Invent. math. 146, 1–33 (2001)
Digital Object Identifier (DOI) 10.1007/s002220100149

Inventiones mathematicae

How to find all roots of complex polynomials by Newton's method

John Hubbard[1,2], **Dierk Schleicher**[3], **Scott Sutherland**[4]

[1] Department of Mathematics, Malott Hall, Cornell University, Ithaca, NY 14853-4201, USA (e-mail: jhh8@cornell.edu)
[2] Centre de Mathématiques et Informatique, Université de Provence, 39 rue F. Joliot-Curie, F-13453 Marseille cedex 13, France
[3] Mathematisches Institut, Ludwig-Maximilians-Universität, Theresienstraße 39, D-80333 München, Germany (e-mail: dierk@rz.mathematik.uni-muenchen.de)
[4] Institute for Mathematical Sciences, State University of New York, Stony Brook, NY 11794-3660, USA (e-mail: scott@math.sunysb.edu)

Oblatum 24-II-2000 & 14-II-2001
Published online: 20 July 2001 – © Springer-Verlag 2001

Abstract. We investigate Newton's method to find roots of polynomials of fixed degree d, appropriately normalized: we construct a finite set of points such that, for every root of every such polynomial, at least one of these points will converge to this root under Newton's map. The cardinality of such a set can be as small as $1.11\,d\log^2 d$; if all the roots of the polynomial are real, it can be $1.30\,d$.

Contents

1 Introduction

This paper concerns the global dynamics of Newton's method for finding roots of a polynomial $p(z)$ in one variable. We show how to find all roots of such a polynomial without recourse to deflation (which has serious

computational difficulties for polynomials of high degree, and which is not available in some problems of interest).

Finding roots of polynomials is a venerable problem of mathematics, and even the dynamics of Newton's method as applied to polynomials has a long history. Our approach gives a picture of the global geometry of the basins of the roots in terms of accesses to infinity; understanding the sizes of these accesses is the key to the proof.

We were not just motivated by the intrinsic interest of the subject. We developed the techniques described in this paper in part because we actually needed to find all the roots of polynomials of rather high degree (a few hundred or a few thousand), in order to compute an approximation to the invariant measure for Hénon mappings. We describe this particular problem in more detail at the end of this introduction; similar problems often appear in holomorphic dynamics.

The dynamics of Newton's method always presents difficult problems, even as applied to polynomials in one variable. For instance, already for cubic polynomials there may be open sets of initial points which do not lead to any root but instead to an attracting cycle of period greater than one, and the boundaries of the basins will usually be complicated fractals whose topology is poorly understood. An example is shown in Fig. 1. C. McMullen [McM1,McM2] has shown that there are no generally convergent purely iterative algorithms for solving polynomials of degrees 4 or greater. Thus, any algorithm analogous to Newton's method must have polynomials with a positive measure set of initial points that do not lead to roots.

Our main result is the following. Let $\mathcal{P}_d$ be the space of polynomials of degree d, normalized so that all their roots are in the open unit disk $\mathbb{D}$. For such a polynomial p, we will be interested in its Newton map $N_p : \mathbb{P}^1 \to \mathbb{P}^1$ defined by $N_p(z) = z - p(z)/p'(z)$. If the sequence

$$z_0,\ z_1 = N_p(z_0),\ z_2 = N_p(z_1),\ \ldots$$

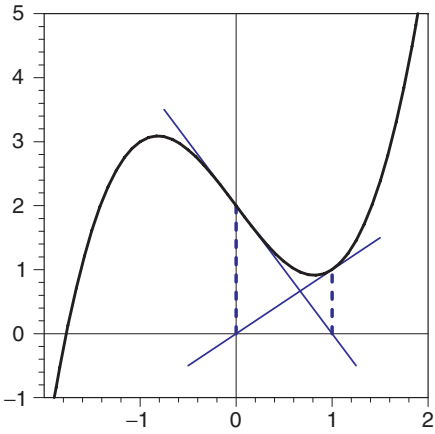
converges to a root ξ of p, we will say that z_0 is in the basin of ξ.

Theorem 1 (Main Theorem)
For every $d \geq 2$, there is a set $\mathcal{S}_d$ consisting of at most $1.11\, d \log^2 d$ points in $\mathbb{C}$ with the property that for every polynomial $p \in \mathcal{P}_d$ and each of its roots, there is a point $s \in \mathcal{S}_d$ in the basin of the chosen root. For polynomials all of whose roots are real, there is an analogous set $\mathcal{S}$ with at most $1.3\, d$ points.

The theorem is constructive: in Sect. 9, we explicitly construct such a set $\mathcal{S}_d$ in the general case (see Fig. 2): it will consist of approximately $0.2663 \log d$ circles, each containing $4.1627\, d \log d$ points at equal distances. If you are more interested in a set of starting points for practical use than in the theory, you may skip ahead directly to Sect. 9.

The restriction to polynomials with all the roots in the unit disk is not severe. For any polynomial, it is easy to estimate the maximal absolute value

of any of its roots in terms of the coefficients: for $|z|$ too large, the highest order term will dominate the rest of the polynomial and there can be no root.



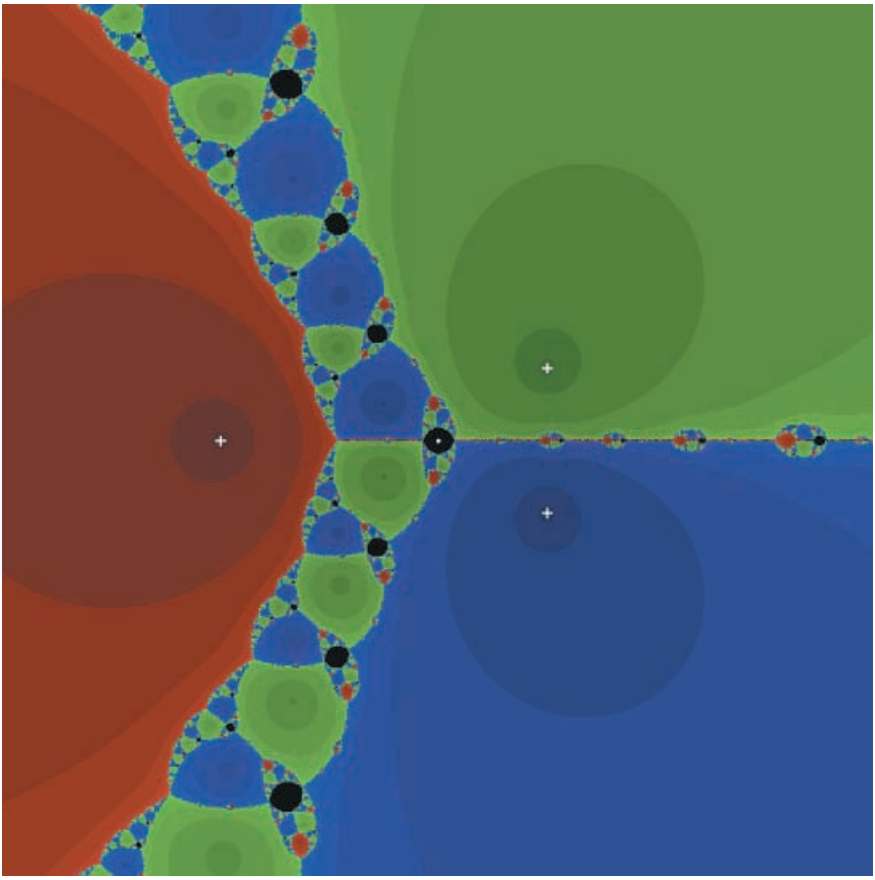


Fig. 1 The Newton map for the polynomial $p : z \mapsto z^3 - 2z + 2$ has a superattracting cycle of period 2. *Left:* the graph of p over the interval $[-2, 2]$, with the superattracting 2-cycle $0 \mapsto 1 \mapsto 0$ of the Newton map indicated. *Bottom:* the same Newton map over the complex numbers. Colors indicate to which of the three roots a given starting point converges; black indicates starting points which converge to no root, but to the superattracting 2-cycle instead

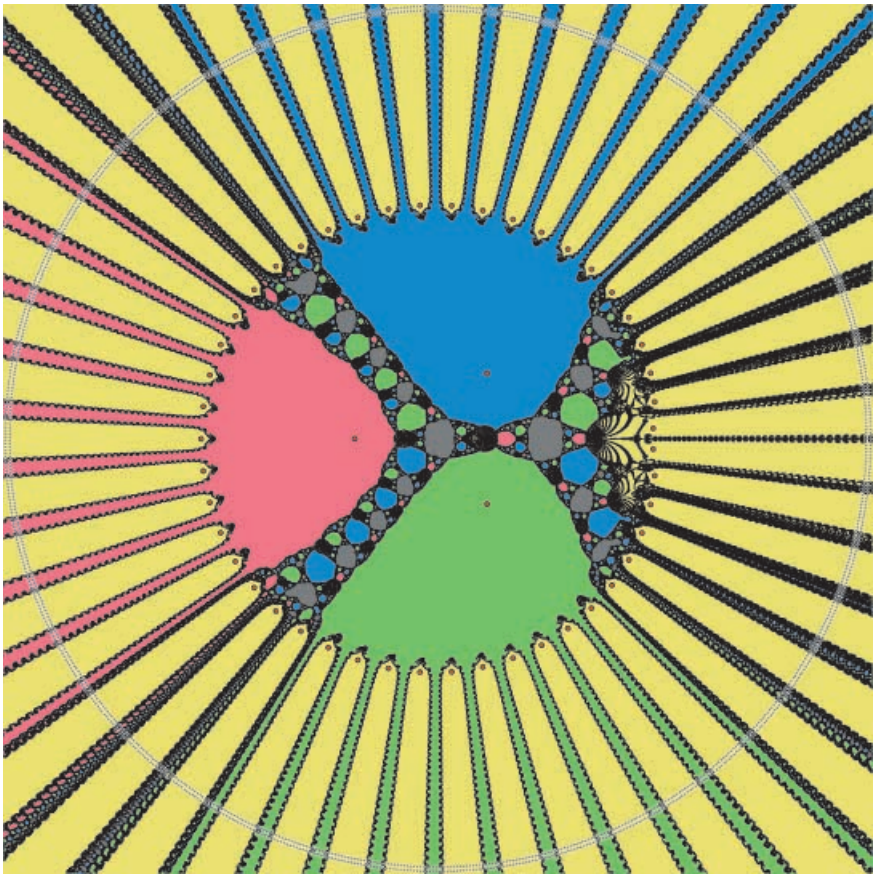
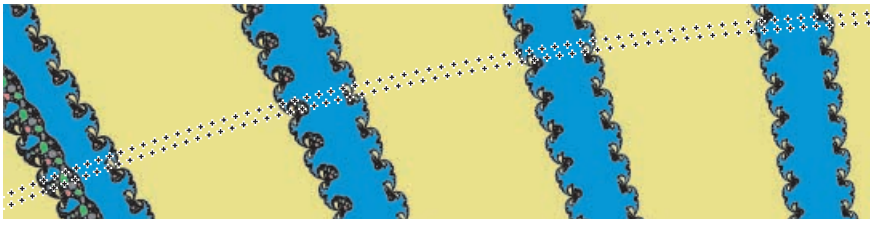


Fig. 2 A set of starting points as specified by our theorem for degree 50, indicated by small crosses distributed on two large circles. Also shown is the Julia set for the Newton map of the degree 50 polynomial $z^{50} + 8z^5 - \frac{80}{3}z^4 + 20z^3 - 2z + 1$ (*black*). There are 47 roots near the unit circle, and 3 roots well inside, all marked by *red disks*. As in Fig. 1, there is an attracting periodic orbit (*basin in grey*). A close-up is shown below



An appropriate affine coordinate change will bring all the roots into the unit disk. Such coordinate changes do not alter the dynamics of Newton's method: the Newton map for $p(az + b)$ is conjugate to that for $p(z)$. Moreover, there is a precise classical criterion based on continued fractions to determine whether or not all the roots of a given polynomial are in the unit

disk; it was brought to our attention by A. Leutbecher and can be found in [JT, Chapter 7.4] or [MS, Sects. 2.3 and 2.4].

We believe that the cardinality of $\mathcal{S}_d$ is not too far from optimal: we anticipate that a universal set $\mathcal{S}_d$ as specified in the theorem must have more than $O(d \log d)$ elements, i.e., $|\mathcal{S}_d|/d \log d \to \infty$. When all roots are real, the number of our points is only a constant times the number of roots.

Concerning the number of iterations, it is not difficult to give a compactness argument which shows that there is a number $N(d, \varepsilon)$ depending only on the degree d and the desired accuracy ε such that for every root there is at least one of the chosen starting points in the basin of this root and whose distance to the root after $N(d, \varepsilon)$ iterations is at most ε. An explicit bound is in [Sch]; it cannot be found from the compactness argument above. It is exponential in d and far from optimal.

The constructions in this paper also apply to the relaxed Newton method, i.e. to $z \mapsto z - h\,p(z)/p'(z)$ with a real constant $h \in (0, 1)$.

Of course, while Newton's method is one of the simplest and most widely-used root-finding methods, there are many others appearing in the literature, with their own advantages and disadvantages. In addition to the methods covered in numerical analysis texts (for example, [He, Chap. 6]), we refer the interested reader to Pan's survey article [Pa].

This paper builds on results and ideas of many other people: F. Przytycki [Pr] has investigated the immediate basins of attraction of roots under Newton's map; A. Manning [M] has shown how to find at least a single root of a polynomial, and he has given an explicit bound how many iterations it takes to find it with prescribed precision. Since we will need some of their results together with their constructions, we include them here.

The organization of the paper and of the proof is as follows: in Sect. 2, we give a number of introductory lemmas which help to set up the algorithm. In Sect. 3, we turn to the immediate basins of the roots and show that they have "channels" which extend all the way out to infinity; these results go back to Przytycki and Manning. We investigate the geometry of the channels in Sect. 4: these channels are not too narrow; following a suggestion of Douady, we measure their widths in terms of the modulus of the quotient of the annulus identified by the dynamics. Using a variant of the residue theorem, we give a lower bound on the width of at least one channel for every root.

Since all the roots and thus all the interesting dynamics of the Newton map are within the unit disk, we have good control sufficiently far away from this disk. In Sect. 5, we show how to find starting points within these channels. This is accomplished using extremal length arguments and a sequence of conformal mappings involving the elliptic modular function; the detailed calculations are collected in Sect. 6. The case when all the roots of the polynomial are real allows a number of improvements which will be discussed in Sect. 7. Placing all the starting points onto a single circle is also possible, but the number of necessary points is substantially larger (and the calculations get more involved): see Sect. 8. Finally, in Sect. 9 we discuss

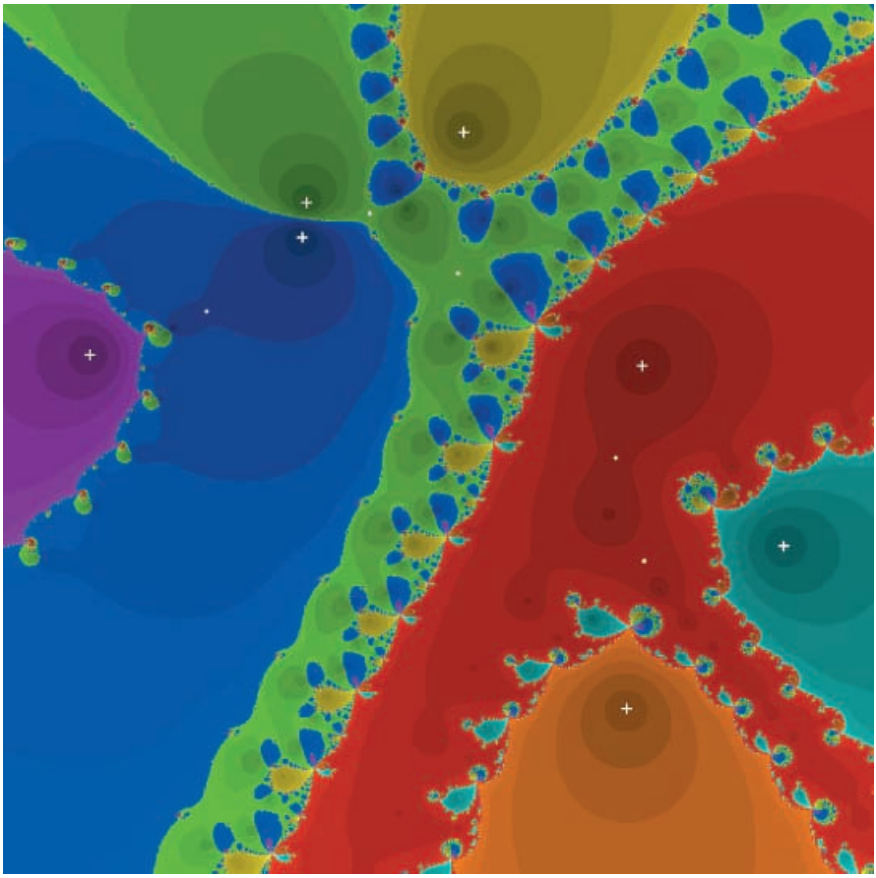
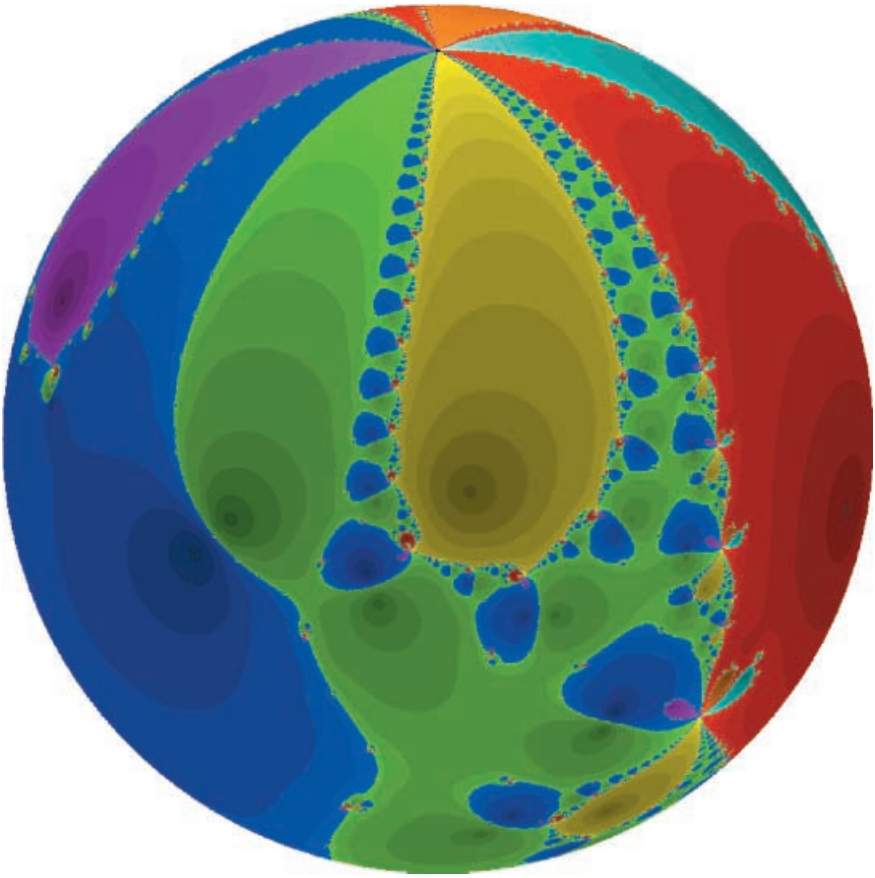


Fig. 3 *Left:* The Newton map for a polynomial of degree 7, showing that every root can be connected to ∞ within its basin of attraction. Indicated are the seven roots (which are all critical points of N_p), as well as the five further critical points of N_p. For every root, the number of channels to ∞ equals the number of critical points in its immediate basin (see Proposition 6). The shading of the colors indicates the speed of convergence to the roots. *Right:* The same picture projected onto the Riemann sphere. The repelling fixed point ∞ is clearly visible near the top of the picture, together with the periodic channels leading to ∞

Newton's method from a practical point of view and describe explicitly a set of starting points for given degree.

A prototypical problem. Below we discuss one of the problems that motivated this paper; see [HP]. A Hénon mapping $H : \mathbb{C}^2 \to \mathbb{C}^2$ is given by the formula $(x, y) \mapsto (x^2 + c - ay, x)$. This map has a unique invariant measure. An approximation to this measure is provided by a normalized sum of δ-functions at the intersections $H^n(\Delta) \cap H^{-n}(\Delta)$ of the nth forward image and the nth backwards image of the diagonal $\Delta \subset \mathbb{C}^2$ of equation $x = y$.



It is fairly easy to see that computing these intersections leads to solving a polynomial equation of degree $d = 2^{2n}$; the higher n is, the better the approximation. We worked with $n = 4 \ldots 7$, i.e., polynomials of degree $d = 256 \ldots 16384$. The roots of these polynomials can be shown to lie in a disk of radius

$$R = \frac{1}{2}\left(|a| + 1 + \sqrt{(|a|+1)^2 + 4|c|}\right)$$

and are usually packed along some sort of fractal curve, with typical spacings something like 10^{-8}; but of course these roots collide and bifurcate as a and c vary.

We know the polynomial we need to solve as an iteration. This is mainly a blessing, as evaluating the polynomial and its derivative requires of the order of $\log d$ operations rather than d. But it means that we don't know the coefficients of the polynomial (and in fact, can't compute them, as they overflow the real numbers of the computer); in particular, deflating the

polynomial is not an option; the temptation to deflate is best resisted in any case, because of numerical instability.

Acknowledgements. Over the years in which we worked on this project, we have benefitted from suggestions by many people, among them Walter Bergweiler, Xavier Buff, Adrien Douady, Myong-Hi Kim, Hartje Kriete, Armin Leutbecher, Anthony Manning, Feliks Przytycki, Steffen Rohde, and Mitsuhiro Shishikura; many of of their contributions are acknowledged throughout the paper. We are grateful to all of them, as well as to two referees for valuable suggestions. Moreover, we would like to thank the Institute for Mathematical Sciences and its director John Milnor for its hospitality, support, and inspiring environment. It has helped bring us together again.

Scott Sutherland is especially grateful to his thesis advisor Paul Blanchard for all his help and advice; this thesis [Su] was a germ of this work.

2 The global geometry

A critical point of a holomorphic map is a point where the derivative vanishes. Critical points of $N_p = \mathrm{id} - p/p'$ are solutions of $N_p' = pp''/(p')^2 = 0$, i.e., zeros and inflection points of p. If a root of p is simple, then it is a critical point of N_p, i.e., a superattracting fixed point. A root of multiplicity $k \geq 2$ is attracting with real multiplier $(k-1)/k$.

In this section, we will assume for convenience that our polynomials are monic. Multiplying a polynomial by a non-zero complex constant has no effect on Newton's method or on our results. We also assume that the degree is at least 2.

In order to study the geometry of the immediate basins, we will need three simple lemmas. The first is known as Lucas' theorem; see the first two chapters in Marden [Ma] and in particular Theorem 6.1.

Lemma 2 (Critical points of polynomial)
The convex hull of the roots of any polynomial contains all its critical points, as well as the zeros of all higher derivatives of the polynomial. In particular, it contains the critical points of the Newton map.

Proof. By induction, it suffices to discuss the relation between the position of the critical points and the roots. Consider any half-plane containing all of the roots; by applying an affine change of coordinates, we may assume the roots all have negative real part. If we write our polynomial as $\prod_i (z - \xi_i)$, the derivative can be written as $\prod_i (z - \xi_i) \cdot \sum_i (1/(z - \xi_i))$. If z is any point with $\mathrm{Re}(z) \geq 0$, the real parts of $z - \xi_i$ and thus of $1/(z - \xi_i)$ are also positive. Consequently, the derivative cannot vanish at z. □

In the subsequent two lemmas, and throughout the rest of the paper, we will always consider a polynomial $p \in \mathcal{P}_d$ and its Newton map N_p.

Lemma 3 (Asymptotic geometry of the Newton map)
For $|z| \geq 1$, we have $\left|N_p(z) - \frac{d-1}{d} z\right| < \frac{1}{d}$ and $|N_p(z)| < |z|$.

Proof. Factoring the polynomial p again as $p(z) = \prod_i (z - \xi_i)$, we can write

$$N_p(z) = z - \frac{p}{p'}(z) = z - \frac{1}{\sum_i \frac{1}{z-\xi_i}} \ .$$

All the $z - \xi_i$ are contained within the open disk D of radius 1 around z. Since $0 \notin D$; the domain $D' := \{w : w^{-1} \in D\}$ is another disk in $\mathbb{C}$ and contains all the $(z - \xi_i)^{-1}$ (if $|z| = 1$, then D' is a half plane). Summing and taking the inverse, we see that $1/\sum_i \frac{1}{z-\xi_i}$ is contained in the open disk of radius $1/d$ around z/d. This proves the first inequality. The second one follows: $|N_p(z)| < (d-1)|z|/d + 1/d \le |z|$. □

Lemma 4 (Domain of linearization near infinity)
Let $U := \{z \in \mathbb{P}^1 : |z| > 1\}$. Then there is a domain $V \subset U$ and a conformal isomorphism $g\colon U \to V$ with $N_p \circ g = \mathrm{id}$ on U. Moreover, there is a univalent map $\varphi : U \to \mathbb{P}^1$ with

$$\varphi \circ N_p = \frac{d-1}{d}\,\varphi$$

on V, and V contains every $z \in \mathbb{P}^1$ with $|z| > (d+1)/(d-1)$.

Proof. Let g be the branch of N_p^{-1} fixing ∞; it is defined in a neighborhood of ∞ and satisfies $|g(z)| > |z|$ there. Analytic continuation along curves in U is possible: we can never have $|g(z)| \le |z|$ by Lemma 3, so by Lemma 2 we never encounter a critical value of N_p along the analytic continuation. Since U is simply connected, this defines g uniquely on U with holomorphic inverse N_p, and $g\colon U \to g(U) =: V$ is a conformal isomorphism. If $|z| > (d+1)/(d-1)$, then $|N_p(z)| > 1$; it follows that $z \in V$.

The map $g\colon U \to V \subset U$ is a contraction of the Poincaré metric in U and fixes ∞, so every point in U converges to ∞ under iteration of g. Near infinity, g has the asymptotic form $z \mapsto (d/(d-1))z$, so the map

$$\varphi : U \to \mathbb{P}^1;\ \varphi(z) := \lim_{n\to\infty} \left(\frac{d-1}{d}\right)^n g^{\circ n}(z)$$

is univalent on U and satisfies the relation $\varphi(g(z)) = (d/(d-1))\varphi(z)$ (see for instance [Mi, Sect. 6]). Therefore, we have $((d-1)/d)\varphi = \varphi \circ N_p$ on V. □

3 The immediate basins

The *basin* of a root ξ of a polynomial p is the set of starting points z which eventually converge to this root. Since every root is a superattracting or at least attracting fixed point of N_p, the basin includes a neighborhood of the

root and is in particular non-empty. The *immediate basin* U_ξ of the root ξ is the connected component of the basin containing the root.

The following proposition is due to Przytycki [Pr] and, in more general form, to Shishikura [Sh] (who proves that the Julia set of a rational map is connected if there is only one repelling fixed point). We give the proof as explained to us by Sebastian Mayer.

Proposition 5 (Immediate basins are simply connected)
Let ξ be an attracting fixed point of a rational map $f\colon \mathbb{P}^1 \to \mathbb{P}^1$ and let U be the immediate basin. If ∂U contains no more than one fixed point of f, then U is simply connected.

In particular, the immediate basin of every root ξ of the Newton map N_p is simply connected and thus conformally isomorphic to the open unit disk (unless all roots of p are identical).

Proof. Let $W_0 \subset U$ be a simply connected neighborhood of ξ, such that ∂W_0 is a simple closed curve, no critical orbit meets ∂W_0, and W_0 is relatively compact in $f^{-1}(W_0)$.

Define $W_k := U \cap f^{-k}(W_0)$ and let V_k be the component of W_k containing ξ. The boundary components of W_k and V_k are simple closed curves. Note that the union $V = \cup V_k$ is an open subset of U whose boundary is contained in the boundary of U, so $V = U$. In particular, if all the V_k are simply connected, then U is also simply connected.

If U is not simply connected, there exists a first k such that V_k is not simply connected; let $A_1, \ldots, A_m$ be the connected components of $\mathbb{P}^1 \setminus V_k$. We will show that each contains a fixed point of f; call such a component A.

Then $V_k \setminus \overline{V}_{k-1}$ is path-connected (we are removing a closed disc from a connected open set), and we can choose an arc γ_0 in $\overline{V_k} \setminus V_{k-1}$ connecting a point $z_1 \in \partial A$ to $z_0 = f(z_1) \in \partial V_{k-1}$, avoiding the orbits of the critical points in U. Let γ_1 be the component of $f^{-1}(\gamma_0)$ starting at z_1; suppose it ends at z_2, and continue this way: let γ_{k+1} be the component of $f^{-1}(\gamma_k)$ starting at z_{k+1}, and let z_{k+2} be the point where it ends. Let γ be the simple arc formed by the union of the γ_i; it lies entirely in A. The sequence $z_1, z_2, \ldots$ forms a sequence in A; we claim it converges to a fixed point of f.

There exists n such that W_n contains all the critical values of f in U. Then

$$f : U \setminus \overline{W}_{n+1} \to U \setminus \overline{W}_n$$

is a covering map, hence a local isometry in the Poincaré metrics of the two spaces; in particular, it is strictly expanding if we use the Poincaré metric of $U \setminus \overline{W}_n$ in both the domain and the range. All the γ_i are in $U \setminus \overline{W}_n$ when $i + k \geq n$, and in the Poincaré metric of $U \setminus \overline{W}_n$ the γ_i are shorter and shorter. Since they accumulate to ∂U, this means that their spherical length tends to 0. It easily follows that the accumulation set of γ is a connected subset of ∂U which is pointwise fixed, hence a single fixed point in A.

If ∂U contains only one fixed point, then $\mathbb{P}^1 \setminus V_k$ must be connected for every k, so all V_k and hence U are simply connected. Since the only fixed

point of the Newton map N_p which is not a root is ∞, all immediate basins of roots are simply connected. □

Let m_ξ be the number of critical points of N_p in U_ξ, counted with multiplicity. For each root ξ we must have $m_\xi \geq 1$ since ξ is an attracting fixed point (of course, when the root is simple, it is itself a critical point of N_p). Moreover, $N_p \colon U_\xi \to U_\xi$ is proper and has a degree d_ξ. By the Riemann-Hurwitz formula and Proposition 5, these numbers are related by the formula $d_\xi = 1 + m_\xi$.

The basic observation which allows us to find roots is that every immediate basin has the point ∞ on its boundary, and there are simple arcs in the immediate basin of each root joining the root to infinity. A homotopy class of such curves is called an *access* to infinity.

Proposition 6 (Accesses to infinity in immediate basins)
Each immediate basin U_ξ has exactly m_ξ distinct accesses to infinity.

This proposition is another instance of a very general phenomenon in complex dynamics: *critical points determine the dynamics*.

Proof. To lighten notation, we omit the index ξ and write $U = U_\xi$ and $m = m_\xi$. Let $\mathbb{D}$ denote the unit disk and set $\varphi : \mathbb{D} \to U$ to be a conformal isomorphism, uniquely normalized by the two conditions $\varphi(0) = \xi$ and $\varphi'(0) > 0$. The map $f := \varphi^{-1} \circ N_p \circ \varphi$ is then a proper holomorphic $f : \mathbb{D} \to \mathbb{D}$, also of degree $m + 1$. This map extends by reflection to a holomorphic self-map of $\mathbb{P}^1$ with degree $m + 1$, i.e., a rational map which we still denote by f. Then f has exactly $m + 2$ fixed points on $\mathbb{P}^1$, counting multiplicities. Among them, there are the attracting fixed points 0 and ∞ (super-attracting if ξ is a simple root of p) which attract all of $\mathbb{D}$ and $\mathbb{P}^1 \setminus \overline{\mathbb{D}}$ respectively, and m additional fixed points $\zeta_1, \ldots, \zeta_m$ which must necessarily be on $\mathbb{S}^1$.

Since $\mathbb{D}$ and $\mathbb{P}^1 \setminus \overline{\mathbb{D}}$ are completely invariant, it follows that f cannot have critical points on $\mathbb{S}^1$, and is a covering map $\mathbb{S}^1 \to \mathbb{S}^1$ of degree $m + 1$. Moreover, all $f'(\zeta_i)$ are positive and real. The ζ_i are repelling fixed points: otherwise, they would have to be either attracting or parabolic, and would attract points in $\mathbb{D}$. In particular, the m fixed points on $\mathbb{S}^1$ are distinct.

If we assume that the conformal isomorphism $\varphi : \mathbb{D} \to U$ extends continuously to the boundary, then the m fixed points of f on $\partial\mathbb{D}$ will map to m fixed points of N_p on ∂U. A fixed point of $N_p = \mathrm{id} - p/p'$ in $\mathbb{C}$ must be either a root of p, which cannot be on the boundary of U, or the only other fixed point of N_p, namely ∞, so the domain U will extend out to ∞ in m different directions (see Fig. 4).

Even if the boundary of U is not locally connected, so that φ does not extend continuously to the boundary, the argument still goes through: *radial limits*, or equivalently, non-tangential limits, exist at all the ζ_i. Let $\gamma \subset \mathbb{D}$ be a non-tangential arc leading to ζ_i with $f(\gamma) \supset \gamma$, for example a segment of straight line in the linearizing coordinate. A standard argument then says

that $f(\gamma)$ leads to a fixed point of N_p in ∂U: the accumulation set of $f(\gamma)$ in ∂U is connected and pointwise fixed [Mi, Sect. 18]. But the point ∞ is the only fixed point which is not a root of p. This provides us with m accesses of U to infinity. We must show that they are distinct, and that they are the only ones.

All curves converging non-tangentially to the same boundary fixed point of $\mathbb{D}$ are obviously homotopic among non-tangential curves within $\mathbb{D}$, so that every boundary fixed point of $\mathbb{D}$ defines a unique access in U to ∞. Different fixed points of f on $\mathbb{S}^1$ lead to non-homotopic curves in U and thus to different accesses. Indeed, let $l_i, l_j \subset \overline{\mathbb{D}}$ be the radii leading to $\zeta_i \neq \zeta_j$ respectively, parametrized by the radius. If $\varphi(l_i)$ and $\varphi(l_j)$ are homotopic in U by a homotopy fixing $\varphi(l_i(1)) = \varphi(l_j(1)) = \infty$, then one of the components bounded by the simple closed curve $\varphi(l_i) \cup \varphi(l_j)$ must be contained in U. Call this component V; then $\varphi^{-1}(V)$ must be one of the sectors bounded by l_i and l_j; call it S. Both V and S are Jordan domains, so φ extends as a homeomorphism to their closures (by Caratheodory's theorem); but there is nowhere for $\mathbb{S}^1 \cap \overline{S}$ to map to.

Conversely, we show that every access in U to ∞ comes from a fixed point of f on $\mathbb{S}^1$; our argument imitates that of [Mi, Lemma 18.3].

First, suppose $\gamma \subset \mathbb{D}$ is a simple arc such that $\varphi(\gamma)$ defines an access to ∞. Then γ leads to a single point on $\mathbb{S}^1$; otherwise it would accumulate on some connected interval I of angles. By the Riesz Theorem, the set of $\vartheta \in I$ for which $\lim_{r\to 1} \varphi(re^{i\vartheta})$ exists has full measure; since the radius at angle ϑ must intersect γ on a subset which accumulates on $e^{i\vartheta}$, we see that the limit above is always ∞. But the radial limits can take on a given value only on a set of measure 0, again by the Riesz Theorem.

Let $\gamma\colon [0, 1] \to U \cup \{\infty\}$ be a curve representing an access with associated angle $\vartheta \in \mathbb{S}^1$; then for every $k \geq 1$, $N_p^{\circ k}(\gamma)$ represents an access and thus an angle ϑ_k. Since every fixed point of f on $\mathbb{S}^1$ gives rise to an access and N_p is a local homeomorphism near ∞, all ϑ_k must be contained in the same connected component of $\mathbb{S}^1$ with the fixed points removed; this component is an interval, say I, on which (ϑ_k) must be a monotone sequence converging under f to a fixed point ϑ of f in $\overline{I}$, i.e., to one of the endpoints. But these are repelling, so no such sequence can exist unless it is constant. □

A *channel of a root* ξ will be an unbounded connected component W of $U_\xi \setminus \overline{\mathbb{D}}$ with the additional property that there is a $w \in W$ which can be connected to $N_p(w)$ by a curve in W. It follows from Lemma 4 that every access to ∞ of U_ξ corresponds to a unique channel of ξ. It is not hard to show that the extra condition on w is unnecessary: every unbounded component of $U_\xi \setminus \overline{\mathbb{D}}$ is a channel.

We observe that every circle outside $\mathbb{D}$ centered at the origin intersects every channel and hence every immediate basin. This will allow us to find starting points for all the roots.

4 Geometry of the channels

In this section, we will investigate the geometry of the channels more closely. Since outside the closed unit disk the Newton map is approximately linear, the shape of the channels will repeat periodically near infinity. We will measure the "width" of any such channel in terms of the conformal modulus of the annulus formed by the channel modulo the dynamics.

The immediate basins are simply connected. Restricting any channel to the exterior of the unit disk and taking the quotient by the dynamics, we obtain a Riemann surface homeomorphic to an annulus, hence conformally equivalent to a standard cylinder of some height h and circumference c. This is a rectangle of sides c and h with the latter pair of sides identified. Associated to such a cylinder is its modulus h/c, which is a conformal invariant. It is not hard to see that the modulus must be a finite positive number in our case, as is shown below. We will call this number the *modulus of the channel*. It will serve to measure the width of the channel. We will now provide a lower bound on these moduli.

Proposition 7 (Widths of the channels)
If the number of critical points of the Newton map within some immediate basin is m, counting multiplicities, then this basin has a channel to infinity with modulus at least $\pi/\log(m+1)$. *In particular, every basin has a channel with modulus at least* $\pi/\log d$, *where d is the degree of the polynomial.*

Proof. We will use the construction and the notation of the proof of Proposition 6. We have an immediate basin U and a conformal isomorphism $\varphi : \mathbb{D} \to U$ sending the origin to the root ξ. Choose a channel of U corresponding to a fixed point $\zeta \in \mathbb{S}^1$ of the map $f = \varphi^{-1} \circ N_p \circ \varphi$ extended to $\mathbb{P}^1$. The point ζ is repelling with positive real multiplier $\lambda > 1$. (Although the point ζ corresponds to the fixed point ∞ of N_p, the multipliers are unrelated, since the conjugation does not exist in a neighborhood of the fixed points; in particular, all the fixed points of f on $\mathbb{S}^1$ will in general have different multipliers.)

Linearizing the repelling fixed point ζ, we obtain a linear map $w \mapsto \lambda w$, and $\mathbb{S}^1$ intersected with the domain of linearization turns into a straight line, which we may take to be the real axis (see Fig. 4). The quotient of the channel corresponding to ζ will then be conformally equivalent to the upper half plane divided by multiplication by λ. This is well known to be an annulus of modulus $\pi/\log\lambda$.

We will now use the holomorphic fixed point formula (see for example Milnor [Mi, Sect. 10]): for any rational map f of degree d with fixed points $\zeta_1, \zeta_2, \ldots, \zeta_{d+1}$ with multipliers $\lambda(\zeta_i) \neq 1$ (which guarantees that the points are distinct), we have

$$\sum_{i=1}^{d+1} \frac{1}{\lambda(\zeta_i) - 1} = -1 \ .$$

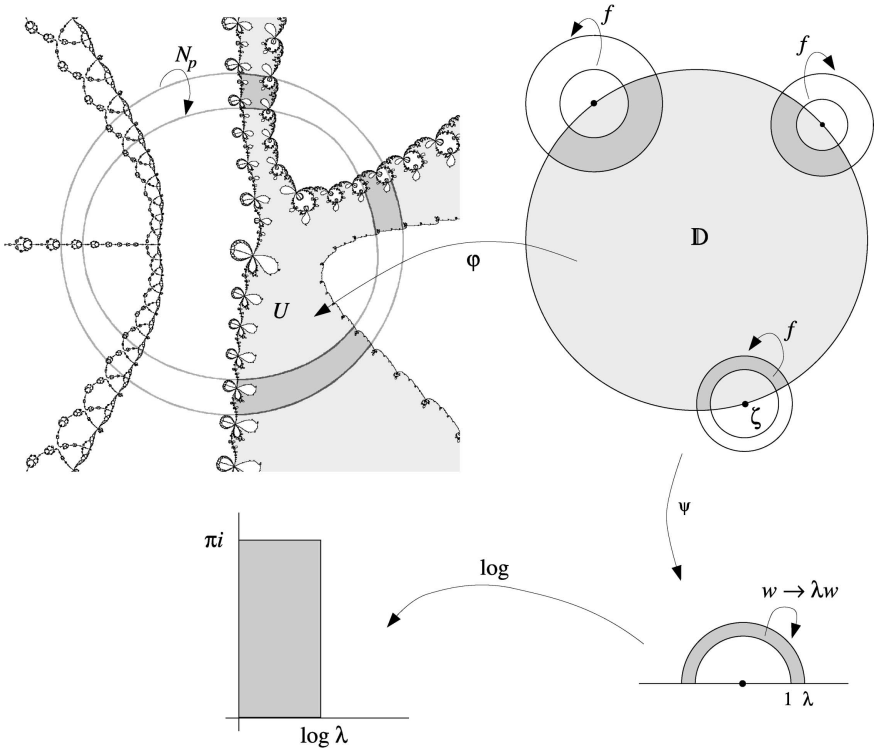


Fig. 4 Estimating the width of the channels, defined as the moduli of the quotient annuli by the dynamics. The figure highlights an immediate basin U of some root, with fundamental domains of its channels shaded; the disk $\mathbb{D}$ as a conformal model of the immediate basin, with one boundary fixed point for every channel of U; a half neighborhood of such a fixed point in local linearizing coordinates ψ; and finally a rectangle representing a fundamental domain of the corresponding channel

This formula is a corollary to the residue theorem for the map $1/(f(z) - z)$ (best evaluated in coordinates for which ∞ is not a fixed point). For the map f we are considering, we have (super)attracting fixed points at 0 and, by symmetry, at ∞. If k is the multiplicity of the root ξ under consideration, then $\lambda(0) = f'(0) = N'_p(\xi) = (k-1)/k$ with $k \geq 1$ and $\lambda(\infty) = \lambda(0)$ by symmetry, so the points 0 and ∞ each contribute $-k$ to the sum. The only further fixed points of f are the m distinct repelling fixed points on $\mathbb{S}^1$. Summing over those, we find

$$\sum_{\zeta_i \in \mathbb{S}^1} \frac{1}{f'(\zeta_i) - 1} = 2k - 1 \geq +1 \ . \tag{1}$$

Since all the denominators are real and positive, there is at least one ζ_i with $1/(f'(\zeta_i) - 1) \geq 1/m$ and hence $f'(\zeta_i) \leq m + 1$, so the modulus of the corresponding channel is at least $\pi/\log(m+1)$.

Recall that $m+1$ is the degree of the restriction of N_p to the immediate basin U. Of course $m+1$ cannot exceed the degree of N_p on $\mathbb{P}^1$, which is at most d (equal to d exactly if all roots are simple). Hence $m \leq d-1$ and the modulus of one channel is at least $\pi/\log(m+1) \geq \pi/\log d$. □

The estimate above is sharp. Consider the polynomial $p(z) = z(z^{d-1}-1)$. There are ordinary critical points of N_p at the roots on the unit circle; all the other critical points are at the origin. It follows that the map $f_0 = \varphi_0^{\circ -1} \circ N_p \circ \varphi_0$ is precisely the map $f_0(\zeta) = \zeta^d$. The channels of U_0 then correspond to the dth roots of unity, and the quotient of each channel by the dynamics is an annulus of modulus $\pi/\log d$.

Remark on the size of the immediate basins. The calculation in Equation (1) allows us to estimate the area of all the immediate basins: all fixed points $\zeta \in \mathbb{S}^1$ are repelling with $f'(\zeta) > 1$. For a simple root ($k = 1$), it follows from (1) that all fixed points $\zeta \in \mathbb{S}^1$ even have $f'(\zeta) \geq 2$, so the moduli of the channels add up to at least

$$\sum_{\zeta \in \mathbb{S}^1} \frac{\pi}{\log f'(\zeta)} \geq \sum_{\zeta \in \mathbb{S}^1} \frac{\pi}{(\log 2)(f'(\zeta)-1)} = \frac{\pi}{\log 2} .$$

For a root with multiplicity $k \geq 2$, we need not have $f'(\zeta) \geq 2$ for fixed points on $\mathbb{S}^1$, but

$$\sum_{\zeta \in \mathbb{S}^1} \frac{\pi}{\log f'(\zeta)} \geq \sum_{\zeta \in \mathbb{S}^1} \frac{\pi}{f'(\zeta)-1} = (2k-1)\pi \geq \frac{k\pi}{\log 2} .$$

The sum of the moduli of all the channels for this multiple root is at least as large as estimated for k simple roots (one could arrive at the same conclusion by perturbing the multiple root into k separate roots and using semicontinuity of the attracting basins). Thus, adding up the moduli of all channels of all roots always yields at least $\pi d/\log 2$.

In linearizing coordinates near ∞, the fundamental domain of the Newton map is bounded by two circles such that the quotient of their radii is $d/(d-1)$, and there is room for modulus $2\pi/\log(d/(d-1)) = 2\pi d + O(1)$. By a standard length-area inequality, the fraction of the area of all the channels within any linearized fundamental domain is at least the fraction of the moduli they use up. In the limit as $d \to \infty$ and for fundamental domains far outside the roots, this fraction is $1/2\log(2) \approx 0.721$. Therefore the portion of area taken up by all the immediate basins within any centered disk of radius R is at least 0.72 for R and d large. A quantitative version of this result has been given in [Su] (with a fraction of 0.09) and in [Kr] (with 0.5).

5 Hitting the channels

At least in principle, it is now clear how to proceed in order to find the roots: outside of a compact neighborhood of the origin, the map N_p is

approximately linear, and we know that every immediate basin must have a channel with a certain minimal width. If we distribute sufficiently many points within some fundamental domain, we can ensure that a channel of this width cannot possibly avoid all these points. One possibility is to put the points onto a single circle centered at the origin. This was first done in [Su]. We will show in Sect. 8 that $(\pi^2/4)d^{3/2}$ points suffice for this purpose.

However, it is geometrically conceivable that a channel with large modulus becomes relatively thin at just one place within every fundamental domain (compare the sketch in Fig. 5). In the worst case, this will happen precisely on the circle containing the starting points. We will thus place the points onto several circles within a single fundamental domain. This has two advantages: the number of points to be placed onto each of these circles may be reduced to such an extent that the total number of necessary points can be as small as $1.11\, d(\log d)^2$. Moreover, the geometry of the problem rescales and becomes independent of the degree. This allows us to derive the result (up to a constant which is independent of the degree) without explicit calculations.

Before giving a precise proof of the Main Theorem, we will outline the argument using various simplifications. The first of them is the assumption that the Newton map outside of $\mathbb{D}$ is exactly the linear map $z \mapsto z(d-1)/d$, so that a fundamental domain for the dynamics is any annulus between two radii $R > 1$ and $Rd/(d-1)$. We are looking for a channel which traverses such a fundamental domain and which has modulus at least $\pi/\log d$. We cut this fundamental annulus into some number s of conformally equivalent circular sub-annuli, each of them bounded by a pair of circles such that the ratio of outer and inner radius is $(d/(d-1))^{1/s}$. Making the further assumption that the inner and outer boundaries of every sub-annulus meet the channel in a connected arc, then the intersection of the channel with a sub-annulus becomes a conformal quadrilateral in such a way that its

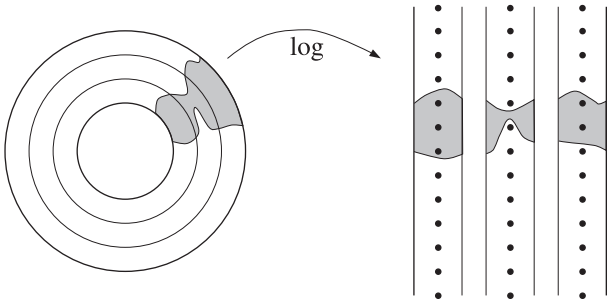


Fig. 5 *Left:* A fundamental domain of the Newton dynamics under the simplifying assumption that it is bounded by two circles, subdivided into $s = 3$ concentric sub-annuli. Shaded is the intersection of the fundamental domain with a channel of some root. *Right:* The same situation in logarithmic coordinates; the three sub-annuli turn into vertical strips (moved apart to show them separately). Each strip contains a vertical sequence of points

boundary parts on the circles form two opposite sides of the quadrilateral. Each quadrilateral has an associated modulus which is a conformal invariant. By the Grötzsch inequality, the inverses of the moduli of the rectangles add up to at most the inverse of the modulus of the channel. Therefore, at least one sub-annulus intersects the channel in a quadrilateral with modulus at least $s\pi/\log d$. If we let $s = \alpha \log d$, the modulus of the quadrilateral can be written as $\alpha\pi$, where α is independent of d. We will distribute some fixed number of points onto a single circle within each of the sub-annuli such that the angles between adjacent points on the various circles differ by the same constant.

We now take logarithms; see Fig. 5. The sub-annuli become infinite vertical strips of width $\log(d/(d-1))/s > 1/ds$ (we take all the branches of the log, so that the exponential maps from the strips to the sub-annuli are universal covers). Stretching by a factor $s/\log(d/(d-1)) < ds$, the strips get width 1; the channel becomes a series of s quadrilaterals connecting the two boundaries of each of the strips, and one of these quadrilaterals has modulus at least $\alpha\pi$ (strictly speaking, there are such quadrilaterals for every branch of the logarithm). We will now use the following universal result from conformal geometry:

Lemma 8 (Universal geometry of points in strip)
Let S be the vertical strip $\{z \in \mathbb{C} : -1/2 < \mathrm{Re}(z) < 1/2\}$ and let Q be a quadrilateral in S connecting the vertical boundaries of S. For every $\alpha > 0$, there is a number $\tau > 0$ with the following property: if the modulus of Q is $\alpha\pi$, then at least one of the points $i\tau\mathbb{Z}$ is contained in Q.

We will justify this in Lemma 12. If τ is small enough, one of the points $i\tau\mathbb{Z}$ hits the logarithm of the channel. Transporting the points $i\tau\mathbb{Z}$ back into the sub-annuli we started with, we obtain a collection of points within every sub-annulus such that the angles of adjacent points differ by more than τ/ds, and every root has a channel which is hit by at least one point on at least one circle. The number of points on such a circle is then less than $2\pi ds/\tau$, and since there are s sub-annuli with one circle each, the total number of points used is less than $2\pi ds^2/\tau = (2\pi\alpha^2/\tau)d(\log d)^2$. The choice of $\alpha > 0$ is arbitrary, and τ depends only on α but not on d. These numbers can be optimized so as to minimize the total number of points.

This gives a proof of the Main Theorem, up to a universal constant $\tau(\alpha)$ yet to be specified, and up to the various simplifying assumptions made in the argument: the Newton map is not exactly linear, and the intersections of the channel with boundaries of the sub-annuli need not be as simple as assumed. We will now provide rigorous arguments for the claim; precise estimates will be given in Sect. 6.

First we recall the definition of the modulus of an annulus via extremal length (see e.g. Ahlfors [A1, Sect. I.D]): if $U \subset \mathbb{C}$ is an annulus, then

$$\frac{1}{\mathrm{mod}\,(U)} = \sup_{\rho} \inf_{\gamma} \frac{\ell^2(\gamma)}{\|\rho^2\|_U} ,$$

where the supremum is taken over all measurable functions $\rho\colon U \to \mathbb{R}_0^+$ with strictly positive norm $\|\rho^2\|_U := \int_U \rho^2(x, y)\, dx\, dy$. The infimum is taken over all rectifiable curves $\gamma \subset U$ which wind once around U (i.e., which generate the fundamental group of U). The length of such a curve is defined with respect to ρ as $\ell(\gamma) := \int_\gamma \rho(z)|dz|$ whenever this integral exists; if it does not exist, we set $\ell(\gamma) := +\infty$. It follows easily that the modulus of an annulus is a conformal invariant: if $\varphi\colon U \to U'$ is a conformal isomorphism, then an optimal function ρ for U turns into an optimal ρ' for U' by the relation $\rho'(\varphi(z)) = \rho(z)/|\varphi'(z)|$.

Closely related to moduli of annuli are moduli of quadrilaterals: roughly speaking, a quadrilateral is a bounded subset of $\mathbb{C}$ with two distinguished connected subsets of the boundary so that the remaining boundary also consists of two connected subsets (for details, see [A1]). A quadrilateral can be mapped biholomorphically to a rectangle so that the distinguished boundaries map to a pair of opposite sides. Admissible curves in a quadrilateral are curves connecting the two distinguished boundaries, and the definition of the conformal modulus is exactly as above. We again obtain a conformal invariant, and the modulus of a rectangle with sides h and c is h/c if the pair of sides of length h is distinguished.

In our case, the restriction of a channel to a fundamental domain is a quadrilateral which becomes an annulus if the two circular boundaries are identified by the dynamics. The modulus of the annulus is at most that of the quadrilateral: requiring that a curve close up imposes an extra condition, which can only increase its length. This inequality will be used below: any allowable function ρ in a quadrilateral gives a lower bound for $\ell(\gamma)$ and thus an upper bound for the modulus, and the modulus of the quotient annulus can not be larger.

Using extremal length, we can rephrase Lemma 8 as follows:

Lemma 9 (Universal geometry of points in strip II)
Let $S := \{z \in \mathbb{C} : -1/2 < \mathrm{Re}(z) < 1/2\}$ be a vertical strip and fix a number $\tau > 0$. Then there is a number $M(\tau) > 0$ and a continuous function $\rho_S : S \to \mathbb{R}_0^+ \cup \infty$ with $\|\rho_S^2\|_S = 1$ such that every curve in S connecting the two boundaries through the interval $(0, i\tau)$ has squared length $\ell^2(\gamma) \geq 1/M(\tau)$. The function ρ_S can be chosen so that $\rho_S \to 0$ as $|\mathrm{Im}(z)| \to \infty$.

As τ tends to zero, $1/M(\tau)$ tends to ∞, and $M(\tau)$ can be chosen to be monotone.

We will provide a proof in Lemma 12, together with an estimate for $M(\tau)$.

For the construction of the point grid, we need to have some space within a fundamental domain.

Lemma 10 (Fundamental domain)
For every radius $R > (d+1)/(d-1)$, there is a number $\kappa \in (0,1)$ such that the round annulus

$$V := \left\{ z \in \mathbb{C} : \quad R\left(\frac{d-1}{d}\right)^{\kappa} < |z| < R \right\}$$

is contained within a single fundamental domain of the dynamics. The allowed values of κ tend to 1 as $R \to \infty$.

The quantity κ measures the size of a round annulus that fits into a fundamental domain, relative to the size expected by the linearized map $z \mapsto (d-1)z/d$. The larger κ, the better will be our estimates.

Proof. By Lemma 4, the image of the circle at radius R is a simple closed curve around the unit disk. The region between the circle and its image curve is an annulus, which is a fundamental domain for the dynamics. By Lemma 3, every point on the image curve has distance to the origin of at most $(d-1)R/d + 1/d = R - R/d + 1/d$. If $\kappa \in (0,1)$ is such that

$$R\left(\frac{d-1}{d}\right)^{\kappa} \geq R - \frac{R}{d} + \frac{1}{d} \qquad \text{i.e.} \qquad \kappa \leq \frac{\left|\log\left(1 - \frac{1}{d} + \frac{1}{Rd}\right)\right|}{\left|\log\left(1 - \frac{1}{d}\right)\right|} ,$$

then the round annulus V is indeed contained within a single fundamental domain. This inequality is obviously satisfied for κ sufficiently close to 0. As R gets large, the admissible values of κ get arbitrarily close to 1. □

Remark. The choice $\kappa = 1/2$ is allowed for $R > 1 + \sqrt{2}$ independently of d.

The construction of the point grid. We start with a round annulus V with outer radius R and inner radius $R((d-1)/d)^{\kappa}$ as specified in Lemma 10. The grid of starting points will be determined by two numbers α and β which will be determined later: the number of circles onto which we place our points is $s = \alpha \log d$, and the number of points per circle is $\beta d \log d$ (the quantity β is related to the constant τ in the heuristic argument above).

For $\nu = 1, 2, \ldots, s$, consider the disjoint sub-annuli of V

$$V_\nu := \left\{ z \in \mathbb{C} : \quad R\left(\frac{d-1}{d}\right)^{\nu\kappa/s} < |z| < R\left(\frac{d-1}{d}\right)^{(\nu-1)\kappa/s} \right\} ;$$

their closures cover the annulus V. The core curve of V_ν is the circle C_ν centered at the origin with radius

$$r_\nu := R\left(\frac{d-1}{d}\right)^{(\nu-1/2)\kappa/s} .$$

All the sub-annuli are conformally equivalent and can be mapped onto each other by multiplication with appropriate powers of $((d-1)/d)^{\kappa/s}$.

We place $\beta d \log d$ points onto each of the s circles C_ν, distributed evenly along each circle. This fixes the points on individual circles up to rotation of the entire circles, and for our estimates it is inessential how the circles are rotated with respect to each other. For example, we may align the grid of circles so that there is one point at the intersection of every circle with the positive real axis. The total number of points used is $\alpha\beta d(\log d)^2$.

We can now conclude that a channel must be narrow if it is to avoid the given grid of points.

Proposition 11 (Channel avoiding grid)
Let $\mathcal{S}$ be any grid of points depending on constants α and β and constructed as above. Then the modulus of any channel which avoids $\mathcal{S}$ is bounded above by $\mu(\alpha, \beta)/\log d$, where $\mu(\alpha, \beta)$ depends only on α and β but not on the degree. As β tends to ∞ with α fixed, the quantity $\mu(\alpha, \beta)$ tends to zero. One possible choice of $\mu(\alpha, \beta)$ is $M(2\pi\alpha/\kappa\beta)/\alpha$, where $M(\cdot)$ is as in Lemma 9.

Proof. It will be convenient to use the number $q := \dfrac{\kappa}{s} \log\left(\dfrac{d}{d-1}\right) > \dfrac{\kappa}{sd}$. For $\nu = 1, 2, \ldots, s$, the maps $f_\nu \colon S \to V_\nu$ with

$$f_\nu(z) := r_\nu \exp(qz)$$

are universal covers from the vertical strip S to the sub-annuli V_ν, and they wrap the imaginary axis over the circles C_ν. The number of points on each circle is $\beta d \log d$, so the angles of adjacent points differ by $2\pi/(\beta d \log d)$. Let $\tau := 2\pi/(q\beta d \log d)$; then points on the imaginary axis mapping to adjacent starting points on any C_ν have vertical spacing τ.

We will construct a measurable function $\rho \colon V \to \mathbb{R}^+$ such that any curve γ running through the channel from the root to ∞ satisfies a lower bound on its length within V. Since all these curves are homotopic within the channel, each of the circles C_ν has a pair of adjacent starting points such that every curve γ must pass the arc segment of C_ν between these points.

By Lemma 9, there is a continuous function $\rho_S \colon S \to \mathbb{R}^+ \cup \infty$ with $\|\rho_S^2\|_S = 1$ such that any curve in S connecting the two boundaries has length at least $1/\sqrt{M(\tau)}$. Transport this ρ_S forward by f_ν to a continuous function $\rho \colon V_\nu \to \mathbb{R}^+$ via

$$\rho(w) := \sup\left\{\rho_S(z)/|f_\nu'(z)| \colon z \in S \text{ and } f_\nu(z) = w\right\} .$$

The map $f_\nu \colon S \to V_\nu$ is a universal covering such that $f(z_1) = f(z_2)$ implies $f'(z_1) = f'(z_2)$. Since $\rho_S(z)$ decays to 0 as $\mathrm{Im}(z)$ gets large, only finitely many points z compete for the evaluation of the supremum in $\rho(w)$, and ρ is continuous.

For every domain $S' \subset S$ on which f_ν is injective, the norm is preserved in the sense that $\|\rho_S^2\|_{S'} = \|\rho^2\|_{f_\nu(S')}$. Then the norm of ρ^2 on V_ν equals the norm of ρ_S^2 on a subset of S, so it is at most 1. Any curve in V_ν connecting

the inner to the outer boundary and passing the arc between the two selected points has length at least $1/\sqrt{M(\tau)}$ as before.

It is not true that any curve within the channel connecting the root to ∞ will connect the inner boundary of V_ν to its outer boundary, passing through the selected arc: it might first traverse all of V_ν and then come back in through the outer boundary, say, traverse the selected arc and disappear again to the outer boundary (compare Fig. 6). However, any curve γ in the channel must connect the selected arc on C_ν twice to the boundary of V_ν within V_ν, and the two corresponding subcurves together will have length at least $1/\sqrt{M(\tau)}$.

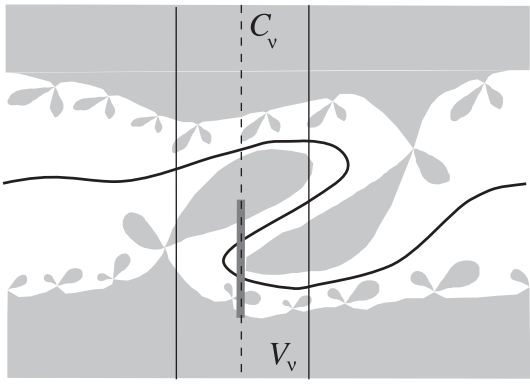


Fig. 6 Sketch in logarithmic coordinates of a fundamental domain of a channel of U (*white*) and one of the sub-annuli V_ν, together with its core circle C_ν and the distinguished arc. Within this channel is a curve which connects the root to ∞, but the connected component in V_ν which crosses the distinguished arc has both ends on the same component of ∂V_ν

The function ρ is now defined on the union of the sets V_ν. We set it equal to 0 elsewhere on $\mathbb{C}$. Then ρ is measurable with total norm at most s, and every curve γ in the channel has length at least $s/\sqrt{M(\tau)}$. Therefore, the modulus of the channel is at most $M(\tau)/s = M(\tau)/\alpha \log d$.

Recall that

$$\tau = \frac{2\pi}{q\beta d \log d} < \frac{2\pi s d}{\kappa\beta d \log d} = \frac{2\pi\alpha}{\kappa\beta} \tag{2}$$

independently of d. If β tends to ∞ with α fixed, then τ tends to 0. By Lemma 9, $M(\tau)$ also tends to 0. By monotonicity of $M(\cdot)$, we have $M(2\pi\alpha/\kappa\beta) > M(\tau)$, and the claim holds for $\mu(\alpha, \beta) = M(2\pi\alpha/\kappa\beta)/\alpha$. □

Remark. It is geometrically clear that increasing α (the number of circles) for fixed β can only decrease $\mu(\alpha, \beta)$. This fact can be extracted from our calculations only by a more precise estimate on $\alpha(\tau)$, but we will not need it. The limit of $\mu(\alpha, \beta)$ for $\alpha \to \infty$ with fixed β is $2\pi/\beta$ (for large d), so it

does not even tend to 0: the limiting point grid will be densely packed onto $\beta d \log d$ radial lines.

We can now prove the theorem up to constants.

Proof of the theorem. By Proposition 7, every root has at least one channel with modulus at least $\pi/\log d$. Choose a number $\alpha > 0$. For every $\beta > 0$, Proposition 11 specifies a number $\mu(\alpha, \beta)$ such that the point grid consisting of $\alpha \log d$ circles having $\beta d \log d$ points each hits every channel with modulus at least $\mu(\alpha, \beta)/\log d$. The number β may be chosen to ensure that $\mu(\alpha, \beta) < \pi$; this choice depends only on α. Consequently, the grid of points will hit at least one channel of every root. The total number of points used in such a grid is $\alpha\beta d \log^2 d$. □

It remains to calculate the number β in terms of α and to minimize their product. We will do this in the next section.

6 Estimating the constants

In order to provide the constants promised in the theorem, we will need the elliptic integral

$$z \mapsto \Psi(z) = \int_1^z \frac{dz'}{\sqrt{z'(z'-1)(P-z')}}\ , \tag{3}$$

depending on a real constant $P > 1$. This integral is defined on the entire complex plane except if z is real and either greater than P or less than 1; where defined, it should be evaluated along a straight line from 1 to z. Of special importance will be the two numbers

$$A(P) = \int_P^\infty \frac{dz'}{\sqrt{z'(z'-1)(z'-P)}} = \int_0^1 \frac{dz'}{\sqrt{z'(1-z')(P-z')}} \tag{4}$$

and

$$B(P) = \int_1^P \frac{dz'}{\sqrt{z'(z'-1)(P-z')}} = \int_{-\infty}^0 \frac{dz'}{\sqrt{(-z')(1-z')(P-z')}}\ . \tag{5}$$

The map Ψ provides a conformal isomorphism of the upper half plane onto the rectangle with vertices 0, $B(P)$, $B(P)+iA(P)$, $iA(P)$; this isomorphism extends as a homeomorphism to the boundaries so that the four points 1, P, ∞, 0 on the extended real line map to the vertices in this order (compare for example Nevanlinna and Paatero [NP]). The reason is that the integrand on the real line is either real or imaginary, so that the image of the real line surrounds a rectangle exactly once. Similarly, the negative half plane is mapped to the reflected rectangle.

The following lemma will be the motor for all the estimates. It is a quantitative version of Lemma 9 and proves Lemma 8 as well.

Lemma 12 (Geometry of points in strip)
Let $S := \{z \in \mathbb{C} : -1/2 < \mathrm{Re}(z) < 1/2\}$ be a vertical strip, fix a number $\tau > 0$ and set $P := e^{2\pi\tau}$. Then there is a continuous function $\rho_S \colon S \to \mathbb{R}_0^+ \cup \infty$ with the following properties:

- *every curve in S connecting the two boundaries through the interval $(0, i\tau)$ has length $\ell(\gamma) \geq \sqrt{2A(P)/B(P)}$ with respect to ρ_S;*
- *for every $r > 0$, there is a $y > 0$ such that $\rho_S(z) > r$ only if $|\mathrm{Im}(z)| < y$;*
- *$\|\rho_S^2\|_S = 1$;*
- *$\rho_S(z) = \infty$ only for $z \in \{0, i\tau\}$.*

It follows that every quadrilateral in S which connects the two vertical boundaries of S avoiding the points $i\tau\mathbb{Z}$ has modulus at most $B(P)/2A(P)$.

The function $\tau \mapsto 2A(P)/B(P)$ is strictly monotonically decreasing and tends to 0 as τ tends to 0.

Proof. Let S^- (resp. S^+) be the parts of S with negative (resp. positive) real parts. The map $E(z) := \exp(-2\pi i z)$ is a conformal isomorphism from S^- to $\mathbb{H}^+$ and from S^+ to $\mathbb{H}^-$, and it sends the interval $I := (0, i\tau)$ onto $(1, P)$ for $P = e^{2\pi\tau}$, while it sends both boundaries of S onto the negative real line.

The map Ψ sends $\mathbb{H}^+$ biholomorphically onto the rectangle

$$R^+ := \left\{z \in \mathbb{C} : 0 < \mathrm{Re}(z) < B(P), \quad 0 < \mathrm{Im}(z) < A(P)\right\}$$

with area $A(P)B(P)$, and it sends $\mathbb{H}^-$ onto the analogous rectangle R^- with imaginary parts between 0 and $-A(P)$.

On the rectangle $R := \overline{R^+ \cup R^-}$, we will use the constant function $\rho_0(z) := 1/\sqrt{2A(P)B(P)}$. It is normalized so that $\|\rho_0^2\|_R = 1$. For any curve within R connecting the upper and lower boundaries (i.e., imaginary parts $A(P)$ and $-A(P)$), the length with respect to ρ_0 is at least $2A(P)/\sqrt{2A(P)B(P)} = \sqrt{2A(P)/B(P)}$. (By Ahlfors [A1, Chapter 1], this choice for ρ_0 is optimal).

Now we transport ρ_0 to a function $\rho_S \colon S^+ \to \mathbb{R}_0^+$ using the conformal isomorphism $\Psi \circ E \colon S^+ \to R^-$ by setting

$$\rho_S(z) := \rho_0\left((\Psi \circ E)(z)\right) \cdot \left|(\Psi \circ E)'(z)\right| .$$

For $\mathrm{Im}(z) < -y$, we have $|E(z)| < e^{-2\pi y}$, $\Psi'(E(z)) = O\left(|E(z)|^{-1/2}\right)$ and $|E'(z)| < 2\pi e^{-2\pi y}$; it follows that $\rho_S(z) = O(e^{-\pi y})$; by symmetry, an analogous estimate holds for $\mathrm{Im}(z) > y$.

The map ρ_S extends continuously to a map from $\overline{S}^+$ to $\mathbb{R}_0^+ \cup \infty$ with value ∞ only at 0 and $i\tau$. Similarly, we get a continuous map from $\overline{S}^-$ to $\mathbb{R}_0^+ \cup \infty$; by symmetry, both maps coincide on $\overline{S}^+ \cap \overline{S}^-$ and extend to a continuous map $\rho_S \colon S \to \mathbb{R}_0^+ \cup \infty$.

We have $\|\rho_S^2\|_S = \|\rho_0^2\|_R = 1$, and any curve connecting the two boundaries of S through the interval $(0, i\tau)$ has length with respect to ρ_S of at least $\sqrt{2A(P)/B(P)}$: under E, the image of the curve connects the negative real

axis to itself through the interval $(1, P)$, running around $[0, 1]$. Under Ψ, we obtain a curve traversing the pair of rectangles from top to bottom, indeed with length at least $\sqrt{2A(P)/B(P)}$ with respect to ρ_0. The statement about moduli of quadrilaterals follows.

As τ decreases, P decreases as well; there is less room for the curves γ, so their lengths can only increase. This shows monotonicity of the function $\tau \mapsto 2A(P)/B(P)$ (see also Ahlfors [A1, Chapter 3]). As τ tends to 0, P tends to 1 and $A(P)$ tends to ∞, while $B(P)$ tends to a well-defined finite value (this can be seen most easily in the second variants of the two integrals (4) and (5)). The limiting behavior for the lengths follows. □

Minimizing the total number of points. Our next task is to minimize the total number of points $\alpha\beta d \log^2 d$, i.e. to minimize $\alpha\beta$. Recall that we use $s = \alpha \log d$ circles containing $\beta d \log d$ points each, and that $\kappa \in (0, 1)$ measures the relative size of the round annulus we put our circles into with respect to the size of a fundamental domain (κ depends on the radius R: the larger R, the larger we can choose κ and the fewer the total number of points needed, but the longer it takes for the starting points to iterate towards the unit disk). In the vertical strip S of width 1, adjacent points have distance

$$\tau = \frac{2\pi s}{\kappa\beta d \log d \log(d/(d-1))} = \frac{2\pi\alpha}{\kappa\beta d \log(d/(d-1))} < \frac{2\pi\alpha}{\kappa\beta} . \qquad (6)$$

The further quantities involved are

$$P = e^{2\pi\tau} \quad \text{auxiliary variable for elliptic integrals} \qquad (7)$$

$$M(\tau) = \frac{B(P)}{2A(P)} \quad \text{the maximal modulus of a quadrilateral in } S \text{ avoiding the points } i\tau\mathbb{Z} \qquad (8)$$

and the fundamental inequality we have to satisfy is

$$M(\tau) < \alpha\pi . \qquad (9)$$

The optimal values of α and β do not depend on κ, but of course the total number of points does. It is convenient to choose a value τ_0 first, then set $P_0 := e^{2\pi\tau_0}$ and calculate $M_0 := B(P_0)/2A(P_0)$. Finally, choose

$$\alpha := \frac{M_0}{\pi} \quad \text{and} \quad \beta := \frac{2\pi\alpha}{\kappa\tau_0} . \qquad (10)$$

Calculating τ as defined in (6) assures that $\tau < \tau_0$, hence $M < M_0$ by monotonicity, so (9) is satisfied. The total number of points is then

$$\alpha\beta = \frac{2\pi\alpha^2}{\kappa\tau_0} = \frac{1}{2\pi\kappa\tau_0}\left(\frac{B(P_0)}{A(P_0)}\right)^2 ;$$

it only depends on τ_0 and κ, not on the degree. We can choose $\tau_0 = 0.40198$ so that $P_0 \approx 12.5$, $\alpha \approx 0.2663$ and $\beta \approx 4.1627/\kappa$, so $\alpha\beta \approx 1.1086/\kappa$. This

gives a valid choice of points for the theorem; the reader is invited to verify that these numbers are indeed close to the minimum (see Fig. 7.)

Of course, the numbers of circles and the numbers of points on each of them must be integers, so we have to round up (see Sect. 9).

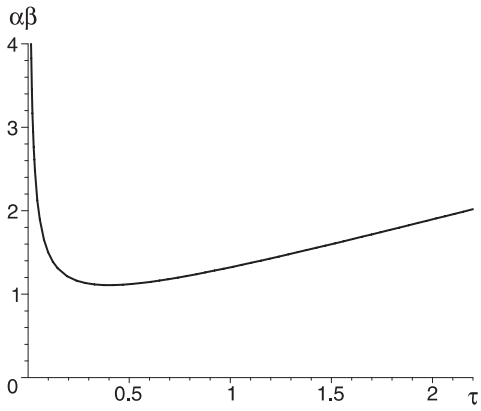


Fig. 7 Graph of $\alpha\beta/\kappa$ as a function of τ. As described in the text, the total number of points needed is $\lceil \alpha \log d \rceil \lceil \beta d \log d/\kappa \rceil$. The minimum $\alpha\beta/\kappa \approx 1.1086$ occurs near $\tau = 0.40198$

Remark. We did not require any correlation between the points on the different circles. This worst case is that all the points are lined up (so that all the points on the various circles sit at equal angles). The bounds can be somewhat improved if we rotate the points on the circles "out of phase", so that the points on adjacent circles are rotated by exactly half the angle between adjacent points. However, we have not been able to give a better upper bound for the moduli of annuli avoiding all these points.

Remark. The elliptic integrals $A(P)$ and $B(P)$ used in the calculation can of course be evaluated by any numerical method. However, many numerical integration techniques can be reluctant when applied to such integrals; for example, releases of Maple prior to Maple VR5 were unable to compute these integrals in a reasonable time. Fortunately, there is a very efficient way to calculate these elliptic integrals using an iterative procedure going back to Gauß. It uses the *arithmetic-geometric mean*.

For two positive real numbers a_0 and b_0, we define two sequences as follows: set recursively $a_{n+1} := (a_n + b_n)/2$ (the arithmetic mean) and $b_{n+1} := \sqrt{a_n b_n}$ (the geometric mean). For any choice of initial values, these two iterations converge extremely rapidly to a common limit, which is denoted $\mathrm{AGM}(a_0, b_0)$. In fact, this convergence is quadratic: as fast as the Newton map near a simple root. Now we can calculate the elliptic integrals as follows:

$$A(P) = \frac{\pi}{\mathrm{AGM}(\sqrt{P-1}, \sqrt{P})} \quad \text{and} \quad B(P) = \frac{\pi}{\mathrm{AGM}(1, \sqrt{P})} \ .$$

This relation can be found in many textbooks on elliptic functions; see for example Hurwitz and Courant [HC]. Another more explicit reference is Bost and Mestre [BM, Sect. 2.1] and in particular their formulas (10) and (11) (their polynomials are normalized so as to have a factor 4 in front, so their formulas differ from ours by a factor of 2).

7 If all the roots are real

The case that all the roots of a given polynomial are real might be of special interest. It allows a particularly simple treatment.

Let p be a polynomial in $\mathcal{P}_d$ all of whose roots are real, and denote the smallest and largest roots by $\xi_{\min}$ and $\xi_{\max}$. Then under iteration of N_p, all of $\{x \in \mathbb{R} : x \le \xi_{\min}\}$ will converge to $\xi_{\min}$, and all of $\{x \in \mathbb{R} : x \ge \xi_{\max}\}$ will converge to $\xi_{\max}$, so both ends of $\mathbb{R}$ will be parts of channels to infinity. Let $d' \le d$ be the geometric number of roots (not counting multiplicities). Then the algebraic degree of N_p is easily seen to be d' and N_p has $2d' - 2$ critical points. By symmetry, all the $d' - 2$ roots other than $\xi_{\min}$ and $\xi_{\max}$ must have at least two channels each, so their immediate basins contain at least two critical points. Since this exhausts the available critical points, it follows that the minimal and maximal roots each have exactly one channel, and the other roots each have exactly two channels. (This can also be shown by the fact that two channels of any root always enclose a further root.)

By Proposition 7, every channel has width at least $\pi/\log 3$, independently of the degree. We will use a single circle, so $s = \alpha \log d = 1$ (since the modulus is uniformly large, it does not help to use more circles). First we find P such that the moduli are

$$\frac{B(P)}{2A(P)} = \frac{\pi}{\log 3} .$$

We obtain $P \approx 3\,972\,132$ and $\log P \approx 15.194\,8$. In the limit $R \to \infty$, we then need $4\pi^2 d/\log P \approx 2.5982\,d$ points on a full circle or, using the symmetry, $1.2991\,d$ points on a semicircle. That is an efficiency of 77% (the ratio of the roots to the number of starting points).

There might be an efficient bound on the number of iterations in the case of real roots: Proposition 2.3 in Manning [M] implies that, when any point z has imaginary part $y > 0$, then its image under the Newton map has imaginary part at most $(1 - 1/d)\,y$ (the imaginary part of the image may be negative; in that case, the bound says nothing at all). This might give a bound on the number of necessary iteration steps needed to get ε-close to the root for points converging to it within the upper half plane.

Since all the critical points converge to the roots, the Newton maps of polynomials with only real roots are hyperbolic, there are no extra attracting orbits, and the measure of the Julia set is zero. Thus almost every starting point will converge to some root. Restricting to the real line, an old result by Barna [Ba] says that the set of starting points not converging to any root has measure zero and is a Cantor set or countable.

For arbitrary polynomials, wide basins are not so untypical, as the following corollary shows.

Corollary 13 (Number versus widths of channels)
More than half of the roots of a complex polynomial of arbitrary degree have a channel of modulus at least $\pi/\log 3$.

Proof. All the channels of any root can be smaller only if the immediate basin contains at least three critical points, but since the number of critical points is less than twice the number of roots, fewer than half of the roots can behave in this way. □

If we distribute points as above in the real case, except using an entire circle rather than its upper half, we can be sure to find at least half the roots. We can then deflate the polynomial and continue with the remainder. We need $2.6\,d$ starting points in the first step, half of that after the first deflation, a quarter after the second deflation, and so on. The total number of starting points used is then less than $5.2\,d$. Of course, deflation is not always possible. Moreover, it introduces numerical errors, but this will occur only $\log d$ times, rather than d times when deflating for every individual root. In a different setting, Kim and Sutherland [KS] have controlled the errors introduced by similar $\log d$ deflation steps.

8 Points on a single circle

In Sect. 5, we have used $\alpha \log d$ circles to place our starting points onto. This had the advantage that the dependence on the degree no longer showed up in our calculations, and it also improved the necessary number of starting points in a substantial way.

In this section, we explore how the number of starting points behaves if we place all the points onto a single circle. This had earlier been done in [Su]: after various improvements, $11d(d-1)$ points were needed. Here, we will show that $c_d(\pi^2/4)d^{3/2} \approx 2.47 c_d\, d^{3/2}$ points are sufficient, where c_d is a constant tending to 1 as $d \to \infty$.

First, we will investigate the asymptotic behavior of $A(P)$ and $B(P)$ when $P \searrow 1$. We should note that the following calculation up to Equation (11) can also be deduced from [A2, Sect. 4-12]. We give a different argument in order to make the paper self-contained. The estimates are particularly easy for $B(P)$ because the integral still exists at $P = 1$. Writing $P = 1 + \varepsilon$, we have

$$\begin{aligned} B(P) &= \int_0^\infty \frac{dx}{\sqrt{x(x+1)(x+1+\varepsilon)}} < \int_0^\infty \frac{dx}{(1+x)\sqrt{x}} \\ &= 2 \arctan \sqrt{x}\Big|_0^\infty = \pi. \end{aligned}$$

The evaluation of $A(P)$ needs more care:

$$A(P) = \int_{1+\varepsilon}^{\infty} \frac{dx}{\sqrt{x(x-1)(x-1-\varepsilon)}} = \int_0^{\infty} \frac{dx}{\sqrt{(x+1+\varepsilon)(x+\varepsilon)x}}$$
$$> \frac{1}{\sqrt{1+\varepsilon}} \int_0^{\infty} \frac{dx}{\sqrt{x(x+1)(x+\varepsilon)}} .$$

Now we split up the integral to the intervals $(0, \sqrt{\varepsilon})$ and $(\sqrt{\varepsilon}, \infty)$. We get

$$\int_0^{\sqrt{\varepsilon}} \frac{dx}{\sqrt{x(x+1)(x+\varepsilon)}} > \frac{1}{\sqrt{1+\sqrt{\varepsilon}}} \int_0^{\sqrt{\varepsilon}} \frac{dx}{\sqrt{x(x+\varepsilon)}}$$
$$= \frac{1}{\sqrt{1+\sqrt{\varepsilon}}} \int_0^{1/\sqrt{\varepsilon}} \frac{dt}{\sqrt{t(t+1)}} = \frac{1}{\sqrt{1+\sqrt{\varepsilon}}} \, 2\log\left(\sqrt{t} + \sqrt{t+1}\right)\Big|_0^{1/\sqrt{\varepsilon}}$$
$$> \frac{2\log\left(2\varepsilon^{-1/4}\right)}{\sqrt{1+\sqrt{\varepsilon}}} = \frac{-\frac{1}{2}\log\varepsilon + 2\log 2}{\sqrt{1+\sqrt{\varepsilon}}} .$$

For the second interval $(\sqrt{\varepsilon}, \infty)$, we have the same value as for the interval $(0, \sqrt{\varepsilon})$ because the Möbius transform $z \mapsto \varepsilon/z$ fixes the four singularities $\{-1, -\varepsilon, 0, \infty\}$ of the integrand and turns one interval into the other. (Quite explicitly, the substitution $x \mapsto \varepsilon/x$ turns one integral into the other.)

Therefore, we obtain

$$A(1+\varepsilon) > \frac{-\log\varepsilon + 4\log 2}{\sqrt{1+\varepsilon}\,\sqrt{1+\sqrt{\varepsilon}}} > \frac{-\log\varepsilon + 4\log 2}{1+\sqrt{\varepsilon}} .$$

Now we can estimate the modulus again as above: ε describes the spacing of the points on a single circle, and any channel avoiding all these points has its modulus bounded above by $B(1+\varepsilon)/2A(1+\varepsilon)$. Since we know that there is a channel with modulus at least $\pi/\log d$, we have to find an ε such that $\pi/\log d > B(P)/2A(P)$. Using our estimate above, we have to find an ε such that

$$\frac{B(1+\varepsilon)}{2A(1+\varepsilon)} < \frac{\pi}{2\dfrac{-\log\varepsilon + 4\log 2}{1+\sqrt{\varepsilon}}} = \frac{\pi(1+\sqrt{\varepsilon})}{2(-\log\varepsilon + 4\log 2)} < \frac{\pi}{\log d} . \quad (11)$$

This yields $\varepsilon^{1/(1+\sqrt{\varepsilon})} < d^{-1/2} \cdot 16^{1/(1+\sqrt{\varepsilon})}$. But since $\sqrt{\varepsilon}^{\sqrt{\varepsilon}}$ and $16^{\sqrt{\varepsilon}}$ both tend to 1 as $\varepsilon \searrow 0$, we obtain the necessary inequality $\varepsilon < 16\, d^{-1/2}/c_d$, where $c_d \geq 1$ is a factor tending to 1 as $d \to \infty$ or $\varepsilon \to 0$. (The estimates of [A2] allow us to conclude that $c_d = 1$ works for all d.)

It follows that we need a value $\varepsilon < 16\, d^{-1/2}/c_d$. In the approximation that $N(z) = ((d-1)/d)z$ (which means that points are placed onto an infinitely large circle), adjacent points should have angles which differ by $\varepsilon/2\pi d$, so the number of points we need is $4\pi^2 d/\varepsilon = (\pi^2/4) d^{3/2} c_d \approx 2.47\, d^{3/2}$.

Note that restricting the points onto a single circle not only increases the number of points substantially, it also requires the estimates to depend on the degree d. (In [Su], the moduli are estimated by a different method, using ellipses.)

9 A recipe for dessert

In this section, we give a recipe for applying the results of this paper in practice, and we suggest how to modify it according to the ingredients available (and according to taste).

Construction of the initial point grid. We have assumed throughout the paper that our polynomial p has all its roots in the unit disk. If instead the roots are in the disk $|z| < r$, you may simply scale all the starting points by a factor of r.

The first step is to select a collection $\mathcal{S}_d$ of initial points so that there is at least one in every immediate basin of every root. Here we choose our points in an annulus whose outer boundary is of radius $R = 1 + \sqrt{2}$, allowing us to use $\kappa = 1/2$ (from Lemma 10). The number of circles needed and the number of points on each circle are a function of a parameter τ (Sect. 6). We take $\tau = 0.40198$, which gives

$$s = \lceil 0.26632 \log d \rceil \text{ circles, with } N = \lceil 8.32547 d \log d \rceil \text{ points on each}$$

(we use $\lceil x \rceil$ to denote the smallest integer $\geq x$). Set

$$r_\nu = \left(1 + \sqrt{2}\right)\left(\frac{d-1}{d}\right)^{\frac{2\nu-1}{4s}} \quad \text{and} \quad \vartheta_j = \frac{2\pi j}{N}$$

for $1 \leq \nu \leq s$ and $0 \leq j \leq N-1$. With these choices, our grid $\mathcal{S}_d$ consists of the collection of Ns points $r_\nu \exp(i\vartheta_j)$. (There may be a slight advantage in rotating the circles with odd ν by π/N.)

Remark. The quantity κ is a bound on how much our fundamental domain at radius R differs from being a round annulus. For any finite R, the quantity κ is smaller than 1. Larger radii R yield larger values of κ and thus fewer points per circle. For example, $R \geq 13.94$ allows $\kappa = 9/10$ instead of $\kappa = 1/2$ as above. However, a price must be paid: increasing R from R_1 to R_2 increases the number of iterations required to arrive near a root by approximately $d \log(R_2/R_1)$.

Rounding and the numbers of starting points. In the above recipe, the number of circles is usually quite small: for degrees $d \leq 42$, the number $0.26632 \log d$ is less than 1; for $d \leq 1825$, it is less than 2; for $d \leq 78\,015$, it is less than 3 and for $d \leq 3\,333\,550$ it is less than 4. Of course, the number of circles and the number of points on them must be integers, so rounding can

have a significant impact on the size of $\mathcal{S}_d$, although this effect diminishes for very large degrees (one always has to round up to be on the safe side).

To get the smallest number of initial points, one can, for each degree d, choose α so that $\alpha \log d$ (the number of circles) is an integer (typically near $0.26632 \log d$), and then solve for τ as a function of α. Since

$$\alpha = \frac{B(e^{2\pi\tau})}{2\pi A(e^{2\pi\tau})}$$

where A and B are the integrals defined in Equations (4) and (5) of Sect. 6, this requires calculating α for a number of values of τ and interpolating the inverse. Then the number of points per circle can be computed as $\lceil 2\pi\alpha d \log d/(\kappa\tau)\rceil$.

If this is done, the total number of points is minimized by using one circle for $d \le 173$, two circles are best for $174 \le d \le 9255$, and so on. For example, for degree 100 the basic recipe above requires two circles with 1918 points each, while one circle with 2722 points will also work and requires only 63% as many starting points; for degree 12 000, the basic recipe calls for 1 407 570 points spread over three circles, while the minimal number is 1 201 893 points, also on three circles (that is, 85%).

In practice, reducing the number of starting points by even as much as a factor of 2 or 3 can have little effect: one often has found all the roots of a polynomial long before all the starting points have been tried, depending on the particular scheme used (see below). For example, in an experiment solving 1000 degree 256 polynomials with randomly chosen coefficients, on average only $758 \approx 0.53\, d \log d$ initial points were tried before all roots were found.

The reasons for this are twofold: $\mathcal{S}_d$ is a universal set of starting values, and must allow for the smallest basin of the worst-case polynomial of degree d. In particular, for a polynomial which has a root with many small channels, the grid must be tight enough to ensure that a starting point lies in the largest of them, but in practice, points will often fall into several of the smaller channels as well. Furthermore, by Corollary 13 at least half of the roots of every polynomial have relatively wide channels. Secondly, $\mathcal{S}_d$ does not take into account the preimages of the immediate basins, which, while smaller than the immediate basins, can still have a significant size.

Taking twice as many starting points on the same circles will ensure that not only is there a starting point within every basin, but also that there is one which is not too close to the basin boundary. This yields a better rate of convergence for such a point, and makes it possible to give an upper bound on the necessary number of iteration steps [Sch].

Locating the roots. By the main theorem of this paper, for each root of any polynomial $p(z) \in \mathcal{P}_d$, there is a point of $\mathcal{S}_d$ that lies in its immediate basin. Although we give no upper bound on the number of iterates of Newton's

method required to approximate the zeroes of p, it is easy to construct an algorithm to approximate the roots.

The goal is to find points $\widehat{\xi}_1, \widehat{\xi}_2, \ldots, \widehat{\xi}_d$ which approximate the d roots ξ_j of $p(z)$ (regarded with multiplicity), so that $|\widehat{\xi}_j - \xi_j| < \varepsilon$.

First, make a reasonable guess K for the number of iterations needed. Any positive value for K will work, but better guesses make the algorithm more efficient. Taking $K = \lceil d \log(R/\varepsilon) \rceil$ will do, which assumes the convergence is always at least linear. Then, for each point $z_0 \in \mathcal{S}_d$, apply Newton's method at most K times, stopping when $|z_n - z_{n-1}| < \varepsilon/d$ (where $z_n = N_p^{on}(z_0)$). This condition guarantees that there is at least one root ξ_j with $|z_n - \xi_j| < \varepsilon$ [He, Cor. 6.4g]. Any enumeration of $\mathcal{S}_d$ will satisfy the theorem, but we find enumerations where the argument of subsequent initial values differ by about $2\pi/d$ to be most efficient.

If the root ξ_j approximated by z_n is different from all previously found roots, let $\widehat{\xi}_j = z_n$ (see below for how to detect this). Notice that roots must be accounted for with multiplicity. If z_n approximates a previously found root, discard this orbit entirely. If Newton's method has been applied K times to z_0 without converging to a root, save the value of z_K in a set $\mathcal{S}_d^1$ for possible future use. In addition, if $|z_k| > R$ for any $k > 1$, the iteration of z_k can safely be deferred by storing z_k in $\mathcal{S}_d^1$. If after trying all the initial points in $\mathcal{S}_d$ the number of roots found so far is less than d, then begin again, using points from $\mathcal{S}_d^1$ as starting values and saving the non-convergent points in a set $\mathcal{S}_d^2$. Continue in this way until all d roots are found.

When $|z_n - z_{n-1}| < \varepsilon/d$, we know that $|z_n - \xi_j| < \varepsilon$ for some root ξ_j, but even if z_n is in the immediate basin of some root ξ_j', we cannot guarantee that $\xi_j' = \xi_j$. The method described here would fail in the following situation: there is a root ξ_0 so that the orbit of every point $z_0 \in \mathcal{S}_d$ in its basin comes within ε of some other root ξ_j with $|\xi_j - \xi_0| > \varepsilon$ (or we would view the roots as a multiple). Furthermore, this orbit must pass ξ_j slowly, so that $|z_n - z_{n-1}| < \varepsilon/d$. We have never observed this, but the best bound we can provide to exclude such a situation is much weaker [Sch].

Detecting duplicate approximations. When we have a point $\widehat{\xi}$ for which $|N_p(\widehat{\xi}) - \widehat{\xi}|$ is sufficiently small, we must decide whether it approximates a previously found root ξ_j or a new one. If the root in question is simple, there are explicit criteria to decide this: see [KS, Lemma 2.7]. These criteria use the Kim-Smale α-function, introduced by Kim in [K1,K2] and sharpened and extended by Smale [Sm]. Unfortunately, computing the α-function requires evaluating (or at least estimating) all of the derivatives of p at the point $\widehat{\xi}$. This can make its use computationally troublesome if the degree is very large.

In most cases, the roots are simple and sufficiently separated so that detection of duplication is not an issue. In the rare cases where there are multiple or clustered roots, the multiplicity can often be inferred from the rate of convergence or the value of N_p'.

The accuracy necessary for calculations. One significant advantage of Newton's method over some other root-finding methods is its great numerical stability. Even if the goal is to approximate the roots of $p(z)$ to very high precision, we need only first find d approximations to moderate precision (say, with $\varepsilon = 10^{-8}$), and then refine these approximations as needed. Since we are finding all of the roots at once, with no intermediate deflation, the fact that we initially use a low precision will not affect the final result.

References

[A1] Lars Ahlfors: Lectures on quasiconformal mappings. Van Nostrand publishers (1966)/Wadsworth&Brooks/Cole (1987)

[A2] Lars Ahlfors: Conformal invariants; Topics in geometric function theory. McGraw-Hill, New York (1973)

[Ba] Béla Barna: Über die Divergenzpunkte des Newtonschen Verfahrens zur Bestimmung von Wurzeln algebraischer Gleichungen IV. Publ. Math. Debrecen **14** (1967), 91–97

[BM] Jean-Benoît Bost, Jean-François Mestre: Moyenne Arithmético-géométrique et Périodes des Courbes de Genre 1 et 2. Preprint, LMENS-88-13, Ecole Normale Superieure (1988)

[HC] Adolf Hurwitz, Richard Courant: Funktionentheorie (4. Auflage). Springer, Berlin (1964)

[He] Peter Henrici: Applied and computational complex analysis. Volume 1: Power series, integration, conformal mapping, location of zeros. Wiley, New York (1974)

[HP] John Hubbard, Karl Papadantonakis: Exploring the parameter space for Hénon mappings. Manuscript (2000), to appear in: Experimental Mathematics

[JT] William B. Jones, Wolfgang H. Thron: Continued fractions (Analytic theory and applications). Addison-Wesley (1980)

[K1] Myong-Hi Kim: Computational complexity of the Euler type algorithms for the roots of complex polynomials. Thesis, City University of New York, New York (1985)

[K2] Myong-Hi Kim: On approximate zeros and rootfinding algorithms for a complex polynomial. Mathematics of Computation **51** (1988), 707–719

[KS] Myong-Hi Kim, Scott Sutherland: Polynomial root-finding algorithms and branched covers. SIAM Journal of Computing **23** (1994), 415–436

[Kr] Hartje Kriete: On the efficiency of relaxed Newton's method. In: Pitman research notes in Mathematics Series **305** (1994), 200–212

[M] Anthony Manning: How to be sure of finding a root of a complex polynomial using Newton's method. Boletim da Sociedade Brasileira de Matematica **22**(2) (1992), 157–177

[Ma] Morris Marden: The geometry of the zeros of a polynomial in a complex variable. Mathematical Surveys III, Amer. Math. Soc. (1949)

[McM1] Curt McMullen: Families of rational maps and iterative root-finding algorithms. Ann. Math. (2) **125**(3) (1987), 467–493

[McM2] Curt McMullen: Braiding of the attractor and the failure of iterative algorithms. Invent. math. **91**(2) (1988), 259–272

[Mi] John Milnor: Iteration in one complex variable: Introductory lectures. Vieweg Verlag (1999)

[MS] Maurice Mignotte, Doru Ştefănescu: Polynomials: An algorithmic approach. Springer, Singapore (1999)

[NP] Rolf Nevanlinna, Veikko Paatero: Einführung in die Funktionentheorie. Birkhäuser Verlag, Basel (1964)

[Pa] Victor Pan: Solving a polynomial equation: some history and recent progress. SIAM Review **39**, (1997) 187–220

[Pr] Feliks Przytycki: Remarks on the simple connectedness of basins of sinks for iterations of rational maps. In: Dynamical Systems and Ergodic Theory, ed. by K. Krzyzewski, Polish Scientific Publishers, Warszawa (1989), 229–235

[Sch] Dierk Schleicher: On the number of iterations of Newton's method for complex polynomials. Preprint (2000), to appear in: Ergodic Theory and Dynamical Systems

[Sh] Mitsuhiro Shishikura: Connectivity of the Julia set and fixed point. Preprint, Institut des Hautes Etudes Scientifiques IHES/M/90/37 (1990)

[Sm] Steven Smale: Newton's method estimates from data at one point. In: The Merging disciplines: New directions in pure, applied, and computational mathematics. Springer (1986), 185–196

[Su] Scott Sutherland: Finding roots of complex polynomials with Newton's method. Thesis, Boston University (1989)