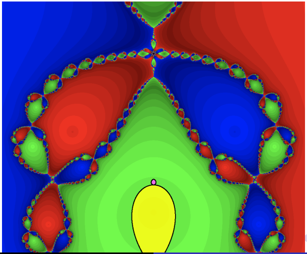
Finding polynomial roots using complex analysis, dynamical systems, and computer algebra

Dierk Schleicher

Aix-Marseille Université

Transcendental Dynamics and Beyond
Barcelona, 19 April 2021



2 The fundamental theorem for algebra

Theorem 1 (Gauss)

Every (univariate) degree d polynomial over $\mathbb{C}$ has exactly d roots, counting multiplicity.

Theorem 2 (Ruffini–Abel)

For degree $d \geq 5$, there is no general formula by finitely many radicals to find all roots of every degree d polynomial.

- **Conclusion:** the d roots have to be found by iterative methods.

Theorem 3 (Curt McMullen; PhD thesis)

For degrees $d > 3$ there is no holomorphic root finding method iterating on $\mathbb{C}$ that is generally convergent (i.e. converges to roots on an open dense subset of $\mathbb{C}$ for all polynomials).

3 Several root-finding methods

Numerical analysis has numerous algorithms to find the roots:

1. Newton's method — finds one root at a time (when it does), well understood as a one-dimensional dynamical system
2. Weierstrass method (Durand–Kerner) — finds all d roots simultaneously (when it does), good in practice, seems to converge fast in most cases, but no theory
3. Ehrlich–Aberth-method — similar to Weierstrass' method, but apparently faster
4. Eigenvalue methods — seem to work well for moderate degrees
5. Victor Pan's algorithm: almost best possible theoretical complexity, but unusable in practice — many further methods

Our focus today: the first three methods (blue); iteration of complex analytic mappings. — *Dynamical systems in heavy practical use — but very hard to understand*

4 Overview

The root-finding methods by Newton, Weierstrass, Ehrlich–Aberth

Motto: what is the difference between theory and practice?

Local properties, global questions

do they always converge? almost?

positive practical evidence — BUT failure of global convergence;

I. New theorem: *Weierstrass is not generally convergent; has orbits that are always defined and converge, but not to roots*

II. (not so) new results: *Newton's method has good theory and works well in practice*

III. Newton as a dynamical system: pioneering method for understanding dynamics of rational maps

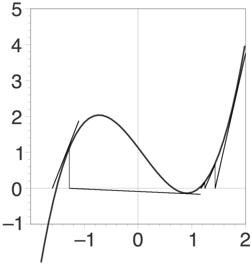
5 The Newton Method

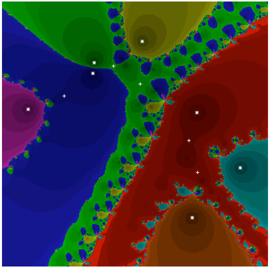
Given differentiable $f\colon \mathbb{R} \to \mathbb{R}$, or $f\colon \mathbb{C} \to \mathbb{C}$, the Newton map is

$$z \mapsto N_f(z) = z - f(z)/f'(z)$$

Idea: the best-approximating linear map to f at z is $w \mapsto (z-w)f'(z) + f(z)$, its zero is at $z - f(z)/f'(z)$.

This is good approximation when w is close to z; otherwise, iterate . . . and hope!





This finds one zero at a time; how can we go for all roots? Starting any d orbits might find the same root many times. . .

6 Derivation of Weierstrass and Ehrlich–Aberth

Let $\mathcal{P}_d$ be the space of monic polynomials of degree d. Given $p(z) = \prod_i (z - \alpha_i) \in \mathcal{P}_d$ and $(z_1, \ldots, z_d) \in \mathbb{C}^d$, consider

$$q_i(z) = \frac{p(z)}{\prod_{j \neq i}(z - z_j)} \in \mathbb{C}(z)$$

Heuristic interpretation: if $p(z) = \prod_i (z - \alpha_i)$ and $z_j \approx \alpha_j$, then $q_i(z) \approx z - \alpha_i$.

(Each coordinate thinks that the others must be correct)

Weierstrass:

$$z_i \mapsto z_i - q_i(z_i)$$

solves q_i as linear polynomial

Ehrlich–Aberth:

$$z_i \mapsto z_i - \frac{q_i(z_i)}{q_i'(z_i)}$$

Newton step for q_i

Update all d components in parallel: dynamics in $\mathbb{C}^d$

How to find all d roots of a degree d polynomial p at once?

Given polynomial
$p(z) = z^d + a_{d-1}z^{d-1} + \cdots + a_1 z + a_0 = \prod_i (z - \alpha_i)$
From a root vector $(\alpha_1, \ldots, \alpha_d)$, the coefficient vector $(a_1, \ldots, a_d)$ is determined through elementary symmetric functions.
Need to compute the inverse $(a_1, \ldots, a_d) \mapsto (\alpha_1, \ldots, \alpha_d)$.

We have to solve d equations in d variables. Can use Newton's method in d variables and iterate. This iteration equals the *Weierstrass method*.

Consequence: invariant hyperplane in which the sum of all components is constant, Weierstrass projects to this hyperplane, hence iteration in $\mathbb{C}^{d-1}$.

8 Relations of Ehrlich–Aberth to Newton

Write Newton in one variable, using roots α_i

$$z \mapsto z - p(z)/p'(z) = z - \left(\sum_i \frac{1}{z-\alpha_i}\right)^{-1}$$

(roots not known ahead of time, but expression helpful anyway).

Electrostatic interpretation: have test charges at positions of the roots α_i; the Newton displacement $N_p(z) - z$ is (the inverse of) the electrostatic field at z.

For d orbits in parallel with approximation vector $(z_1, \ldots, z_d)$, produce next generation approximation vector $(z_1', \ldots, z_d') = EA(z_1, \ldots, z_d)$ via improved Newton step:

$$z_k \mapsto z_k - q_k(z_k)/q_k'(z_k) = z_k - \left(\sum_i \frac{1}{z_k-\alpha_i} - \sum_{i\neq k} \frac{1}{z_k - z_i}\right)^{-1}$$

Attracting charges at roots, repelling charges at other approximations!

9 Local Properties, Global Problems

For all three methods, *points* $z \in \mathbb{C}$ *(for Newton) or vectors* $(z_1, \ldots, z_d) \in \mathbb{C}^d$ *(for the others) are fixed if and only if one root, resp. all roots, are found*.

All these fixed points are *attracting*: there is a neighborhood in $\mathbb{C}$ resp. $\mathbb{C}^d$ so that all these points converge to the root(s).

The convergence is always very fast: quadratic (number of valid digits doubles in each step); Ehrlich–Aberth is even cubic.

Global properties difficult: where does one have to start? how many iterations required (what is the complexity)?

Known problems:

a) symmetry; for real polynomials with non-real roots, real starting vectors, convergence must fail
b) basin boundaries: different root (vectors) have disjoint open basins, $\Longrightarrow$ $\exists$ basin boundaries; may have positive measure!
c) orbits may jump to ∞ (Newton) or fail to be defined (points of indeterminacy) (Weierstrass / E–A)

9 Local Properties, Global Problems

For all three methods, *points* $z \in \mathbb{C}$ *(for Newton) or vectors* $(z_1, \ldots, z_d) \in \mathbb{C}^d$ *(for the others) are fixed if and only if one root, resp. all roots, are found*.

All these fixed points are *attracting*: there is a neighborhood in $\mathbb{C}$ resp. $\mathbb{C}^d$ so that all these points converge to the root(s).

The convergence is always very fast: quadratic (number of valid digits doubles in each step); Ehrlich–Aberth is even cubic.

Global properties difficult: where does one have to start? how many iterations required (what is the complexity)?

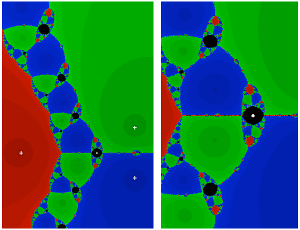
Known problems:
a) symmetry; for real polynomials with non-real roots, real starting vectors, convergence must fail
b) basin boundaries: different root (vectors) have disjoint open basins, $\implies \exists$ basin boundaries; may have positive measure!
c) orbits may jump to ∞ (Newton) or fail to be defined (points of indeterminacy) (Weierstrass / E–A)

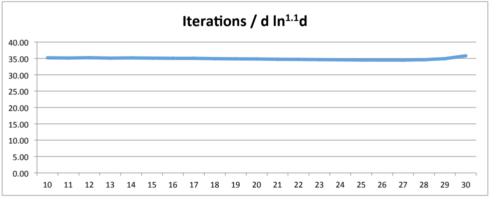
10 Empirical evidence: Newton

Newton: for as simple polynomials as $z \mapsto z^3 - 2z + 2$, there are attracting orbits: open sets in $\mathbb{C}$ that converge to periodic cycles but not roots. *Newton's method is not generally convergent!*

Difficult to correlate the orbits of d starting points, hoping to find d different roots (deflation is usually not an option!). [But theory available; see below].

Attracting cycles are a problem, but polynomials with degrees in the millions and billions have been factorized successfully!





attracting cycle (left) and complexity up to period $2^{30} > 10^9$

Weierstrass / Ehrlich–Aberth: are “known to converge in all cases, except the obvious cases with symmetry”; have passed the test of time over decades in uncounted experiments. Successful in standard implementations such as MPSolve, degrees up to millions.

Recent (heuristic) improvement by Dario Bini (Pisa) to MPSolve for special (recursive) polynomials: Ehrlich–Aberth successful in selected cases for degrees up to billions.

12 What can go wrong? Complex manifolds and eigenvalues

Main (known) problem are attracting cycles: periodic orbits for which a neighborhood converges to these, rather than to roots.

Lemma: *If $f\colon \mathbb{C}^k \to \mathbb{C}^k$ differentiable with $f(w) = w$ and $Df|_w$ has all eigenvalues in $\mathbb{D}$, then w is attracting.*

For us, $k = 1$ for Newton and $k = d$ for Weierstrass / Ehrlich–Aberth.

If f is Newton, Weierstrass, E-A, then this is what happens at root. But if it is an iterate, we obtain an attracting cycle. This situation is stable under perturbations:
attracting cycles spoil general convergence.

If $k > 1$ and only $m < k$ eigenvalues are in $\mathbb{D}$, then there is an invariant m-dimensional complex manifold (the stable manifold) on which orbits converge to the cycle (tangent to the appropriate eigenvectors). *Semi-attracting cycles!*

But does this happen?

13 Parameter space and dimension count

Space of monic complex polynomials of degree d is $\mathbb{C}^d$ (parametrized e.g. by coefficients).

All three methods respect affine coordinate changes: these lead to conjugate dynamics with same (semi-)attracting dynamics. Hence space of non-equivalent polynomials (parameter space) is complex $d-2$-dimensional.

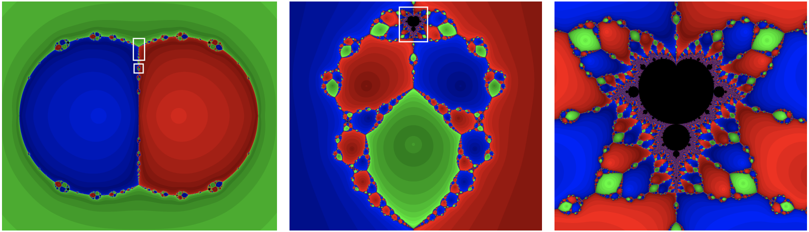
Easiest case: cubic polynomials, up to conjugation can suppose roots are at 0, 1, $\lambda \in \mathbb{C}$: parameter space is $\mathbb{C}$ for all three methods. *Fundamental case!*

Equivalently, parametrize as $p(z) = z^3 + \lambda z + 1$ or in other ways.

Every periodic point satisfies an algebraic equation in λ, so all eigenvalues are algebraic functions. Can they all be in $\mathbb{D}$?
$\Longrightarrow$ real algebraic question!

Dimension count: for cubic Newton: one parameter, one eigenvalue:
$\Longrightarrow$ there must (generically) be attracting cycles of all periods!

Theorem [Douady–Hubbard, 1980's]
The set of parameters λ so that the Newton method for $p(z) = z(z-1)(z-\lambda)$ has an attracting cycle of period $n \geq 2$ is contained in finitely many homeomorphic copies of the Mandelbrot set.



15 Cubic Weierstrass

Cubic Weierstrass: one parameter, three eigenvalues (iteration in $\mathbb{C}^3$): generically, expect semi-attracting cycles.

Reminder: dynamics projects to invariant $d-1$-dim. hyperplane, so one eigenvalue always zero, two eigenvalues remain.

There is *no a priori reason* why both eigenvalues can or cannot be in $\mathbb{D}$. (Separate question for each period.)

Conjecture (Steven Smale): in some cases, there should be attracting cycles

Common wisdom (numerical analysis community): attracting cycles would have been found experimentally (decades ago!) if they existed.

Since both eigenvalues are algebraic functions on the same surface, they must satisfy an algebraic (polynomial) relation.

So question moves from numerical analysis via dynamical systems to algebra; computer algebra brings new tools !

16 Dynamical results about Weierstrass method

Theorem A (Reinke–S.–Stoll 2020)

There is an open set of polynomials p of every degree $d \geq 3$ such that the (partially defined) Weierstrass iteration $W_p \colon \mathbb{C}^d \to \mathbb{C}^d$ associated to p has attracting cycles of period 4. Period 4 is minimal with this property.

- $\Rightarrow$ Weierstrass's method is not generally convergent!
- Existence of attracting cycles was already conjectured by Smale, has never been observed in practice.
- Additional result by Bernhard: *Weierstrass (and Ehrlich–Aberth) have orbits that are defined forever that converge, but not to a root but to ∞ (or elsewhere).*
- This work brings together dynamical systems, numerical analysis, algebra, and computer algebra.
- $\Longrightarrow$ See talk of Bernhard Reinke this Thursday, 17:00, *The Weierstrass root finder is not generally convergent*

17 Is Ehrlich–Aberth generally convergent?

In principle, the question can be asked. Similar methods can be used for Ehrlich–Aberth. Three additional difficulties:

$\star$ The Ehrlich–Aberth-map is more complicated than Weierstrass;
$\star$ Ehrlich–Aberth does not project to a hyperplane, so have one more non-trivial eigenvalue to consider;
$\star$ Ehrlich–Abert not only invariant under affine maps, but even under all Möbius maps.

This seems like a good property, however:

Up to equivalence, there are only three different cubic Ehrlich–Aberth-cases: all roots distinct, or one double/one simple root, or one triple root. Cubic parameter space trivial, so interesting case is degree 4 with one complex degree of freedom: even more complicated.

Seems out of reach of current computer algebra systems :-(

18 Towards Global Results

All these are local results: periodic points and their eigenvalues. But what are the global properties of the dynamics?

In particular: where do we need to start the iteration to find the roots?

No global theory for Weierstrass and Ehrlich–Aberth: *they seem to work and nobody knows why*.
Experience suggests to start with d equidistributed points on large circle.

However, for Newton there is meanwhile quite a bit of global theory.

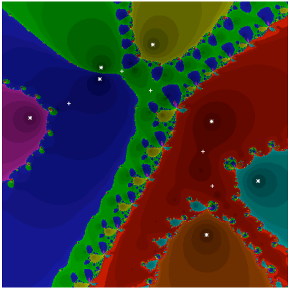
My “mathematical home”: dynamical systems. Interesting interaction with root finding experts (Dario Bini, Victor Pan)

19 Good Starting Points for Newton's Method

Definition. The *immediate basin* of a root α is the connected component of the basin containing α.

Theorem. Every immediate basin is simply connected and unbounded, extends to ∞ in $k \geq 1$ directions "channels". (Przytycki, 1980's)

Let d be the degree of the polynomial.



Theorem 4 (Hubbard, S., Sutherland, 2001)

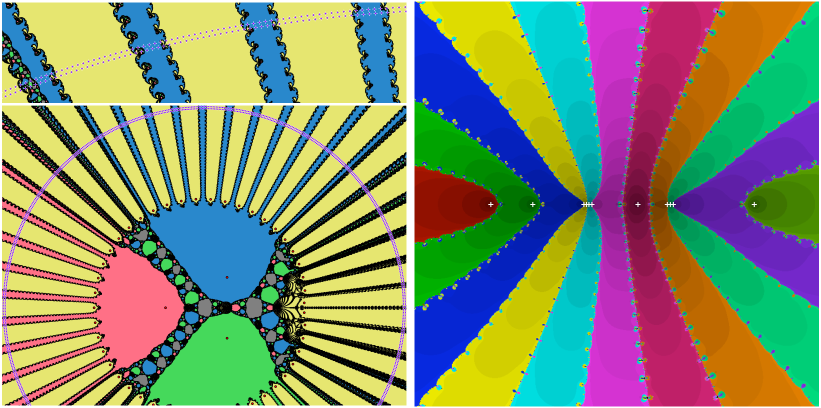
Every root has at least one channel with "thickness" $\pi / \log d$.
Hence $1.1 d \log^2 d$ starting points suffice to find all roots.
If all roots are real, then $1.3\, d$ starting points suffice.

(First paper with color figures in *Inventiones.*)

Theorem 5 (Bollobás, Lackmann, S., 2011)

With high probability, $O(d(\log \log d)^2)$ starting points suffice.

20 Good starting points



Left: the guaranteed starting points for a degree 50 polynomial. Note the extra attracting cycles (grey).
Right: complex basins help even when all roots are real. Near-multiple roots are easily distinguished.

21 Theory on Number of Iterations

Let S_d be the guaranteed set of starting points with $|S_d| \approx 2d \log^2 d$.

Theorem 6 (Fast convergence; S, 2002–2015)

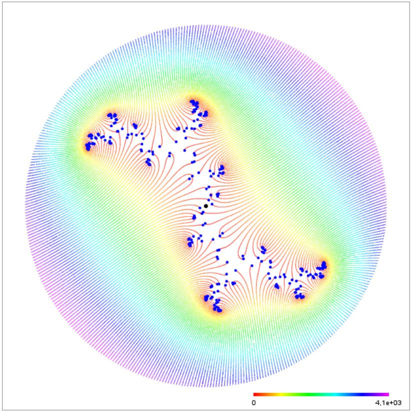
For every degree d and every $\varepsilon > 0$ there is an $M_d > 0$ so that for each $p \in P_d$ and each root α_j of p, at least one starting point $z_j \in S_d$ converges to α_j under N_p, and $|N_p^{\circ m_j}(z_j) - \alpha_j| < \varepsilon$ with $\sum_j m_j < M_d$.
We have $M_d \in O(d^4 + d \log \varepsilon)$ in the worst case, and $M_d \in O(d^2 \log^4 d + d \log|\log \varepsilon|)$ in the expected case.

Refinements in Bachelor thesis of T. Bilarev; Math Comp 2016

Roughly speaking, some of the guaranteed starting points suffice to find all roots in $O(d^2 \log^4 d)$ iterations.

Note. This is essentially best possible for the guaranteed set of starting points.

22 Why is this bound best possible? And Wasteful!

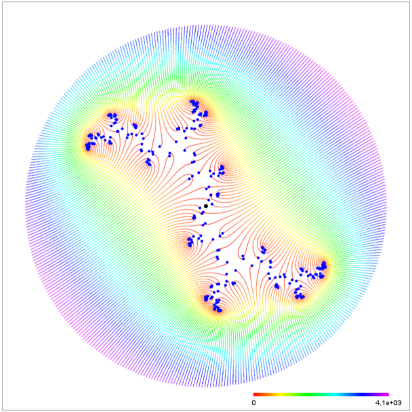


All orbits have to start in area of controlled dynamics, away from disk containing roots.
On this domain $N_p(z) \approx \frac{d-1}{d} z$.
To move a factor of r towards $\mathbb{D}$, each orbit needs $\approx d \log r$ iterations.
Need at least d orbits for d roots: complexity $\geq \Omega(d^2)$.

Note. Upper and lower bounds coincide up to log-factors.

... but it does work in practice!

23 Newton's Method in Practice

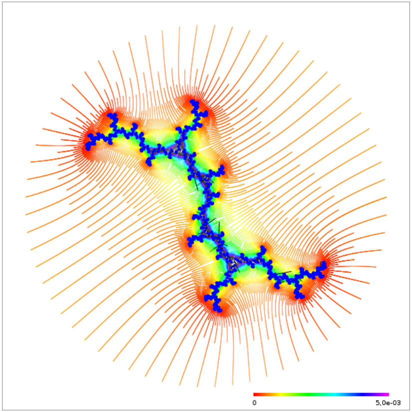


Heuristically, $4d$ equidistributed starting points on single circle find "most" roots;
if some remain missing, refine to $8d$ ...
In practice, several families of polynomials *of degree one million* completely factored (Robin Stoll, high school student, on old laptop).

A posteriori guarantee that all proofs found: if $|N(z) - z| < \varepsilon$, then at least one root α with $|z - \alpha| < d\varepsilon$, need to have d disjoint disks.

Clear refinement strategy while roots are missing: add more initial orbits.

24 The Iterated Refinement Method



Natural heuristics: in area where dynamics "parallel", use only few orbits and refine as dynamics becomes non-linear. **Danger:** no longer clear that all roots are found. If roots missing, no obvious place for refining Newton dynamics.

"Experimental" improvements, no complete theory yet.

But successful experiments: again Robin Stoll, now young university student — and Marvin Randig, another high school student. Up to degrees $2^{30} > 10^9$ on a student laptop!

25 Root finding: overview

⋆ root finding is a classical and important problem

⋆ there are various methods that work well in practice (at least for moderate degrees), but GAP between theory and practice

⋆ Weierstrass and Ehrlich–Aberth: basis of standard software package, have very good reputation, but little theory

⋆ failure of general converge (Weierstrass) and diverging orbits (both) are new phenomena in global dynamics

⋆ Newton used to have poor reputation (except locally), but meanwhile unique good combination of theory and practice

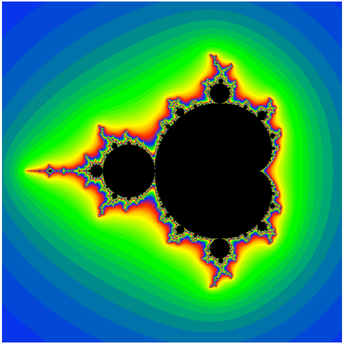
⋆ (but most efficient implementation not yet supported by theory)

⋆ For a systematic comparison on Newton and Ehrlich–Aberth as root finders, see talk by Sergey Shemyakov tomorrow 18:00

⋆ many questions on all topics still wide open

$\Longrightarrow$ these root finding methods give rise to most interesting dynamical systems

The BIGGEST question in (holomorphic) dynamics:
Is the Mandelbrot set locally connected? (Space of iterated quadratic polynomials



Question with long history: Douady-Hubbard 80's, Yoccoz 90's, Lyubich&coauthors since then: deep results, still not resolved.

Why relevant? Provides topological model for Mandelbrot set; *shows that all quadratic polynomials have combinatorially distinct dynamics.*

Question is far more general: can all holomorphic dynamical systems be combinatorially distinguished? *Rigidity conjecture!*

27 Rigidity in holomorphic dynamics

Rigidity conjecture: Can all holomorphic dynamical systems be combinatorially distinguished?

Big and fundamental question, very deep results — and still wide open!

Fundamental principle in holomorphic dynamics: many features determined by dynamics of critical points (points where the derivative vanishes). The more critical points, the more complications!

A rational map of degree d has $2d - 2$ critical points. For a polynomial, half of them are fixed at ∞.

$\implies$ easiest case: quadratic polynomial, only 1 “active” critical point! Thus pioneering and deepest results about quadratics

Second easiest case: unicritical polynomials: 1 critical point of high multiplicity: combinatorics same as for quadratics, analytic difficulties resolved by Kahn&Lyubich early 2000’s.

Several “active” critical points MUCH HARDER!

28 Rigidity beyond the unicritical case

Holomorphic dynamics in the 2020's: move from special cases with one active critical point to natural generality. Live up to the promise that "one active critical point" features are universal.

However, with several active critical points, structure is *much more difficult!*

Fundamental ingredient: *box mappings* by Kozlovski–van Strien (recently rejuvenated jointly with Kostya Drach).

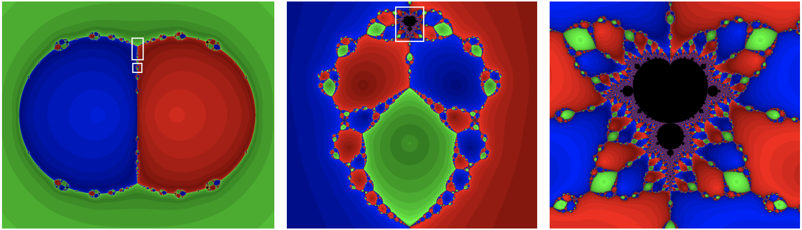
Yields progress for *polynomials* of higher degrees.

Common wisdom: non-polynomial rational maps yet MUCH more difficult. Polynomials have very good and useful coordinates around infinity, not available for rational maps.

29 Newton's method as a rational map

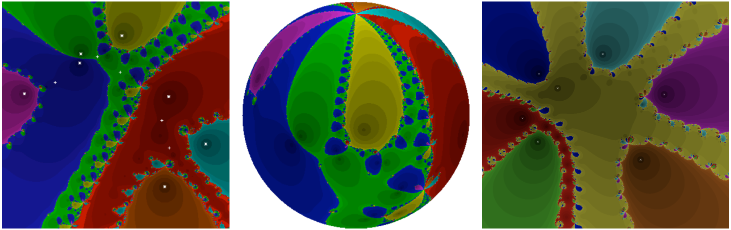
For polynomial p of degree d, Newton's method $N_p(z) = z - p(z)/p'(z)$ is a rational map of degree d (in general). Of its $2d-2$ critical points, d are fixed at the roots, the remaining $d-2$ are "active". Parameter space has complex dimension $d-2$.

Classical case: $d = 3$, one active critical point, parameter space complex 1-dimensional. Satisfactory work by Douady-Hubbard, Tan Lei, Pascale Roesch & others: well understood "modulo embedded Mandelbrot sets".



Higher dimensional parameter spaces much bigger challenge!

30 Newton's method as a rational map



Various Newton dynamical systems on the complex plane and the Riemann sphere

31 Newton's method as a rational map, general degree

Global structure for arbitrary rational maps much more difficult, no general theory. Assumed to be much more complicated than polynomials. Even polynomials of degree $d \geq 3$ are not rigid in general (Buff&Henriksen, ...)

Rational Rigidity Principle: every iterated rational map is rigid (can be distinguished combinatorially from other rational maps), *except where* embedded polynomial dynamics interferes.

However, Newton maps are well behaved dynamical systems:

Theorem 7 (Rigidy of Newton dynamics: Drach–S., 2020)

Every Newton map of every polynomial of every degree is rigid, except when embedded dynamics of iterated polynomials interferes.

(See talk by Kostiantyn Drach this Friday morning.)

There is hope to extend these methods: from Newton maps to "Newton-like" and more...

32 Conclusion

$\star$ Newton maps are dynamical systems, well motivated by root finding

$\star$ are known to be very good for polishing up "approximate roots", but globally chaotic, not generally convergent, have bad reputation

$\star$ but are unique as root finders that have good theory and work well for degrees up to billions

$\star$ Weierstrass and Ehrlich–Aberth have very good reputation, converge "always in practice"

$\star$ but have no global theory whatsoever, and *(W.) not generally convergent either*

$\star$ Moreover, Newton is a pioneering dynamical system: the only non-polynomial rational maps for which we have satisfactory theory! "Rational rigidity" works for Newton maps and probably beyond.

Thank you!

33 Conclusion

⋆ Newton maps are dynamical systems, well motivated by root finding

⋆ are known to be very good for polishing up “approximate roots”, but globally chaotic, not generally convergent, have bad reputation

⋆ but are unique as root finders that have good theory and work well for degrees up to billions

⋆ Weierstrass and Ehrlich–Aberth have very good reputation, converge “always in practice”

⋆ but have no global theory whatsoever, and *(W.) not generally convergent either*

⋆ Moreover, Newton is a pioneering dynamical system: the only non-polynomial rational maps for which we have satisfactory theory! “Rational rigidity” works for Newton maps and probably beyond.

Thank you!