| \le (1-\epsilon)|z - r|$, as well as $|T_f^n(z) - r| \le (1-\epsilon)^n |z - r|$ and thus $T_f^n(z) \to r$.

■

Lemma A.2: *Let $y_1, ..., y_m$ lie in the wedge W of angle $\alpha \le \pi$ with vertex 0. Then*

$$\frac{|y_1^2 + \cdots y_m^2|}{|y_1 + \cdots + y_m|^2} \le \cos^2(\frac{\alpha}{2}).$$

Proof: By a rotation we may assume that $W = \{z : |\arg(z)| \le \alpha/2\}$. Then we have

$$\begin{aligned}|y_1 + \cdots + y_m|^2 &\ge |\Re(y_1 + \cdots + y_m)|^2 \\ &\ge (|y_1| + \cdots + |y_m|)^2 \cos^2(\frac{\alpha}{2}) \\ &\ge (|y_1|^2 + \cdots + |y_m|^2) \cos^2(\frac{\alpha}{2})\end{aligned} \ .$$

■

Lemma A.3: *Let l be the line through r and perpendicular to the bisector of the interior angle at r with respect to the polygon $\langle r_j \rangle$. Let H_r be the set (half-plane) of points lying to side of l opposite from $\langle r_j \rangle$. Then if $z \in H_r$ we have*

$$|T_f'(z)| \le 1 + \frac{1}{\cos^2(\frac{\pi+\alpha}{4})}$$

where $\alpha = \angle r$.

Figure 6

Proof: (See figure 6) We have

$$T'(z) = 1 - \frac{\sum \left(\frac{1}{z-r_i} \right)^2}{\left[\sum \left(\frac{1}{z-r_i} \right) \right]^2}. \qquad (A.1)$$

We can assume that l is parallel to the imaginary axis.

If z lies above the bisector of $\angle r$, then the $z - r_i$'s lie in the wedge $W = \{y : -\pi/2 \le \arg(y) \le \alpha/2\}$, a wedge of angle $\frac{\pi}{2} + \frac{\alpha}{2}$ (see figure 7). If z lies below the bisector of $\angle r$, then similarly the $z - r_i$'s lie in a wedge of angle $\frac{\pi}{2} + \frac{\alpha}{2}$. It then follows that the $1/(z - r_i)$'s

Figure 7

also lie in a wedge of angle $\frac{\pi}{2} + \frac{\alpha}{2}$. Combining equation (A.1) with lemma A.3 yields the lemma.

∎

Lemma A.4: *Let W be a wedge of angle $\alpha < \pi$ and let $w \in W$. Then*

$$|w + w'| \geq |w| \left\{ \begin{array}{r} \sin\alpha \ \textit{if} \ \alpha \geq \pi/2 \\ 1 \ \textit{if} \ \alpha < \pi/2 \end{array} \right\}$$

for any $w' \in W$.

Proof: We can assume $w = 1$. Let $w' \in W$ and let w' make an angle of β with the x-axis. Then $|\beta| \leq \alpha$. If $|\beta| \leq \pi/2$ then clearly $|1 + w'| \geq 1$.

Figure 8

Otherwise, from figure 8 we can see that

$$|1 + w'| \geq \sin(\pi - \beta) = \sin\beta \geq \sin\alpha.$$

∎

Lemma A.5: *Let z lie in the exterior wedge at r, let $\alpha = \angle r$, and let $r' \neq r$ be a root of f with*

$$\left|\frac{r'-r}{z-r}\right| \leq \nu.$$

Then

$$\left|\frac{T_f(z)-r}{z-r}\right| \geq \frac{c}{\nu+1}\left\{\begin{array}{r}\sin\alpha \text{ if } \alpha \geq \pi/2 \\ 1 \text{ if } \alpha < \pi/2\end{array}\right\}$$

for some absolute constant c.

Proof: We can assume $z=0$, $r=1$. We have all the r_i's lie in the wedge determined by the interior angle at r,

$$W = \{w : \theta_1 \leq \arg(w-1) \leq \theta_2\},$$

with $\theta_2 - \theta_1 = \alpha$. Since z is in the exterior wedge we have

$$W \subset W' = \{w : \theta_1 \leq \arg(w) \leq \theta_2\}.$$

It follows that all the $1/r_i$'s lie in

$$W'' = \{w : -\theta_2 \leq \arg(w) \leq -\theta_1\}.$$

Hence

$$\sum \frac{1}{r_i} = \frac{1}{r} + \frac{1}{r'} + \cdots = 1 + \frac{1}{r'} + w$$

where $w \in W''$. Since $|r'| \leq \nu + 1$ we have by lemma A.4 that

$$\left|\frac{1}{r'} + w\right| \geq \frac{1}{\nu+1}\left\{\begin{array}{r}\sin\alpha \text{ if } \alpha \geq \pi/2 \\ 1 \text{ if } \alpha < \pi/2\end{array}\right\}. \tag{A.2}$$

Since

$$T_f(z) = \frac{1}{\sum \frac{1}{r_i}} = \frac{1}{1 + (\frac{1}{r'} + w)}$$

and since the quantity on the right hand side of equation (A.2) is always ≤ 1, it follows that

$$|T_f(z) - 1| \leq \frac{c}{\nu+1}\left\{\begin{array}{r}\sin\alpha \text{ if } \alpha \geq \pi/2 \\ 1 \text{ if } \alpha < \pi/2\end{array}\right\}$$

for some absolute constant c.

■

Corollary A.6: *Let ϕ be the curve defined just before lemma 3.6. Then $\phi(t) \geq c\delta/d^2$ for $t \leq 1$ for some absolute constant c.*

Proof: Since $|\phi(0) - r| = \delta/2$ we can apply lemma A.5 with $\nu = 4d$ and $\alpha \le \pi(1 - \frac{2}{d})$ to obtain $|\phi(1) - r| \ge c\delta/d^2$ for some absolute constant c. It then follows that $|\phi(t) - r| \ge c\delta/d^2$ for any $t \in [0, 1]$, and therefore also for any $t < 0$ since $|T_f(z) - r| \le |z - r|$ for $z \in B_\delta(r)$.

∎

Lemma 3.3: *There exist constants μ and c, independent of d, such that if $s \le 0$ and $\theta(\phi(s)) < \mu/d$ then*

$$\frac{\theta(\phi(s))}{\theta(\phi(s+1))} \ge \frac{c}{d^6}.$$

Proof: Let $z = \phi(s)$. The mean-value theorem implies

$$\rho(T_f(z)) \le \rho(z) \max_{\zeta \in B_{\rho(z)}(z)} |T_f'(\zeta)|.$$

For μ sufficiently small (independent of d) we have $\zeta \in B_{\rho(z)}(z)$, $\rho(z) \le \frac{\mu}{d}|z - r|$, and the fact that z lies in the exterior wedge at r together imply $\zeta \in H_r$ and thus

$$\max_{\zeta \in B_{\bar\rho}(z)} |T_f'(\zeta)| \le 1 + \frac{1}{\cos^2(\frac{\pi + \alpha}{4})} \le cd^2.$$

Since $z = \phi(s)$ with $s \le 0$ we have by lemma A.6 that $|z - r| \ge c\delta/d^2$ and thus

$$\left|\frac{r' - r}{z - r}\right| \le \frac{2d\delta}{c\delta/d^2} = c'd^3$$

for some root $r' \ne r$. Hence by lemma A.5

$$\left|\frac{T_f(z) - r}{z - r}\right| \ge \frac{c}{d^4}.$$

Thus

$$\frac{\theta(z)}{\theta(T_f(z))} \ge \frac{\rho(z)}{\rho(T_f(z))} \frac{|T_f(z) - r|}{|z - r|} \ge \frac{c}{d^6}.$$

∎

Lemma A.7: *Let z lie on the bisector of the exterior angle. Then T_f lies in the exterior angle.*

Proof: It suffices to prove it for $z = 0$, $r = 1$. The map $y \to 1/y$ maps the wedge $W = \{z : |\arg(z - 1)| \le \alpha/2\}$ to the lune (intersection of two circles) L with vertices at 0 and 1 and of angle α (see figure 9).

It follows that

$$\sum \frac{1}{r_i}$$

Figure 9

lies in the wedge $1+(d-1)L$, since one of the r_i's is 1. Since $1+(d-1)L$ is a subset of W, its reciprocal, $T_f(z)$, lies in L. Since L lies in the exterior wedge, we are done.

∎

Lemma 3.4: *There exist μ and c independent of d such that the following is true. Let m and M be positive integers with $m < M$. Assume that for each $n \in [m, M]$, n integer, we have*

$$\frac{|r_i - r|}{|\phi(t-n) - r|} \left\{ \begin{array}{ll} < \dfrac{c}{d^2} & \text{for } i = 1, \ldots, k \\ > \dfrac{d^2}{c} & \text{for } i = k+1, \ldots, d \end{array} \right\}$$

and

$$\theta(\phi(t-n)) < \frac{\mu}{d}.$$

Then

$$\theta(\phi(t-M)) \geq \frac{1}{2}\theta(\phi(t-m)).$$

Proof: Let z be such that $\frac{|r-r_i|}{|z-r|} \leq \epsilon$ for $i = 1, \ldots, k$ and $\frac{|r-r_i|}{|z-r|} \geq B$ for $i = k+1, \ldots, d$. For $i = 1, \ldots, k$ we have

$$\left| \frac{1}{r_i - z} - \frac{1}{r - z} \right| = \left| \frac{r_i - r}{(r-z)(r_i - z)} \right| \leq \frac{1}{|r-z|} \frac{\epsilon}{1-\epsilon}$$

since

$$|r_i - z| \geq |r - z| - |r - r_i| \geq |r - z|(1-\epsilon).$$

For $i = k+1, \ldots, d$ we have

$$\left| \frac{1}{r_i - z} \right| \leq \frac{1}{|r - r_i| - |r - z|} \leq \frac{1}{|r-z|} \frac{1}{B-1}.$$

Thus

$$\sum_{i=1}^{d} \frac{1}{r_i - z} = \frac{k}{r-z} + \frac{d}{|r-z|} O(\frac{1}{B} + \epsilon) = \frac{k}{r-z}[1 + O(\frac{d}{B} + d\epsilon)]$$

and so

$$\frac{T_f(z)-r}{z-r} = \frac{z - \frac{r-z}{k}[1+O(\frac{d}{B}+d\epsilon)] - r}{z-r} = 1 - \frac{1}{k} + O(\frac{d}{B}+d\epsilon). \qquad (A.3)$$

For

$$y \in B_{|z-r|\mu/d}(z)$$

we have

$$|z-r|(1-\frac{\mu}{d}) \le |y-r| \le |z-r|(1+\frac{\mu}{d}).$$

Assuming, say, $\mu/d \le 1/2$, we get for $i=1,\ldots,k$

$$\frac{1}{r_i-y} = \frac{1}{r-y}(1+O(\epsilon)),$$

$$\left(\frac{1}{r_i-y}\right)^2 = \left(\frac{1}{r-y}\right)^2(1+O(\epsilon)),$$

and for $i=k+1,\ldots,d$

$$\frac{1}{r_i-y} = \frac{1}{r-y}O(1/B),$$

$$\left(\frac{1}{r_i-y}\right)^2 = \left(\frac{1}{r-y}\right)^2 O(1/B^2).$$

Thus

$$\begin{aligned} T'(y) &= 1 - \frac{\sum\left(\frac{1}{y-r_i}\right)^2}{\left[\sum\left(\frac{1}{y-r_i}\right)\right]^2}. \\ &= 1 - \frac{\frac{k}{(y-r)^2}(1+O(\frac{d}{B^2}+d\epsilon))}{(\frac{k}{y-r})^2(1+O(\frac{d}{B}+d\epsilon))^2} \\ &= 1 - \frac{1}{k} + O\left(\frac{d}{B}+d\epsilon\right) \end{aligned}$$

Hence, assuming $\rho(z) \le \frac{\mu}{d}|z-r|$ we have

$$\begin{aligned} \frac{\theta(z)}{\theta(T_f(z))} &= \frac{\rho(z)}{\rho(T_f(z))}\frac{|T_f(z)-r|}{|z-r|} \\ &\ge \frac{1}{\max_{y\in B_{\rho(z)}}|T_f'(y)|}\frac{|T_f(z)-r|}{|z-r|} \\ &= 1 + O\left(\frac{d}{B}+d\epsilon\right), \end{aligned}$$

where the constant in $O()$ is absolute.

If d/B and $d\epsilon$ are sufficiently small, say $\leq 1/2d$, then from equation (A.3)

$$\frac{|T_f(z) - r|}{|z - r|} \leq 1 - \frac{1}{k} + \frac{1}{2d} \leq 1 - \frac{1}{2d}.$$

So given $m < M$ and k integers with

$$\frac{|r_i - r|}{|\phi(t-n) - r|} \left\{ \begin{array}{ll} < \dfrac{c}{d^2} & \text{for } i = 1, \ldots, k \\ > \dfrac{d^2}{c} & \text{for } i = k+1, \ldots, d \end{array} \right\}$$

for all $n \in [m, M]$ and for some c sufficiently small, we have

$$\begin{aligned} &|\phi(t-M) - r| + |\phi(t-M+1) - r| + \cdots + |\phi(t-m) - r| \\ &\leq |\phi(t-M) - r| \left(1 + \left(1 - \frac{1}{2d}\right) + \left(1 - \frac{1}{2d}\right)^2 + \cdots \right) \\ &\leq |\phi(t-M) - r| 2d \end{aligned}$$

and

$$\frac{1}{|\phi(t-m) - r|} + \cdots + \frac{1}{|\phi(t-M) - r|} \leq \frac{1}{|\phi(t-m) - r|} 2d.$$

Letting $s = \max_{i \leq k} |r_i - r|$ and $S = \min_{i > k} |r_i - r|$ we have, assuming $\theta(\phi(t-n)) \leq \mu/d$ for $n \in [n+1, M]$,

$$\begin{aligned} \frac{\theta(\phi(t-M))}{\theta(\phi(t-m))} &= \frac{\theta(\phi(t-M))}{\theta(\phi(t-M+1))} \cdots \frac{\theta(\phi(t-m-1))}{\theta(\phi(t-m))} \\ &\geq \prod_{i=m}^{M-1} 1 - O\left(\frac{|\phi(t-i) - r|}{S} d + d \frac{s}{|\phi(t-i) - r|} \right) \\ &\geq 1 - c_1 \sum_{i=m}^{M-1} \frac{|\phi(t-i) - r|}{S} d + d \frac{s}{|\phi(t-i) - r|} \\ &\geq 1 - c_1 2d \left(\frac{|\phi(t-M) - r|}{S} d + d \frac{s}{|\phi(t-m) - r|} \right) \\ &\geq 1 - c_1 \left(\frac{c}{d^2} 2d^2 + \frac{c}{d^2} 2d^2 \right) \geq 1 - c_1(2c + 2c) \end{aligned}$$

where c_1 is an absolute constant. So for c sufficiently small and $\leq \frac{1}{8c_1}$ the lemma is proven.

■

Bibliography

[F] J. Friedman, *On Newton's Method for Polynomials*, in 27th Annual Symposium on Foundations of Computer Science (1986), 153-161.

[M] A. Manning, *How to be Sure of Solving a Complex Polynomial using Newton's Method*, University of Warwick, preprint.

[S] S. Smale, *On the efficiency of algorithms of analysis*, Bull. Amer. Math. Soc. (N.S.) **13** (1985), 87–121.