On the efficient global dynamics of Newton’s method for complex polynomials

Dierk Schleicher

Aix-Marseille Université and CNRS, UMR 7373, Institut de Mathématiques de Marseille, 163 Avenue de Luminy Case 901, 13009 Marseille, France

E-mail: dierk.schleicher@univ-amu.fr

Received 26 April 2021; revised 16 August 2022
Accepted for publication 9 January 2023
Published 23 January 2023

Abstract
We investigate Newton’s method as a root finder for complex polynomials of arbitrary degrees. While polynomial root finding continues to be one of the fundamental tasks of computing, with essential use in all areas of theoretical mathematics, numerics, computer graphics and physics, known methods may have excellent theoretical complexity but cannot be used in practice, or are practically efficient but lack a successful theory behind them. We provide precise and strong upper bounds for the theoretical complexity of Newton’s method and show that it is near-optimal with respect to the known set of starting points that find all roots. This theoretical result is complemented by a recent implementation of Newton’s method that finds all roots of various polynomials of degree more than a billion, significantly faster than our upper bounds on the complexity indicate, and often much faster than established fast root finders. Newton’s method thus stands out as a method that has strong merits both from the theoretical and from the practical point of view. Our study is based on the known explicit set of universal starting points, for each degree $d$, that are guaranteed to find all roots of polynomials of degree $d$ (appropriately normalized). We show that this set contains $d$ points that converge very quickly to the $d$ roots: the expected total number of Newton iterations required to find all $d$ roots with precision $\varepsilon$ is $O(d^3 \log^2 d + d \log |\log \varepsilon|)$, which can be further improved to $O(d^2 \log^4 d + d \log |\log \varepsilon|)$. The key argument shows that many root finding orbits are ‘$R$-central’ in the sense that they stay forever in a disk of radius $R$, and each iteration ‘uses up’ an explicit amount $A_{n,k}(\ell)$ of area within this disk.
Keywords: root finding, Newton’s method, polynomial, complexity, complex dynamics 2022
Mathematics Subject Classification numbers: 49M15, 37F10, 37N30, 68Q25
Secondary: 65H04, 65-04

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

1. Introduction

Finding roots of polynomials is one of the oldest problems in mathematics, and it is of significant interest today, in all areas of theoretical mathematics as well as in applications such as computer algebra and computer geometry. Especially in statistical physics and dynamical systems, there is a natural need to find all roots of polynomials of very high degrees. Surprisingly, theory and practice are not as well understood as one might expect: there are root finding methods that have near-optimal theoretical complexity, and others are very successful to find roots in practice, some to very large degrees; but apparently none so far that combines theory and practice. We describe several known methods below.

Newton’s root-finding method is as old as analysis, and has been known for a long time to be a very efficient method for locally finding approximate roots of smooth equations such as polynomials: once a reasonable approximation to a simple root is known, every iteration of the Newton method doubles the number of valid digits (this was surely known in the 19th century, if not earlier; see e.g. Kantorovich [K] for an explicit result from the 1940’s). However, Newton’s method has a reputation as being difficult to understand as a global dynamical system due to its ‘chaotic’ nature and various further problems described below; see for instance [Rü].

The purpose of our work is to control this ‘chaotic’ dynamics by methods from holomorphic dynamics and complex geometry, and to show that the classical simple and stable (and elegant) Newton method is far more efficient than anticipated: this method is supported by theory and works well in practice. It is known from [HSS] that for every degree $d$ there is an explicit small universal set $S_d$ of starting points that finds all roots of all complex polynomials of degree $d$ (appropriately normalized) when Newton’s method is started at these points; we have $|S_d| = 1.11d \log^d d$. The main result of the present paper (with a subsequent refinement in [BAS]) is a strong estimate on the complexity of this process: there is a subset of $d$ of these points that together only need $O(d^2 \log^3 d + d \log \log \frac{1}{\varepsilon})$ Newton iterations to find all roots with precision $\varepsilon$ at least in the absence of near-multiple roots (if there are near-multiple roots, then all the isolated roots are found with this speed; the others are found too but with complexity $O(d^3 \log^2 d(d + \log \frac{1}{\varepsilon}))$, so no a priori knowledge is required). These complexity bounds are near-optimal (i.e. up to logarithmic factors in $d$) for methods that start the Newton iteration at points where good control can currently be established, that is at uniform distance from the disk containing all roots. We should also point out that the output is an unordered set of approximate roots, in the sense that there is a bijection to the true set of roots that changes each component by no more than $\varepsilon$ (elsewhere, the quality of approximation is sometimes measured by the values of the polynomial).

Our theoretical estimates are complemented by practical experiments [SS, RaSS] that show that Newton’s method routinely finds all roots of complex polynomials of degrees exceeding one billion ($2^{30} > 10^9$), and in some cases with experimental complexity of $O(d \log^{1.1} d)$ iterations, so indeed Newton has the potential to combine theoretical analysis with practical success in a unique way.
Here are some of the challenges that Newton’s method faces in the case of a polynomial $p$ in a single complex variable:

- orbits of the Newton map that get close to zeroes of the derivative $p'$ will, under the Newton dynamics, jump near $\infty$ and will take a long time until they can get close to roots (if ever);
- there may be open sets in $\mathbb{C}$ in which the Newton dynamics does not converge to any root of $p$ (even for as simple polynomials as $p(z) = z^3 - 2z + 2$);
- the boundary of the attracting basins of the various roots (the Julia set) may have positive measure, so that the set of ‘bad’ starting points has positive measure;
- the Newton dynamics is ‘chaotic’ in a precise sense on its Julia set, and globally unpredictable in the convex hull between the roots;
- even if almost all starting points in $\mathbb{C}$ converge to some root of $p$, it is not clear how to find starting points for all roots: it is conceivable that some roots are ‘hidden’ and can be found only from small sets of starting points;
- the obvious possibility of ‘deflation’ should be avoided (dividing the polynomial by polynomial factors representing roots already found, thus reducing the degree): deflation is numerically unstable unless the roots are found in a certain order, and it destroys specific forms of the polynomial that are easy to evaluate;
- finally, one needs efficient estimates on the required number of iterations.

This manuscript addresses all these issues: we specify, for arbitrary polynomials in a single complex variable, a universal set of starting points (depending only on the degree and some normalization) from which all roots of all such polynomials are found, and so that the required number of iterations (or arithmetic complexity) is small: for fixed accuracy $\varepsilon$, it is $O(d^2)$ in the expected case (incorporating refinements from [BAS] that build on the present paper), and $O(d^4)$ in the worst case when there are multiple roots (all up to factors of log $d$). More precisely, we prove the following theorem (the first half of which is not new, but required to state the main result).

**Main Theorem (Newton efficiency).** For every degree $d \geq 2$, let $\mathcal{P}_d$ be the set of complex polynomials of degree $d$, normalized so that all roots are in the complex unit disk $\mathbb{D}$, and denote the $n$th iterate of $N_p$ by $N_p^n$. Then there is an explicit and finite universal set $\mathcal{S}_d$ consisting of $3.33d \log^2 d (1 + o(1))$ points in $\mathbb{C}$ with the following property:

- for every $p \in \mathcal{P}_d$, written as $p(z) = c \prod (z - \alpha_j)$, there are $d$ points $z^{(1)}, \ldots, z^{(d)} \in \mathcal{S}_d$ with $N_p^n(z^{(j)}) \to \alpha_j$ as $n \to \infty$.

Given $\varepsilon > 0$, let $n_j \in \mathbb{N}$ be so that $|N_p^n(z^{(j)}) - \alpha_j| < \varepsilon$ for all $n \geq n_j$. Then the required number of iterations is bounded as follows.

- **Worst-case complexity:** we always have
  \[ \sum_j n_j \in O \left( d^4 \log^2 d + d^3 \log^2 d \log \varepsilon \right). \]

- **Average complexity:** if the roots $\alpha_j$ of $p$ are all simple and have mutual distance at least $d^{-k}$ for some $k \in \mathbb{N}$, then

\[
\sum_j n_j \leq O \left( d^3 (\log^2 d)(\log d + k) + d \log \log \varepsilon \right).
\]  

In particular, if the roots are randomly distributed in \( \mathbb{D} \) (independently with respect to Euclidean area), or the coefficients are chosen randomly (subject to the condition that the roots are in \( \mathbb{D} \)), then the expected number of iterations is \( O \left( d^3 \log^3 d + d \log \log \varepsilon \right) \).

In [BAS], we refine this theorem so that for randomly distributed roots in \( \mathbb{D} \), one can expect \( \sum_j n_j \leq O(d^3 \log^3 d + d \log \log \varepsilon) \) (and similar bounds hold if the coefficients are distributed randomly, subject to the restriction that all roots are in \( \mathbb{D} \)). Our current results build upon earlier work [Sch2] that established convergence in polynomial time, but with rather sub-optimal degree. In turn, [BAS] builds on the current paper. It is reasonable to expect that the worst-case complexity is really one degree lower than stated (see the remark at the end of section 7).

This result measures the complexity in terms of Newton iterations. Each Newton iteration needs to evaluate both \( p \) and \( p' \), so requires \( O(d) \) arithmetic operations (for instance if \( p \) is given in coefficient form, then at least the \( d + 1 \) coefficients have to be processed). However, the simultaneous evaluation of a degree \( d \) polynomial at \( d \) different points is possible in complexity \( O(d) \) up to logarithmic factors, for instance using interpolation (compare for instance [MB], [AHU, section 8.5]). Therefore, the arithmetic complexity differs from the complexity in terms of Newton iterations only by a log factor.

The far more restrictive issue is whether high degree polynomials can be computed accurately in terms of coefficients. This is a serious issue for polynomial evaluation. Our point of view is that we can contribute to root finding but not to polynomial evaluation, so we treat polynomial evaluation as a ‘black box’ and assume exact arithmetic for evaluation. Note that root finding by Newton is stable, so small arithmetic errors are self-correcting; this can be quantified precisely (see e.g. [SSt, theorem 5]). In any case, all root finding experiments in [SSt, RaSS] come with a-posteriori confirmation that all roots have indeed been found with given precision.

Note on parallelization. A parallel or multi-core computer can take advantage of the inherent parallel structure of the independent Newton iterations, so the algorithm is almost ideally parallelizable. On the other hand, as mentioned earlier, each processor can compute \( d \) independent orbits almost as fast as a single orbit, no matter whether the computer has multiple cores or a single one (in classical von Neumann architecture).

1.1. Polynomial root-finding

There is an enormous literature on polynomial root-finding; see for instance the surveys by McNamee and Pan [McN1, McN2, MP] and the references therein, as well as Pan [P3] for a more recent overview.

Two of the most established root finding methods that are known to work well in practice and that have passed the test of time are the methods by Weierstrass (also known as Durand–Kerner) and by Ehrlich–Aberth (also known as Börsh–Sapan). Both approximate the \( d \) roots of a degree \( d \) polynomial by iterating in \( \mathbb{C}^d \), and locally they converge quadratically resp. cubically to a vector of roots in the absence of multiple roots. Ehrlich–Aberth is the basis for the successful implementation of the MPSolve root finding software package (see Bini and Robol [BR] for a description of MPSolve 3.0), and it has successfully found roots of polynomials of degrees in the millions and recently even in the billions.

However, in the words of Pan [P1], ‘Theoretically, the weak point of these algorithms is their heuristic character. … Moreover, in spite of intensive effort of many researchers, convergence
of these algorithms has been proved only in the cases where the initial point is already close to a zero or where another similar condition is satisfied. In other words, these methods lack global theory and have neither a proof of convergence in general, nor an estimate on their speed of convergence.

It was recently discovered that the Weierstrass (Durand–Kerner) method is not generally convergent \([\text{ReSS}]\); for an open set of polynomials of every degree \(d \geq 3\) there are open sets of starting points that converge to periodic cycles away from the roots (even for simple polynomials such as \(z^2 + z + 175\)). There are reasons to expect that a similar result is true for Ehrlich–Aberth. It is well known that both methods have orbits that fail to be defined after finitely many iterations (because they reach the locus of indeterminacy). Moreover, it was recently established that they all may have orbits that are defined forever but converge to \(\infty\) or some other non-root fixed point \([\text{ReSS}, \text{Re}]\).

Another well known approach to root finding are eigenvalue methods: these perform well in practice (see for instance \([\text{BBEGGI}]\)) especially for moderate degrees, but they are not supported by theory.

On the other end of the spectrum, there are root finding methods that have near-optimal performance in theory but that cannot, or have not, been used successfully in practice. In particular, Shub and Smale \([\text{ShSm}]\) have an algorithm that may be viewed as a higher-order variant of Newton’s method (where the order depends on the degree of the polynomial), and for which success is guaranteed with explicit complexity bounds. Kim \textit{et al} \([\text{KMS}]\) have a path-lifting algorithm based on Newton’s method that also comes with good complexity bounds. Finally, we would like to mention Pan’s algorithm \([\text{P2}]\) with near-optimal complexity bounds, but it has stability issues. All these methods have, to our knowledge, never shown their strength and feasibility in practice through systematic or large-scale practical implementations.

1.2. Notation

Throughout this text, we will fix a polynomial \(p \in \mathcal{P}_d\) of degree \(d \geq 2\), and we write \(p(z) = c \prod_j (z - \alpha_j)\) and \(N_p(z) = z - p(z)/p'(z)\) (of course, the roots \(\alpha_i\) are not assumed to be known ahead of time). The leading coefficient \(c\) cancels for Newton’s method and will be omitted. We denote the \(n\)th iterate of \(N_p\) by \(N_p^n\), and similarly \(f^n\) for the \(n\)th iterate of a function \(f\).

Each root \(\alpha_j\) has its basin \(U_j \subset \mathbb{C}\): this is the set of points that converge to \(\alpha_j\) under iteration of \(N_p\). The immediate basin \(U_j\) is the connected component of \(U_j\) containing the root \(\alpha_j\). As long as we focus attention on a single root, we call it \(\alpha\) and its immediate basin \(U\). It is well known that \(U\) is simply connected and unbounded \([\text{Pr}]\) (see also \([\text{Sh}, \text{HSS}]\)). Denote by \(d_U\) the distance with respect to the unique hyperbolic metric on \(U\) with constant curvature \(-1\).

Let \(D_r(a) := \{z \in \mathbb{C} : |z - a| < r\}\) for \(a \in \mathbb{C}\) and \(r > 0\), and let \(D := D_1(0)\) be the complex unit disk. For a set \(K \subset \mathbb{C}\) and \(z \in \mathbb{C}\), we denote by \(\text{dist}(z, K)\) Euclidean distance of \(z\) to \(K\), that is \(\text{dist}(z, K) = \sup\{r > 0 : D_r(z) \cap K = \emptyset\}\). The Euclidean straight line segment connecting two points \(a, b \in \mathbb{C}\) is denoted \([a, b]\).

We denote the natural logarithm by \(\log\) and the dyadic logarithm by \(\log_2\).

1.3. Overview of the arguments and structure of the paper

The first new ingredient in this paper is the concept of ‘\(R\)-central orbits’: these are orbits for Newton’s method that stay in the disk \(D_R(0)\), so we can maintain control. We will show how to find starting points of Newton’s method that are in immediate basins and that have \(R\)-central orbits. In order to estimate the possible number of iterations, the fundamental idea is the \textit{area used per iteration step}. We partition \(D_R(0)\) into domains \(S_k\) so that for \(z \in S_k\), the nearest root
has distance approximately $2^{-k}$ from $z$. This will imply that $|z - N_p(z)| \geq \ell := 2^{-k}/d$. We find orbits $(z_n)$ in the immediate basin $U$ of $\alpha$ for which the hyperbolic distance $d_U(z_n, z_{n+1}) = O(\log d)$. Roughly speaking, Euclidean distance bounded below by $\ell$ and hyperbolic distance bounded above by $\log d$ means that $z_n$ and $z_{n+1}$ can be connected by a hyperbolic geodesic segment $\gamma \subset U$ that has Euclidean distance at least $\ell / \log d$ from the boundary, so this curve ‘uses up’ an area of approximately $|A_{n,k}(\ell)| \geq \ell^2 / \log d = 4^{-k}/d^2 \log d$ (length times width of the neighbourhood of the curve). But $S_k$ is contained in the union of $d$ disks of radius $2^{-k}$ and hence with total area at most $\pi d 4^{-k}$, so there is room for no more than $(d4^{-k})(4^{-k}/d^2 \log d) = d^3 \log d$ iterations in each $S_k$ (always up to bounded factors). In the worst case, when there are multiple or near-multiple roots, we will show that we need to consider $k \leq O(d)$, hence a total of $O(d^3 \log d)$ iterations is required. If the roots are well separated from each other, for instance if they are randomly distributed, it turns out that it $k \leq \log d$ will usually suffice until the domain of quadratic convergence is reached where $\log \ell / \log \epsilon$ further iterations yield precision $\epsilon$, so we need $O(d^3 \log d + \log \log \epsilon)$ iterations. All these count the number of iterations required to find a single root. But since all roots are competing for the area, the number of iterations to find all roots satisfies the same bounds (except that the small $\log \ell / \log \epsilon$ term acquires a factor $d$).

Of course, all these estimates have to be made precise, and we have to make sure that the domains of area do not overlap, which will introduce additional factors of $\log d$.

The paper is structured as follows. In section 2 we introduce $R$-central orbits and show how to find them. In section 3, we construct an explicit finite set of starting points that contains, for each root, at least one $R$-central orbit $(z_n)$ with $d_U(z_n, z_{n+1}) = O(\log d)$ as required. In section 4 we estimate the area needed per single iteration step. In section 5, we estimate the area needed for each orbit: the main point is to make sure that the pieces of area assigned to each orbit point are disjoint; this will introduce another factor $\log d$.

It remains to discuss when to stop the Newton iteration. If the roots are well-separated from each other, then the iteration reaches the domain of quadratic convergence, and the necessary stopping criterion will be given in section 6. We then bring all arguments together and describe the required number of iterations for ‘good’ starting points. The worst case of roots that are not well separated, or possibly even multiple, is treated in section 7.

In a brief final section 8, we report some of the aforementioned numerical experiments that support our claim that Newton’s method stands out as a root finding method that combines good theory with remarkable practical usefulness, and we compare it with the performance of the Ehrlich–Aberth-method.

In appendix, we prove a general lemma on the area of certain neighbourhoods of hyperbolic geodesics in Riemann domains; a major concern is to make sure that these neighbourhoods do not overlap.

Note that in general we are not interested in optimizing constant factors. At a number of places, we specify explicit constants when it is easy to do so, rather than referring to unspecified values. It is certainly possible to improve most constants significantly, but we want to make clear that all of them have very feasible values.

2. Channels and $R$-central orbits

After a brief review of the geometry of immediate basins outside of $\mathbb{D}$, our main goal in this section is to give a condition on orbits that always stay within a certain Euclidean disk around the origin; such orbits will be called ‘$R$-central’.
Consider the immediate basin $U$ of a root $\alpha$ (see figure 1); as mentioned above, it is simply connected. The geometry of these basins outside of $\mathbb{D}$ has been studied in [HSS, section 3]; in particular, $U$ is unbounded. A channel of $U$ is an unbounded connected component of $U \setminus \overline{D}$, and an access to $\infty$ of $U$ is a homotopy class (with endpoints fixed) of curves in $U \cup \{\infty\}$ connecting $\alpha$ to $\infty$. Every access to $\infty$ is fixed by $N_p$, and so is every channel: if $B$ is a channel of $U$, then $N_p(B) \setminus \overline{D} = B$. Each channel contains one access to $\infty$, and each access to $\infty$ defines one channel through which it runs to $\infty$.

For a channel $B$, we denote by $A := B/(N_p)$ its quotient by the Newton dynamics (identifying points on the same orbit). This is a conformal annulus and as such has a modulus $\mu = \mu(A) =: \mu(B)$. One of the main results in [HSS] is that each root of a polynomial of degree $d$ has a finite positive number of channels, and at least one of them has modulus at least $\pi/\log d$ [HSS, propositions 6 and 7].

On the set $B$, we will use three different hyperbolic metrics (all with constant curvature $-1$): $d_B$ is the hyperbolic metric on $B$, while $d_U$ is the hyperbolic metric on $U$ restricted to $B$. Finally, the quotient annulus $A = B/(N_p)$ has a hyperbolic metric, and the infinitesimal metric on $A$ lifts to an infinitesimal metric on $B$, called $d_A$. (Only for $d_B$ is $B$ a complete metric space; for $d_A$ and $d_U$, certain boundary points can be reached in finite distance.) Asymptotically near $\infty$, all three metrics coincide: it is a general principle that if $U_0$ is a hyperbolic Riemann surface and $U_2 \subset U_1 \subset U_0$ are open subsets, then restricted to $\overline{U_2}$ the hyperbolic metrics of $U_0$ and $U_1$ differ little if $U_0 \setminus U_1$ is far away from $\overline{U_2}$ (with respect to the hyperbolic metric of $U_0$); see for instance [MR, proposition 3.4]. This implies that on all of $B$ we have $d_B > d_A > d_U$: the map $N_p$ is an isometry on $A$ and a contraction on $U$, hence $d_A > d_U$ (for a point $z \in B$, there is a $z_n \in U_0 \setminus U_1$ far away from $\overline{U_2}$, and $N_p(z_n) = z$, and asymptotic equality of metrics near $z_0$ implies the claimed inequality). Similarly, there exists a branch $N_p^{-1}: B \to B$; this branch contracts $d_B$ (inclusion of hyperbolic domains) and preserves $d_A$, hence $d_B > d_A$ (for $z \in B$, there is a
Definition (R-central orbits). An orbit \((z_n)\) will be called R-central if 
\[ |z_n| \leq 1 \implies |z_{n+1}| \leq R \text{ for all } n' \geq n. \]

In view of the remark just above, for an orbit to be R-central it is sufficient to show that \(|z_0| \leq 1\) implies \(|z_{n+1}| \leq R\). Figure 2 illustrates that while orbits that are not R-central are a substantial conceptual problem for the Newton dynamics, ‘most’ orbits are R-central even for moderate values of \(R\) such as \(R = 2\). However, for every \(R\) the orbits that fail to be R-central form an open dense subset of the Julia set of \(N_p\) (the backwards orbit of \(\infty\) is dense in the Julia set, and a neighbourhood of these points is not R-central).

To locate R-central orbits, we need the following definition.

Definition 2.2 (Central subannulus and central channel). Let \(B\) be a channel of an immediate basin \(U\), let \(A := B/(N_p)\) be the quotient annulus, and let \(\mu := \text{mod}A\) be its modulus (so that the core curve of \(A\) has hyperbolic length \(\pi/\mu\)). We define the central subannulus of \(A\) as the set of points \(z \in A\) such that every smooth non-contractible simple closed curve in \(A\) through \(z\) has hyperbolic length less than \(2\pi/\mu\).

Let the central subchannel of \(B\) be the preimage in \(B\) (under the quotient map) of the central subannulus: this is the set of points \(z \in B\) with \(d_A(z, N_p(z)) < 2\pi/\mu\).
Lemma 2.3 (Central subchannel). If \( A \) has modulus \( \mu \), then the central subannulus of \( A \) is a parallel subannulus of \( A \) with modulus greater than \( 2\mu/3 \).

Proof. Setting \( h = \mu/2 \), the quotient annulus \( A \) is conformally equivalent to the horizontal strip \( \{ z \in \mathbb{C} : |\text{Im} z| < h \} \) modulo \( z \sim z + 1 \). The infinitesimal hyperbolic metric on the strip is given by

\[
ds = \frac{\pi|dz|}{2h \cos(\pi|y|/2h)},
\]

so the unique simple closed geodesic in \( A \) is the projection of the real axis, and it has length \( \pi/\mu \). A parallel subannulus is the set of points \( x + iy \) for which \( |y| \) satisfies a certain upper bound. In particular, if \( \cos(\pi|y|/2h) > 1/2 \), then \( A \) is conformally equivalent to the horizontal curves of Euclidean length 1 at imaginary part \( y \) has hyperbolic length less than \( 2\pi/\mu \), and the corresponding geodesic is shorter than this, but longer than \( \pi/\mu \). The condition \( \cos(\pi|y|/2h) > 1/2 \) is satisfied if \( |y| < 2h/3 \), so the central subannulus has modulus greater than \( 2\mu/3 \).

We can now state the main result of this section.

Theorem 2.4 (Large central subchannels have \( R \)-central orbits). Let \( U \) be the immediate basin of a root \( \alpha \) and let \( B \subset U \) be a channel of \( U \) with largest modulus. Then all points in the central subchannel of \( B \) have \( R \)-central orbits, for a value of \( R \) that is specified in proposition 2.7.

The idea of the proof is simple: if \( z_0 \) is in the central subchannel of \( B \), then for its orbit \( (z_n) \) we have the estimate

\[d_U(z_n, N_p(z_n)) \leq d_U(z_0, N_p(z_0)) < d_A(z_0, N_p(z_0)) < 2\pi/\mu.\]

If \( |z_n| \leq 1 \), then the bound on \( d_U(z_n, N_p(z_n)) \) implies an upper bound on \( |N_p(z_n)| \) and thus makes the orbit \( R \)-central. We believe that \( R = 2 \) works for all but a few low values of \( \text{deg} p \), but to prove this would require more control on the possible shapes of the channels than we can currently provide. The estimates that follow in this section are relatively weak because they have to account for all possible shapes of channels.

It might be interesting to observe that \( R \)-central orbits, while being helpful for finding roots of a polynomial \( p \) (i.e. attracting fixed points of \( N_p \)), do not seem to be related to attracting cycles of \( N_p \) of periods greater than 1. For instance, consider the family of degree 4 polynomials \( p(z) = (z^2 - 1)(z^2 + \alpha^2) \) with roots at \( \pm 1 \) and \( \pm i\alpha \): the two roots of \( p'' \) are critical points of \( N_p \) but not roots of \( p \), and they are contained in the convex hull of all the roots of \( p \) by Lucas’ theorem. One can choose \( \alpha \) so that the two roots of \( p'' \) are exchanged by \( N_p \), so they form a superattracting 2-cycle within the convex hull of the roots and are thus \( R \)-central for every \( R > |\alpha| \). On the other hand, consider the polynomial \( p(z) = z^5 - z^3 + 1 \). Here the point \( z = 0 \) is simultaneously a root of \( p' \), hence a pole of \( N_p \), and a root of \( p'' \), hence a critical point of \( N_p \), so we have a critical point of \( N_p \) that maps to \( \infty \). Adding a perturbation \( \varepsilon z^2 \) with small (complex) \( \varepsilon \), the perturbed Newton map will have a critical point near \( z = 0 \) with image close to \( \infty \), and by an appropriate choice of \( \varepsilon \) one can assure that this critical point is periodic (with large period), so it is part of a superattracting cycle that contains a point close to \( \infty \). During this perturbation, the roots of \( p \) change continuously, so for any \( |\varepsilon| > 1 \) one can make sure that there is a superattracting cycle that is not \( R \)-central. This observation is not relevant to the goals of this paper: the existence of attracting cycles is one of the major problems of Newton’s method, but of course finding attracting cycles of \( N_p \) is not the purpose of (rather an obstacle to) finding roots of \( p \).
We start proving theorem 2.4 by stating a simple and well-known worst-case estimate of hyperbolic distances in a hyperbolic domain.

Lemma 2.5 (Standard bound on hyperbolic arc length). Let \( V \) be a Riemann domain, \( p, q \in V \), and \( a \in \partial V \). If \( s = |p - a| \), then

\[
d_{V}(p, q) \geq \frac{1}{2} \int_{0}^{[p-q]} \frac{dt}{s + t} = \frac{1}{2} \log \left( 1 + \frac{|p-q|}{s} \right).
\]  

(2)

Proof. The key ingredient in this proof is to compare infinitesimal hyperbolic distance in \( V \), denoted \( ds \), and infinitesimal Euclidean distance, denoted \( |dz| \). At a point \( z \in V \) with \( r := \text{dist}(z, \partial V) \), they satisfy the well-known bound \( ds \geq |dz|/2r \).

If \( \gamma : [0, T] \to V \) is a smooth curve parametrized by Euclidean arc length, then its hyperbolic length in \( V \) is at least

\[
\frac{1}{2} \int_{0}^{T} \frac{dt}{\text{dist}(\gamma(t), \partial V)} \geq \frac{1}{2} \int_{0}^{T} \frac{dt}{\text{dist}(\gamma(0), \partial V) + t} = \frac{1}{2} \log \frac{\text{dist}(\gamma(0), \partial V) + T}{\text{dist}(\gamma(0), \partial V)}.
\]

In particular, if \( \gamma : [0, T] \to V \) is the hyperbolic geodesic segment connecting \( p \) to \( q \), parametrized by Euclidean arc length \( T \geq |p-q| \), then \( \text{dist}(\gamma(0), \partial V) \leq |p-a| = s \) and the claim follows.

\( \square \)

Lemma 2.6 (Hyperbolic distance across fundamental domain). Let \( U \) be an immediate basin and \( B \subset U \) be a channel of modulus \( \mu \). Then all \( w, \tilde{w} \in B \) with \( |w|(d-1)/d \geq |\tilde{w}| > 5 \) satisfy \( d_{V}(w, \tilde{w}) > 2/5(\mu + \pi) \).

This result is somewhat weaker than expected: one would expect approximately \( d_{V}(w, \tilde{w}) \geq \pi / \mu \) (and perhaps a simpler proof), but channels may have complicated geometry; and while some intermediate estimates become less elegant, the final result will be affected only marginally.

Proof. By \([\text{HSS}, \text{lemma 3}]\), we have \( |N_{\mu}(w) - w(d-1)/d| < 1/d \) whenever \( |w| > 1 \). Let \( w' := N_{\mu}(w); \) we have \( d_{B}(w, w') > d_{K}(w, w') \geq \pi / \mu \) and \( |w-w'| \leq (|w|+1)/d \). If all points on \( [w, w'] \) had Euclidean distance to \( \partial V \) at least \( 2\mu(|w|+1)/\pi d \), then the standard estimate on hyperbolic distance would imply \( d_{B}(w, w') \leq 2|w-w'|/(2\mu(|w|+1)/\pi d) < \pi / \mu \), a contradiction. Therefore, there are points \( a \in \partial B \) and \( w'' \in [w, w'] \) with \( |a-w''| < 2\mu(|w|+1)/\pi d \). This implies

\[
|a - w| \leq |a - w''| + |w - w'| < 2(\mu / \pi + 1)(|w| + 1)/d.
\]

The point \( a \in \partial B \) is on \( \partial U \) or on \( \partial B \cap U \). Suppose first that \( a \in \partial U \). Then by lemma 2.5, and using the well-known bound \( \log (1 + x) > x/(1 + x) \) for \( x \in (0, 1), \)

\[
d_{U}(w, \tilde{w}) \geq \frac{1}{2} \log \left( 1 + \frac{|w - \tilde{w}|}{(2\mu / \pi + 1)(|w| + 1)/d} \right)
\]

\[
> \frac{1}{2} \log \left( 1 + \frac{|w|/d}{(2\mu / \pi + 1)(|w| + 1)/d} \right) = \frac{5}{2} \log \left( 1 + \frac{5}{6(2\mu / \pi + 1)} \right)
\]

\[
> \frac{5\pi}{24\mu + 22 \pi} > \frac{2}{5(\mu + \pi)}.
\]
Now we discuss the case that \( a \in \partial B \cap U \), i.e. \( |a| = 1 \) and \( |w| - 1 \leq |a - w| < (2\mu/\pi + 1)(|w| + 1)/d \), hence

\[
\mu > \frac{\pi}{2} \left( d \frac{|w| - 1}{|w| + 1} - 1 \right) > \frac{\pi}{2} \left( \frac{4d - 6}{6} \right) = \frac{\pi(2d - 3)}{6}
\]

\( (3) \)

or \( d < 3(\mu/\pi) + 3/2 \). Every point \( a' \in \mathbb{D} \) satisfies \( |w - a'| < |w| + 1 \), and if we also assume \( a' \in \partial U \) we get

\[
d_U(w, \tilde{w}) \geq \frac{1}{2} \log \left( 1 + \frac{|w - \tilde{w}|}{|w| + 1} \right) \geq \frac{1}{2} \log \left( 1 + \frac{5}{d} \right) > \frac{5}{6d} \left( \frac{1}{2} + \frac{5}{6d} \right) \]

\[
= \frac{5}{12d + 10} \geq \frac{5}{36\mu + 28} = \frac{2}{36\mu + 28\pi} > \frac{2}{5(\mu + \pi)},
\]

so the claimed inequality holds in both cases.  

For the following proposition, recall that every immediate basin has at least one channel with modulus \( \pi/\log d \) or larger, while there may always be channels with arbitrarily small moduli. We show that an upper bound for the modulus of all channels of an immediate basin yields an existence criterion of \( R \)-central orbits in this immediate basin.

**Proposition 2.7 (Existence criterion for \( R \)-central orbits).** Let \( U \) be the immediate basin of some root (simple or multiple), and let \( \mu \geq \pi/\log d \) be the maximum of moduli of all its channels. If a point \( z \in U \) has \( d_U(z, N_p(z)) \leq 2\pi/\mu \), then the orbit of \( z \) is \( R \)-central for \( R \leq 5(d/(d - 1))^{[5\pi(\log d + 1)]} \), and the same holds for all points on the hyperbolic geodesic in \( U \) connecting \( z \) to \( N_p(z) \).

**Proof.** Any two points \( w, \tilde{w} \in U \) with \( |w|/(d - 1)/d \geq |\tilde{w}| > 5 \) have \( d_U(w, \tilde{w}) > 2/5(\mu + \pi) \) by lemma 2.6, so going in \( U \) from radius \( r \geq 5 \) to radius \( r(d/(d - 1))^k \) (for some integer \( k \geq 1 \)) one needs to traverse at least \( k \) disjoint concentric annuli with boundary radii differing by a factor of \( d/(d - 1) \), so the hyperbolic distance is at least \( 2k/5(\mu + \pi) \). In particular for \( k = [5\pi(\log d + 1)] \) the required hyperbolic distance is at least

\[
[5\pi(\log d + 1)] \cdot \frac{2}{5(\mu + \pi)} \geq 2\pi \frac{\log d + 1}{\mu + \pi} > \frac{2\pi}{\mu} > d_U(z, N_p(z)),
\]

which exceeds the available hyperbolic distance along the orbit of \( z \), so this orbit is indeed \( R \)-central for the choice of \( R \) specified in the claim.

Now consider any \( z' \) on the hyperbolic geodesic in \( U \) connecting \( z \) to \( N_p(z) \). Then

\[
d_U(z', N_p(z')) \leq d_U(z', N_p(z)) + d_U(N_p(z), N_p(z')) \\
\leq d_U(z, N_p(z)) + d_U(z, z') = d_U(z, N_p(z)) < 2\pi/\mu
\]

and the arguments given above also apply to \( z' \). 

**Remark.** This result provides an upper bound for \( R \) that is uniform in \( d \): we have \( R \leq 5(d/(d - 1))^{[5\pi(\log d + 1)]} \leq 5\pi^{[5\pi(\log d + 1)]}/d \). More precisely, we have \( R < 50 \) for \( d > 30 \), and \( R < 10 \) for \( d > 133 \). Ultimately, the precise value of \( R \) is not of too large importance as it will enter our estimates only logarithmically: the number of iterations scales with \( M \in \mathbb{N} \) such that \( 2^M - 1 \geq R \) (see proposition 6.2); we can use \( M = 5 \) for \( d \geq 41 \) and \( M = 3 \) for \( d \geq 316 \). Presumably, \( R \leq 3 \) and \( M = 2 \) work for all but very few values small of \( d \).
Proof of theorem 2.4. Let \( \mu \) be the modulus of \( B \). If \( z \) is in the central subchannel of \( B \), then \( d_\mu(z, N_\mu(z)) < d_\alpha(z, N_\alpha(z)) < 2 \pi / \mu \). By proposition 2.7, the orbit of \( z \) is \( R \)-central. \( \square \)

3. Good starting points

In this section, we construct an explicit finite set of starting points \( \mathcal{S}_d \) depending only on the degree \( d \) and the normalization of \( p \) so that for each root \( \alpha \) of \( p \), one of the points \( z \in \mathcal{S}_d \) is in the immediate basin \( U_\alpha \) and satisfies \( d_\mu(z, N_\mu(z)) \leq 2 \log d \), and so that the orbit of \( z \) is \( R \)-central for \( R \) as in proposition 2.7.

All we need to do is specify a finite set of starting points that will intersect, for every root, the central subchannel of the channels with largest modulus. Since every root has a channel with modulus at least \( \pi / \log d \), by lemma 2.3 it is sufficient to specify a finite set of starting points that intersects all subchannels of all channels with moduli at least \( 2 \pi / 3 \log d \). This can be accomplished by the methods in [HSS], so we can now construct an explicit point grid (in that paper, we used \( 0.2663 \log d \) concentric circles that each contain \( 33.33d \log d \) points; here we use \( 3/2 \)) as many circles because we want to hit the channel within the central subchannel with \( 2/3 \) the modulus.

Definition 3.1 (Efficient grid of starting points). For each degree \( d \), we construct a circular grid \( \mathcal{S}_d \) of starting points as follows (as sketched in figure 3). For \( \nu = 1, 2, \ldots, s = \lfloor 0.4 \log d \rfloor \), set

\[
r_\nu := (1 + \sqrt{2}) \left( \frac{d - 1}{d} \right)^{(\nu - 1)/2} \]

and for each circle around 0 of radius \( r_\nu \), choose \( [8.33d \log d] \) equidistant points (independently for all the circles).

Theorem 3.2 (Efficient grid of starting points). For each degree \( d \), the set \( \mathcal{S}_d \) contains \( 3.333d \log^2 d (1 + o(1)) \) points, and it has the following universal property. If \( p \) is any complex polynomial of degree \( d \), normalized so that all its roots are in \( \mathbb{D} \), then there are \( d \) points in \( \mathcal{S}_d \) that converge to the \( d \) roots of \( p \), so that for each root, one of these points is in the central subchannel of a channel with largest modulus. The corresponding orbits are all \( R \)-central for a uniform value of \( R \) (as in proposition 2.7). More precisely, if \( \mu \) is the largest modulus of all channels of a root \( \alpha \) with immediate basin \( U \), then the corresponding orbit converging to \( \alpha \) satisfies \( d_\mu(z, N_\mu(z)) < 2 \pi / \mu < 2 \log d \).

Proof. The annulus

\[
V := \left\{ z \in \mathbb{C} : (1 + \sqrt{2}) \sqrt{(d - 1)/d} < |z| < (1 + \sqrt{2}) \right\}
\]

is contained in a fundamental domain of the Newton dynamics by [HSS, lemma 10]. For \( \nu = 1, 2, \ldots, s = \lfloor 0.4 \log d \rfloor \), subdivide \( B \) into \( s \) subannuli

\[
V_\nu := \left\{ z \in \mathbb{C} : (1 + \sqrt{2}) \left( \frac{d - 1}{d} \right)^{\nu/2} < |z| < (1 + \sqrt{2}) \left( \frac{d - 1}{d} \right)^{(\nu - 1)/2} \right\}.
\]

Since \( V \) is contained in a fundamental domain of the dynamics, each subchannel of any root with modulus \( \mu \) intersects \( V \) in a quadrilateral of modulus at least \( \mu \), and by the Grötzsch inequality, it intersects at least one \( V_\nu \) in a quadrilateral with modulus at least \( \mu \). Each root has a channel with modulus at least \( \pi / \log d \), so the central subchannel has modulus at
least $2\pi/3 \log d$, and this central subchannel intersects some $V_\nu$ in a quadrilateral with modulus at least $2\pi/3 \log d > 0.2663\pi$, independent of $d$. In [HSS, section 6], it is shown that $[8.3254d \log d]$ equidistributed points on each of these circles will find all quadrilaterals connecting the boundaries of the $B_\nu$ with modulus at least $0.2663\pi$, as in our case.

Therefore, the grid $S_d$ intersects all central subchannels of all largest channels, and the claim follows.

**Remark.** The number of starting points of $O(d \log d)$ from [HSS] has been further reduced to $O(d \log^2 d)$ in [BLS], by using a probabilistic set of starting points. This approach can also be used in our case.

Now we have a good set of starting points leading to $R$-central orbits. We proceed to estimate the number of required iterations.

### 4. Area per iteration step

In this section, we show that every iteration step ‘uses up’ a certain area in the plane; since $R$-central orbits remain within some disk $D_R(0)$, this will provide an upper bound on the possible number of iterations.

Consider some point $z \in U$, set $\tau := d_U(z, N_\rho(z))$, and let $\gamma : [0, T] \to U$ be the hyperbolic geodesic connecting $z$ to $N_\rho(z)$, parametrized by Euclidean arc length. For $t \in [0, T]$, let $\eta(t)$ be the Euclidean distance from $\gamma(t)$ to $\partial U$, and let $X(t)$ be the Euclidean straight line segment of length $\eta(t)$ (not containing the endpoints) with centre at $\gamma(t)$ and perpendicular to $\gamma$ at $\gamma(t)$, so that $\gamma(t)$ disconnects $X(t)$ into two open segments of length $\eta(t)/2$; see figure 4. All the segments $X(t)$ are disjoint (proposition A.2 in the appendix).

For $k \in \mathbb{Z}$, let $X_k(t)$ be the restriction of $X(t)$ to lengths at most $2^{-k+1}$ (i.e. $X_k$ is the perpendicular line segment to $\gamma(t)$ centred at $\gamma(t)$ and extending in both directions for a length of $\min(\eta(t)/2, 2^{-k})$). Let $A_k := \bigcup_{t \in [0, T]} X_k(t)$ be the subset of $U$ covered by the $X_k(t)$ for

![Figure 3](image-url)
Figure 4. A line segment $X(t)$ perpendicular to the geodesic $\gamma$; in both directions from the point $\gamma(t)$, it extends to a length of $\eta(t)/2$, where $\eta(t)$ is the distance from $\gamma(t)$ to $\partial U$ (Adapted from [HSS], with permission from Springer Nature).

$t \in [0, T]$, and for $\ell \leq T$ let $A_k(\ell) := \bigcup_{t \in [0, \ell]} X_k(t)$ be the analogous set for $t \in [0, \ell]$. This set of course depends on $z$, so a more explicit description of this set would be $A_k(\ell, z)$. We denote the Euclidean area of $A_k(\ell)$ by $|A_k(\ell)|$. Clearly $A_k(\ell) \subset A_{k-1}(\ell)$, and the limit as $k \to -\infty$ is $A_{-\infty}(\ell) = \bigcup_{t \in [0, \ell]} X(t)$.

**Lemma 4.1 (Area per iteration step).** We have $|A_k(\ell)| \geq \ell^2/(2\tau + 2^{k-1}\ell)$. Similarly, the set $\bigcup_{t \in [0, \ell]} X(t)$ has Euclidean area at least $\ell^2/2\tau$.

**Proof.** The hyperbolic length of $\gamma$ between $z$ and $\mathcal{N}_p(z)$ satisfies

$$\tau \geq \frac{1}{2} \int_0^T \frac{dt}{\eta(t)} \geq \frac{1}{2} \int_0^\ell \frac{dt}{\eta(t)}.$$

This implies

$$\int_0^\ell \frac{dt}{\min(\eta(t), 2^{-k+1})} \leq \int_0^\ell \frac{dt}{\eta(t)} + \int_0^\ell \frac{dt}{2^{-k+1}} \leq 2\tau + 2^{k-1}\ell.$$

Since all $X_k(t)$ are disjoint, an exercise in elementary differential geometry shows that

$$|A_k(\ell)| = \int_0^\ell 2 \min(\eta(t)/2, 2^{-k}) \, dt = \int_0^\ell \min(\eta(t), 2^{-k+1}) \, dt.$$

The Cauchy-Schwarz inequality for the functions

$$\sqrt{\min(\eta(t), 2^{-k+1})} \quad \text{and} \quad 1/\sqrt{\min(\eta(t), 2^{-k+1})}$$
implies
\[ |A_k(\ell)| = \int_0^\ell \min (\eta(t), 2^{-k+1}) \, dt \geq \frac{\ell^2}{\int_0^\ell \min (\eta(t), 2^{-k+1}) \, dt} \geq \frac{\ell^2}{2\tau + 2^{k-1}\ell}. \]

The computation for \( \bigcup_{\ell \in [0, \eta]} X(\ell) \) is even simpler, replacing the min by \( \eta(\ell) \).

Define the following subsets of \( \mathbb{C} \), for \( k \in \mathbb{Z} \):
\[ S_k^+ := \left\{ z \in \mathbb{C} : \min_j |z - \alpha_j| \leq 2^{-k} \right\} \]
and
\[ S_k := \left\{ z \in \mathbb{C} : \min_j |z - \alpha_j| \in \left( 2^{-(k+1)}, 2^{-k} \right) \right\} = S_k^+ \setminus S_{k+1}^+. \]

Every \( S_k^+ \) is the union of closed disks of radius \( 2^{-k} \) around all the roots, and for every \( M \in \mathbb{Z} \) the disk \( D_{2M-1}(0) \) is covered and thus partitioned by the \( S_k \) for \( k \geq -M \) (appropriately restricted).

**Lemma 4.2 (Distance from \( A_k(\ell) \) to roots).** If \( z \in S_k^+ \) and \( |z - N_p(z)| \geq \ell \), then each point in \( A_k(\ell) \) has Euclidean distance at most \( 2^{-k+1} + \ell \) from a root of \( p \).

**Proof.** Every point \( w \in A_k(\ell) \) has distance less than \( 2^{-k} \) from a point \( w' \) on the geodesic connecting \( z \) to \( N_p(z) \), and by construction \( w' \) has distance at most \( \ell \) from \( z \). Since \( z \in S_k^+ \), the point \( z \) has distance at most \( 2^{-k} \) from a root, and the claim follows by combining these three distances. \( \square \)

**5. Area along an orbit**

In this section, we investigate how much area is ‘used up’ along an orbit \((z_n)\); this involves specifying under which conditions the area used for different iteration steps is disjoint, and determining a geometric condition on the hyperbolic displacement in each Newton step before the orbit enters the domain of quadratic convergence near a root; this condition is used for estimating how many iterations the orbit can spend within each \( S_k \).

We continue to consider the immediate basin \( U \) of some root \( \alpha \); in this section, we will assume this root is simple. Let \( \varphi : U \to \mathbb{D} \) be a Riemann map with \( \varphi(\alpha) = 0 \); then \( f := \varphi \circ N_p \circ \varphi^{-1} : \mathbb{D} \to \mathbb{D} \) is holomorphic with \( f(0) = 0 = f'(0) \). Hence \( f(w)/w^2 \) is still holomorphic and sends \( \mathbb{D} \) to itself, and setting \( w := \varphi(z) \) it follows that \( |\varphi(N_p(z))| = |f(w)| \leq |w|^2 = |\varphi(z)|^2 < |\varphi(z)| \).

Smale [Sm] has the concept of ‘approximate zero’; that is a point sufficiently near a simple root from which the convergence is quadratic (see [Sm] for a precise formulation). We begin with a lemma that gives useful dynamic consequences both for the case when a finite orbit point is an approximate zero (parts 2 and 3), or when it is not yet (part 1): for us, a practical criterion is whether or not \( |\varphi(z)| < e^{1/2} - 1 \).

Consider the orbit \( z_n = N_p^m(z_0) \) for a point \( z_0 \in U \) and let \( \tau := d_U(z_0, z_1) \); then \( d_U(z_n, z_{n+1}) \leq \tau \) for all \( n \) (Eventually, we will have \( \tau = O(\log d) \)).
Lemma 5.1 (Hyperbolic distance to root and Newton dynamics). (a) If \( n' > n \) so that \( |\varphi(z_n)| > |\varphi(z_{n'})| \geq \varepsilon^{1/2} = 1 \approx 0.649 \), then
\[
d_H(z_n, z_{n'}) \geq d_H(|\varphi(z_n)|, |\varphi(z_{n'})|) > (n' - n)/2;
\]
(b) if \( |\varphi(z_n)| < 1/2 \), then \( |z_n - \alpha| < \varepsilon \) for all \( n' > n + \log 2 |\log 2 \varepsilon - 5| \); and
(c) if \( |\varphi(z_n)| < \varepsilon^{1/2} - 1 \), then \( |z_n - \alpha| < \varepsilon \) for all \( n' > n + 1 + \log 2 |\log 2 \varepsilon - 5| \).

**Proof.** Define sets \( A_n(z) := A_k(z, z_n) \): these are the sets \( A_k(z) \) based at the points \( z_n \), as defined at the beginning of section 4.

Lemma 5.2 (Disjointness of areas). The sets \( A_{n,k}(z) \) and \( A_{n',k}(z) \) are disjoint if \( n' - n > 2\tau + 4 \log 3 \) and \( |\varphi(z_{n'})| \geq \varepsilon^{1/2} - 1 \).

**Proof.** Since \( A_{n,k}(z) \) and \( A_{n',k}(z) \) intersect only the case \( A_{n,-\infty}(z) \), i.e. with perpendicular segments \( X(t) \) of length \( \eta(t) \) independent of \( k \).

It is a generality that if \( V \subset \mathbb{C} \) is a Riemann domain and \( D_r(z) \subset V \), then all points \( z \in D_r(z) \) have hyperbolic distance \( d_H(z, a) \leq \log 3 \); this follows by comparing hyperbolic distance in \( D_r(z) \) as follows: \( d_H(z, a) \leq d_{D_r(0)}(z, a) \leq d_{D}(1/2, 0) = \log 3 \). This implies that each point on \( X(t) \) has hyperbolic distance from its midpoint \( \gamma(t) \) of less than \( \log 3 \). The claim follows.

The following result gives an explicit upper bound how many points on an orbit \( (z_n) \) with \( d_H(z_n, z_{n+1}) \leq \tau \) can be contained in \( S_k^p \) until quadratic convergence sets in or the Newton
displacement \(|z_{n+1} - z_n|\) becomes small (for near-multiple roots, we cannot expect quadratic convergence except very close to the roots).

**Proposition 5.3 (Number of points in \(S_k^*\)).** For every \(k\) and every \(\ell > 0\), the set \(S_k^*\) contains at most

\[
\pi d(2^{-k+1} + \ell)^2 (2\tau + 2^{k-1}\ell)[2\tau + 4 \log 3] \ell^{-2}
\]

orbit points \(z_n\) with \(|z_n - z_{n+1}| \geq \ell\) and \(|\varphi(z_n)| \geq e^{1/2} - 1\).

**Proof.** If \(z_n \in S_k^*\) and \(|z_n - z_{n+1}| \geq \ell\), then each point in \(A_{n,k}(\ell)\) has Euclidean distance at most \(2^{-k+1} + \ell\) from some root by lemma 4.2, so \(A_{n,k}(\ell)\) is contained in a set of total area at most \(\pi d(2^{-k+1} + \ell)^2\). Each \(A_{n,k}(\ell)\) has area at least \(\ell^2 / (2\tau + 2^{k-1}\ell)\) by lemma 4.1, and by lemma 5.2 the sets \(A_{n,k}(\ell)\) and \(A_{n',k}(\ell)\) are disjoint if \(n' - n > 2\tau + 4 \log 3\) and \(|\varphi(z_n)| \geq e^{1/2} - 1\). Therefore, there can be at most

\[
\frac{\pi d(2^{-k+1} + \ell)^2 (2\tau + 2^{k-1}\ell)[2\tau + 4 \log 3]}{\ell^2}
\]

such points, for any choice of \(\ell\).

\(\square\)

**Remark.** The sets \(A_{n,k}(\ell)\) are contained in \(U\), so in the end the various orbits in the different immediate basins \(U_\alpha\) for different roots \(\alpha\) will compete for the area. The last result can thus be sharpened as follows. For a root \(\alpha\), let

\[U_{\alpha,k}(\ell) := \{z \in U_\alpha : |z - \alpha_j| < 2^{-k+1} + \ell\text{ for some root }\alpha_j\}\]

(this is the \(2^{-k} + \ell\)-neighborhood of \(S_k^*\) restricted to \(U_\alpha\)). Then the set \(S_k^*\) contains at most

\[
|U_{\alpha,k}(\ell)| (2\tau + 2^{k-1}\ell)[2\tau + 4 \log 3] \ell^{-2}
\]

points on the orbit \((z_n) \subset U_\alpha\) with \(|z_n - z_{n+1}| \geq \ell\) and \(|\varphi(z_n)| \geq e^{1/2} - 1\), and of course we have

\[
\sum_j |U_{\alpha_j,k}(\ell)| \leq \pi d(2^{-k+1} + \ell)^2.
\]

**Lemma 5.4 (Newton displacement and nearest root).** For any \(z \in \mathbb{C}\), the nearest root \(\alpha\) satisfies \(|z - \alpha| \leq d|z - N_p(z)|\).

**Proof.** This is easy and well known:

\[
z - N_p(z) = \frac{1}{p'(z)/p(z)} = \frac{1}{\sum_{\alpha_i} \frac{1}{z - \alpha_i}},
\]

hence

\[
\frac{1}{|z - N_p(z)|} \leq \sum_{\alpha_i} \frac{1}{|z - \alpha_i|} \leq \frac{1}{\inf_{\alpha_i} |z - \alpha_i|}.
\]

\(\square\)

**Corollary 5.5 (Number of points in \(S_k\)).** For any \(k\), the set \(S_k\) contains at most

\[
\pi (4d + 1)^2 (2\tau d + 1/4)[2\tau + 4 \log 3] \in O(d^3 \tau^2)
\]

points on any orbit \((z_n) \subset U\) with \(|\varphi(z_n)| > e^{1/2} - 1\).
Proof. If \( z_n \in S_k \), then \( |z_n - z_{n+1}| > 1/d^2 \) by lemma 5.4, so we use \( \ell = 1/d^2 \) in proposition 5.3 and obtain the estimate
\[
\pi d^2 4^{k+1} \left( 2^{-k+1} + \frac{2^{-k-1}}{d} \right)^2 \left( 2\tau + \frac{1}{4d} \right) [2\tau + 4 \log 3]
\]
\[
= \pi (4d + 1)^2 (2\tau d + 1/4)[2\tau + 4 \log 3]
\]
as claimed. \( \square \)

Remark. As before (see equation (5)), the different roots have to compete for the total area available, and the set \( S_k \) can contain at most
\[
|U_{\alpha,k} (1/d^2) | \cdot (2\tau + 1/4d)[2\tau + 4 \log 3] d^2 4^{k+1}
\]
points on the orbit \( (z_n) \subset U_\alpha \) with \( |\varphi(z_n)| > e^{1/2} - 1 \).

6. Uniformly separated roots

Now that we have good bounds on how many iterations any of our selected orbits can spend within each \( S_k \), we have to discuss the possible values of \( k \). Any disk \( D_k(0) \) is partitioned by \( S_k \cap D_k(0) \) for \( k \geq - \log_2 (R + 1) \), and we gave an upper bound for \( R \), hence a lower bound for \( k \), in proposition 2.7. We also need an upper bound for \( k \), that is a ‘stopping criterion’ when the orbit is sufficiently close to a root.

We need two kinds of stopping criteria: a worst-case estimate that applies especially when there are multiple or near-multiple roots, and a better estimate in case the roots are reasonably well separated from each other, so the orbit is already an approximate zero. We first investigate well-separated roots: we say that the roots are \( \delta \)-separated if they are all simple and have mutual distance at least \( \delta \). If roots are randomly distributed in \( D \), with high probability they will be \( \delta \)-separated with \( \delta = O(1/d) \) (see the remark at the end of this section). Multiple or near-multiple roots will be treated in section 7.

Lemma 6.1 (Stopping criterion). (a) If \( |z - \alpha| < |z - \alpha'|/2d \) for all roots \( \alpha' \neq \alpha \), then the Newton orbit of \( z \) converges to \( \alpha \).

(b) If \( \varepsilon > |z - \alpha'|/(4d + 3) \) for all \( \alpha' \neq \alpha \), then \( |N_n^\alpha(z) - \alpha| < \varepsilon \) for all \( n > \log_2 \varepsilon - 5 \).

Proof. (a) We may rescale coordinates by an automorphism of \( \mathbb{C} \) so that \( z = 0 \) and \( \alpha = 1 \).

By hypothesis, we have \( |z - \alpha| = 1 \) and \( |z - a_j| > 2d \) for all \( a_j \neq \alpha \). As in the proof of lemma 5.4, this implies \( |\sum_{a_j \neq \alpha} 1/(z - a_j)| < (d - 1)/2d < 1/2 \), so \( \sum_{a_j} 1/(z - a_j) \in D_{1/2}(1) \), the open disk of radius 1/2 around 1. Thus \( z - N_\alpha(z) = \left( \sum_{a_j} 1/(z - a_j) \right) \) \( \in D_{1/3}(4/3) \) (the image of a circle under \( z \mapsto 1/z \) is a circle that in this case is real symmetric, and it is easy to compute the points where it intersects the real line). Therefore \( N_\alpha(z) \in D_{2/3}(4/3) \) and \( |N_\alpha(z) - \alpha| < 1 = |z - \alpha| \), so by induction the orbit of \( z \) converges to \( \alpha \).

(b) If \( |z - \alpha| < |z - \alpha'|/(4d + 3) \), then again we choose coordinates with \( \alpha = 1 \) and \( z = 0 \), so all \( |a'| > 4d + 3 \). All \( a' \in D_1 \) have \( |a' - \alpha| < 2 |z' - \alpha'|/2d \), so by part (1) the disk \( D_1(1) \) is contained in the (rescaled) immediate basin of \( \alpha \) and we have \( d_1(z, \alpha) < d_1(0, 1) = d_1(1/2, 0) \), hence \( |\varphi(z)| < 1/2 \). The claim thus follows from lemma 5.1 (2). \( \square \)

We would like to point out that the following result does not require that all roots are \( \delta \)-separated, but only that we have some root \( \alpha \) that has distance at least \( \delta \) from all other roots
Suppose \( |\alpha' - \alpha| > \delta \) for all roots \( \alpha' \neq \alpha \). Again, the hypotheses of proposition 5.5 \( = \delta \) and will be at most
\[
\pi(4d + 1)^2(2\pi d + 1)4(2\pi d + 4)^2/\delta^2 + M + 1
\]
\[
= (64\pi d^3\tau^2(\log 2d + \log 2\delta^2) + \log 2 \log \varepsilon)(1 + o(1))
\]
\[
\leq O(d^3\tau^2(\log 2d + \log 2\delta^2) + \log 2 \log \varepsilon).
\]
\textbf{Proof.}\ If \( |z - \alpha| < \delta/(4d + 4) \), then all roots \( \alpha' \neq \alpha \) satisfy \( |z - \alpha'| > \delta - \delta/(4d + 4) \).\( = \delta/(4d + 4) \). So the orbit of \( z \) satisfies the hypothesis of lemma 6.1 (2) and will be \( \epsilon \)-close to \( \alpha \) after at most \( \log 2 \) \( \log 2 \epsilon - 5 \) iterations.

Therefore, it is sufficient to consider the number of iterations that the orbit stays in \( S_k \) with \( k \leq K \), where \( K \in \mathbb{N} \) is such that \( 2 - K \leq \delta/(4d + 4) \); we thus set \( K := \lceil \log 2((4d + 4)/\delta) \rceil \). Since the orbit is contained within \( D_{2^{\tau} - 1}(0) \) by hypothesis, we have \( k \geq -M \), so we need to consider \( k \in \{-M, -M + 1, \ldots, K\} \).

By corollary 5.5, any orbit \( (z_0) \) with \( dU(z_0, z_1) < \tau \) has at most \( \pi(4d + 1)^2(2\tau d + 4) \log 2 \leq 5 \) points within each \( S_k \), so the total number of iterations required for the orbit \( z_0 \) is at most
\[
\pi(4d + 1)^2(2\tau d + 4) \log 2 \leq 5 \times \log 2 \leq 5
\]
\[
\leq O(d^3\tau^2(\log 2d + \log 2\delta^2) + \log 2 \log \varepsilon).
\]

Note that here and elsewhere when we give asymptotic complexity results in \( O \)-notation we always have explicit constants, and these are small (so we are not hiding large constants behind this notation).

\textbf{Remark.}\ Again, the \( d \) roots have to compete for the available area within \( D \). If all roots are simple and \( \delta \)-separated, and there are \( d \) orbits (one in each immediate basin) that satisfy the hypotheses of proposition 6.2, then the combined number of iterations required to reach \( \epsilon \)-precision for all \( d \) roots is at most
\[
\pi(4d + 1)^2(2\tau d + 4) \times 2 \times \log 3 \times \log 2 \left( \frac{4d + 4}{\delta} \right) + M + 1 \times d \log 2 \times \log 2 \epsilon - 5
\]
\[
\leq O(d^3\tau^2(\log 2d + \log 2\delta^2) + M \times d \log 2 \times \log 2 \epsilon).
\]

this is almost the same bound as for a single orbit (each area element can be used for only one root), except that the estimate \( \log 2 \times \log 2 \epsilon - 5 \) (which takes care of approximate zeroes and does not involve area) applies for each root separately.

If only some of the roots are \( \delta \)-separated, then we get in a similar way a combined bound for the number of iterations to find all those roots that are \( \delta \)-separated from all other roots.

We are now in a position to prove that the explicit finite point grid \( S_d \) from Definition 3.1 contains efficient starting points for finding all roots.

\textbf{Theorem 6.3 (Efficient grid of starting points).}\ For each degree \( d \), the set \( S_d \) has the following universal property: if \( p \) is a complex polynomial of degree \( d \), normalized so that all its roots are in \( D \), and so that all roots are simple and have mutual distance at least \( \delta \), then
\(S_d\) contains \(d\) points that converge to the \(d\) roots of \(p\) and so that the combined number of iterations required to reach \(\varepsilon\)-precision is at most
\[
(256 \pi d^2 \log^2 d (\log d + | \log \delta |) + d \log \varepsilon) (1 + o(1))
\]
\[
\leq O \left( d^3 \log^2 d (\log d + | \log \delta |) + d \log | \log \varepsilon | \right).
\]
If not all roots are \(\delta\)-separated, then a subset of these \(d\) points finds all those roots that are \(\delta\)-separated from all other roots with \(\varepsilon\)-precision in the given number of iterations.

**Proof.** Let \(R\) be as in proposition 2.7 (depending only on \(d\)). Then by theorem 3.2, for each root \(\alpha_i\), there is a point \(z^{(i)} \in S_d\) in the central subchannel of the largest channel of the immediate basin, and its orbit is \(R\)-central and satisfies \(d_0(z^{(i)}), N_p(z^{(i)}) \leq \tau < 2 \log d\) (theorem 3.2).

Therefore, by proposition 6.2 and the remark thereafter, for all roots \(\alpha_i\) that are \(\delta\)-separated from all other roots the orbits of the points \(z^{(i)}\) combined need at most
\[
\pi (4d + 1)^2 \left( 4d \log d + \frac{1}{4} \right) \left[ 4 \log d + 4 \log 3 \right] \cdot \left[ \frac{\log 2 (4d + 4)}{\delta} + M + 1 \right]
\]
\[
+ d \log \varepsilon - 5
\]
iterations to reach \(\alpha_i\) within \(\varepsilon\)-precision. The claim follows. \(\Box\)

Recall from the remark after proposition 2.7 that \(R\) satisfies an explicit bound for every \(d\), and is universally bounded for all \(d\). In particular, we can use \(M = 5\) for \(d \geq 41\) and \(M = 3\) for \(d \geq 316\), and for all but very low values of \(d\) the term \(\log 2 ((4d + 4)/\delta) + M + 1\) is dominated by the term \(\log \varepsilon (4d + 4)\).

**Remark (Expected mutual distance between roots).** The results in this section are under the assumption that all (or at least some) roots were \(\delta\)-separated for some \(\delta > 0\). If \(d\) roots are placed independently and randomly into \(\mathbb{D}\) (with respect to planar Lebesgue measure), then the expected mutual distance between any two roots is easily seen to be at least \(O(1/d)\). Theorem 3.2 thus applies and yields, for \(\delta \approx 1/d\), a number of iterations of at most \(O(d^3 \log^2 d + d \log | \log \varepsilon |)\).

If not the locations of the roots are chosen randomly, but for instance the coefficients, then the roots may no longer be equidistributed with respect to area; they tend to distribute uniformly along a circle [ET], and the expected mutual distance is at least \(O(1/d^2)\). In any case, the relation between coefficients and roots is algebraic, so the expected mutual distance \(\delta\) between roots is bounded by a power law in \(d\), say \(\delta \geq 1/d^\beta\) with some \(\beta \geq 1\), but since our estimates only involve \(\log | \delta |\), this still becomes only a constant factor \(\beta\) in the number of iterations.

**Remark (Further improvements).** The greatest loss in our estimates is in the most basic of our estimates, in lemma 5.4: if \(|z - N_p(z)| < s\), then \(|z - \alpha| < ds\) for some root \(\alpha\). This bound is sharp only if all roots form a single multiple root, and then indeed the distance to the root is multiplied by \((d - 1)/\delta\) in each Newton iteration. If the roots are randomly distributed, then the bound is much better, and this leads to significant improvements. Refining our methods in this direction, the following is shown in [BAS]: If the \(d\) roots are distributed independently in \(\mathbb{D}\) and randomly with respect to Lebesgue measure of \(\mathbb{D}\), then the number of iterations for the same grid \(S_d\) as before is at most \(O(d^2 \log^2 d + d \log | \log \varepsilon |)\), with high probability. This improves our bound by a factor of \(d / \log d\) and it is optimal except for some powers of \(\log\); if we have \(d\) starting points outside of \(\mathbb{D}\) at radius \(r_0 > e\), then the simple estimate \((d - 1)/\delta \approx 1/e\) implies that each of them takes approximately \(d\) iterations to move from any radius \(r > e\) to radius \(r/e\), so the \(d\) points together need \(O(d^2)\) iterations even to get close
to \( \mathbb{D} \) (and if \( r_0 = e^\beta \) with \( \beta \in [0, 1] \), then only a constant factor \( \beta \) is gained). Our universal set of starting points requires us to place the starting points uniformly outside of \( \mathbb{D} \), and under this assumption the number of iterations is essentially best possible. This remark also applies when the roots are randomly distributed along a circle, for instance when the coefficients are independently randomly distributed.

7. Non-uniformly separated roots

If the roots are not uniformly \( \delta \)-separated, then they may be multiple, and the local rate of convergence will be linear rather than quadratic. For practical purposes, multiple roots are indistinguishable from simple roots at mutual distance smaller than the required precision \( \varepsilon \). Our previous estimate on the required number of iterations scales with \( \delta \) essentially as \( O(d^3 \log \delta) \): this is of course unbounded, but diverges slowly as \( \delta \to 0 \); for random distributions of roots (for instance with respect to Lebesgue measure of the plane) the expected value is finite and of moderate size.

However, there are of course important polynomials with multiple or near-multiple roots. Thus we will now provide a uniform bound on the required number of iterations for all polynomials in \( \mathcal{P}_d \). We will still assume that all roots are simple, but since we do not assume a lower bound on their mutual distance, the estimates hold for multiple roots as well, by continuity.

The point grid that we will use is the same as before.

The issue of ‘clusters of roots’ is relevant from many points of view: from a distance, such clusters look like multiple roots (resulting in slowing down the Newton dynamics), and only near such a cluster does the dynamics begin to see the roots separately (in fact, sufficiently far outside any disk containing all roots of a degree \( d \) polynomial, the roots look like a single root of multiplicity \( d \), which explains the linear convergence with the factor \( (d - 1)/d \) near \( \infty \)). For a systematic study of clusters of roots, as well as further references on this topic, see [GLSY] (however, in this reference the assumption is made that the number of roots within any cluster is known ahead of time).

Our estimates will be based on the following stopping criterion.

**Lemma 7.1 (Worst-case stopping criterion).** If \( z \in U_\alpha \), the immediate basin of a root \( \alpha \), then \( |z - \alpha| < f_d |z - N_p(z)| \), where

\[
 f_d = \frac{d^2(d - 1)}{2(2d - 1)} \left( \frac{2d}{d} \right) < d^2 4^{d - 1}
\]

depends only on \( d \) and satisfies \( \log f_d < 2(d - 1) + 2 \log d \).

**Proof.** This result is proved in [Sch1, lemma 5], using an iterated ‘cluster of roots’ argument. \( \square \)

The difficulty in this result is the following: if \( |z - N_p(z)| < \varepsilon /d \), then \( z \) is \( \varepsilon \)-close to some root \( \alpha' \) by lemma 5.4; but even if \( z \) is in the immediate basin of a root \( \alpha \), this does not mean that \( \alpha' = \alpha \). Stopping the iteration at \( z \) and declaring \( z \) as an approximation to a nearby root, which is necessarily \( \alpha' \), runs the danger that \( z \) was the only starting point guaranteed to find \( \alpha \), and the algorithm might miss \( \alpha \) altogether. The iterated cluster of roots argument in [Sch1] argues that either \( \alpha' = \alpha \), or some other roots \( \alpha'' \) must be reasonably close to \( z \) (with constants depending on \( d \)). Then either \( \alpha'' = \alpha \), or a further root \( \alpha''' \) must be close, etc. In order to assure that \( z \) is indeed \( \varepsilon \)-close to the ‘correct’ root \( \alpha \), this argument might have to be repeated \( d - 1 \) times, and on average a factor of 4 is lost. This explains the exponentially large factor \( f_d \) in the (unlikely) iterated worst case in lemma 7.1.
The following result is the ‘worst-case’ version of proposition 6.2.

**Proposition 7.2 (Number of iterations on orbit, worst case).** Suppose \( \alpha \) is a simple root without lower bound on the distance from other roots, and \( R > 1 \). If \( (z_n) \) is an \( R \)-central orbit in the immediate basin \( U \) with \( |z_0| \leq R \leq 2^M - 1 \) and \( d_U(z_0, z_1) < \tau \), then we have \( |z_N - \alpha| < \varepsilon \) for all \( N \) at least

\[
\pi (4d + 1)^2 (2\tau d + 1/4) (2\tau + 4 \log 3) (2(d - 1) + \log 2d - \log 2(\varepsilon) + M + 1) + \log 2 | \log 2 \varepsilon - 5 | + 1
\]

\[
= (128\pi d^2 \tau^2 + 64\pi d^3 \tau^2 | \log 2 \varepsilon |)(1 + o(1)) + \log 2 | \log 2 \varepsilon |
\]

\[
\in O (d^4 \tau^2 + d^3 \tau^2 (| \log \varepsilon |) + \log | \log \varepsilon |).
\]

**Proof.** We iterate the orbit \( (z_n) \) while \( |z_n - z_{n+1}| \geq \varepsilon / f_d \) and \( |\varphi(z_n)| \geq e^{1/2} - 1 \). If at some time \( |z_n - z_{n+1}| < \varepsilon / f_d \), then we can stop by lemma 7.1, and if \( |\varphi(z_N)| < e^{1/2} - 1 \), then lemma 5.1 (2) applies and only \( \log 2 |\log 2 \varepsilon - 5| \) further iterates are required until \( \varepsilon \)-precision is reached.

Now we estimate how many iterates are necessary until \( |z_N - z_{N+1}| < \varepsilon / f_d \); we may suppose that along the way, we always have \( |\varphi(z_n)| > e^{1/2} - 1 \).

We will use corollary 5.5 for \( k = -M, 0, 1, 2, \ldots, K \), where \( K \) is the least integer such that \( 1/d^{2K+1} \leq \varepsilon / f_d \), i.e. \( K = \left\lceil \log 2 (f_d / 2\varepsilon) \right\rceil = 2(d - 1) + \log 2d - \log 2(\varepsilon) + 1 \). The number of orbit points in \( S_k \) is at most \( \pi (4d + 1)^2 (2\tau d + 1/4) (2\tau + 4 \log 3) \). For the final value \( k = K \), the total number of points in \( S_k \) with \( |z_n - z_{n+1}| \geq \varepsilon / f_d \geq 1/d^{2K+1} \) satisfies the same bound (proposition 5.3). Moreover, we have \( D_{2\varepsilon_{n-1}}(0) \subset S_{-M} \cup S_{-M+1} \cup \ldots \cup S_{K-2} \cup \ldots \cup S_k \cup S_{K-1} \cup S_k \), so the total number of iterations with \( |z_n - z_{n+1}| > \varepsilon / f_d \) is at most \( K + M + 1 \) times the number for each \( S_k \), and by hypothesis the orbit never leaves the disk \( D_\varepsilon(0) \subset D_{2\varepsilon_{n-1}}(0) \).

**Theorem (Worst case number of iterations).** For each degree \( d \), the set \( S_d \) constructed in Definition 3.1 has the following universal property: if \( p \) is any complex polynomial, normalized so that all its roots are in \( \mathbb{D} \), and with simple or multiple roots at arbitrary mutual distances, then \( S_d \) contains \( d \) points that converge to the \( d \) roots of \( p \) and so that the combined number of iterations required to reach \( \varepsilon \)-precision is at most

\[
\pi (4d + 1)^2 (2\tau d + 1/4) (2\tau + 4 \log 3) (2(d - 1) + \log 2d - \log 2(\varepsilon) + M + 1)
\]

\[
+ d \log 2 | \log \varepsilon - 5 | + 1
\]

\[
= (512\pi d^2 \log^2 d + 256\pi d^3 \log^2 d | \log \varepsilon |)(1 + o(1))
\]

\[
\in O (d^4 \log^2 d + d^3 \log^2 d | \log \varepsilon |).
\]

**Proof.** It suffices to prove that the case that all roots are simple; the case of multiple roots follows by continuity (multiple roots will be found by multiple orbits starting at different points in \( S_d \)). The set \( S_d \) intersects the central subannulus of the channel with largest modulus of each root in at least one point, so the corresponding orbits are \( R \)-central, for the choice of \( R \) specified in proposition 2.7, and all their orbit points \( \varepsilon \) satisfy \( d(z, N_R(z)) < \tau := 2 \log d \).

By proposition 7.2, each of these points needs at most

\[
\pi (4d + 1)^2 (2\tau d + 1/4) (2\tau + 4 \log 3) (2(d - 1) + \log 2d - \log 2(\varepsilon) + M + 2)
\]

\[
+ d \log 2 | \log 2 \varepsilon - 5 | + 1
\]

iterations to be \( \varepsilon \)-close to the corresponding root. Since all roots again have to compete for the area within \( \mathbb{D} \), the total number of iterations combined to get \( \varepsilon \)-close to all \( d \) roots, for one starting point per root, satisfies the same bound, except that a factor \( d \) comes in in the part of the estimate where the roots do not compete for area, and this is the term with \( \log 2 | \log 2 \varepsilon | \) (which is subordinate to the \( d \) \( | \log \varepsilon | \) term). This proves the claim.

\[ \square \]
Remark. We believe that this result in the worst case can be improved at least by a factor of $d$: the factor $f_d$ in lemma 7.1 is exponential in $d$, and the worst case leading to this estimate seems very unrealistic. Indeed, in the recursive arguments leading to the exponential factor $f_d$, every recursive step forgets much of what had been concluded about the mutual location of the roots considered so far. It seems plausible that a more elaborate recursive argument should replace $f_d$ by a polynomial in $d$. Even though $f_d$ enters only logarithmically, $\log f_d$ still contributes a factor of $d$. If $f_d$ could be replaced by a polynomial in $d$, this would gain a factor of $d/\log d$.

In contrast, the complexity in $\epsilon$ really is $|\log \epsilon|$, rather than $\log |\log \epsilon|$, in the presence of multiple roots because Newton’s method at multiple roots converges only linearly, not quadratically.

8. Root finding in theory and practice: towards optimal bounds

Since one can argue that any theoretical result on root finding is only as useful as its success in practice, we give a brief overview on practical performance.

As mentioned in the introduction, there are numerous root finding methods that are in practical use. Typically they have a well known analysis of local convergence: that means complexity as the accuracy $\epsilon$ tends 0 for fixed degree. Just as typically, they are not supported by global theory, even when they work well in practice: it is not clear where to start, what the likelihood of success is, and how long it takes. When it comes to polynomials of very large degrees (measuring complexity in terms of degrees $d \to \infty$), we are aware of only two methods that have been implemented successfully: Newton and Ehrlich–Aberth. It is interesting to compare their performance, and this was done systematically in [SCRSSS].

One of the principal results is that Newton’s method finds those roots particularly efficiently that are on the boundary of the convex hull formed by all roots, so Newton is able to factorize polynomials particularly efficiently when most roots are on this boundary. Another obvious advantage, not in the number of iterations but in overall performance, is when the polynomials can be evaluated very fast (for instance, when they are given by iterative methods) because essentially all computing time goes into evaluation of the polynomial and its derivative.

For Ehrlich–Aberth, the slow part is not evaluation of the polynomials, but evaluation of a multipole term that relates the approximations to the roots. Therefore, this method cannot really take advantage of polynomials that are fast to evaluate, such as when given by recursion.

One set of polynomials that were evaluated both using MPSo1ve 3.0 (by Bini and co-authors) and using Newton (in [SSt, RaSS]) are known as ‘Mandelbrot center polynomials’: this is a recursively defined sequence of polynomials of degrees $2^n$, for $n \in \mathbb{N}$: we have $p_1(c) = c$ and $p_{n+1}(c) = p_n^2(c) + c$. These are very ill-conditioned, and they were selected independently as ‘demo cases’ for MPSo1ve 3.0 by the authors of this package, and in the studies of Newton’s method in practice [SSt, RaSS] (even though they are not of the form that all or most roots are on the boundary of the convex hull, and thus not of the form where Newton seems to work best; in reality it works well anyway).

Newton (in the implementation of [SSt] that is based on the present paper) and Ehrlich–Aberth (in the MPSo1ve implementation) have been used quite successfully for polynomials up to degrees one million and more. But both have recent substantial improvements that leave the realm that is supported by theory and thus, with some sort of optimism, lead to drastic acceleration that make, in certain cases, degrees exceeding one billion feasible.

On the Newton side, this is the ‘iterated refinement method’ [RaSS] that is based on the observation that the guaranteed success of Newton is based on starting points at a certain
Table 1. Finding periodic points of $q_2(z) = z^2 + 2$ and of $q_i(z) = z^2 + i$. The first two columns show period $n$ and degree $d = 2^n$. The third column shows the number of Newton iterations (times $d$) required to find all $d$ roots of $q_2^n(z) - z$; the next column show the computing time on a standard PC computer from about 2012 (single core). The final two columns show the same for the polynomials $q_i^n(z) - z$. For comparison: 4 d have 345 600 s, comparable to the computing time for the largest experiment in both series.

| Period $n$ | Degree $d = 2^n$ | $z^2 + 2$ Iterations $\times d$ | Computing time (s) | $z^2 + i$ Iterations $\times d$ | Computing time (s) |
|-----------|------------------|-------------------------------|------------------|-------------------------------|------------------|
| 12        | 4096             | 756                           | 1                | 362                           | 0                |
| 15        | 32 768           | 1220                          | 16               | 461                           | 3                |
| 18        | 262 144          | 1786                          | 201              | 561                           | 29               |
| 21        | 2097 152         | 2437                          | 2407             | 661                           | 320              |
| 24        | 16 777 216       | 3204                          | 32 401           | 762                           | 3320             |
| 27        | 134 217 728      | 4044                          | 320 567          | 865                           | 33 194           |
| 30        | 1073 741 824     | 1008                          |                   | 335 431                       |                  |

Figure 5. Complexity for finding all roots of $q^n(z) - z$, where $q(z) = z^2 + i$, for $n \leq 30$ (this polynomial has degree $2^n$). Blue: the number of iterations, scaled by $50d \log^{0.11} d$; red: the computing time in $\mu$s, scaled by $d \log^2 d$. (The deviations for small degrees are granularity errors due to the fact that the computing time was reported in whole seconds.) Note that the vertical axis does not start at 0.

distance away from the disk containing all roots. Convergence towards this disk is linear, so each of the $d$ or more orbits need $\Omega(d)$ iterations to get closer to the disk, which alone leads to at least quadratic complexity. However, since the dynamics away from this disk is quite predictable, one can start with many fewer orbits on a large disk and refine these orbits only once the dynamics of adjacent orbits differs significantly. In practice, log-linear complexity to find all $d$ roots has been achieved for certain families of polynomials of very high degrees (see table 1 and figure 5).
For Ehrlich–Aberth, we recently learned from Dario Bini, senior author of the MPSolve implementation, that they successfully implemented a fast approximation to the multipole term that usually dominates computing time. This has led to a drastic speed-up especially for polynomials that can be evaluated quickly by recursion, leading to very fast success even for very high degrees (so this is no longer Ehrlich–Aberth proper, but an approximation to this approximative method).

Since this paper is about root finding by Newton, let us conclude with a few remarks on our Newton implementation. All computations were done on a standard PC with a single active core (multiple cores would reduce the time almost linearly). Even for these extremely large degrees, no special arithmetic was required, nor did the compiler software have to be adapted. It seems that the limiting factor for the computations are still not numerical issues but RAM memory constraints (for degree $2^{27}$ the output file alone required about 9 Gigabytes of memory); quite possibly the limit can be pushed substantially further by optimizing the software.

It is a noteworthy and perhaps amusing fact that for $q^n(z) - z$ the constant coefficient has magnitude greater than $2^{2^n}$, far greater than can be stored using any standard arithmetic (for $n = 27$ this quantity has more than 40 million decimal digits!; while all roots are clustered within a disk of radius 2 and are thus very close to each other) — but this caused no problem for the computations at all. Our iterative evaluation of the polynomials does not require the coefficients; these results would be impossible for many reasons (and by many orders of magnitude) had one tried to work on the polynomials in coefficient form.

Acknowledgment

I would like to thank especially Todor Bilarev, John Hubbard, Malte Lackmann, Scott Sutherland, Robin Stoll, and Michael Thon for many helpful discussions; I am also grateful to Magnus Aspenberg, Walter Bergweiler, Edward Crane, Björn Gustafsson, Curtis McMullen, Mary Rees, Marcel Oliver, Steffen Rohde, Michael Stoll, and anonymous referees.

This research was partially supported by the EU Research and Training Network CODY, by the ESF network HCAA, by the German Research Council DFG, and by the Advanced Grant HOLOGRAM of the European Research Council.

Appendix . Geometry of hyperbolic geodesics

In this appendix, we will prove the claim that the line segments $X(t)$ as introduced in section 4 are disjoint. We will repeatedly use Ahlfors’ theorem that every Euclidean disk $D \subset U$ is convex with respect to hyperbolic geodesics in $U$ [1]. Hence for every geodesic $\gamma$ the set $\gamma \cap \overline{D}$ is connected.

Lemma A.1 (Euclidean curvature radius). Let $U \subset \mathbb{C}$ be a Riemann domain and $\gamma: \mathbb{R} \to U$ a geodesic with respect to the hyperbolic metric of $U$. For some $t_0 \in \mathbb{R}$, let $\eta := \text{dist}(\gamma(t_0), \partial U)$. Then the Euclidean curvature radius of $\gamma$ at $\gamma(t_0)$ is at least $\eta/2$.

Proof. Let $D_1$ and $D_2$ be the two open disks of radius $\eta/2$ tangent to $\gamma$ at $\gamma_0$ (on the two sides of $\gamma$); by definition of $\eta$, they are both contained in $U$ (see figure 6). If the curvature radius of $\gamma$ at $\gamma(t_0)$ is less than $\eta/2$, then $\gamma$ must enter the interior of a single one of the two disks $D_1$ and $D_2$ for $t > t_0$ and $t < t_0$ near $t_0$. This is a contradiction to Ahlfors’ theorem that disks are hyperbolically convex.
This bound might well be sharp. Björn Gustafsson [GS, corollary 8.3] observed that it is sharp for domains \( U \subset \mathbb{C} \) that are allowed to contain the point at \( \infty \), and Edward Crane observed that it is not far from being sharp for domains \( U = \mathbb{C} \setminus \mathbb{R}_0 \).

Recall that for a Riemann domain \( U \subset \mathbb{C} \) and a hyperbolic geodesic \( \gamma : \mathbb{R} \to U \) parametrized by Euclidean arc length, we defined \( \eta(t) \) as the Euclidean distance of \( \gamma(t) \) to \( \partial U \), and \( X(t) \) as the straight line segment (without endpoints) of length \( \eta(t) \) with centre at \( \gamma(t) \) that intersects \( \gamma \) perpendicularly at \( \gamma(t) \).

**Proposition A.2.** All \( X(t) \) are disjoint.

**Proof.** (0) Suppose \( X(t_0) \) and \( X(t_1) \) intersect; without loss of generality, suppose that \( \eta(t_0) \geq \eta(t_1) \) and \( t_1 > t_0 \). Let \( D_0 \) be the open disk centred at \( \gamma(t_0) \) and with radius \( \eta(t_0) \), and let \( D_1 \) and \( D_2 \) be the two open disks of radius \( \eta(t_0)/2 \) tangent to \( \gamma'(t_0) \); then both disk boundaries are tangent (from the inside) to \( \partial D_0 \), and \( X(t_0) \) is exactly the open straight line segment connecting their centres. Let \( \ell \) be the straight line that contains \( X(t_0) \). Since \( X(t_0) \subset D_1 \cup D_2 \cup \gamma(t_0) \), it is no loss of generality to assume that \( X(t_0) \) and \( X(t_1) \) intersect within \( \overline{D_1} \).

(a) First observe that the geodesic \( \gamma \) is disjoint from \( D_1 \) and \( D_2 \): this follows from Ahlfors’ theorem mentioned above: if not, there is some point \( \gamma(t_2) \in D_1 \), say. Consider a disk \( D'_1 \subset D_1 \) tangent to \( D_1 \) at \( \gamma(t_0) \) and so that \( \gamma(t_2) \in \partial D'_1 \). Then \( \gamma(t_0) \) and \( \gamma(t_2) \) are in \( \partial D'_1 \), so by Ahlfors’ theorem we must have \( \gamma(t_0, t_2) \subset \overline{D'_1} \). But the radius of \( D'_1 \) is smaller than the curvature radius of \( \gamma \) at \( t_0 \) (lemma A.1), and this is a contradiction.

(b) Let \( E \) be the convex hull of \( D_1 \cup D_2 \). We claim that \( \gamma \cap \overline{E} \) is connected. Indeed, suppose \( \gamma(t_2) \not\in \overline{E} \), but \( \gamma(t_1) \in \partial E \), with \( t_0 < t_2 < t_3 \), say. Then there is a disk \( D' \subset U \) with radius...
\[ \eta(t_0)/2 \] with centre on \( X(t_0) \), translated so that \( \gamma(t_0) \in \partial D' \). Since \( \gamma(t_0) \in D' \), it follows that \( \gamma \cap D' \) is not connected, a contradiction.

(c) Our next claim is that along \( \gamma \cap E \), no tangent vector is perpendicular to \( \gamma'(t_0) \); this means that, if we call the direction of the tangent vector \( \gamma'(t_0) \) ‘horizontal’, then \( \gamma \cap E \) is a graph over the horizontal direction. For a proof by contradiction, suppose there is a \( t > t_0 \), say, with \( \gamma(t_2) \in E \) and \( \gamma'(t_2) \) perpendicular to \( \gamma'(t_0) \). Let \( D'_1 \) and \( D'_2 \) be the two disks with centres on \( \ell \) and with radii \( \eta(t_0)/2 \) and so that \( \gamma(t_2) \in \partial D'_1 \cap \partial D'_2 \). Since \( \gamma(t_2) \in E \setminus (D_1 \cup D_2) \) by step (1), it follows that \( \gamma(t_0) \in D'_1 \cap D'_2 \) (every disk with radius \( \eta(t_0)/2 \) and centre on \( \ell \) is a ‘vertical’ translate of \( D_1 \) and \( D_2 \), and if it contains a point in \( E \setminus (D_1 \cup D_2) \) it must contain \( \gamma(t_0) \)). Therefore \( D'_i \subset E \) for \( i = 1, 2 \).

One of the two disks \( D'_1 \) and \( D'_2 \) thus has the property that \( \gamma(t_2 + \epsilon) \in D'_i \) for small \( \epsilon > 0 \), but not for small \( \epsilon < 0 \). Since \( \gamma(t_0) \in D'_i \) and \( t_0 < t_2 - \epsilon < t_2 \), this contradicts Ahlfors’ theorem once again.

(d) For the final step, recall that \( t_0 \) and \( t_1 > t_0 \) are such that \( X(t_0) \) and \( X(t_1) \) intersect, and \( \eta(t_0) \geq \eta(t_1) \). Consider the point \( \gamma(t_1) \) and let \( D \) be the unique disk of radius \( \eta(t_0)/2 \) with centre on \( \ell \) and so that \( \gamma(t_1) \in \partial D \) (this leaves two choices, and we take the disk with centre closest to the centre of \( D_1 \)). Since \( \gamma(t_1) \notin D_i \), the centre of \( D \) is strictly between the centres of \( D_1 \) and \( D_2 \), and hence \( \gamma(t_0) \in D \). By Ahlfors’ theorem again, we have \( \gamma([t_0, t_1]) \subset D \), and \( \gamma(t) \notin D \) for \( t > t_1 \).

Since \( \gamma \) must leave \( D \) in the direction of increasing \( t \), the tangent vector \( \gamma'(t_1) \) must either be parallel to the tangent vector of \( \partial D \) at \( \gamma(t_1) \), or its slope must be smaller. If they are parallel, then the distance from \( \gamma(t_1) \) to \( \ell \) along \( X(t_1) \) is exactly \( \eta(t_0)/2 \) (the radius of \( D \)), while the length of \( X(t_1) \) is \( \eta(t_1)/2 \leq \eta(t_0)/2 \) (from the centre point \( \gamma(t_1) \) in both directions), so \( X(t_1) \) cannot intersect \( X(t_0) \) (by definition, the \( X(t) \) do not contain the endpoints). If the slope of \( \gamma'(t_1) \) is smaller than that of the circle, then the distance from \( \gamma(t_1) \) to \( \ell \) in the direction of \( X(t_1) \) is even greater. Therefore, \( X(t_0) \) and \( X(t_1) \) cannot intersect.

\[ \square \]

**ORCID ID**

Dierk Schleicher ☞ https://orcid.org/0000-0002-7586-6189

**References**

[AHU] Aho A V, Hopcroft J E and Ullman J D 1974 *The Design and Analysis of Computer Algorithms* (Reading, MA: Addison Wesley)

[BAS] Bilarev T, Aspengren M and Schleicher D 2016 On the speed of convergence of Newton’s method for complex polynomials *Math. Comput.* **85** 693–705

[BR] Bini DA and Robol L 2014 Solving secular and polynomial equations: a multiprecision algorithm *J. Comput. Appl. Math.* **272** 276–92

[BBEGG] Bini D A, Boito P, Eidelman Y, Gemignani L and Gohberg I 2010 A fast implicit QR eigenvalue algorithm for companion matrices *Linear Algebra Appl.* **432** 2006–31

[BLS] Bollobás B, Lackmann M and Schleicher D 2013 A small probabilistic universal set of starting points for finding roots of complex polynomials by Newton’s method *Math. Comput.* **82** 443–57

[ET] Erdős P and Turán P 1950 On the distribution of roots of polynomials *Ann. Math.* **51** 105–19

[GLSY] Giusti M, Lecerf G, Salvy B and Yakoubsohn J-C 2005 Location and approximation of clusters of zeros of analytic functions *Found. Comput. Math.* **5** 257–311

[GS] Gustafsson B and Sebar A 2012 Critical points of Green’s function and geometric function theory *Indiana Univ. Math. J.* **61** 939–1017
[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

[J] Jørgensen V 1956 On an inequality for the hyperbolic measure and its applications in the theory of functions *Math. Scand.* **4** 113–24

[KMS] Kim M-H, Martens M and Sutherland S 2018 Geometry of polynomials and root-finding via path-lifting *Nonlinearity* **31** 414–57

[K] Kantorovic L V 1949 On Newton’s method (Russian) *Proc. Steklov Inst. Math.* **28** 104–44

[MS] Mayer S and Schleicher D 2006 Immediate and virtual basins of Newton’s method for entire functions *Ann. Inst. Fourier, Grenoble* **56** 325–36

[McN1] McNamee J M 2002 A 2002 update of the supplementary bibliography on roots of polynomials *J. Comput. Appl. Math.* **142** 433–4

[McN2] McNamee J M 2007 *Numerical Methods for Roots of Polynomials. Part I. Studies in Computational Mathematics* vol 14 (Amsterdam: Elsevier B. V.)

[MP] McNamee J M and Pan V 2013 *Numerical Methods for Roots of Polynomials. Part II. Studies in Computational Mathematics* vol 16 (Amsterdam: Elsevier B. V.)

[MR] Mihaljević-Brandt H and Rempe-Gillen L 2013 Absence of wandering domains for some real entire functions with bounded singular sets *Math. Ann.* **357** 1577–604

[MB] Moenck R T and Borodin A B 1972 Fast modular transform via division *Proc. 13th Annual Symp. on Switching and Automata Theory* (Washington, DC: IEEE Computer Society Press) pp 90–96

[P1] Pan V 2000 Approximating complex polynomial zeros: modified Weyl’s quadtree construction and improved Newton’s iteration *J. Complexity* **16** 213–64

[P2] Pan V 2002 Univariate polynomials: nearly optimal algorithms for numerical factorization and root-finding *J. Symb. Comput.* **33** 701–33

[P3] Pan V 2021 New progress in polynomial root-finding Manuscript v19 (arXiv:1805.12042)

[Pr] Przytycki F 1989 Remarks on the simple connectedness of basins of sinks for iterations of rational maps *Dynamical Systems and Ergodic Theory* ed K Krzyzewski (Warszawa: Polish Scientific Publishers) pp 229–35

[RaSS] Randig M, Schleicher D and Stoll R 2017 Newton’s method in practice II: the iterated refinement Newton method and near-optimal complexity for finding all roots of some polynomials of very large degrees (arXiv:1703.05847) submitted

[Re] Reinke B 2022 Diverging orbits for the Ehrlich–Aberth and the Weierstrass root finders *Proc. Am. Math. Soc.* **150** 1287–300

[ReSS] Reinke B, Schleicher D and Stoll M 2023 The Weierstrass-Durand-Kerner root finder is not generally convergent *Math. Comput.* submitted (arXiv:2004.04777)

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

[Rü] Rückert J 2008 Rational and transcendental Newton maps *Holomorphic Dynamics and Renormalization. A Volume in Honour of John Milnor’s 75th Birthday* vol 53, ed M Lyubich and M Yampolsky (Toronto: Fields Institute Communications) pp 197–212

[RS] Rückert J and Schleicher D 2007 On Newton’s method for entire functions *J. London Math. Soc.* **75** 659–76

[Sch1] Schleicher D 2002 On the number of iterations of Newton’s method for complex polynomials *Ergod. Theor. Dyn. Syst.* **22** 935–45

[Sch2] 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 Honour of John Milnor’s 75th Birthday* vol 53, ed M Lyubich and M Yampolsky (Toronto: Fields Institute Communications) pp 213–24

[SSi] Schleicher D and Stoll R 2017 Newton’s method in practice: finding all roots of polynomials of degree one million efficiently *J. Theor. Comput. Sci.* **681** 146–66

[SCRSSS] Shemyakov S, Chernov R, Rumiantsev D, Schleicher D, Schmitt S and Shemyakov A 2020 Finding polynomial roots by dynamical systems – a case study *Discrete Continuous Dyn. Syst.* **40** 6045–65
[Sh] Shishikura M 2009 The connectivity of the Julia set and fixed points *Complex Dynamics: Families and Friends*, ed D Schleicher (Wellesley, MA: AK Peters) pp 257–76

[ShSm] 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

[Sm] Smale S 1986 Newton’s method estimates from data at one point *The Merging of Disciplines: new Directions in Pure, Applied and Computational Mathematics* (Laramie, Wyo., 1985) (New York: Springer) pp 185–96