Geometry of polynomials and root-finding via path-lifting

Myong-Hi Kim\textsuperscript{1}, Marco Martens\textsuperscript{2} and Scott Sutherland\textsuperscript{2,3}

\textsuperscript{1} Mathematics, Computers & Information Science, SUNY at Old Westbury, Old Westbury, NY 11568, United States of America
\textsuperscript{2} Institute for Mathematical Sciences, Stony Brook University, Stony Brook, NY 11794, United States of America

E-mail: kimm@oldwestbury.edu, marco@math.stonybrook.edu and scott@math.stonybrook.edu

Received 22 August 2015, revised 30 August 2016
Accepted for publication 14 September 2017
Published 5 January 2018

Recommended by Professor Konstantin M Khanin

Abstract

Using the interplay between topological, combinatorial, and geometric properties of polynomials and analytic results (primarily the covering structure and distortion estimates), we analyze a path-lifting method for finding approximate zeros, similar to those studied by Smale, Shub, Kim, and others. Given any polynomial, this simple algorithm always converges to a root, except on a finite set of initial points lying on a circle of a given radius.

Specifically, the algorithm we analyze consists of iterating

\[ z = \frac{f(z) - t_k f(z_0)}{f'(z)} \]

where the \( t_k \) form a decreasing sequence of real numbers and \( z_0 \) is chosen on a circle containing all the roots. We show that the number of iterates required to locate an approximate zero of a polynomial \( f \) depends only on \( \log |f(z_0)/\rho_\zeta| \) (where \( \rho_\zeta \) is the radius of convergence of the branch of \( f^{-1} \) taking 0 to a root \( \zeta \)) and the logarithm of the angle between \( f(z_0) \) and certain critical values. Previous complexity results for related algorithms depend linearly on the reciprocals of these angles. Note that the complexity of the algorithm does not depend directly on the degree of \( f \), but only on the geometry of the critical values.

Furthermore, for any polynomial \( f \) with distinct roots, the average number of steps required over all starting points taken on a circle containing all the roots is bounded by a constant times the average of \( \log(1/\rho_\zeta) \). The average of \( \log(1/\rho_\zeta) \) over all polynomials \( f \) with \( d \) roots in the unit disk is \( \Theta(d) \). This algorithm readily generalizes to finding all roots of a polynomial (without
deflation); doing so increases the complexity by a factor of at most $d$.

Keywords: root-finding, Newton’s method, Voronoi region, approximate zeros, branched cover, homotopy method, alpha theory
Mathematics Subject Classification numbers: Primary 65H05; Secondary 30C15, 37F10, 52C20, 57M12, 68Q25

(Some figures may appear in colour only in the online journal)

1. Introduction

We analyze a path-lifting method called the $\alpha$-step method (see section 3 for specifics), which locates an approximate zero (see definition 3.1) for a complex polynomial $f(z)$; from an approximate zero, Newton’s method converges quadratically to a root. For any polynomial, the $\alpha$-step method converges everywhere except on a finite set of starting points lying on a circle of given radius. This is established in this paper, but also follows from [K88, theorems 5A, 5B].

We consider monic polynomials of degree $d$ with distinct roots in the unit disk, and denote the set of all such polynomials by $\mathcal{P}_{d,1}$. Our main results bound the number of iterations required to locate an approximate zero in three contexts: we bound the number of steps needed to locate an approximate zero starting from any point $z_0$ on a circle containing all the roots; we compute the average number of steps over the circle of initial points; we average this quantity over all polynomials in $\mathcal{P}_{d,1}$ to get a bound in terms of the degree. These bounds apply to all roots of a given polynomial, and can be applied to locate all of the roots with a $d$-fold increase in effort.

While we analyze the complexity of the $\alpha$-step method, it is not our primary goal to demonstrate that this achieves the optimal bound. Indeed, there are certainly other algorithms with a lower worst-case arithmetic complexity (at least for finding $\epsilon$-roots) such as that of Pan [P02] which achieves the nearly optimal bound, or of Renegar [Ren] or Kim and Sutherland [KS]. Some further remarks discussing the arithmetic complexity of these and other related methods appear toward the end of this section.

Rather, our goal is to examine how the underlying geometry of a polynomial can be exploited in root-finding methods. Tight upper and lower bounds on the radius of convergence of the inverse of an analytic map are given by $\alpha$-theory; these are useful in understanding the geometry of the polynomial. Since the $\alpha$-theory also applies in the multivariate case, it is our belief that a better understanding of the univariate case will be aid in understanding the case of several variables.

1.1. Background

We now discuss some background related to path-lifting methods in general.

Path-lifting methods are a class of homotopy methods, and are also refered to as ‘modified Newton’s method’ or ‘guided Newton’s method’. In such methods, it is often useful to distinguish between the domain and range, so we have

$$f : \mathbb{C}_{\text{source}} \to \mathbb{C}_{\text{target}};$$

the goal is to lift a path $\gamma$ lying in $\mathbb{C}_{\text{target}}$ to one in $\mathbb{C}_{\text{source}}$ leading from an initial point $z_0$ to a root $\zeta$. Numerically, this is accomplished by constructing a sequence of points $z_j \in \mathbb{C}_{\text{source}}$ via analytic continuation, in such a way that each $f(z_j)$ approximates the path $\gamma$ in $\mathbb{C}_{\text{target}}$ and gives an approximation of the lift $f^{-1}(\gamma)$ in $\mathbb{C}_{\text{source}}$. 

415
In this form, such methods were introduced by Shub and Smale (see, for example, [SS86] or [Sm85]), although one could argue (as Smale points out in [Sm81]) that in some sense this idea goes back to Gauss. See [Ren] and the references therein, as well as [KS]. The series [SS93a, SS93b, SS93c, SS96, SS94, Sh09, BS09] discusses related methods for systems of polynomial equations, as does [BP]. A survey of complexity results for solving polynomial equations in one variable can be found in [P97]; see also [B08].

The difficulty of computing a local branch of $f^{-1}$ along a path $\gamma$ in the target space is related to how close $\gamma$ comes to a critical value of $f$. However, not all critical values of $f$ are relevant: if we fix a branch of $f^{-1}$, then for points $y \in \gamma$ the only critical points that have an impact are those $c$ for which $f(c)$ lies on the boundary of the largest disk where $f^{-1}(y)$ is analytic. Consequently, it is useful to factor $f$ through the (branched) Riemann surface $S$ for $f^{-1}$, giving

$$\text{C}_{\text{source}} \xrightarrow{\hat{f}} S \xrightarrow{\pi} \text{C}_{\text{target}}.$$  

Denoting the set of critical points $c_j$ of $f$ by $\mathcal{C}_f$ and the branch points of $S$ by $\mathcal{Y}_f$, we require the map $\hat{f}$ to be a biholomorphism from $\mathbb{C} \setminus \mathcal{C}_f$ to $S \setminus \mathcal{Y}_f$ and a bijection from $\mathcal{C}_f$ to $\mathcal{Y}_f$. Furthermore, the projection map $\pi : S \to \text{C}_{\text{target}}$ identifies a point in one of the sheets of $S$ with all other points directly above and below it; $\text{C}_{\text{target}}$ is not shown in this figure.

In this form, such methods were introduced by Shub and Smale (see, for example, [SS86] or [Sm85]), although one could argue (as Smale points out in [Sm81]) that in some sense this idea goes back to Gauss. See [Ren] and the references therein, as well as [KS]. The series [SS93a, SS93b, SS93c, SS96, SS94, Sh09, BS09] discusses related methods for systems of polynomial equations, as does [BP]. A survey of complexity results for solving polynomial equations in one variable can be found in [P97]; see also [B08].

The construction of the branched Riemann surface $S$ for $f^{-1}$ is quite standard, going back to Riemann’s dissertation [Rie], although often it is presented somewhat abstractly. Many readers will be familiar with the corresponding surfaces for the logarithm and square root; the explicit view taken here of $S$ as a collection of copies of $\mathbb{C}$ identified along slits is similar.
to the one in [GK, section 10.4] or [MH, section 6.1], to which we refer the interested reader. Note that each point of \( \mathcal{S} \) corresponds to a pair \((z, w)\) with \( z \in C_{\text{source}} \) and \( w \in C_{\text{target}} \), and \( w = f(z) \). It is often helpful to think of the path \( \gamma \) as lying in \( \mathcal{S} \) rather than in \( C_{\text{target}} \); this is possible since for any ray which avoids \( \mathcal{Y} \) there is a neighborhood \( U \) containing it which is isometric to its projection \( \pi(U) \) in \( C_{\text{target}} \).

In order to explicitly describe which critical values are relevant for the path-lifting process, it is helpful to introduce the Voronoi decomposition of \( \mathcal{S} \) relative to the branch points \( \gamma \). That is, for each branch point \( v \) of \( \mathcal{S} \), the Voronoi domain \( \text{Vor}(v) \) is the set of points in \( \mathcal{S} \) which are closer to \( v \) than any other branch point of \( \mathcal{S} \). See figure 4. Note that \( y \in \text{Vor}(v) \) exactly when \( \| v - y \| \) is the radius of convergence of \( \tilde{f}^{-1} \) at \( y \). We show in section 4 that the projection map \( \pi \) restricted to any single \( \text{Vor}(v) \) is at most \((m + 1)\)-to-one, where \( m \) is the multiplicity of the critical point of \( f \) corresponding to \( v \) (hence the projection \( \pi \) is generically at most 2-to-one on \( \text{Vor}(v) \)). When lifting a path \( \gamma \), the number of steps required depends directly on the size of a neighborhood of \( \gamma \) on which a branch of \( f^{-1} \) can be defined. If we think of \( \gamma \) as lying in \( \mathcal{S} \), then the size of this neighborhood is the distance between \( \gamma \) and branch points \( y \) for which \( \gamma \) intersects \( \text{Vor}(v) \). We refer to such a critical value \( f(\epsilon_j) = \pi(\epsilon_j) \in C_{\text{target}} \) as relevant or say that \( \epsilon \) influences the points on \( \gamma \).

As noted earlier, in a path-lifting method we choose a path \( \gamma \) in the target space which connects 0 to a point \( w_0 \) for which we know a point \( z_0 \) with \( f(z_0) = w_0 \). Path-lifting methods attempt to calculate a sequence of points \( \{z_j\} \) so that \( f(z_j) \approx w_j \), and terminate when a point \( z_n \) is sufficiently close to a root of \( f \).

Typically the chosen path \( \gamma \in C_{\text{target}} \) is a segment of a ray, and we use such paths here. It is common (e.g. [SS86] and [KS]) to choose the guide points \( w_j \) to be of the form \( h^jw_0 \) for some \( h < 1 \), and then use one step of Newton’s method to obtain \( z_{j+1} \) from \( z_j \) as \( z_{j+1} = z_j - \frac{f(z_j) - w_{j+1}}{f'(z_j)} \). To ensure convergence, one must choose the \( w_j \) sufficiently closely spaced along \( \gamma \); exactly how close depends strongly on the size of a neighborhood around \( \gamma \) on which a branch of \( f^{-1} \) can be defined via analytic continuation.

While the ultimate goal of root-finding is typically to find a point that lies within an \( \epsilon \)-ball of some root \( \zeta \) of \( f \) (called an \( \epsilon \)-root of \( f \)), we instead focus on the problem of locating an approximate zero of \( f \). This notion is was introduced by Smale (see [Sm81]): a point \( z^* \) is an approximate zero for \( f \) if Newton iteration converges at a definite, rapid rate to a root of \( f \) when begun at \( z^* \). (See definition 3.1 for a precise statement.) From an approximate zero an \( \epsilon \)-root for any desired value of \( \epsilon \) can be produced rapidly, with \( O(\log(1/\epsilon)) \) iterations of Newton’s method (see [Sm85]).

Unlike \( \epsilon \)-roots, the set of approximate zeros is an intrinsic feature of a polynomial and does not depend on an externally imposed quantity \( \epsilon \). We restrict our attention to polynomials with distinct roots, so approximate zeros always exist for each root \( \zeta \). See also remark 11.3 concerning locating \( \epsilon \)-roots.

Rather than using a regular spacing for the target points \( w_j \) in the path-lifting process, the \( \alpha \)-step method considered here selects the points \( w_j \) adaptively, spacing them as far apart as possible while ensuring that at each step \( z_j \) is an approximate zero for the function \( f(z) - w_{j+1} \) (and hence \( z_{j+1} \) is a good approximation for \( w_{j+1} \) with known error bounds). The algorithm terminates when \( z_n \) is an approximate zero for \( f(z) \). The tool we use to detect approximate zeros is the Kim-Smale \( \alpha \) function: if \( \alpha_f(z) < 3 - \sqrt{8} \), then \( z \) is an approximate zero for \( f \). See the beginning of section 3 for further details regarding the \( \alpha \) function and approximate zeros, as well as the specifics of the \( \alpha \)-step method.
1.2. Main results

Our first main result gives an upper bound on the number of steps required by the α-step algorithm to converge to an approximate zero of some root ζ of f, starting from an initial point z₀ ∈ Basin (ζ). The set Basin (ζ) is the collection of all points which converge to the root ζ under the Newton flow (see section 2). The union of these basins over all roots has full measure; in fact, the complement is a collection of d − 1 curves.

The upper bound in the theorem depends on several quantities closely related to the geometry of the critical values of f. Specifically, the number of steps required depends on the radius of convergence ρζ of the branch of f⁻¹ taking 0 to ζ (that is, the norm of some critical value f(cζ)); this is closely related to f′(ζ), as well as on the angle that the path γ makes with the relevant critical values f(cj) (these angles are denoted θj in the statement below) and on the length of the path (which is |f(z₀)|). As noted earlier, a critical value f(cj) is relevant if the corresponding lift of the path γ to F intersects the Voronoi domain of f(cj). The appearance of ρζ in our estimates is not surprising, since the radius of a disk of approximate zeros about a root ζ is at most ρζ.

Let f ∈ ℛd,1 and let ζ be an initial point for the α-step path lifting algorithm with |z₀| > 1. Then the maximum number of steps required for the algorithm to produce an approximate zero in Basin (ζ) starting from z₀ is

\[ \#_f(z₀) \leq 67 \cdot \left( \log \frac{|f(z₀)|}{|\rho_ζ|} + \log 40 + \sum_{j=1}^{β^+(z₀)} (3 - 2 \log |θ_j|) \right) . \]

Observe that theorem 1' implies that for f ∈ ℛd,1, the α-step algorithm converges to a root ζ for every initial point z₀ as long as θj ≠ 0. Thus, the algorithm can only fail for at most 2d − 2 initial points z₀ on a circle of fixed radius larger than 1. See also remark 7.3.

The details of this theorem are established in section 7. It is worth noting that for every polynomial, the expected number of relevant critical values (β⁺(z₀)) is no more than 2 (as shown in proposition 8.3); a relation between ρζ and f′(ζ) is given in lemma 9.1.

We should emphasize that in the literature the dependence on the reciprocal of the angle |θj| is linear (see [Sm97] for an overview), while in theorem 1' the dependence is logarithmic. Beltrán and Shub have recently shown (see section 7 of [BS13] or [BS10]) the existence of homotopy methods whose number of steps depends logarithmically on a quantity comparable to our θj (in projective space), but currently there is no known constructive method to produce the necessary path. Since our paths are line segments in the target space, this is a significant improvement.

For any fixed polynomial f, our second main result gives a bound on the expected value of the number of steps required when an initial point is taken on the circle of radius 1 + 1/d (with uniform measure on the circle). This is established in section 8.

Theorem 2'. Let f : ℂ → ℂ be a monic polynomial with distinct roots ζ in the unit disk. Let \( \#_f \) be the average number of steps required by the α-step algorithm to locate an approximate zero of f starting from an initial point on the circle of radius 1 + 1/d (with uniform measure on the circle). Then

\[ \#_f \leq 67 \cdot \left( \log \frac{|f(z₀)|}{|\rho_ζ|} + \log 40 + \sum_{j=1}^{β^+(z₀)} (3 - 2 \log |θ_j|) \right) . \]
zero for \( f \), where the average is taken over starting points on the circle of radius \( 1 + 1/d \) with uniform measure. Then
\[
\overline{f} 
\leq 134 \left( \frac{1}{d} \sum_{i=1}^{d} \log \frac{1}{\rho_i} + 6.2 \right).
\]

We wish to emphasize that for a specific polynomial \( f \), this bound does not depend directly on the degree, but only on the arrangement of the critical values (or, more precisely, on the geometry of the branched surface \( S \)). While \( \log \frac{1}{\rho_i} \) is not bounded above or below for \( f \in \mathcal{P}_{d,1} \), its average value grows no more than linearly in the degree of \( f \) (as stated in theorem 3, established in section 9).

As is apparent in theorem 2, the sum of the logarithms of \( \rho_i \) plays a crucial role in the estimates. Indeed, this quantity is a direct measurement of the difficulty of solving \( f(z) = 0 \).

We let \( K_f = \sum_{f(\zeta) = 0} \log \frac{1}{\rho_\zeta} \), and consider its average over all polynomials of a given degree (including those with multiple roots).

**Theorem 3’.** Let \( \Lambda \) be the average value of \( K_f / d \) over \( f \in \mathcal{P}_{d,1} \) where \( \mathcal{P}_{d,1} \) is parameterized by the polydisk of the roots endowed with Lebesgue measure. Then
\[
\Lambda < \frac{3}{d}.
\]

Consequently, the average of \( \overline{f} \) over \( \mathcal{P}_{d,1} \) is \( O(d) \).

**Remark 1.2.** The cost of each step of the \( \alpha \)-step algorithm is dominated by the calculation of \( \alpha_f(z) \) (defined in equation (3.1)), which can be done with \( \Theta(d \log^2 d) \) arithmetic operations (see [BM], for example). Consequently, theorem 2’ implies that for a specific polynomial \( f \), the expected arithmetic complexity to locate an approximate zero via the \( \alpha \)-step algorithm is less than \( \Theta \left( K_f \log^2 d \right) \). Combining this with theorem 3 gives an expected arithmetic complexity of \( \Theta \left( d^2 \log^2 d \right) \) to locate a root for a polynomial in \( \mathcal{P}_{d,1} \).

**Remark 1.3.** For \( f \in \mathcal{P}_{d,1} \), by choosing \( d \) appropriate starting values, an approximate zero can be found for each root \( \zeta_j \) in \( O(K_f) \) steps of the \( \alpha \)-step algorithm. This has an average arithmetic complexity of \( \Theta \left( d^3 \log^2 d \right) \). An explicit method for choosing initial points is given in section 10.

In addition to the theorems above, we wish to highlight several surprising intermediate results which appear in section 5. Specifically, let \( |z_r| = r \) with \( r > 1 \). Then a bound on the rate of change of \( \text{Arg} \ f(z_r) \) is given by our angular speed lemma (lemma 5.1); applying this improves proposition 2 of [SS86] regarding the measure of ‘good starting points’ from 1/6 to 5/6 (see remark 5.5).

Also worth noting are corollary 5.10, which shows that the average value of \( |f(z_r)| \) is \( d \log r \), and proposition 5.13, which states that \( |f(z_r)| \) is bounded below by a constant times \( \rho_\zeta \).

### 1.3. Related work

In [Ren], Renegar gives an algorithm which approximates all \( d \) roots of a polynomial with an arithmetic complexity of \( \Theta(d^3 \log d + d^2 \log d \log |\log \epsilon|) \) in the worst case. However, this algorithm includes a component requiring exact computation. Pan’s algorithm [P97] achieves the nearly optimal bound with a complexity of \( \Theta(d^2 \log d \log |\log \epsilon|) \),
but implementation requires high precision computations (of the order exceeding the degree of the input polynomial).

In practice, the software package MPSolve [BF] is widely used and empirical data indicates good global convergence properties; the software uses the Aberth–Ehrlich method (see [Ab] and [Ehr]) to locate the roots of the given polynomial. There is not a lot of theoretical support, however: to our knowledge the global behavior of the Aberth–Ehrlich method is not understood.

In [KS], a worst-case complexity of \(O(d^2 \log^2 d + d \log d |\log \epsilon|)\) yields an \(\epsilon\)-factorization for a polynomial \(f\). This relies on a path-lifting algorithm which finds half the roots, then deflates the polynomial (that is, divides out by the approximations).

Recent work of Schleicher ([Sch, BAS] and his co-authors have extended the results of [HSS] to obtain bounds for the complexity of finding \(\epsilon\)-roots. In [HSS], it is shown that there is a universal set of \(1.1d \log^2 d\) points on a circle containing all the roots; if the roots are uniformly and independently distributed, [BAS] shows that \(O(d^2 \log^2 d)\) iterations of Newton’s method will locate all of the roots (an arithmetic complexity of \(O(d^2 \log^2 d + d^2 \log d |\log \epsilon|)\) with a high probability, comparable with the average arithmetic complexity of \(O(d^2 \log^2 d + d^2 \log d |\log \epsilon|)\) for the \(\alpha\)-step method in this paper (here the \(|\log \epsilon|\) term is added to account for the cost of refining an approximate zero to an \(\epsilon\)-root).

One significant advantage of path-lifting methods over other methods is that of stability: as a consequence of estimates in [K85], as long as \(f\) and its derivatives are computed with a relative error of \(10^{-3}\), the algorithm will converge to an approximate zero in the same way.

1.4. Organization

The paper is organized as follows. In section 2, we set out notation and preliminary notions. Section 3 describes the \(\alpha\)-step path-lifting algorithm explicitly. In section 4, we discuss the branched surface \(\mathcal{K}\) and the corresponding Voronoi partition. This section discusses underlying topological and geometric properties, and may be of interest independent to the question of root-finding.

Section 5 computes several estimates related to how the polynomial \(f\) behaves on the initial circle. In section 6, we bound the distance between \(w_n\) and \(w_{n+1}\), and use this in section 7 to estimate the number of steps needed for the algorithm to locate an approximate zero from a given starting point \(z_0\), proving theorem 1.

In section 8, we combine the topological and geometric results of section 4 with the more analytical results from section 7 to calculate an average upper bound over all starting points for a given polynomial, proving theorem 2. In section 9, we discuss the relation between the number of steps required and the degree of \(f\) and proves theorem 3. This is followed by section 10 where we describe how to use this method to locate all roots of a polynomial \(f\). We conclude in section 11 with some remarks and comments regarding extensions of these results.

2. Preliminaries

We will use the following general notions and notations throughout.

An open disk of radius \(r > 0\) centered around \(z \in \mathbb{C}\) is denoted by \(D_r(z)\).

Let \(S_r(z)\) denote the circle of radius \(r\) and center \(z\); if the circle is centered at the origin, we will denote it by \(S_r\).
The function \( \text{Arg} \) denotes the argument of a complex number (in the interval \((-\pi, \pi] \) unless otherwise noted).

The **ray** \( \ell_w \subset \mathbb{C} \) of a point \( w \in \mathbb{C} \setminus \{0\} \) is
\[
\ell_w = (0, \infty) \cdot w = \{ z \in \mathbb{C} \mid \text{Arg} \ w = \text{Arg} \ z \},
\]
and the **slit** of this point is the part of the ray extending outward from \( w \), that is
\[
\sigma_w = [1, \infty) \cdot w = \{ z \in \ell_w \mid |z| \geq |w| \}.
\]

For a polynomial \( f : \mathbb{C} \to \mathbb{C} \), denote the critical points of \( f \) by
\[
\mathcal{C}_f = \{ z \mid f'(z) = 0 \}.
\]
For a regular point \( z_0 \), we shall use \( f^{-1}_{z_0} \) to denote a holomorphic branch of the inverse of \( f \) for which \( f^{-1}(f(z_0)) = z_0 \).

We now discuss the Newton flow, and some notation related to it. Consider the following vector field on \( \mathbb{C} \),
\[
X(z) = \frac{-f(z)}{f'(z)}.
\]

The corresponding flow is called the **Newton flow**. This vector field blows up near the critical points of \( f \). By rescaling the length of the vector \( X(z) \) by \( 2|f'(z)|^2 \), the critical points of \( f \) become well-defined singular points of the rescaled vector field. This rescaled vector field is the gradient vector field \( \hat{z} = -\nabla|f(z)|^2 \); the solution curves of the former coincide with the latter, and we will use the two interchangeably. The equilibria of the Newton flow are exactly the roots and critical points of \( f \). Each root \( \zeta \) is a sink; we shall denote its basin of attraction by **Basin** (\( \zeta \)). Critical points are saddles for the flow. Furthermore, we can extend the flow to infinity, which is the only source. Each boundary component of **Basin** (\( \zeta \)) contains critical points \( c \in \mathcal{C}_f \): generically, each critical point \( c \) has an unstable orbit leaving from \( c \) and converging to \( \zeta \), as well as stable orbits from infinity to \( c \), which are separatrices for the flow. Generically, there is a unique critical point in each boundary component; in the degenerate cases, there could be saddle connections resulting in multiple critical points on one boundary component. A general discussion regarding the Newton flow can be found in [STW] and [JIT], as well as [KoS]. See figure 2.

It is important to note that if \( \varphi(t) \) is a solution curve for the Newton flow, \( f(\varphi(t)) \) lies along a ray. To see this, observe that
\[
\frac{d}{dt} f(\varphi(t)) = f'(\varphi(t)) \cdot \left( \frac{-f(\varphi(t))}{f'(\varphi(t))} \right) = -f(\varphi(t)),
\]
and hence \( f(\varphi(t)) = e^{-t} f(z_0) \) for some \( z_0 = \varphi(t_0) \), provided \( f'(\varphi(t)) \) is never zero. (If \( \varphi(t) \) contains a critical point of \( f \), the result follows by continuity.)

Since \( f \) has distinct roots, \( f'(\zeta) \neq 0 \) for each root \( \zeta \), and so \( f \) is a local diffeomorphism in a neighborhood of \( \zeta \). Thus, for every angle \( \theta \) there will be a solution \( \varphi_\theta(t) \) in this neighborhood with \( \text{Arg} \ f(\varphi_\theta(t)) = \theta \). Noting that the ray \( f(\varphi_\theta(t)) \) extends to infinity unless \( \varphi_\theta(t) \) encounters a critical point \( c \), we obtain the following lemma.

**Lemma 2.1.** For each root \( \zeta \), \( f \) is a biholomorphism
\[
f : \text{Basin} (\zeta) \to \mathbb{C} \setminus \bigcup_{c \in \mathcal{C}_f} \sigma_f(c),
\]
where the union is taken over the critical points \( c \) which lie on the boundary of **Basin** (\( \zeta \)).
Remark 2.2. Observe that iteration of Newton’s method beginning at a point \( z_0 \) corresponds to construction of an approximate solution to the Newton flow with initial condition \( \varphi(0) = z_0 \) using Euler’s method with stepsize \( h = 1 \). When the path \( \gamma \) is a ray in the target space, a path-lifting method corresponds to constructing approximate solutions of the Newton flow via a method that self-corrects to always follow a solution curve that contains the initial condition.

Throughout the paper, we will consider polynomials \( f \in \mathcal{P}_{d,1} \), that is, \( f : \mathbb{C} \to \mathbb{C} \) given by

\[
f(z) = \prod_{j=1}^{d} (z - \zeta_j) \quad \text{with} \ |\zeta_j| \leq 1,
\]

with distinct roots \( \zeta_j \). The set of roots of \( f \) will be denoted by

\[
\mathcal{R}_f = \{ \zeta_j \mid j = 1, \ldots, d \}.
\]

The restriction to \( \mathcal{P}_{d,1} \) is not severe; provided its roots are simple, an affine change of coordinates depending only on the coefficients will transform any polynomial into one in \( \mathcal{P}_{d,1} \) (see [Mar], for example). The space \( \mathcal{P}_{d,1} \) is somewhat different from that considered in other works (such as \( \mathcal{P}_1 \) of [Sm81], \( \mathcal{P}_d(1) \) of [KS], etc), where the space of polynomials is represented as
\( \{ f(z) = \sum a_j z^j \mid |a_j| \leq 1 \} \). In this case, all the roots lie in the disk of radius 2, and our results are readily adapted to any set of polynomials where the roots lie in any disk of a known radius.

We shall use the following standard result several times.

**Lemma 2.3 (Koebe distortion theorem).** Let \( g : D_r(0) \rightarrow \mathbb{C} \) be univalent with \( g(0) = 0 \) and \( g'(0) = 1 \). For \( z \in D_r(0) \) with \( s = |z|/r \), we have

\[
\frac{1 - s}{(1 + s)^3} \leq |g'(z)| \leq \frac{1 + s}{(1 - s)^3}
\]

and

\[
\frac{|z|}{(1 + s)^2} \leq |g(z)| \leq \frac{|z|}{(1 - s)^2}
\]

Consequently,

\[
D_{s/r}(0) \subset g(D_r(0)).
\]

**Remark 2.4.** The statement in equation (2.3) is known as the Koebe \( \frac{1}{4} \)-theorem. The proof can be found in [Ko, Po], or [Du], among others. See also corollary 2.6 of [K88].

### 3. The path-lifting algorithm

In this section, we present the path-lifting algorithm that we use to find an approximate zero of \( f \in \mathcal{P}_{d,1} \). First, we discuss approximate zeros and the Kim–Smale \( \alpha \) function.

**Definition 3.1.** Let \( z_n \in \mathbb{C} \) be the \( n \)th iterate under Newton’s method of the point \( z^* \in \mathbb{C} \), that is,

\[ z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)}, \quad z_0 = z^* \]

The point \( z_n \) is called an **approximate zero** of \( f \) if

\[ |z_{n+1} - z_n| \leq \left( \frac{1}{2} \right)^{n-1} |z_1 - z^*| \quad \text{for all } n > 0. \]

Newton’s method converges quadratically to a root when started from an approximate zero (see [Sm85] for example).

Approximate zeros are an intrinsic, dynamical feature of a polynomial. They form disjoint connected neighborhoods of the roots \( \zeta_i \) on which the Newton map \( N_f(z) = z - f(z)/f'(z) \) converges quadratically to the root, which is a super-attracting fixed point for the rational map \( N_f \).

A sufficient condition for a point to be an approximate zero is developed in [K85] and [Sm86]. We will use the criterion formulated by Smale in [Sm86] to locate approximate zeros. It uses \( \alpha : \mathbb{C} \setminus \mathcal{C}_f \rightarrow \mathbb{R} \) defined by

\[
\alpha(z) = \max_{j>1} \left| \frac{f(z)}{f'(z)} \right|^{\frac{1}{j-1}} \left| \frac{f^{(j)}(z)}{j! f'(z)} \right|^{\frac{1}{j-1}}.
\]
It is sometimes useful to use the related function $\gamma(z)$ instead, where

$$
\gamma(z) = \max_{j \geq 1} \left| \frac{f^{(j)}(z)}{f'(z)} \right|^1.
$$

While we will primarily use $\alpha(z)$, we make use of $\gamma(z)$ in corollary 5.3, sections 6 and 9.

**Theorem 3.2 ([K85, Sm86]).** There is a number $\alpha_0$ such that if $\alpha(z) < \alpha_0$, the point $z$ is an approximate zero.

**Remark 3.3.** It has been shown that $\alpha_0 \geq 3 - \sqrt{8} \approx 0.17157$ (see [WH] or [WZ], for example).

**Remark 3.4.** The number $\alpha_0$ is given in [Sm86] and in many places throughout the literature as $\alpha_0 \approx 0.130707$. However, this specific value is very likely the result of a typographic error in the fifth decimal place. Smale’s bound for $\alpha_0$ is stated as a solution to $(2r^2 - 4r + 1)^2 - 2r = 0$ [Sm86, section 4]; the relevant root of this equation is 0.130716944.

We shall analyze the following algorithm to find an approximate zero for $f \in \mathcal{P}_{d,1}$.

**The $\alpha$-step path-lifting algorithm.**

Input a polynomial $f \in \mathcal{P}_{d,1}$.

**Step 0:** Choose $z_0 \in \mathbb{C}$ with $|z_0| = 1 + \frac{1}{d}$. Let $w_0 = f(z_0)$ and $w = \frac{w_0}{w_0}$.

**Step 1:** Stop if $\alpha(z_0) \leq 3 - \sqrt{8}$; Output $z_0$, an approximate zero for $f$.

**Step 2:** Let

$$
w_{n+1} = w_n - \frac{1}{\pi} \cdot \frac{|f'(z_n)|}{|f(z_n)|} \cdot w
$$

and

$$
z_{n+1} = z_n - \frac{f(z_n) - w_{n+1}}{f'(z_n)}.
$$

Continue with Step 1.

Sometimes we shall refer to the points $w_n$ generated by the algorithm above as guide points or target points.

If $z_0 \in \text{Basin } (\zeta)$ then the $\alpha$-step algorithm will terminate with an approximate zero for $\zeta$. This follows from the fact that $\text{Arg } w_n = \text{Arg } f(z_0)$ for all $n$, and, by the estimates in section 6, $w_{n+1}$ is close enough to $w_n$ to ensure that $f^{-1}$ is univalent on a region containing $w_n$, $w_{n+1}$, $f(z_n)$, and $f(z_{n+1})$. Since $z_0 \in \text{Basin } (\zeta)$, the entire ray $\ell_{w_0}$ lifts to a curve lying in Basin (\zeta) since $\ell_{w_0}$ does not contain a critical value $f(c)$ with $c$ in the closure of Basin (\zeta).

**Remark 3.5.** There may be some values of $n$ for which $z_n \notin \text{Basin } (\zeta)$; even if this occurs, there is a neighborhood $U \subset \mathbb{C}$ of the ray $\ell_{w_0}$ which contains $f(z_j)$ for all $j$ and on which there exists a univalent inverse branch of $f^{-1}$ mapping $w_0$ to $z_0$. As noted in the previous paragraph, $w_{n+1}$ and $f(z_n)$ both lie in a neighborhood of $\ell_{w_0}$ on which $f^{-1}$ is univalent, even if $z_n$ is outside Basin (\zeta). In this case, Basin (\zeta) can be enlarged to a neighborhood $U$ of $\gamma$ which contains all the $z_j$. See figure 3. A more detailed description and explicit construction of $U$ can be found in section 7. Denote this inverse branch by $f^{-1}: U \rightarrow \mathbb{C}$.

**Definition 3.6.** For every zero $\zeta \in \mathcal{R}_f$, let

$$
\rho(\zeta) = \min_{c \in \mathcal{G}_f(\zeta)} |f(c)|
$$

where $\mathcal{G}_f(\zeta) = \mathcal{G}_f \cap \text{Basin } (\zeta)$.
Remark 3.7. Note that \( \rho_\zeta \) is the radius of convergence of \( f^{-1} \) at 0, and is the distance in the surface \( S \) between \( \hat{f}(\zeta) \) and the nearest branch point of \( S \). This follows from the fact that \( \hat{f} : \text{Basin}(\zeta) \to S \setminus V_f \) is a biholomorphism and \( \pi \) is an isometry (see lemma 4.1) from the disk \( D_{\rho_\zeta} \) about \( \hat{f}(\zeta) \) into \( C_{\text{target}} \). Hence, \( f^{-1} : D_{\rho_\zeta}(0) \to C_{\text{source}} \) is a univalent analytic function.

Definition 3.8. For any polynomial \( f \), we define \( K_f = \sum_{\zeta \in \mathbb{C}} \log \frac{1}{\rho_\zeta} \).

Remark 3.9. Notice that \( K_f < \infty \) if and only if the set of roots \( \mathcal{R}_f \) and critical points \( \mathcal{C}_f \) are disjoint. This holds generically for polynomials \( f \), and \( K_f = \infty \) exactly when \( f \) has a multiple zero. Root-finding problems for which there is a multiple zero are typically called ill-conditioned or ill-posed.

Remark 3.10. One can introduce a measure of difficulty \( K_{f,\zeta} = \log 1/\rho_\zeta \) for a specific given root \( \zeta \in \mathbb{C} \). Then theorem 1 describes the cost of reaching an approximate zero for \( \zeta \) in terms of \( K_{f,\zeta} \), theorem 2 gives the cost of finding any approximate zero in terms of the average value of \( K_{f,\zeta} \), and theorem 3 averages \( K_{f,\zeta} \) over all polynomials \( f \) of a given degree.

4. The Voronoi partition in the branched cover

Given a polynomial \( f : \mathbb{C} \to \mathbb{C} \) of degree \( d \), recall from section 2 that we denote its critical points by \( \mathcal{C}_f = \{ z \mid f'(z) = 0 \} \). For any such \( f \), we can express it as a composition \( f = \pi \circ \hat{f} \).
where $\hat{f}$ is a biholomorphism except on $\mathscr{C}_f$ (on which it is merely a bijection), and $\pi$ is a $d$-fold branched cover, ramified at points of $\mathscr{V}_f = \hat{f}(\mathscr{C}_f)$. We denote the metric on $\mathscr{S}$ by $\|\cdot\|$; this metric is such that away from points in $\mathscr{V}_f$, $\pi$ is a local isometry into $\mathbb{C}_{\text{target}}$ (with the standard metric). See also figure 1 and the corresponding discussion in section 1.1.

The multiplicity of a critical point $c \in \mathscr{C}_f$ is

$$m_c = \min \left\{ k \mid f^{(k)}(c) \neq 0 \right\}.$$ 

Notice that

$$\sum_{c \in \mathscr{C}_f} m_c = d - 1.$$ 

The points in $\mathscr{V}_f$ are called critical values in $\mathscr{S}$, and we define the multiplicity $m_v$ of $v = \hat{f}(c) \in \mathscr{V}_f$ to be the multiplicity of $c$; this is also the local degree of the projection $\pi$ in a neighborhood of $v$.

Note that for each root $\zeta \in \mathcal{R}_f$,

$$\pi : \hat{f}(\text{Basin}(\zeta)) \to \mathbb{C} \setminus \bigcup_{y \in \mathcal{V}_f(\zeta)} \sigma_y$$

is an isometry (where $\mathcal{V}_f(\zeta) = f(\mathcal{V}_f(\zeta))$, and $\sigma_y$ is the ray outward from $y$ as defined in section 2).

The Voronoi domain of a point $v \in \mathscr{V}_f$ is

$$\text{Vor}_f(v) = \{ u \in \mathscr{S} \mid \|u, v\| \leq \|u, w\|, \forall w \in \mathscr{V}_f \};$$

this is exactly the set of points $u \in \mathscr{S}$ such that the critical value $\pi(v)$ lies on the boundary of the disk about $\pi(u)$ on which the inverse $f_x^{-1}$ will be analytic ($x$ satisfies $\hat{f}(x) = u$). See also remark 3.7. We will refer to such critical values $\pi(v)$ as relevant to the construction of $f_x^{-1}$.

Recall from section 2 that $D_r(u) = \{ y \mid \|u, y\| < r \}$ denotes the open disk of radius $r$ about $u$. For $u \in \mathscr{S}$, such disks will be isometric to their projections (i.e. be ‘Euclidean disks’) exactly when they avoid the branch points of $\mathscr{S}$.

**Lemma 4.1.** A point $u \in \mathscr{S}$ is in $\text{Vor}_f(v)$ if and only if $\pi : D_{\|u, v\|}(u) \to D_{\|u - v\|}(\pi(u))$ is an isometry. In particular, if $u \in \text{Vor}_f(v)$ then

$$D_{\|u, v\|}(u) \cap \mathscr{V}_f = \emptyset.$$

**Proof.** If $u \in \text{Vor}_f(v)$ then $D_{\|u, v\|}(u) \cap \mathscr{V}_f = \emptyset$. Thus, $\pi$ is a local isometry on all of $D_{\|u, v\|}(u)$, and in particular, $\pi$ is a global isometry on this disk. Conversely, If $\pi$ is an isometry on all of $D_{\|u, v\|}(u)$, there can be no critical values in the disk, and so $u \in \text{Vor}_f(v)$. 

Let $u_1, u_2 \in \mathscr{S}$. If the line segment $[\pi(u_1), \pi(u_2)] \subset \mathbb{C}$ has a lift in $\mathscr{S}$ which connects $u_1$ with $u_2$, we denote this lifted line segment by $[u_1, u_2]$. Observe that many pairs $u_1, u_2$ do not
have such a connecting line segment. In this case we write \[ \left[ u_1, u_2 \right] = \emptyset. \] When \( \left[ u_1, u_2 \right] \) is non-empty, we say that \( u_1 \) is visible from \( u_2 \) in \( S \). Also observe, if \( v \in Vf \) then \( \left[ u, v \right] \neq \emptyset \) for all \( u \in Vr \).

We can form the visibility graph for \( S \) as follows. The vertices of the graph are the critical values \( Vf \), and there is an edge from \( v \) to \( w \) if and only if \( \left[ v, w \right] \) is non-empty. We can identify the visibility graph with the subset of \( S \) given by

\[ G = \bigcup_{v, w \in Vf} \left[ v, w \right]. \]

Since \( \hat{f} \) is a bijection between \( C_{\text{source}} \) and \( S \), \( \hat{f}^{-1}(G) \) is well-defined, so we can also view \( G \) as a graph immersed in \( C_{\text{source}} \), with the critical points of \( f \) as vertices.

**Question 4.2.** Characterize the graphs which occur as a visibility graph \( G \) for a polynomial.

Recall from section 2 that the ray \( \ell_y \subset \mathbb{C} \) of a point \( y \in \mathbb{C} \setminus \{0\} \) is the set of points which have the same argument as \( y \).

If \( \hat{0} \in S \) projects onto \( 0 \) and \( \left[ \hat{0}, u \right] \neq \emptyset \), the geodesic starting at \( \hat{0} \) and containing \( \left[ \hat{0}, u \right] \) is the ray through \( u \in S \), which we denote by \( \ell_u \). Observe that if \( \ell_u \cap Vf = \emptyset \) then \( \pi : \ell_u \to \ell_{\pi(u)} \) is a surjective isometry.
Let \( y = \pi(u) \). If \( \ell_y \cap f(Cf) = \emptyset \), then
\[
\pi^{-1}(\ell_y) = \hat{\ell}_{y_1} \cup \hat{\ell}_{y_2} \cup \cdots \cup \hat{\ell}_{y_d},
\]
where the points \( y_i \in \mathcal{S} \) are the \( d \) different preimages of \( y \).

**Proposition 4.3.** Given \( v \in \mathcal{V} \) and \( y \in C \setminus f(Cf) \). Then
\[
\text{card}\{ i \mid \hat{\ell}_{y_i} \cap \text{Vor}(v) \neq \emptyset \} \leq m_v + 1.
\]
Furthermore, each \( \hat{\ell}_{y_i} \cap \text{Vor}(v) \) is a connected set.

**Proof.** Suppose \( \hat{\ell}_{y_1}, \hat{\ell}_{y_2}, \ldots, \hat{\ell}_{y_d} \) intersect \( \text{Vor}(v) \), with \( v = \hat{f}(c), c \in \mathcal{C} \). Pick a point \( u_i \) in each of these intersections, that is,
\[
u_i \in \hat{\ell}_{y_i} \cap \text{Vor}(v).
\]
Let \( D_i = D_{\|v-p_i\|}(u_i) \). According to lemma 4.1, we know that \( \pi : D_i \rightarrow \pi(D_i) \) is an isometry. Let \( p_i \in \hat{\ell}_{y_i} \) be the perpendicular projection of \( v \) onto \( \hat{\ell}_{y_i} \) and let \( p \) be the projection of \( f(c) = \pi(v) \) onto \( \ell_y \) (see figure 6). Then for all \( i \leq k \), \( \pi(p_i) = p \),
\[
\emptyset \neq [v,p_i] \subset D_i \quad \text{and} \quad \emptyset \neq [\pi(v),p] \subset \bigcap_{i \leq k} \pi(D_i).
\]
Since for each $i$ between 1 and $k$, $\pi$ is a surjective isometry from $[v, p_i]$ to $[\pi(v), p]$, $k$ can be no larger than the degree of $\pi$ on a neighbourhood of $v$. That is,
\[ k \leq 1 + m_v. \]

The connectedness of $\hat{\ell}_v \cap \text{Vor}(v)$ follows from the triangle inequality. □

**Corollary 4.4.** Each projection $\pi : \text{Vor}(v) \to \mathbb{C}$ is at most $(m_v + 1)$-to-one.

Let $z \in \mathbb{C}$. We will say that a critical point $c \in \mathbb{C}$ *influences the orbit of* $z$ if the segment $[0, \hat{f}(z)]$ passes through $\text{Vor}(\hat{f}(c))$.

We are interested in the critical points which influence the starting points for our algorithm, and, conversely, the starting points which are influenced by a given critical point.

**Definition 4.5.** For starting points $z$ on the circle of radius $r$, we define the following sets:

\[ \mathcal{S} = \left\{ (t, c) \in [0, 1] \times \mathbb{C} \mid [0, \hat{f}(r e^{2\pi i t})] \cap \text{Vor}(\hat{f}(c)) \neq \emptyset \right\} \]

\[ \mathcal{I} = \{ c \in \mathcal{S} \mid (t, c) \in \mathcal{S} \} \quad \mathcal{J} = \{ t \in [0, 1] \mid (t, c) \in \mathcal{S} \} \]

Notice that, for $z = re^{2\pi it}$ fixed, we have $c \in \mathcal{I}$ precisely when, for some $y \in \ell_{\hat{f}(c)}$, $D_{\hat{f}(c) - y}(y)$ is the largest ball on which $f_c^{-1}$ is defined. Similarly, for this pair $(t, c)$, we also have $t \in \mathcal{J}$. 

---

**Figure 6.** As proven in proposition 4.3, the projection $\pi$ is at most $(m_v + 1)$-to-one on $\text{Vor}(v)$.
5. The behavior of \( f \) on the initial circle

Consider the function \( a_r : [0, 1) \rightarrow \mathbb{R} \) defined by

\[
a_r(t) = \text{Arg}(re^{2\pi it}),
\]

with \( r > 0 \). We can easily bound the rate of change of \( a_r(t) \); while elementary, these bounds play a crucial role for us.

**Lemma 5.1 (Angular speed lemma).** Let \( r > 1 \). Then for all \( t \in [0, 1) \), we have

\[
2\pi d \cdot \frac{r}{r+1} \leq a_r'(t) \leq 2\pi d \cdot \frac{r}{r-1}.
\]

**Proof.** Let \( z = re^{2\pi it} \), with \( r > 1 \). Since \( |z| \leq 1 \), we have \( \frac{z}{\xi} \in D_1(0) = \{ w \mid |w| \leq \frac{1}{r} \} \).

A calculation shows

\[
a_r'(t) = \text{Im} \left( \frac{d}{dz} \log(f(re^{2\pi it})) \right) \cdot 2\pi i
\]

\[
= 2\pi \cdot \text{Re} \left( \frac{f'(z)}{f(z)} \right) = 2\pi \cdot \text{Re} \sum_{j=1}^{d} \frac{z}{z-\xi_j}.
\]

(5.1)

For each root \( \xi_i \), we have

\[
\frac{r}{r+1} \leq \text{Re} \left( \frac{1}{1-\xi_i/z} \right) \leq \frac{r}{r-1}.
\]

Summing this inequality over the \( d \) roots and applying it to equation (5.1) gives the desired result. \( \square \)

**Remark 5.2.** The estimates in lemma 5.1 are sharp.

The following bounds \( \alpha(z) \) for points on the initial circle. This will be of use in proving lemma 6.10, used in selecting starting points to locate all \( d \) roots of \( f \) in section 10.

**Corollary 5.3.** For \( z \) with \( |z| = 1 + 1/d \), we have

\[
\left| \frac{f(z)}{f'(z)} \right| < \frac{3}{d}, \quad \gamma(z) \leq \frac{d(d-1)}{2}, \quad \text{and} \quad \alpha(z) < \frac{3}{2}(d-1).
\]

**Proof.** Since \( r = |z| = 1 + 1/d \), lemma 5.1 gives us \( \pi d < a_r' \). From this and the observation that \( \text{Re}(w) \leq |w| \), we have

\[
\pi d < \left| \frac{f'(z)}{f(z)} \right| \cdot 2\pi(1 + \frac{1}{d}) < 3\pi \left| \frac{f'(z)}{f(z)} \right|, \quad \text{and so} \quad \left| \frac{f(z)}{f'(z)} \right| < \frac{3}{d}.
\]

Note that if \( \xi_i \) are the \( k \) solutions to \( f^{(k-1)}(\xi_j) = 0 \) (with multiplicity), then by Lucas’ theorem [Lu], we have each \( \xi_i \) in the unit disk and so \( |z - \xi_i| \geq 1/d \). Thus

\[
\left| \frac{f^{(k)}(z)}{f^{(k-1)}(z)} \right| = \sum_{j=1}^{k} \frac{1}{|z - \xi_i|} \leq d(d-k).
\]

Observe that
Since $\gamma(z)$ is the maximum of the above expression over $f$, we have $\gamma \leq \frac{1}{2}d(d-1)$; combining the two estimates also gives $\alpha(z) < \frac{1}{2}(d-1)$. \hfill $\square$

The corollary below has direct implications for path-following methods that use a constant ratio step-size (such as [Sm85] or [KS]), which need a cone of a given angular width about $\ell_w$ containing no (relevant) critical values in order to set the stepsize that ensures convergence. The $\alpha$-step algorithm considered here adjusts for the presence of critical values (unless they fall on $\ell_w$) and does not need a constant width cone, although a starting value lying in $\text{Bad}_\theta$ will have a contribution of at least $\log(1/\theta)$ to the arithmetic complexity caused by the corresponding critical point $c$. Recall from definition 4.5 that $c \in \mathcal{F}_t$ means that the segment $[\emptyset, \widehat{f}(re^{2\pi i})] \in \mathcal{F}$ intersects $\text{Vo}(c)$.

**Corollary 5.4.** Let $r = 1 + 1/d$, and define

$$\text{Bad}_\theta = \left\{ t \in [0, 1) \mid \left| \text{Arg} \frac{f(re^{2\pi i})}{f(c)} \right| < \theta, \text{ for some critical point } c \in \mathcal{F}_t \right\}.$$ 

Then

$$\text{measure}(\text{Bad}_\theta) \leq \frac{2\theta}{\pi} \cdot \frac{d-1}{d}.$$ 

**Proof.** For fixed $r$, the set $\text{Bad}_\theta$ consists of the inverse image by $\widehat{f}$ of $d-1$ arcs of angle $2\theta$ in $\mathcal{F}$ (one for each critical point). Each of these will grow by no more than $1/\min a'_t(t)$, so by lemma 5.1, when $r = 1 + 1/d$ we have

$$\text{measure}(\text{Bad}_\theta) \leq \sum_{c \in \Theta_t} \frac{2\theta}{\max a'_t(t)} \leq (d-1) \frac{r+1}{2\pi rd} = (d-1) \frac{\theta(2d+1)}{\pi d(d+1)} \leq \frac{2\theta(d-1)}{\pi d}.$$ 

$\square$

Recall that here we are using the convention that the circle has measure 1.

**Remark 5.5.** Let $\text{Good}_\theta$ be the complementary notion to $\text{Bad}_\theta$, that is,

$$\text{Good}_\theta = \left\{ t \in [0, 1) \mid \left| \text{Arg} \frac{f(re^{2\pi i})}{f(c)} \right| \geq \theta, \text{ for all critical points } c \in \mathcal{F}_t \right\}.$$ 

For each $t \in \text{Good}_\theta$, $f_{re^{2\pi i}}^{-1} : \mathbb{C}_{\text{target}} \to \mathbb{C}_{\text{source}}$ will be analytic in a cone

$$\left\{ w \in \mathbb{C}_{\text{target}} \mid \left| \text{Arg} (w) - \text{Arg} (f(re^{2\pi i})) \right| \geq \theta \right\},$$

and consequently such $t$ correspond to ‘good starting points’ for a path-lifting algorithm: in a method with a fixed-ratio stepsize, the convergence is assured, and for the $\alpha$-step algorithm, convergence is rapid.

This is essentially Condition $\Theta$ of [Sm85] and [SS86], with $\theta = \pi/12$. Both these works use $V_f$ to denote our $\text{Good}_{\pi/12}$ (also taking $r = 3/2$), and show in proposition 2 that $\text{Good}_{\pi/12}$
has measure at least $1/6$. Above in corollary 5.4, we show that the measure of $\text{Good}_{1/12}$ is at least $5/6$.

Recall from section 2 that the circle of radius $r$ is denoted by $S_r = \{ z \mid |z| = r \}$.

**Lemma 5.6.** Let $c$ be a critical point on the boundary of $\text{Basin} (\zeta)$, and let $\gamma_c$ be the solution to the Newton flow emanating from $c$ whose interior lies in $\text{Basin} (\zeta)$. Then if $r > 1$, $\gamma_c \cap S_r = \emptyset$. 

**Proof.** Note that the Newton flow points inward on $S_r$ for $r > 1$, which follows from the observation that

$$\frac{f(z)}{f'(z)} = \sum_{z \in \zeta} \frac{1}{z - c}. $$

The uniqueness of $\gamma_c$ follows from lemma 2.1 (which says that $f$ is a biholomorphism from $\text{Basin} (\zeta)$ onto a slit plane) and the observation that $f$ sends solutions into rays; if there were two solutions $\gamma_c$ and $\varphi_c$ both emanating from $c$ and lying in $\text{Basin} (\zeta)$, $f(\gamma_c)$ and $f(\varphi_c)$ would coincide near 0, and thus $\gamma_c = \varphi_c$.

The transversality and uniqueness facts immediately imply lemma 5.6.

The transversality of the Newton flow to $S_r$ appears in many places (e.g. [STW]), but we include a justification here. Observe that since $|z| > 1$ and $|\zeta| \leq 1$, the vectors $z - \zeta$ all lie in a half-plane $\mathcal{H}$ which does not include the origin. Consequently, their inverses and hence their sum $\sum 1/(z - \zeta)$ lie in a (possibly different) half-plane $\mathcal{H}'$. Inverting again gives $f(z)/f'(z) \in \mathcal{H}'$. Since $f(z)/f'(z)$ lies in the original half-plane $\mathcal{H}$, it is transverse to $S_r$. □

Observe that $\text{Basin} (\zeta) \cap D_1(0)$ will consist of one or more connected components. The following lemma enables us to estimate the width of these.

**Lemma 5.7.** Let $r > 1$, $\zeta \in \mathcal{R}_f$, and let $\nu$ be a connected component of $S_r \cap \overline{\text{Basin} (\zeta)}$. Then

$$\text{length} (\nu) \cdot \min \alpha'_1(t) \leq 2\pi r,$$

where the minimum is taken over points with $\text{re}^{2\pi i} \in \nu$.

**Proof.** Let $B \subset \text{Basin} (\zeta)$ be a boundary component of $\text{Basin} (\zeta)$ which does not intersect $\nu$, and let $c$ be a critical point of $f$ contained in $B$. Let $\gamma_c$ be the orbit of the Newton flow which begins at $c$ and ends at the root $\zeta$; then $\gamma_c \setminus \{ c \}$ will be contained in $\text{Basin} (\zeta)$ since $f(\gamma_c) \subset \mathbb{C}_{\text{target}}$ is the segment $(0, f(c))$.

Observe that $f(\gamma_c \cup B)$ is exactly the ray through $f(c)$. From the definition of $\nu$ and lemma 5.6 we get $\text{int} (\nu) \cap (B \cup \gamma_c) = \emptyset$. Hence,

$$\text{Arg} (f(\text{int} (\nu))) \cap \text{Arg} (f(c)) = \emptyset,$$

that is, the image of $\nu$ cannot make more than a full turn in the target space. The lemma follows. □

The following corollary follows immediately from the proof.

**Corollary 5.8.** Let $z_1$ and $z_2$ satisfy $|z_1| = |z_2| = r$ with $r > 1$, and suppose also that they lie in the same connected component of $S_r \cap \text{Basin} (\zeta)$. Then there is a well-defined branch of the argument $\text{Arg}$ which is continuous on $S_r \cap \text{Basin} (\zeta)$ and such that

$$|\text{Arg} f(z_1) - \text{Arg} f(z_2)| \leq 2\pi.$$
In the sequel we will consider integrals over the circle $S_r = \{ z \in \mathbb{C} \mid |z| = r \}$, which, for all $r > 0$, carries Lebesgue measure with unit mass.

We require the following lemma and its corollary in the proofs of lemmas 9.4 and 8.4.

**Lemma 5.9.** Let $r > 0$ and $|\zeta| < r$ then
\[ \int_0^1 \log |re^{2\pi i t} - \zeta| dt = \log r. \]

**Proof.** Define
\[ S(\zeta) = \int_0^1 \log |re^{2\pi i t} - \zeta| dt = \int_{S_r} \text{Re} (\log(z - \zeta)) \cdot \frac{1}{2\pi i} \frac{dz}{z}, \]

Note that
\[ \frac{ds}{d\zeta} = \text{Re} \frac{1}{2\pi i} \int_{S_r} \frac{1}{z - \zeta} dz = \text{Re} \frac{1}{2\pi i} \int_{S_r} \left( \frac{1}{z - \zeta} - \frac{1}{\zeta} \right) dz = 0. \]

Hence,
\[ S(\zeta) = S(0) = \log r. \]

The following corollary is needed in the proof of lemma 8.4, but is also interesting in its own right.

**Corollary 5.10.** Let $f(z) = \prod_{j=1}^d (z - \zeta_j)$, with $|\zeta_j| < r$. Then
\[ \int_0^1 \log |f(re^{2\pi i t})| dt = d \log r. \]

**Remark 5.11.** Notice that if $r = 1 + 1/d$, we have $d \log r < 1$.

**Proof.**
\[ \int_0^1 \log |f(re^{2\pi i t})| dt = \int_0^1 \log \left| \prod_{j=1}^d (re^{2\pi i t} - \zeta_j) \right| dt = \sum_{j=1}^d \int_0^1 \log |re^{2\pi i t} - \zeta_j| dt = d \log r, \]

where the last equality follows from lemma 5.9.

**Question 5.12.** The previous corollary shows that the average value of $\log |f(z)|$ on $S_r$ is $d \log r$. Is there a constant $c_r$ independent of $d$ so that
\[ \text{measure} \{ t \mid \log |f(re^{2\pi i t})| < d \log r \} > c_r? \]

We now establish a lower bound on $|w_0| = |f(z_0)|$ for starting points $z_0$ on the circle $S_r$ with $r > 1$. We shall use this in lemma 6.9 to give a lower bound on the size of our final point $w_N$. The existence of such a bound should be expected, since $z_0$ is taken outside the disk containing all the roots; we need this result in the proof of theorem 2 to handle the case where $z_0$ is already an approximate zero of $f$. 

433
Proposition 5.13. Let $z \in \text{Basin } (\zeta)$ with $|z| = r > 1$. Then

$$|f(z)| \geq s_r \cdot \rho_{\zeta},$$

where $\rho_{\zeta}$ is the radius of convergence of the branch of $f^{-1}$ taking $0$ to $\zeta$, and $s_r < 1$.

If $r > 1 + \frac{2\pi}{\rho_{\zeta}}$, $s_r = \frac{1}{2}$. Otherwise, for $r = 1 + \frac{\zeta}{\rho_{\zeta}}$, $s_r$ is the smallest positive solution of

$$C = 8\pi \frac{s}{(1-s)^2},$$

Remark 5.14. For $0 < C \leq 2\pi$, we have $0 < s_r \leq 3 - \sqrt{8}$. For $C = 1$, we have $s_r \approx 0.0369 > \frac{1}{2\pi}$.

Proof. Without loss of generality, we may assume that $\zeta$ is a non-negative real number. Define $l$ to be the radius of the largest disk centered at $\zeta$ which is mapped univalently into $D_{\rho_{\zeta}}(0)$, that is,

$$D_l(\zeta) \subset f_{\zeta}^{-1}(D_{\rho_{\zeta}}(0)).$$

Observe that these lie entirely inside $\text{Basin } (\zeta)$.

Applying the Koebe $\frac{1}{4}$-lemma (equation (2.3)) to $f_{\zeta}^{-1}$, we then obtain

$$l \geq \frac{1}{|f'(\zeta)|} \cdot \rho_{\zeta} \cdot \frac{1}{4}. \tag{5.2}$$

Let $z$ be a point in $\text{Basin } (\zeta)$ with $|z| = r$.

First consider the case $|z - \zeta| \geq l$. Here, we must have $|f(z)| \geq \rho_{\zeta}/4$. If not, the Koebe $\frac{1}{4}$-lemma is violated: by definition of $l$, the map $f$ is univalent on $D_l(\zeta)$ and so $f(D_l(\zeta))$ contains a disk of radius $\rho_{\zeta}/4$ about $0$. Thus, we need only consider the case when $|z - \zeta| < l$.

Observe that the function $g(w) = (f_{\zeta}^{-1}(w) - \zeta)f'(\zeta)$ satisfies the hypotheses of the Koebe distortion theorem (lemma 2.3) on the disk of radius $\rho_{\zeta}$. Take $w = f(z)$ to obtain

$$|z - \zeta||f'(\zeta)| \leq \frac{|f(z)|}{(1-s)^2} \quad \text{or, equivalently} \quad |z - \zeta| \leq \frac{1}{|f'(\zeta)|} \cdot \rho_{\zeta} \cdot \frac{s}{(1-s)^2}, \tag{5.3}$$

where $s = |f(z)|/\rho_{\zeta}$.

We now look for a lower bound on $|z - \zeta|$ by estimating $\frac{|z - \zeta|}{r}$ for $z \in S_r \cap D_l(\zeta)$.

Since we have $z \in D_l(\zeta)$ and also $|z| > 1$, there is a point $A \in S_1 \cap D_l(\zeta)$; let $\phi$ be the angle of the sector connecting $0, A$, and $1$. See figure 7.

Notice that

$$l = \sqrt{\zeta^2 - 2\zeta \cos(\phi) + 1}, \quad \text{since} \quad (\cos \phi - \zeta)^2 + \sin^2 \phi = l^2$$

where $(\cos(\phi), \sin(\phi))$ is the coordinate of the point $A$ on $S_l(\zeta) \cap S_l$.

From corollary 5.8, we have $|\text{Arg } (f(A)) - \text{Arg } (f(A))| \leq 2\pi$, and by the angular speed lemma (lemma 5.1), we have
\[ \phi = \text{Arg}(A) \leq \frac{\pi}{d} \cdot \frac{r + 1}{r} \leq \frac{2\pi}{d}, \text{ for all } r > 1. \]

Since \( r = 1 + \frac{C}{d} \) and \( 0 < \phi \leq \pi \), we have
\[ |z - \zeta| \geq \frac{1 + \zeta - \zeta}{\sqrt{\zeta^2 - 2\zeta \cos(\phi) + 1}} \geq \frac{1 + \zeta - \zeta}{\sqrt{\zeta^2 - 2\zeta \cos(\frac{2\pi}{d}) + 1}}. \]

Since we are only considering \( 0 < C < 2\pi \) and \( |\zeta| \leq 1 \), the above expression is minimized when \( \zeta = 1 \). Hence, we have
\[ \frac{|z - \zeta|}{l} \geq \frac{\zeta}{\sqrt{1 - 2\cos(\frac{2\pi}{d})} + 1} \geq \frac{C}{2\pi}, \]
for all \( d \). Using this with equation (5.2), we obtain
\[ |z - \zeta| \geq \frac{C l}{2\pi} \geq \frac{C}{2\pi} \cdot \frac{\rho_\zeta}{4|f'\zeta|}. \]

This, together with the estimate from equation (5.3), gives the lower bound on \( s \) as the solution to
\[ \frac{C}{2\pi} \cdot \frac{\rho_\zeta}{4|f'\zeta|} \leq \frac{s}{(1 - s)^2} \cdot \frac{\rho_\zeta}{|f'\zeta|}, \]
which simplifies as
\[ C \leq 8\pi - \frac{s}{(1 - s)^2}. \]

Denote the smaller positive solution of the above by \( s_r \). Since \( s \) was defined by \( s = |f(z)|/\rho_\zeta \), this gives us \( |f(z)| \geq s_r \cdot \rho_\zeta \), as desired.
6. The size of the step

Recall that the \( \alpha \)-step algorithm (see section 3) generates a sequence of points \( z_n \) with

\[
\zeta_{n+1} = \zeta_n - \frac{f(z_n) - w_{n+1}}{f'(z_n)},
\]

where the \( w_n \) are a sequence of points tending towards 0 with the same argument as \( w_0 = f(z_0) \).

In this section, for notational convenience we will sometimes write \( f_n \) for \( f(z_n) \), \( f_n' \) for \( f'(z_n) \), \( \alpha_n \) for \( \alpha(z_n) \), and so on.

We call the distance between \( w_{n+1} \) and \( w_n \) the \textit{nth-jump} and denote it by

\[
\delta_n = |w_{n+1} - w_n| = A \cdot \frac{|f(z_n)|}{\alpha(z_n)}. \]

The coefficient \( A \) (and hence \( w_{n+1} \)) must be chosen so that \( f(z_n) \) will lie close enough to \( w_n \) to ensure that the algorithm efficiently follows the ray \( \ell_{wn} \). In particular, we show in proposition 6.7 that taking \( A = \frac{1}{15} \) gives us \( J_n \geq r_n/66 \), where \( r_n \) is the radius of convergence of the appropriate branch of \( f^{-1} \) centered at \( w_n \). The proof of this uses induction; the inductive hypothesis is established in proposition 6.1.

If \( f \) were linear, the algorithm would follow \( w_n \) exactly, and \( f(z_n) \equiv w_n \). When the degree of \( f \) is at least 2, there will be a small error which we denote by

\[
\delta_n = |f(z_n) - w_n|.
\]

While the algorithm is described in terms of \( C_{\text{source}} \) (the \( z_n \)) and \( C_{\text{target}} \) (\( f(z_n) \)) and the \( w_n \), it is more straightforward to think of it in terms of the branched surface \( \mathcal{S} \).

Let \( r \geq 0 \) be maximal such that

\[
f^{-1}_n : D_r(w_n) \to U
\]

is univalent, where \( U \) is a neighborhood of \( z_n \). This is the distance between \( \hat{w}_n \in \mathcal{S} \) and the critical value \( v \in \mathcal{V} \) for which \( \hat{w}_n \in \text{Vor}(v) \). Also, let \( R_n \geq 0 \) be maximal such that

\[
f^{-1}_n : D_{R_n}(f_n) \to V
\]

is univalent, where \( V \) is a neighborhood of \( z_n \). Note that \( \hat{f}_n \) could be in \( \text{Vor}(v') \) for a critical value different from that used for \( \hat{w}_n \); in this case, we still use \( R_n = |v' - f_n| \).

We introduce the following notation, used throughout this section.

\[
\epsilon_n = z_n - z_{n+1} \quad \text{and} \quad h_n = (z_n - z_{n+1}) \cdot \frac{f_n'}{f_n} = \epsilon_n \cdot \frac{f_n'}{f_n}.
\]

As noted earlier, we use \( f_n = f(z_n) \), \( f_n' = f'(z_n) \), \( f_n'' = f''(z_n) \), and \( f_n^{(j)} = f^{(j)}(z_n) \) as notation for the derivatives of \( f \) at \( z_n \), and use \( \alpha_n = \alpha(z_n) \). Let \( \gamma_n = \gamma(z_n) \), where

\[
\gamma(z) = \max_{j > 1} \frac{|f^{(j)}(z)|}{|f'(z)|},
\]

as defined in section 3; hence \( \alpha_n = \gamma_n |f_n|/|f_n'| \).

**Proposition 6.1.** Using the preceding notation, suppose we have \( A > 0 \) and \( c > 0 \) given by

\[
\delta_n < c \cdot \frac{|f_n'|}{\gamma_n} \quad \text{and} \quad |w_{n+1} - w_n| = A \cdot \frac{|f_n|}{\alpha_n},
\]

436
Let $\psi(u) = 1 - 4u + 2u^2$. Then if $A + c$ satisfies $(A + c)^2 < c\psi(A + c)^2$, we have

$$\delta_{n+1} < c \cdot \frac{|f'_{n+1}|}{\gamma_{n+1}}.$$

In order to establish this, we need some preparatory lemmas.

**Lemma 6.2.** If $|\alpha_nh_n| < 1$ then

$$\delta_{n+1} = |f_{n+1} - w_{n+1}| \leq |h_nf_n| \cdot \frac{|\alpha_nh_n|}{1 - \alpha_nh_n}.$$

**Proof.** Note that since $z_{n+1} = z_n - \frac{f_n - w_{n+1}}{f_n}$, we have $w_{n+1} = f_n - (z_n - z_{n+1})f'_{n} = (1 - h_n)f_n$.

Thus,

$$\delta_{n+1} = |f_{n+1} - (1 - h_n)f_n| = |f(z_n + \epsilon_n) - (1 - h_n)f_n|
= \left| f_n + f'_{n}\epsilon_n + \frac{f''_{n}}{2!}\epsilon_n^2 + \cdots + f_n + h_nf_n \right|
= \left| h_nf_n \right| \cdot \left| f_n + f'_{n}\epsilon_n + \frac{f''_{n}}{3!}\epsilon_n^3 + \cdots \right|
\leq |h_nf_n| \cdot \left| \alpha_nh_n + (\alpha_nh_n)^2 + \cdots \right|
\leq |h_nf_n| \cdot |\alpha_nh_n| \cdot \frac{|\alpha_nh_n|}{1 - \alpha_nh_n}.
\square

The proof of the following lemma can be found in [BCSS] (lemma 8.2b and proposition 8.3b).

**Lemma 6.3.** Let $u_n = \alpha_nh_n$ and $\psi(u) = 1 - 4u + 2u^2$. Then if $u_n < 1 - 1/\sqrt{2}$, we have

$$\frac{f''_{n}}{f'_{n+1}} \leq \frac{(1 - u_n)^2}{\psi(u_n)} \quad \text{and} \quad \frac{\gamma_{n+1}}{\gamma_n} \leq \frac{1}{(1 - u_n)\psi(u_n)}.$$

**Remark 6.4.** In [BCSS], $u_n$ is defined as $(z_n - z_{n+1})\gamma_n$. We use

$$h_n = \frac{f_n - w_{n+1}}{f_n} = (z_n - z_{n+1})\frac{f_n}{f_n},$$

and so our usage and that of [BCSS] agree.

We are now ready for the proof of proposition 6.1.

**Proof of proposition 6.1.** First, observe that if $A$ and $c$ satisfy

$$\delta_n < c \cdot \frac{|f'_n|}{\gamma_n} \quad \text{and} \quad |w_{n+1} - w_n| = A \cdot \frac{|f_n|}{\alpha_n},$$

(6.1)
we have $|\alpha_n h_n| \leq A + c$.

\[
|h_n f_n| = |f_n - w_{n+1}| \\
\leq |w_n - w_{n+1}| + |f_n - w_n| \\
\leq J_n + \delta_n \\
\leq A \cdot \frac{|h_n|}{\alpha_n} + c \cdot \frac{|f_n|}{\alpha_n} = (A + c) \cdot \frac{|h_n|}{\alpha_n}.
\]

(6.2)

We impose the further condition

\[
A + c < 1 - \frac{1}{\sqrt{2}}
\]

which allows us to apply lemma 6.3; this also ensures that the hypothesis of lemma 6.2 is satisfied.

Since $\alpha_n = \gamma_n \cdot \frac{|f'|_{n}}{f_n}$, by equation (6.2) we have

\[
|h_n f_n| \leq (A + c) \frac{|f'|_{n}}{\gamma_n}.
\]

In lemma 6.2, we obtained

\[
\delta_{n+1} \leq \frac{h_n h_n}{\alpha_n} \frac{\alpha_n h_n}{1 - \alpha_n h_n} \leq (A + c) \frac{|f'|_{n}}{\gamma_n} \cdot \frac{\alpha_n h_n}{1 - \alpha_n h_n}.
\]

Thus, it is sufficient to impose the condition

\[
(A + c) \frac{|f'|_{n}}{\gamma_n} \cdot \frac{\alpha_n h_n}{1 - \alpha_n h_n} \leq c \cdot \frac{|f'|_{n+1}}{\gamma_{n+1}},
\]

or equivalently,

\[
(A + c) \cdot \frac{\gamma_{n+1}}{\gamma_n} \cdot \frac{|f'|_{n}}{|f'|_{n+1}} \cdot \frac{1}{c} \cdot \frac{\alpha_n h_n}{1 - \alpha_n h_n} < 1.
\]

From lemma 6.3, after simplification we obtain

\[
(A + c) \frac{\alpha_n h_n}{\psi(\alpha_n h_n)^2} \cdot \frac{1}{c} < 1.
\]

Since $\alpha_n h_n \leq A + c$ and $u / \psi(u)$ increases monotonically for $u \in [0, 1 - 1/\sqrt{2}]$, we must have

\[
\frac{(A + c)^2}{\psi(A + c)^2} \cdot \frac{1}{c} < 1.
\]

(6.3)

Thus, if $A$ and $c$ satisfy the hypotheses of the proposition, the conclusion follows.

\[
\square
\]

**Remark 6.5.** To optimize the speed of the algorithm, we need to find the largest $A > 0$ for which there is a $c > 0$ such that the pair $(A, c)$ satisfies inequality Equation (6.3). Numerics show that such solutions exist for $A < 0.0703039 < 1/14.22396$; one can readily check that taking $A = 1/15$ and $c = 1/74$ satisfies the conditions. We will use these values of $A$ and $c$ henceforth.
In order to prove proposition 6.7, we need the following lemma, which is essentially corollary 4.3 of [K88]; the lower bound of $\frac{1}{4}$ follows from the extended Löwner’s theorem in [Sm81]. See also [DKST], where the same constant is obtained for the inverse of an analytic map between Hilbert spaces.

**Lemma 6.6.**

$$\frac{1}{4} \cdot R_n \leq \left| \frac{f_n}{\alpha_n} \right| \leq \frac{R_n}{3 - \sqrt{8}}$$

**Proposition 6.7.** If in the $\alpha$-step algorithm, we choose $w_{n+1}$ along $\ell_{w_0}$ so that

$$J_n = \left| w_n - w_{n+1} \right| = \frac{1}{15} \cdot \left| \frac{f_n}{\alpha_n} \right|,$$

we have $J_n \geq \frac{1}{66} \cdot r_n$ for all $n$.

**Proof.** First, observe that since $w_0 = f_0$, we have $\delta_0 = 0$.

Applying proposition 6.1 with $A = 1/15$ and $c = 1/74$ then gives us

$$\delta_n \leq \frac{1}{74} \cdot \left| \frac{f_n}{\alpha_n} \right|$$

for all $n \geq 0$.

From lemma 6.6, we get

$$J_n = A \cdot \frac{\left| f_n \right|}{\alpha_n} \geq \frac{1}{15} \cdot \left| \frac{f_n}{\alpha_n} \right| \geq \frac{R_n}{4 |f_n|} = \frac{1}{60} \cdot \frac{R_n}{r_n} \cdot r_n.$$

The radius of convergence at $w_n$ is

$$r_n = \left| w_n - v_n \right|,$$

where $v_n$ is the critical value for which $\hat{w}_n \in S$ lies in $\text{Vo}(v_n)$. It might be that the radius at $f_n$ is determined by another critical value, say

$$R_n = \left| f_n - v_n' \right|.$$

Let $r'_n = \left| w_n - v_n' \right|$. Then we have

$$r_n \leq r'_n \leq \left| v_n' - f_n \right| + \left| f_n - w_n \right| = R_n + \delta_n.$$

In the case when $v_n = v_n'$ we get the same estimate for $r_n$. Notice, by using equation (6.4) and lemma 6.6,

$$r_n \leq R_n + \delta_n \leq R_n + \frac{1}{74} \cdot \left| \frac{f_n}{\alpha_n} \right| \leq R_n + \frac{1/74}{3 - \sqrt{8}} \cdot R_n = \frac{3 - \sqrt{8}}{3 - \sqrt{8}} \cdot R_n.$$

Consequently, we have

$$J_n \geq \frac{3 - \sqrt{8}}{3 - \sqrt{8} + 1/74} \cdot \frac{r_n}{60} > \frac{r_n}{66},$$

as desired. \[\square\]
The following corollary tells us how well \( f_n \) tracks \( w_n \) and how \( w_{n+1} \) relates to \( w_n \) as the algorithm progresses. We use this below in order to estimate the size of our final guide point \( w_N \).

**Corollary 6.8.** If \( \alpha_n > 3 - \sqrt{8} \), then
\[
|f_n| \leq \frac{35}{38} |w_n| \quad \text{and} \quad |w_{n+1}| \geq \frac{30}{49} |w_n|.
\]

**Proof.** Observe,
\[
|f_n| \leq w_n + \delta_n \leq |w_n| + \frac{1}{74} \cdot \frac{|f_n|}{\alpha_n}.
\]

Hence,
\[
|f_n| \leq \frac{1}{1 - \frac{1}{15}} \cdot |w_n| = \frac{\alpha_n}{\alpha_n - \frac{1}{74}} \cdot |w_n| \leq \frac{38}{35} \cdot |w_n|,
\]

where we used \( \alpha_n > 3 - \sqrt{8} \) to finish the estimate.

For the second estimate, we have
\[
|w_{n+1}| = |w_n| - \frac{1}{15} \cdot \frac{|f_n|}{\alpha_n} \geq |w_n| - \frac{1}{15} \cdot \frac{\alpha_n}{\alpha_n - \frac{1}{74}} \cdot |w_n|
\]
\[
\geq |w_n| \cdot \left( 1 - \frac{1}{15} \cdot \frac{1}{3 - \sqrt{8}} \right) \geq \frac{30}{49} \cdot |w_n|.
\]

Using this corollary, we can also obtain a relationship between the guide point \( w_N \) where the algorithm terminates and \( \rho_\zeta \), the norm of the closest critical value to 0. Since the algorithm halts when \( w_N \) is an approximate zero for \( f \), we have \( \alpha_N \leq 3 - \sqrt{8} \) but \( \alpha_{N-1} > 3 - \sqrt{8} \).

**Lemma 6.9.** For \( r \geq 1 + \frac{1}{d} \)
\[
|w_N| \geq \frac{1}{40} \cdot \rho_\zeta.
\]

**Proof.** From proposition 5.13 and remark 5.14, we have
\[
|w_0| \geq s_r \cdot \rho_\zeta \geq \frac{\rho_\zeta}{28}.
\]

If \( w_N = w_0 \), the lemma holds trivially.

If \( N > 0 \), then \( \alpha_{N-1} \geq 3 - \sqrt{8} \) (and \( \alpha_N \leq 3 - \sqrt{8} \)).

From lemma 6.6, we get
\[
|f_{N-1}| \geq \frac{1}{4} \cdot \alpha_{N-1} \cdot R_{N-1} \geq \frac{3 - \sqrt{8}}{4} \cdot R_{N-1} \geq \frac{3 - \sqrt{8}}{4} \cdot (\rho_\zeta - |f_{N-1}|).
\]

This last inequality follows from the triangle inequality: if \( v \) is the critical value with \( |v| = \rho_\zeta \), then \( 0, v, \) and \( f_{N-1} \) form a triangle with side lengths \( \rho_\zeta, R_{N-1}, \) and \( |f_{N-1}| \). Rewriting the above yields
\[ |f_{N-1}| \geq \frac{3 - \sqrt{8}}{4 + 3 - \sqrt{8}} \cdot \rho_\zeta. \] (6.5)

We now apply corollary 6.8 to obtain
\[ |w_N| \geq \frac{30}{49} \cdot |w_{N-1}| \geq \frac{30}{49} \cdot \frac{f_{N-1}}{35/38}. \] (6.6)

Combining equations equation (6.5) and equation (6.6) gives
\[ |w_N| \geq \frac{30 \cdot 38 \cdot (3 - \sqrt{8})}{38 \cdot 49 \cdot (4 + 3 - \sqrt{8})} \cdot \rho_\zeta \cdot \rho_\zeta > 40. \]

Finally, we give a lemma which allows us to measure the size of an angular neighborhood about a point \( z_0 \) on the initial circle for which the \( \alpha \)-step algorithm will lift \( \ell_{w_0} \). We use this in section 10.

**Lemma 6.10.** For \( |z_0| > 1 + \frac{1}{d} \), if
\[ \frac{\delta_0}{|f_0|} \leq \frac{1}{111d}, \quad \text{then} \quad \delta_0 < \frac{1}{74} \frac{|f_0|}{\gamma_0}, \]
and the hypotheses of proposition 6.1 are satisfied at \( z_0 \).

**Proof.** Since \( |z_0| > 1 + \frac{1}{d} \), corollary 5.3 gives us \( |f_0^\prime|/|f_0| > d/3 \) and \( \gamma_0 < d^2/2 \). Hence,
\[ \frac{\delta_0}{|f_0|} \leq \frac{1}{111d} = \frac{1}{37d^2} \cdot \frac{d}{3} < \frac{1}{37d^2} \frac{|f_0^\prime|}{|f_0|}. \]

Thus
\[ \delta_0 \leq \frac{|f_0^\prime|}{37d^2} \leq \frac{|f_0^\prime|}{74} \cdot \frac{2}{d^2} < \frac{1}{74} \frac{|f_0^\prime|}{\gamma_0}. \]

7. The pointwise cost

In this section we will estimate the number \( \#f(z_0) \) of iterates needed to find an approximate zero starting at \( z_0 \). We need some preparation to be able to state the estimate. To simplify notation and without loss of generality, throughout this section we shall assume that \( \ell_{w_0} \) lies along the positive real axis; this can be ensured by an appropriate change of variables. Furthermore, we shall assume that no relevant critical values of \( f \) lie on \( \ell_{w_0} \) (that is, \( \hat{\ell}_f(z_0) \in \mathcal{S} \) is disjoint from \( \mathcal{V}_f \)); otherwise, \( \#f(z_0) \) will be infinite.

As before, let \( w_0 = f(z_0) \) and the let the \( w_n \) be the guide points along \( \ell_{w_0} \) as produced by the algorithm. Also let \( \hat{w}_0 = \hat{f}(z_0) \) and \( \hat{w}_n \) be the corresponding points in the surface \( \mathcal{S} \), lying along the ray \( \hat{\ell}_{w_0} \) (see figure 8).

We divide \( \hat{\ell}_{w_0} \) into subintervals as follows: as noted in proposition 4.3, for each \( v \in \mathcal{V}_f \) the intersection of \( \ell_{w_0} \) with \( \text{Vor}(v) \) will either be an interval or the empty set. Set \( \hat{q}_0 = \hat{w}_0 \) and denote the first interval by \( [\hat{q}_0, \hat{q}_1] \) with corresponding critical value \( v_1 \). In general, set
\[ [\hat{q}_{j-1}, \hat{q}_j] = \text{Vor}(v_j) \cap \hat{\ell}_{w_0}. \]
Let $\beta = \beta(z_0)$ denote the total number of such intervals. Note that for a point $z_0 = re^{2\pi i \theta}$ on our initial circle, we have

$$\beta(z_0) = \text{card } \mathcal{J}_0,$$

where $\mathcal{J}_0$ is the set of critical points which influence the orbit of $z_0$, as in definition 4.5.

So that we may work in the target space $\mathbb{C}$ rather than in the surface $\mathcal{S}$, we make the following observation. The projection $\pi$ is an isometry in a neighbourhood of $\hat{\ell}_{w_0}$, since $\mathcal{U}_f \cap \hat{\ell}_{w_0} = \emptyset$. We define a set $U(\hat{\ell}_{w_0}) \subset \mathcal{S}$ as

$$U(\hat{\ell}_{w_0}) = \{ \hat{y} | [\hat{y}, \hat{y}_\perp] \neq \emptyset \},$$

where for $y \in \mathbb{C}$, $y_\perp$ denotes the orthogonal projection of $y$ onto $\ell_{w_0}$ (or its extension $\ell_{-w_0}$).

That is, for each critical point $c_i$ which influences the orbit of $w_0$, we remove the ray perpendicular to $\ell_{w_0}$ starting at the critical value $f(c_i)$. Lifting the result to $\mathcal{S}$ via the branch of $\pi^{-1}$ taking $\ell_{w_0}$ to $\hat{\ell}_{w_0}$ yields the set $U(\hat{\ell}_{w_0})$.

Observe that $\pi$ is an isometry on $U(\hat{\ell}_{w_0})$, and furthermore, $U(\hat{\ell}_{w_0})$ contains $\hat{\ell}_{w_0}$ and a unique lift of each of the points $f(z_0)$ produced by the algorithm. Consequently, we have a well-defined correspondence between the target space $\mathbb{C}$ (minus finitely many rays) and a subset of $\mathcal{S}$ most relevant to the $\alpha$-step algorithm starting at $z_0$. In what follows, we shall use the notation

$$\text{vo}(v_i) = \pi(\text{Vo}(v_i) \cap U(\hat{\ell}_{w_0})), $$

and shall slightly abuse notation by using $v_i$ for $f(c_i)$.

Note that the branch of $f^{-1}$ which takes $w_0$ to $z_0$ is well-defined throughout all of $\pi(U(\hat{\ell}_{w_0}))$; in particular, it coincides with analytic continuation of $f^{-1}$ along $\ell_{w_0}$.

Let $p_j$ be the orthogonal projection of $v_j$ onto the ray $\ell_{w_0}$ (or its extension, $\ell_{-w_0}$), and let $x_j = |v_j - p_j|$. See figure 9. Also, let $\theta_j \in (-\pi, \pi]$ be the angle between $v_j$ and the ray $\ell_{w_0}$, that is,

$$\theta_j = \text{Arg } (v_j/w_0).$$

Figure 8. The various notations used throughout this section, shown in the target space.

Let $\beta = \beta(z_0)$ denote the total number of such intervals. Note that for a point $z_0 = re^{2\pi i \theta}$ on our initial circle, we have

$$\beta(z_0) = \text{card } \mathcal{J}_0,$$

where $\mathcal{J}_0$ is the set of critical points which influence the orbit of $z_0$, as in definition 4.5.

So that we may work in the target space $\mathbb{C}$ rather than in the surface $\mathcal{S}$, we make the following observation. The projection $\pi$ is an isometry in a neighbourhood of $\hat{\ell}_{w_0}$, since $\mathcal{U}_f \cap \hat{\ell}_{w_0} = \emptyset$. We define a set $U(\hat{\ell}_{w_0}) \subset \mathcal{S}$ as

$$U(\hat{\ell}_{w_0}) = \{ \hat{y} | [\hat{y}, \hat{y}_\perp] \neq \emptyset \},$$

where for $y \in \mathbb{C}$, $y_\perp$ denotes the orthogonal projection of $y$ onto $\ell_{w_0}$ (or its extension, $\ell_{-w_0}$).

That is, for each critical point $c_i$ which influences the orbit of $w_0$, we remove the ray perpendicular to $\ell_{w_0}$ starting at the critical value $f(c_i)$. Lifting the result to $\mathcal{S}$ via the branch of $\pi^{-1}$ taking $\ell_{w_0}$ to $\hat{\ell}_{w_0}$ yields the set $U(\hat{\ell}_{w_0})$.

Observe that $\pi$ is an isometry on $U(\hat{\ell}_{w_0})$, and furthermore, $U(\hat{\ell}_{w_0})$ contains $\hat{\ell}_{w_0}$ and a unique lift of each of the points $f(z_0)$ produced by the algorithm. Consequently, we have a well-defined correspondence between the target space $\mathbb{C}$ (minus finitely many rays) and a subset of $\mathcal{S}$ most relevant to the $\alpha$-step algorithm starting at $z_0$. In what follows, we shall use the notation

$$\text{vo}(v_i) = \pi(\text{Vo}(v_i) \cap U(\hat{\ell}_{w_0})), $$

and shall slightly abuse notation by using $v_i$ for $f(c_i)$.

Note that the branch of $f^{-1}$ which takes $w_0$ to $z_0$ is well-defined throughout all of $\pi(U(\hat{\ell}_{w_0}))$; in particular, it coincides with analytic continuation of $f^{-1}$ along $\ell_{w_0}$.

Let $p_j$ be the orthogonal projection of $v_j$ onto the ray $\ell_{w_0}$ (or its extension, $\ell_{-w_0}$), and let $x_j = |v_j - p_j|$. See figure 9. Also, let $\theta_j \in (-\pi, \pi]$ be the angle between $v_j$ and the ray $\ell_{w_0}$, that is,

$$\theta_j = \text{Arg } (v_j/w_0).$$
Furthermore, use $\beta^+(z_0)$ to denote the number of $\theta_j$ for which $|\theta_j| \leq \pi/2$ (or, equivalently, for which $p_j$ lies on $\ell_{w_0}$).

With this notation in hand, we can state an upper bound on the cost of finding an approximate zero starting from $z_0$.

**Theorem 1.** Let $f \in P_{d,1}$ and let $z_0$ be an initial point for the $\alpha$-step path-lifting algorithm with $|z_0| > 1$. Denote $f(z_0)$ by $w_0$. Then the maximum number of steps required for the algorithm to produce an approximate zero starting from $z_0$ is

$$\#_f(z_0) \leq 67 \cdot \left( \log \frac{|w_0|}{|w_N|} + \sum_{j=1}^{\beta^+(z_0)} (3 - 2 \log |\theta_j|) \right)$$

$$\leq 67 \cdot \left( \log \frac{|f(z_0)|}{\rho_\zeta} + \log 40 + \sum_{j=1}^{\beta^+(z_0)} (3 - 2 \log |\theta_j|) \right),$$

where $\beta^+(z_0)$ is the number of relevant critical values along $\ell_{w_0}$ with angle $|\theta_j| < \pi/2$, and $w_N$ is the final ‘guide point’ for the algorithm.

**Remark 7.1.** The second inequality follows from the fact that $\rho_\zeta/40 \leq |w_N| < \rho_\zeta$, as established in lemma 6.9. We shall use this fact in the proving theorem 2.

**Remark 7.2.** As is shown in proposition 8.3 below, for a typical starting point, $\beta^+(z_0) \leq 2$ and there are no more than two angles $\theta_j$ which are relevant.

**Remark 7.3.** In theorem 1, the algorithm converges to a root $\zeta$ as long as $\theta_j \neq 0$. If $\theta_j = 0$, there is a relevant critical value on $\ell_{w_0}$ and the algorithm converges to the corresponding critical point; in this case, $z_0 \not\in \text{Basin } (\zeta)$ for any root $\zeta$ because $z_0$ lies on the stable manifold of a critical point. If $\rho_\zeta = 0$, the algorithm will converge to a root $\zeta$ but the number of steps $\#_f$ will be infinite; in this case $\zeta$ is a multiple root. This remark is a restatement of [K88, theorem 5B] in the current context.
In order to establish theorem 1, we estimate the number of steps required to pass each Voronoi domain, and then sum over the $\beta(z_0)$ domains that $\ell_{u_0}$ passes through.

If $w_j$ and $w_k$ are two guide points lying on $\ell_{u_0}$ with $k > j$, we can define the rather trivial function $\text{Cost}(w_j, w_k) = k - j$. This measures the number of iterations required by the $\alpha$-step algorithm beginning at a point $z_j$ near $f_n^{-1}(w_j)$ to obtain a point $z_k$ near $f_n^{-1}(w_k)$. We extend this function to all pairs of points $y_1$ and $y_2$ lying on $\ell_{u_0}$ by linear interpolation. It is our goal in this section to estimate $N = \text{Cost}(w_0, w_N)$ where $w_N$ corresponds to an approximate zero of $f$.

Rather than count the number of steps directly (which is possible, but tedious), instead we follow a suggestion of Mike Shub and integrate the reciprocal of the stepsize along $\ell_{u_0}$.

**Lemma 7.4.** Let $y_1$ and $y_2$ be two points of $\ell_{u_0}$. Then

$$\text{Cost}(y_1, y_2) \leq 67 \int_{y_2}^{y_1} \frac{dy}{r_y},$$

where $r_y = |y - v|$ for each $y \in \text{vor}(v) \cap \ell_{u_0}$.

**Proof.** Recall that in section 6, we used $J_n$ to denote the $n$th jump, that is, $J_n = |w_n - w_{n+1}|$ where $w_n$ is a guide point for the algorithm. Set $J(w_n) = J_n$, and extend the function $J(y)$ to all of $\ell_{u_0}$ by linear interpolation. Now consider the differential equation along $\ell_{u_0}$ given by

$$\frac{dy}{dt} = -J(y) \quad y(0) = w_0. \tag{7.1}$$

Since $J(y)$ is Lipschitz, equation (7.1) has a unique solution. Observe that the points $w_n$ are exactly the values given by using Euler’s method with stepsize 1 to solve equation (7.1) numerically.

Now consider instead the differential equation given by

$$\frac{dy}{dt} = -\frac{r_y}{67} \quad y(0) = w_0. \tag{7.2}$$

We wish to compare the solution of equation (7.2) to the Euler method for equation (7.1). We will show that for every $y$ in any interval $[w_{n+1}, w_n]$, we have $r_y/67 \leq J(y)$. Consequently, if $\varphi(t)$ is the solution to equation (7.2) and $\varphi(t_1) = y_1$, $\varphi(t_2) = y_2$, then we will have $t_2 - t_1 \geq \text{Cost}(y_1, y_2)$.

To see that $r_y/67 \leq J_y$ for all $y \in [w_{n+1}, w_n]$, we must examine a few cases. First, note that if $y \in \text{vor}(v_j)$, we have

$$r_y^2 = (y - p_j)^2 + x_j^2.$$

Also, recall that by virtue of proposition 6.7, we have $J(w_n) \geq r_{w_n}/66$.

First consider the case where the interval $[w_{n+1}, w_n]$ lies entirely in $\text{vor}(v_i)$. If $w_{n+1} \geq p_i$, then since $r_y$ is decreasing on the interval $[p_i, w_n]$, we have $J(y) \geq r_y/66$. If $p_i \geq w_{n+1}$, $r_y$ will be non-decreasing. However, we can apply the triangle inequality (recalling that $J(w_n) = w_n - w_{n+1}$) to see that

$$r_y \leq J(w_n) + r_{w_n} \leq J(w_n) + 66J(w_n),$$

and so $J(w_n) \geq r_y/67$ for all $y$ in the interval.
In the case where the interval intersects more than one Voronoi region, we proceed as follows. First, observe that for all \( y \in [q_i, w_n] \), we have already established that \( J(y) \geq r_y/67 \) holds (where \( q_i \) is the smallest point of \([w_{n+1}, w_n] \cap \text{vor}(vi)\)). Since \( |vi - q_i| = |q_i - vi+1| \), we have \( J(q_i) \geq r_{q_i}/67 \), and we continue as above.

Finally, equation (7.2) is separable; elementary calculus yields
\[
t(y) = 67 \int_y^{w_0} \frac{dy}{r_y}. \]

Let \( y \) be a point on \( \ell_{w_0} \), and let \( c \) be a critical point which influences \( w_0 \); as before, let \( p \) be the orthogonal projection of \( f(c) \) onto \( \ell_{w_0} \), and let \( x \) denote the distance between \( f(c) \) and \( p \).

For each \( y \) and a fixed critical point \( c \), we also define the angle \( A_y \), which is the angle that the segment from \( y \) to \( f(c) \) makes with the segment between \( f(c) \) and \( p \). Notice that \( r_y = |f(c) - y| \). As before, use \( \theta_c \) to denote the angle between \( f(c) \) and \( \ell_{w_0} \). See figure 10.

We now define the following function, related to \( \text{Cost}(y_1, y_2) \):
\[
\mathcal{L}(y_1, y_2, c) = \log \left( \frac{(y_1 - p) + r_{y_1}}{(y_2 - p) + r_{y_2}} \right).
\]

By virtue of lemma 7.4, if \( y_1 \) and \( y_2 \) are both in \( \text{vor}(f(c)) \), we have
\[
\text{Cost}(y_1, y_2) \leq 67 \int_{y_2}^{y_1} \frac{dy}{r_y} = 67 \mathcal{L}(y_1, y_2, c). \tag{7.3}
\]

However, \( \mathcal{L} \) will still be useful even when one or both of its first two arguments are not in \( \text{vor}(f(c)) \). We establish some bounds on the value of \( \mathcal{L} \) in the next few lemmas.

**Lemma 7.5.**
\[
r_y + (y - p) \leq \begin{cases} 
3(y - p) & \text{if } A_y = \frac{\pi}{6} \\
\frac{x\sqrt{3}}{6} & \text{if } A_y \leq \frac{\pi}{6} 
\end{cases}
\]

**Proof.** Note that \( r_y + (y - p) = x(\tan A_y + \sec A_y) \). If \( A_y > \pi/6 \), we have \( x(\tan A_y + \sec A_y) \leq 3x \tan A_y = 3(y - p) \). When \( A_y \leq \pi/6 \), note that \( \tan A_y + \sec A_y \) is increasing in \( A_y \); at \( A_y = \pi/6 \), \( r_y + (y - p) = x\sqrt{3} \).

We remark that this holds even if \( p < 0 \). \( \square \)
**Lemma 7.6.** Let \( y_1, y_2 \in \ell_{w_0} \) with \( y_1 > y_2 \geq 3p > 0 \). Then
\[
\ell(y_1, y_2, c) < \log \frac{y_1}{y_2} + \log \frac{9}{4}.
\]

**Proof.** We consider two cases: when the angle \( A_y \) is large and when it is small.
If \( A_y \leq \pi/6 \), since \( y_2 > p \)
\[
\ell(y_1, y_2, c) < \ell(y_1, p, c) \leq \log \frac{x\sqrt{3}}{x} = \log \sqrt{3},
\]
where we have used lemma 7.5 in the second inequality.
If \( A_y > \pi/6 \), we have (using lemma 7.5 again)
\[
\ell(y_1, y_2, c) \leq \log \frac{3(y_1 - p)}{2(y_2 - p)} = \log \frac{3y_1(1 - p/y_1)}{2y_2(1 - p/y_2)}.
\]
Since \( y_2 \geq 3p \), we have \( (1 - p/y_1)/(1 - p/y_2) < 3/2 \), and so
\[
\ell(y_1, y_2, c) \leq \log \frac{y_1}{y_2} + \log \frac{9}{4}.
\]
Since \( \sqrt{3} < 9/4 \), the above bound holds in either case. \( \square \)

**Lemma 7.7.** If \( p > 0 \),
\[
\ell(3p, 0, c) \leq \log \frac{4 + \tan |\theta_c|}{\sec |\theta_c| - 1}.
\]

We note that since \( p > 0 \), we have \(-\pi/2 < \theta_c < \pi/2\). Consequently, \( \frac{4 + \tan |\theta_c|}{\sec |\theta_c| - 1} > 1 \).

**Proof.** We have
\[
\ell(3p, 0, c) = \log \frac{3p - p}{r_0 - p} \leq \log \frac{2p + (2p + p \tan |\theta_c|)}{p \sec |\theta_c| - p} = \log \frac{4 + \tan |\theta_c|}{\sec |\theta_c| - 1}.
\]

Finally, we handle the case where \( |\theta_c| \geq \pi/2 \).

**Lemma 7.8.** If \( y_1 > y_2 \geq 0 \geq p \), \( \ell(y_1, y_2, c) < \log(y_1/y_2) \).

**Proof.** Observe that \( r_{y_1} \geq y_2 - p \), since \( r_{y_1} \) is the hypotenuse of the right triangle with a leg of length \( y_2 - p \). Also, by the triangle inequality, \( r_{y_1} - r_{y_2} \leq y_1 - y_2 \).

Using this, we have
\[
\frac{r_{y_1} + (y_1 - p)}{r_{y_2} + (y_2 - p)} \leq \frac{r_{y_2} + (y_1 - y_2) + (y_1 - p)}{2(y_2 - p)} = \frac{2y_1 - p + r_{y_2}}{2(y_2 - p)} \leq \frac{2(y_1 - p) + r_{y_2} - (y_2 - p)}{2(y_2 - p)} \leq \frac{2y_1 - p}{y_2 - p} < \frac{y_1}{y_2}.
\]

Consequently, \( \ell(y_1, y_2, c) = \log \frac{r_{y_1} + (y_1 - p)}{r_{y_2} + (y_2 - p)} < \log(y_1/y_2) \) as desired. \( \square \)

The next lemma enables us to complete the proof of theorem 1.
Lemma 7.9. Let \( z_0 \) be an initial point for the \( \alpha \)-step path lifting algorithm, with \( |z_0| > 1 \), let \( f \in \mathcal{P}_{dL} \), \( w_0 = f(z_0) \). Then the maximum number of steps required for the algorithm to produce an approximate zero starting from \( z_0 \) is

\[
\#f(z_0) \leq 67 \cdot \left( \log \frac{|w_0|}{|w_N|} + \beta^+(z_0) \log \frac{9}{4} + \sum_{j=1}^{\beta^+(z_0)} \log \left( \frac{4 + \tan |\theta_j|}{\sec |\theta_j| - 1} \right) \right),
\]

where \( \beta^+(z_0) \) is the number of relevant critical values along \( \ell_{w_0} \) with angle \( |\theta_j| < \pi/2 \), and \( w_N \) is the final ‘guide-point’ for the algorithm.

Proof. First, divide \( \ell_{w_0} \) into segments where it intersects each of the \( \beta(z_0) \) Voronoi regions \( \text{vor}(v_j) \); the \( j \)th segment will be bounded by points \( q_{j-1} \) and \( q_j \) (we set \( q_0 = w_0 \), and \( q_{\beta(z_0)} = w_N \)). See figure 9.

Now, we have

\[
N = \text{Cost}(w_0, w_N) = \sum_{j=1}^{\beta(z_0)} \text{Cost}(q_{j-1}, q_j) \leq 67 \sum_{j=1}^{\beta(z_0)} \mathcal{L}(q_{j-1}, q_j, c_j),
\]

where the inequality follows from lemma 7.4 and equation (7.3). Applying lemmas 7.6 and 7.7 gives us

\[
\beta^+(z_0) \sum_{j=1}^{\beta^+(z_0)} \mathcal{L}(q_{j-1}, q_j, c_j) \leq \beta^+(z_0) \sum_{j=1}^{\beta^+(z_0)} \log \left| \frac{q_{j-1}}{q_j} \right| + \beta^+(z_0) \log \frac{9}{4} + \sum_{j=1}^{\beta^+(z_0)} \log \frac{4 + \tan |\theta_j|}{\sec |\theta_j| - 1},
\]

where \( q_j^* = \max(|q_j|, |3p_j|) \).

Note that since \( q_j^* \geq |q_j| \), replacing \( q_j^* \) with \( q_j \) will still give us an upper bound; furthermore, since \( |q_{j-1}| > |q_j| \), the logarithm of their ratio is positive. Thus, we have

\[
\beta^+(z_0) \sum_{j=1}^{\beta^+(z_0)} \mathcal{L}(q_{j-1}, q_j, c_j) \leq \sum_{j=1}^{\beta^+(z_0)} \log \left| \frac{q_{j-1}}{q_j} \right| + \beta^+(z_0) \log \frac{9}{4} + \sum_{j=1}^{\beta^+(z_0)} \log \frac{4 + \tan |\theta_j|}{\sec |\theta_j| - 1}. \quad (7.5)
\]

Now we apply lemma 7.8 to the remaining intervals (if any).

\[
\sum_{j=\beta^+(z_0)+1}^{\beta(z_0)} \mathcal{L}(q_{j-1}, q_j, c_j) \leq \sum_{j=\beta^+(z_0)+1}^{\beta(z_0)} \log \left| \frac{q_{j-1}}{q_j} \right|^\frac{\beta(z_0)}{\beta^+(z_0)+1}. \quad (7.6)
\]

Combining equation (7.5) and equation (7.6) with equation (7.4) and recalling that \( q_0 = w_0 \), \( q_{\beta} = w_N \) gives the desired result. \( \square \)

Proof of theorem 1. The proof of the main result of this section now follows immediately from lemma 7.9. First combine the term \( \beta^+(z_0) \log \frac{9}{4} \) with the sum, and then observe that for \( |\theta| < \pi/2 \), we have

\[
\log \frac{9(4 + \tan |\theta|)}{4(\sec |\theta| - 1)} \leq \log \frac{1}{\theta^2} + 3.
\]

This can be readily seen via the series expansion, which is \( \log(18) - 2 \log(\theta) + \theta/4 + O(\theta^2) \). \( \square \)
8. The average cost

In this section we shall prove theorem 2, which follows from averaging the bound found in section 7 over the starting points on the circle of radius \( r = 1 + C/d. \)

Recall from definition 4.5 that \( \mathcal{I} \) is the set of pairs \((t, c)\) for which the critical points \( c \in C \) influence the starting points \( z_0 = re^{it} \) on the initial circle of radius \( r \). \( \mathcal{I}_t \) is the set of critical points which influence a given \( t \), and \( \mathcal{I}_c \) are the \( t \in S_r \) which are influenced by \( c \).

For each pair in \((t, c) \in \mathcal{I}\), we use \( \theta(t, c) = \theta(c, t) = \text{Arg} \left( \frac{f(re^{2\pi it})}{f(c)} \right) \).

In the notation of section 7, \( \theta(t, c) = \theta_j \) where \( v_j = \hat{f}(c_j) \) and \((t, c_j) \in \mathcal{I}\).

Note that for each fixed \( c \), \( \mathcal{I}_c \) is a collection of finitely many intervals: \( \mathcal{I}_c \) consists of for those \( t \) such that \( \hat{f}(re^{it}) \) intersects \( \text{Vor}(\hat{f}(c)) \).

Define for every critical point \( c \in C \) the function \( \theta_c : \mathcal{I}_c \rightarrow \mathbb{R} \) by

\[
\theta_c(t) = \theta(t, c) = \text{Arg} \left( \frac{f(re^{2\pi it})}{f(c)} \right).
\]

Lemma 8.1. For each \( c \in C \), the map \( \theta_c \) is at most \((m_c + 1)\)-to-one.

Proof. For every \( \theta \in (-\pi, \pi] \) there are at most \((m_c + 1)\) rays \( \hat{e} \subset S_r \) for which the angle between \([0, f(c)] \) and \( \pi(\hat{e}) \) is \( \theta \) and which also intersect \( \text{Vor}(\hat{f}(c)) \). This is a consequence of proposition 4.3.

As an immediate consequence of the angular speed lemma (lemma 5.1), we have

\[
2\pi d \cdot \frac{r}{r + 1} \leq \frac{d}{dt} \theta_c(t) \leq 2\pi d \cdot \frac{r}{r - 1}.
\]  

(8.1)

Proposition 8.2. Let \( f \in P_d \) be of degree \( d \) and \( r > 1 \). Then

\[
\int_0^1 \sum_{(t, c) \in \mathcal{I}_c} \log \left( \frac{4 + \tan |\theta(t, c)|}{\sec |\theta(t, c)|} \right) |\theta(t, c)|^{-1} \, dt \leq 3 \cdot \frac{r + 1}{r}.
\]

Proof. Throughout the proof, let \( \psi(\theta) = \frac{4 + \tan |\theta|}{\sec |\theta| + 1} \). From lemma 8.1 and equation (8.1), we see that for fixed values of \( c \), we have

\[
\int_{|\theta(\ell)| < \pi/2} \log \psi(\theta) \, d\theta \leq (m_c + 1) \int_{-\pi/2}^{\pi/2} \log \psi(\theta) \, d\theta \leq (m_c + 1) \frac{r + 1}{2\pi d} \int_{-\pi/2}^{\pi/2} \log \psi(\theta) \, d\theta.
\]

Thus
Recall from section 7 that $\beta^+(z)$ denotes the number of critical points that influence the orbit of $z = re^{2\pi i t}$ with the critical value in the same half-plane, i.e.

$$\beta^+(r e^{2\pi i t}) = \text{card} \{ c \in \mathcal{I} \mid -\pi/2 < \theta(t, c) < \pi/2 \}.$$ 

The next proposition bounds the number of such Voronoi domains a starting point encounters, on average.

**Proposition 8.3.**

$$\int_0^1 \beta^+(r e^{2\pi i t}) dt \leq \frac{1 + r}{r}.$$

**Proof.** Note that

$$\int_0^1 \beta^+(r e^{2\pi i t}) dt = \int_0^1 \sum_{c \in \mathcal{C}_f} 1 dt = \int_0^1 \sum_{c \in \mathcal{C}_f} 1 dt.$$

As in the proof of proposition 8.2, we transport the calculation from the source space to the target space using the bound on $\theta_c'(t)$ in equation (8.1) and the fact that for fixed $c$, $\theta_c(t)$ is at most $(m_c + 1)$-to-one (lemma 8.1). This gives us

$$\int_0^1 \beta^+(r e^{2\pi i t}) dt \leq \sum_{c \in \mathcal{C}_f} \int_{-\pi/2}^{\pi/2} \frac{d\theta}{\theta_c'(t)} \leq \sum_{c \in \mathcal{C}_f} (m_c + 1) \frac{r + 1}{2\pi rd} \cdot \pi \leq 2(d - 1) \frac{r + 1}{2\pi d} < \frac{1 + r}{r}.$$

Above, we used the fact that $\sum_{c \in \mathcal{C}_f} m_c = d - 1$. 

**Lemma 8.4.** If $r \geq 1 + \frac{1}{d}$

$$\int_0^1 \log \frac{|w_0|}{|w_N|} dt \leq d \log r + \log 40 + \frac{1}{d} \cdot \frac{1 + r}{r} \cdot K_f.$$

**Proof.** Corollary 5.10, proposition 5.13, lemma 5.1, and lemma 6.9 are used in the following calculation.
\[
\int_0^1 \log \left| \frac{w_0}{w_N} \right| \, dt = \int_0^1 \log |w_0| \, dt - \int_0^1 \log |w_N| \, dt \\
\leq d \log r - \int_0^1 \frac{\log \rho_c}{40} \, dt \\
\leq d \log r + \log 40 + \sum_{\zeta \in \mathcal{S}} |\log \rho_c| \cdot \frac{1 + r}{d} \\
\leq d \log r + \log 40 + \frac{1 + r}{d} \cdot K_f. 
\]

**Remark 8.5.** If \( r = 1 + \frac{1}{d} \), then \( d \log r < 1 \), giving \( \int_0^1 \log \left| \frac{w_0}{w_N} \right| \, dt \leq 1 + \log 40 + \frac{2K_f}{d} \).

Now we are ready to provide a proof of the following

**Theorem 2.** Let \( f : \mathbb{C} \to \mathbb{C} \) be a monic polynomial with distinct roots \( \zeta_i \) in the unit disk. Let \( \bar{\#}_f \) be the average number of steps required by the \( \alpha \)-step algorithm to locate an approximate zero for \( f \). Then

\[
\bar{\#}_f \leq 67 \left( 12.4 + \frac{2K_f}{d} \right).
\]

where the average is taken over starting points on the circle of radius \( 1 + 1/d \) endowed with uniform measure.

**Proof.** Let \( r = 1 + 1/d \). Lemma 7.9, proposition 8.3, lemma 8.4, and proposition 8.2 imply

\[
\bar{\#}_f = \int_0^1 \#_f(re^{2\pi i t}) \, dt \\
\leq \int_0^1 67 \left[ \log \left| \frac{w_0}{w_N} \right| + \frac{\beta + (re^{2\pi i t}) \log 9}{4} + \sum_{\zeta \in \mathcal{S}} \log \frac{4 + \tan |\theta(t, c)|}{\sec |\theta(t, c)| - 1} \right] \, dt \\
\leq 67 \left[ 1 + \log 40 + \frac{2K_f}{d} \right] + 1.622 + 6 \\
\leq 67 \left[ 12.4 + \frac{2K_f}{d} \right]. \quad \square
\]

**9. The relation between cost and degree**

In the previous section, we showed that the expected number of steps required for the algorithm to converge to an approximate zero is bounded by \( \bar{\#}_f \), which depends directly on \( K_f/d \). For every degree \( d \), this is neither bounded above nor below, even if we restrict \( f \) to monic polynomials with distinct roots in the unit disk. As noted in remark 3.9, \( K_f \) (and hence \( \bar{\#}_f \)) is infinite precisely when \( f \) has a multiple zero. Since distinct roots of \( f \in \mathcal{P}_{d,1} \) may be arbitrarily close together, \( K_f \) cannot be bounded above.

We can, however, estimate the average value of \( K_f/d \) as \( f \) ranges over \( \mathcal{P}_{d,1} \) (in fact, its closure \( \overline{\mathcal{P}_{d,1}} \)). We shall see in this section that this average value grows no faster than linearly in \( d \), using the product measure on the distribution of roots on \( \overline{\mathcal{P}_{d,1}} \).
The value of $\rho_\zeta$ is closely related to the function $\gamma(\zeta)$ mentioned in section 3. Indeed, we have the following relationship, which enables us to bound $K_f$ and $K_f$ from $\gamma(\zeta)$ and $f'(\zeta)$ at each of the roots $\zeta$.

**Lemma 9.1.** Let $\gamma(\zeta) = \max_{j>1} \left| \frac{f^{(j)}(\zeta)}{f'(\zeta)} \right|^{\frac{1}{j-1}}$ and let $\zeta$ be a nondegenerate root of $f$. Then

$$(3 - \sqrt{8}) \frac{|f'(\zeta)|}{\gamma(\zeta)} \leq \rho_\zeta \leq 4 \frac{|f'(\zeta)|}{\gamma(\zeta)}.$$ 

**Proof.** This follows immediately from [K88, theorem 4.1].

It is not hard to show by induction that

$$f^{(j)}(z) = \sum_{k_1} \sum_{k_2 \neq k_1} \sum_{k_3 \notin \{k_1, k_2\}} \cdots \sum_{k_j \notin \{k_1, k_2, \ldots, k_{j-1}\}} \prod_{i \notin \{k_1, k_2, \ldots, k_j\}} (z - \zeta_i),$$

and so

$$f^{(j)}(\zeta_m) = \sum_{k_1 \notin \zeta_m} \sum_{k_2 \notin \{k_1, k_3\}} \cdots \sum_{k_j \notin \{k_1, k_2, \ldots, k_{j-1}\}} \prod_{i \notin \{k_1, k_2, \ldots, k_j\}} (\zeta_m - \zeta_i),$$

(9.1)

that is, a sum of $\binom{d-j+1}{j}$ terms, each of which is a product of $d-j$ factors. Using this observation, we obtain the following. (Compare to [Ded, proposition 5.1].)

**Lemma 9.2.** $\gamma(\zeta_m) \leq \frac{d-1}{2} \min_{\zeta_m(\zeta_m - \zeta_i)} \frac{1}{\min_{\zeta_m(\zeta_m - \zeta_i)}}$

**Proof.** Using equation (9.1) above and cancelling common factors between $f'$ and $f^{(j)}$ yields

$$\left| \frac{f^{(j)}(\zeta_m)}{f'(\zeta_m)} \right| = \left| \frac{1}{h_1} \sum_{k_1 \notin \zeta_m} \sum_{k_2 \notin \{k_1, k_3\}} \cdots \sum_{k_j \notin \{k_1, k_2, \ldots, k_{j-1}\}} \prod_{i \notin \{k_1, k_2, \ldots, k_j\}} \frac{1}{(\zeta_m - \zeta_i)} \right| \leq \frac{1}{h_1} \sum_{k_1 \notin \zeta_m} \sum_{k_2 \notin \{k_1, k_3\}} \cdots \sum_{k_j \notin \{k_1, k_2, \ldots, k_{j-1}\}} \frac{1}{\min_{\zeta_m(\zeta_m - \zeta_i))}^{j-1}} \frac{1}{\min_{\zeta_m(\zeta_m - \zeta_i))}^{j-1}}$$

$$= \frac{1}{d} \left( \frac{d}{j} \right) \left[ \frac{1}{\min_{\zeta_m(\zeta_m - \zeta_i))} \right]^{j-1}.$$ 

Consequently,

$$\gamma(\zeta_m) = \max_{j>1} \left| \frac{f^{(j)}(\zeta)}{f'(\zeta)} \right|^{\frac{1}{j-1}} \leq \max_{j>1} \frac{1}{d} \left( \frac{d}{j} \right) \left[ \frac{1}{\min_{\zeta_m(\zeta_m - \zeta_i))} \right]^{j-1} \leq \frac{d-1}{2} \min_{\zeta_m(\zeta_m - \zeta_i)} \frac{1}{\min_{\zeta_m(\zeta_m - \zeta_i)}}.$$ 

We now turn to estimating the average value of the components which control $K_f$; the derivative at each root and the minimal inter-root distance. Identify a polynomial $f(z) = \prod_{i=1}^d (z - \zeta_i) in \mathcal{D}_{d,1}$ with the $d$-tuple of its roots, and thus we can view its closure $\overline{\mathcal{D}}_{d,1}$ as the polydisk $D^d$. Using Lebesgue measure on $D^d$ gives $\overline{\mathcal{D}}_{d,1}$ a volume of $\pi^d$.

**Lemma 9.3.** For each $m$, we have

$$\int_{(\zeta_1, \ldots, \zeta_d) \in D^d} \log \frac{1}{\min_{\zeta \notin \zeta_m(\zeta_m - \zeta_i)} d\zeta_1 d\zeta_2 \cdots d\zeta_d} \leq 2(d-1)\pi^d.$$
Figure 11. The two cases for $E_k$ in lemma 9.3: when $r_k + R_1 \leq 1$ (left), and when $r_k + R_1 > 1$ (right). All the roots except for $\zeta_1$ lie in the shaded region $E_k$.

**Proof.** Without loss of generality, we may take $m = 1$.

Let $|\zeta_1| = R_1$, and let $\zeta_k$ be a root for which $|\zeta_1 - \zeta_k|$ is minimized. Set $\zeta_k = \zeta_1 + r_k e^{i\theta_k}$. Let $D_r(\zeta_1)$ be the disk centered at $\zeta_1$ with radius $r_k$, and let $E_k = \mathbb{D} \setminus D_{r_k}(\zeta_1)$ denote the part of the unit disk exterior to it. See figure 11. There are two possibilities: either $E_k$ is a disk of radius $r_k$ which occurs when $R_1 + r_k < 1$, or $R_1 + r_k > 1$ and $E_k$ is a crescent. Let $s_k$ represent the arc length of the part of boundary of $E_k$ which contains $\zeta_k$.

Observe that for fixed $\zeta_1$, we have $(\zeta_2, \ldots, \zeta_d) \in E_k^{d-1}$ (with $\zeta_k$ on the interior boundary). So we have

$$ L_1 = \int_{(\zeta_1, \ldots, \zeta_d) \in \mathbb{D}^d} \log \frac{1}{\min_{k \neq 1} |\zeta_k - \zeta_1|} \, d\zeta_1 \, d\zeta_2 \cdots d\zeta_d = \int_{\zeta_1 \in \mathbb{D}} \int_{(\zeta_2, \ldots, \zeta_d) \in E_k^{d-1}} \log \frac{1}{r_k} \, d\zeta_1 \, d\zeta_2 \cdots d\zeta_d. $$

The closest root to $\zeta_1$ could be any of remaining $d - 1$ roots; we shall do the calculation for $\zeta_k$; by symmetry, the remaining cases will have the same value.

Observe that all roots except $\zeta_1$ lie in in $E_k$. The area of $E_k$ is always less than $\pi$ (since it is a subset of the unit disk), and we always have $s_k \leq 2\pi r_k$ (since $s_k$ is part of the circumference of a disk of radius $r_k$).

If we also write $\zeta_1 = R_1 e^{i\phi_1}$ and $\zeta_k = r_k e^{i\theta_k}$, and note that integrating $\phi$ and $\theta_k$ give factors of $2\pi R_1$ and $s_k$. Calculating the integral for each $k$ and summing gives

$$ L_1 \leq \pi^{d-2}(d - 1) \int_0^1 \int_0^{1 + R_1} (2\pi R_1)(s_k) \log \frac{1}{r_k} \, dr_k \, dR_1. $$

Observe that the integrand $\log(1/r_k)$ is positive only for $0 < r_k < 1$. Thus, we can give an upper bound on the integral by ignoring the contribution when $r_k > 1$.

This gives us the following bound on the integral.

$$ L_1 \leq 4\pi^d(d - 1) \int_0^1 \int_0^{1 + R_1} R_1 r_k \log \frac{1}{r_k} \, dr_k \, dR_1 = 2(d - 1)\pi^d. $$

**Lemma 9.4.** For $f(z) = \prod(z - \zeta_k)$ with $|\zeta_k| \leq 1$, we have

$$ \int_{(\zeta_1, \ldots, \zeta_d) \in \mathbb{D}^d} \prod_{m=1}^d \log \frac{1}{|f'(\zeta_m)|} \, d\zeta_1 \cdots d\zeta_d = \frac{d(d - 1)}{4} \pi^d. $$
Proof. From equation (9.1) in the case $j = 1$, we obtain $\prod_{m=1}^{d} f'(f_m) = \prod_{m=1}^{d} P'(z_m - \zeta)$, and so
\[
\int_{(z_1, \ldots, z_d) \in \mathbb{D}^d} \log \prod_{m=1}^{d} |f'(f_m)| \, dz_1 \cdots dz_d = \sum_{m=1}^{d} \sum_{k \neq m} \int_{(z_1, \ldots, z_d) \in \mathbb{D}^d} \log |z_m - \zeta| \, dz_1 \cdots dz_d
\]
\[= -\pi^{d-2} \sum_{m=1}^{d} \int_{\mathbb{D} \setminus \mathbb{D}_m} \int_{\mathbb{D}_m} \log |z_m - \zeta| \, dz_m \, d\zeta.
\]

For each of the integrals in the sum we divide $\mathbb{D}^2$ into two parts: those where $|z_m| \leq |\zeta|$ and the complement where $|z_m| > |\zeta|$. When $|z_m| > |\zeta|$, we let $\zeta = e^{2\pi i (z)}$ and apply lemma 5.9:
\[
\int_{\mathbb{D} \setminus \mathbb{D}_m} \int_{|z_m| > |\zeta|} \log |z_m - \zeta| \, dz_m \, d\zeta = 2\pi \int_{\mathbb{D} \setminus \mathbb{D}_m} \int_{|z_m| > |\zeta|} \log |e^{2\pi i (z)} - \zeta| \, R \, dR \, d\zeta
\]
\[= 2\pi \int_{\mathbb{D} \setminus \mathbb{D}_m} \int_{|z_m| > |\zeta|} R \log R \, dR \, d\zeta = -\frac{\pi^2}{8}.
\]

Similarly, the value of the integral when $|z_m| \leq |\zeta|$ is also $-\pi^2/8$. Summing the $(d-1)$ integrals, each of which contributes $\pi^2/4$, gives the desired result. \hfill \Box

Theorem 3. For $f \in \mathcal{F}_{d, 1}$, let $\Lambda_f$ be the average value of $\log(1/\rho_f)$, that is, $\Lambda_f = K_f/d$. Define $\overline{\Lambda}$ to be the average value of $\Lambda_f$ over $f \in \mathcal{F}_{d, 1}$, where we parameterize $\mathcal{F}_{d, 1}$ by the polydisk of the roots with Lebesgue measure. Then
\[
\overline{\Lambda} < \frac{3d}{2}.
\]

Proof. Applying lemmas 9.1 and 9.2 and using the fact that $3 - \sqrt{8} < 1/6$, we have
\[
\Lambda_f = \frac{K_f}{d} = \frac{1}{d} \sum_{\zeta \in \mathcal{A}_f} \log \frac{1}{\rho_f}
\leq \frac{1}{d} \sum_{\zeta \in \mathcal{A}_f} \log \left(\frac{6\gamma(\zeta)}{|f'(\zeta)|}\right)
\leq \log 6 + \frac{1}{d} \sum_{\zeta \in \mathcal{A}_f} \log \gamma(\zeta) + \frac{1}{d} \sum_{\zeta \in \mathcal{A}_f} \log \frac{1}{|f'(\zeta)|}
\leq \log 6 + \log \frac{d-1}{2} + \frac{1}{d} \sum_{\zeta \in \mathcal{A}_f} \log \frac{1}{\min_{\zeta \neq \zeta} |\zeta - \zeta|} + \frac{1}{d} \sum_{\zeta \in \mathcal{A}_f} \log \frac{1}{|f'(\zeta)|}.
\]
Integrating over $f \in \mathcal{F}_{d, 1}$ and applying lemmas 9.3 and 9.4 yields
\[
\int_{f(\zeta) \in \mathcal{A}_f} \Lambda_f \leq \pi^d \left(\log 3 + \log(d-1) + \frac{2(d-1)}{d} + \frac{d-1}{4}\right) < \pi^d \frac{3d}{2}.
\]
Since the volume of $\mathcal{F}_{d, 1}$ is $\pi^d$, we obtain $\overline{\Lambda} \leq 3d/2$ for all $d$ (and is asymptotic to $d/4$). \hfill \Box

Corollary 9.5. For $f \in \mathcal{F}_{d, 1}$, the average number of steps required to locate an approximate zero is $\Theta(d)$. 

453
**Question 9.6.** How does the bound in theorem 3 change if we average with respect to a measure on the coefficients of f rather than uniform measure on the roots of f?

**10. How to find all roots of a polynomial**

The focus of the paper has been on the question of locating a single approximate zero for a given polynomial, but these results can easily be used to locate all d roots of a polynomial $f \in \mathcal{R}_{d,1}$.

To do so, we need to locate d initial points, one in Basin ($\zeta_j$) for each root $\zeta_j$. Then we apply the $\alpha$-step algorithm starting at each of these, and as long as $f \in \mathcal{R}_{d,1}$, the algorithm will produce an approximate zero for each root. Our estimates do not rely on roots with special properties (such as being ‘exposed’ as in [Man], or having a large sector in the target space which is free of critical values as in [KS] or [Sm85]); consequently they apply equally well to each of the roots $\zeta_j$.

To choose these initial points, we can do the following.

1. Choose $[111\pi d^2]$ points $y_j$ equally spaced around the circle of radius $1 + \frac{1}{d}$. Let $z_0 = y_0$.
2. Let $k = 1$. For each $j > 0$, evaluate $f(y_j)$.

If $\text{Arg } f(y_j) \geq \text{Arg } f(y_0)$ but $\text{Arg } f(y_{j-1}) < \text{Arg } f(y_0)$, set $z_k = y_j$ and increment $k$.

At the end of step (2), there will be exactly $d$ points $z_k$ with $\text{Arg } f(z_k) = \text{Arg } f(y_0) \leq \frac{1}{111\pi d}$. This holds as a result of the angular speed lemma (lemma 5.1) and the fact that the image of the circle winds exactly $d$ times around the origin.

Now we use the $d$ points $z_k$ to lift $d$ copies of the same ray $\ell_{f(y_0)}$, one in each basin, by using a slight modification of the $\alpha$-step algorithm from section 3. Specifically, we modify Step 0 to set

$$w_{0,k} = |f(z_k)| \frac{f(y_0)}{|f(y_0)|},$$

that is, for each $k$ we choose initial target points on the ray $\ell_{f(y_0)}$ with norm $|f(z_k)|$. Then the $\alpha$-step algorithm proceeds as usual.

While there could be some $k$ for which $z_k \notin \text{Basin } (\zeta_k)$, as a consequence of lemma 6.10, each of the points $z_k$ are close enough to some point $z_{0,k} \in \text{Basin } (\zeta_k)$ (and with $f(z_{0,k}) \in \ell_{f(y_0)}$) so that the $\alpha$-step algorithm will converge to an approximate zero for the root $\zeta_k$.

The above method for determining the points $z_k$ requires $O(d^2)$ evaluations of $f$, at an arithmetic complexity of $O\left(d^3 \log^2 d\right)$; the number of steps required to find all $d$ roots is $O(K_y) = O(\sum \log (1/\rho_f))$. Applying corollary 9.5, the average complexity to find approximate zeros for all $d$ roots of $f$ will be $O\left(d^3 \log^2 d\right)$.

**Remark 10.1.** For $f \in \mathcal{R}_{d,1}$, by using the method given above, $d$ approximate zeros can be found (one for each root $\zeta_j$) in $O(K_y)$ steps of the $\alpha$-step algorithm. This has an average arithmetic complexity of $O\left(d^3 \log^2 d\right)$.

11. Concluding remarks and extensions

**Remark 11.1.** Our major goal in this work was to bound the number of iterations of the $\alpha$-step algorithm and examine the relationship to the underlying geometry of the polynomial, rather than to optimize the arithmetic complexity. Since each step of the algorithm requires computing
of all of the derivatives of $f$, one could use a higher-order method instead of Newton’s method (as in [K88, Ho], [SS86]) in the algorithm without a significant increase in cost. In this case, we calculate $z_{n+1}$ by a single step of a method using higher derivatives of $f$ to approximate the zero of $f(z) = w_{n+1}$ from $z_n$. Use of such a method results in a larger stepsize (and consequently fewer steps). For example, the stepsize is nearly doubled by a method using the first three derivatives of $f$. The interested reader should see [K88], where such methods are examined in depth.

**Remark 11.2.** Alternatively, the use of $\alpha$ could be curtailed (or even entirely removed) by dynamically adjusting the guide points $w_n$ as follows. At each step, set $w_{n+1}$ to be $(1 - h_n)|f(z_n)|w_n$. Initially, take $h_n = h_0$, but if $f(z_n)$ is not sufficiently close to $w_{n+1}$, divide $h_n$ by 2 and try again until it is. At the next step, start with $h_{n+1} = \min(h_0, 2h_n)$. Note that this approach, while similar in spirit, is somewhat different from the variable stepsize methods explored in [HS]. One can still use $\alpha$ to detect whether an approximate zero has been located, or, if evaluating higher derivatives of $f$ is impractical, other methods such as those in [B02] or [O] can be used.

**Remark 11.3.** The $\alpha$-step algorithm could easily be adapted to locate $\epsilon$-roots with no significant increase in complexity. In addition to stopping the iteration when an approximate zero is found, the algorithm could also stop if $z_n$ is an $\epsilon$-root for a pre-determined $\epsilon$. This can be checked at essentially no cost merely by determining if $|f(z_n)/f'(z_n)| < \epsilon/d$ (this follows from the well-known fact that there is always a root within the disk of radius $d$ times the Newton step at $z$).

**Remark 11.4.** Using some of the ideas in [GLSY], the results here can be extended to deal more directly with multiple roots.

**Remark 11.5.** The selection of initial points in section 10 can almost certainly be improved from $O(d^2)$ evaluations of $f$, most likely to $O(d \log d)$ evaluations. However, this does not affect the overall complexity of the algorithm.

**Acknowledgments**

Part of this work was done while Myong-Hi Kim was visiting Stony Brook University; we are grateful for the support and hospitality of the Department of Mathematics and the Institute for Mathematical Sciences. Marco Martens was supported in part by the National Science Foundation under the grant DMS-1600554, which he gratefully acknowledges.

A previous version of this article was circulated as ‘Bounds for the Cost of Root Finding’. The authors would like to thank Araceli Bonifant, Michael Shub, and the anonymous referees for their input and suggestions which have improved this paper.

**References**

[Ab] Aberth O 1973 Iteration methods for finding all zeros of a polynomial simultaneously Math. Comput. 27 339–44

[B02] Batra P 2002 Simultaneous point estimates for Newton’s method BIT: Numer. Math. 42 467–76

[B08] Batra P 2008 Newton’s method and the computational complexity of the fundamental theorem of algebra Electron. Not. Theor. Comput. Sci. 202 201–18

[BP] Beltrán C and Pardo L M 2011 Fast linear homotopy to find approximate zeros of polynomial systems Found. Comput. Math. 11 95–129
[BS09] Beltrán C and Shub M 2009 Complexity of Bezout’s theorem VII: distance estimates in the condition metric Found. Comput. Math. 9 179–95
[BS10] Beltrán C and Shub M 2010 A note on the finite variance of the averaging function of polynomial system solving Found. Comput. Math. 10 115–25
[BS13] Beltrán C and Shub M 2013 The complexity and geometry of numerically solving polynomial systems Recent Advances in Real Complexity and Computation (Contemporary Mathematics vol 604) ed J L Montaña and L Pardo (Providence, RI: American Mathematical Society) pp 71–104
[BAS] Bilarev T, Aspenberg M and Schleicher D 2016 On the speed of convergence of Newton’s method for complex polynomials Math. Comput. 85 693–705
[BF] Bini D A and Fiorentino G 2000 Design, analysis, and implementation of a multiprecision polynomial rootfinder Numer. Algorithms 23 127–73
[BCSS] Blum L, Cucker F, Shub M and Smale S 1997 Complexity and Real Computation (Berlin: Springer)
[BM] Borodin A and Munro J 1975 The Computational Complexity of Algebraic and Numeric Problems (Amsterdam: Elsevier)
[Ded] Dedieu J 1997 Estimations for the separation number of a polynomial system J. Symb. Comput. 24 683–93
[DKST] Dedieu J, Kim M, Shub M and Tisseur F 2003 Implicit gamma theorems (I): pseudoroots and pseudospectra Found. Comput. Math. 3 1–31
[Du] Duren P 1983 Univalent Functions (Berlin: Springer)
[HSS] Hubbard J, Schleicher D and Sutherland S 2001 How to find all roots of complex polynomials by Newton’s method Inventiones Math. 146 1–33
[JJT] Jongen H T, Jonker P and Twilt F 1988 The continuous, desingularized Newton method for meromorphic functions Acta Appl. Math. 13 81–121
[K85] Kim M 1985 Computation complexity of the Euler algorithms for the roots of complex polynomials PhD Thesis City University of New York, New York
[K88] Kim M 1988 On approximate zeros and rootfinding algorithms for a complex polynomial Math. Comput. 51 707–19
[KS] Kim M and Sutherland S 1994 Polynomial root-finding and branched covers SIAM J. Comput. 23 415–36
[Ko] Koebe P 1907 Über die Uniformisierung bebliebiger analytischer kurven Nach. Kgl. Ges. Wiss. Göttingen Math.—Phys. Kl. 1907 191–210
[KoS] Kozen D and Stefánsson K 1997 Computing the Newtonian graph J. Symb. Comput. 24 125–36
[Lu] Lucas F 1874 Propriétés géométriques des fractions rationnelles C. R. Acad. Sci. Paris 77 431–3
[Man] Manning A 1992 Bol. Soc. Bras. Mat. 22 157–77
[Mar] Marden M 1966 The Geometry of the Zeros of a Polynomial in a Complex Variable (Mathematical Surveys and Monographs vol 3) 2nd edn (Providence, RI: American Mathematical Society)
[MH] Marsden J and Hoffman M 1997 Basic Complex Analysis 2nd edn (New York: W H Freeman)
[O] Ocken S 1998 Convergence criteria for attracting cycles of Newton’s method SIAM J. Appl. Math. 58 235–44
[P97] Pan V 1997 Solving a polynomial equation: some history and recent progress SIAM Rev. 39 187–220
[P02] Pan V 2002 Univariate polynomials: nearly optimal algorithms for numerical factorization and root-finding J. Symb. Comput. 33 701–33
[Po] Pommerenke C 1975 Univalent Functions (Göttingen: van den Hoeck & Rupert)

[Ren] Renegar J 1987 On the worst-case arithmetic complexity of approximating zeros of polynomials J. Complexity 3 90–113

[Rie] Riemann B 1851 Grundlagen für eine allgemeine Theorie der Funktionen einer veränderlichen komplexen Grösse Inaugural Dissertation Göttingen

[Sch] Schleicher D 2008 Newton’s method as a dynamical system: efficient root finding of polynomials and the Riemann ζ function Holomorphic Dynamics and Renormalization: a Volume in Honor of John Milnor’s 75th Birthday (Fields Institute Communications vol 53) ed M Lyubich and M Yampolsky (Providence, RI: American Mathematical Society) pp 213–24

[Sh09] Shub M 2007 Complexity of Bezout’s theorem VI: geodesics in the condition (number) metric Found. Comput. Math. 9 171–8

[SS86] Shub M and Smale S 1986 Computational complexity: on the geometry of polynomials and a theory of cost: II SIAM J. Comput. 15 145–61

[SS93a] Shub M and Smale S 1993 Complexity of Bezout’s theorem I: geometrical aspects J. AMS 6 459–501

[SS93b] Shub M and Smale S 1993 Complexity of Bezout’s theorem II: volumes and probabilities Computational Algebraic Geometry (Progress in Mathematics vol 109) ed F Eyssette and A Galligo (Basel: Birkhauser) pp 267–85

[SS93c] Shub M and Smale S 1993 Complexity of Bezout’s theorem III: condition number and packing J. Complexity 9 4–14

[SS94] Shub M and Smale S 1994 Complexity and Bezout’s theorem V: polynomial time Theor. Comput. Sci. 133 141–64

[SS96] Shub M and Smale S 1996 Complexity of Bezout’s theorem IV: probability of success; extensions SIAM J. Numer. Anal. 33 128–48

[STW] Shub M, Tischler D and Williams R 1988 The Newtonian graph of a complex polynomial SIAM J. Math. Anal. 19 246–56

[Sm81] Smale S 1981 The fundamental theorem of algebra and complexity theory Bull. Am. Math. Soc. 4 1–36

[Sm85] Smale S 1985 On the efficiency of algorithms of analysis Bull. Am. Math. Soc. 13 87–121

[Sm86] Smale S 1986 Newton’s method estimates from data at one point The Merging Disciplines: New Directions in Pure, Applied and Computational Mathematics (Berlin: Springer) pp 185–96

[Sm97] Smale S 1997 Complexity theory and numerical analysis Acta Numer. 6 523–51

[WZ] Wang D and Zhao F 1995 The theory of Smale’s point estimation and its application J. Comput. Appl. Math. 60 253–69

[WH] Wang X and Han D 1990 On dominating sequence method in the point estimate and Smale theorem Sci. China Ser. A 33 135–44