On the stochastic bifurcations of random holomorphic dynamical systems

Takayuki Watanabe*

January 14, 2023

Abstract

We give some numerical results on Random Relaxed Newton's Methods by Sumi. This randomized algorithm was proposed in his paper [Sumi21] to compute an approximate root of a given polynomial in one variable. He proved that the randomized algorithm almost surely works well if large noise is inserted in the original Newton methods. In this article, we try to confirm by numerical experiments how large the amplitude of the noise should be. The experiments demonstrate that even small noise can make the randomized algorithm successful. Based on these numerical results, we discusses the mathematical conjecture that the optimal noise amplitude is related to the bifurcation of the relaxed Newton's map.

Key words: random relaxed Newton method, random algorithm, stochastic bifurcation, the Mandelbrot set,

2020 Mathematics Subject Classification: 65C20 (Primary) 37F46 (Secondary)

1 Introduction

1.1 Background

In various fields of science, there is huge demand for finding the roots x^* of a given function P such that $P(x^*) = 0$. Mathematically, a polynomial $P(x) = a_d x^d + \cdots + a_1 x + a_0$ of degree d has d complex roots, but for practical purposes the approximated values are more valuable than the algebraic representation of the roots. Before 1600 B.C., the Babylonians had already calculated a very accurate approximation of $\sqrt{2}$, the positive root of $x^2 - 2$. See [FR]. This shows that approximate calculation is an important activity that we humans have been engaged in for a long time.

Newton's method is the most important example of root-finding algorithms. This iterative method, also known as the Newton-Raphson method, calculates the recurrence formula $x_{n+1} = x_n - P(x_n)/P'(x_n)$ starting from some initial guess x_0. It is known that the sequence $\{x_n\}_{n=0}^{\infty}$ converges to a root x^* under some conditions if x_0 is close to

*Faculty of Science, Kyoto University. Kitashirakawa Oiwake-cho, Sakyo-ku, 606-8502, Kyoto, Japan. E-mail: watanabe.takayuki.3m@kyoto-u.ac.jp

x^*. However, we cannot expect *global convergence* in the sense that $\{x_n\}_{n=0}^{\infty}$ does not converge to any root if the initial guess x_0 is apart from the roots. We further discuss this in Section 2.

It is natural to extend the range of roots to complex numbers from real numbers since we are working on polynomials as a class of target functions. Then the Newton method can be considered as dynamical system under the Newton map $N_P(z) = z - P(z)/P'(z)$ on the complex number plane $\mathbb{C}$ or the extended plane $\mathbb{C} \cup \{\infty\}$.

In [Sm85] Smale asked the question: when does there exist a purely iterative algorithm which is globally convergent for generic choice of a target polynomial of degree d? McMullen proved that there is NO generally convergent purely iterative algorithm if the degree d is 4 or more [Mc87]. See [MNTU] for details of the Newton method from the complex-analytic viewpoint.

Surprisingly, in his paper [Sumi21] Sumi showed that global convergence is typically valid if we *randomize* the original Newton method. More precisely, for example, choose λ randomly following the uniform distribution on a disk $\{\lambda \in \mathbb{C}\colon |\lambda - 1| \leq r\}$, and consider the relaxed Newton map $N_{P,\lambda}(z) = z - \lambda P(z)/P'(z)$. If $1/2 < r < 1$, then for all polynomial P, and for all initial point z_0 except finite points, almost every i.i.d. choice of $\lambda_1, \lambda_2, \ldots$ gives us the random orbit $z_n = N_{P,\lambda_n} \circ \cdots \circ N_{P,\lambda_2} \circ N_{P,\lambda_1}(z_0)$ which converges to some root z^* of P. That is, if large noise is inserted in the original Newton method, then for every polynomial, we can find a numerical root of it with probability one.

Note aside that Sumi's random algorithm follows the same philosophy as the famous algorithm known as stochastic gradient descent which is used to minimize an objective function. The latter algorithm replaces the actual gradient of the objective function with an estimate calculated from randomly sampled data. Randomness helps us to avoid local minima and find the global minimum. See [Bishop] for example.

In this article, we try to confirm by numerical experiments how large the amplitude of the noise should be. Sumi's theorem suggests to us that we should take the noise amplitude r larger than a magic number $1/2$. However, this may not be optimal. Namely, if we take a large r, then the speed of convergence gets slower. In order to find a better choice of noise amplitude, the author conducts several numerical experiments.

These experiments suggests that small amplitude, say $r = 0.01$, is sometimes enough to have global convergence. The author conjectures that this number is related to the bifurcation of the relaxed Newton maps $\{N_{P,\lambda}\}_{\lambda \in \mathbb{C}}$. Mathematical verification is not yet given, but we illustrate the reasons why the author think so.

2 Preliminaries

In this section we give the theory of deterministic iterations of rational maps. First, we consider a general rational map N, and then we discuss specific (relaxed) Newton maps $N_{P,\lambda}$. Second, we give a reason that it is suitable for us to take $P(z) = z^3 - 2z + 2$ as the target of the (randomized) Newton method. Third, we give a "bifurcation diagram" regarding the family of the relaxed Newton maps $\{N_{P,\lambda}\}_{\{|\lambda - 1| < 1\}}$ and give some numerical estimates about bifurcation.

Let N be a rational map which acts on the Riemann sphere $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. Then N can be written as the quotient of two coprime polynomials: $N(z) = a(z)/b(z)$. The degree of N is defined as the maximum of the degrees of these two polynomials: $\deg N = \max\{\deg a, \deg b\}$.

To study the dynamics of iterations of N, the critical points play a crucial role. Here, a critical point is the point where the derivative N' vanishes. The following lemma is well known. See [Mi] for the proofs.

Lemma 2.1. The following hold.

(i) A periodic point $z^\dagger \neq \infty$ of N with period p is attracting if and only if $|(N^{\circ p})'(z^\dagger)| < 1$. Here $N^{\circ p}$ is the p-th iterates of N.

(ii) For every attracting periodic point $z^\dagger$ of N with period p, the orbit of at least one critical point converges to $z^\dagger$.

(iii) The number of critical points of N is at most $2 \deg N - 2$.

(iv) If the orbit of every critical point converges to some attracting cycle, then for every z_0 in an open dense subset of $\mathbb{C}$, the orbit $N^{\circ n}(z_0)$ converges to some attracting cycle as $n \to \infty$.

Using the lemma above, we analyze the iteration of the relaxed Newton map $N_{P,\lambda}(z) = z - \lambda P(z)/P'(z)$ for a polynomial P.

Definition 2.2. A rational map $N_{P,\lambda}$ is *convergent* for P if for every z_0 in an open dense subset of $\mathbb{C}$, the orbit $N_{P,\lambda}^{\circ n}(z_0)$ converges to a root of P.

As we pointed out in Section 1, there is a polynomial P for which the original Newton map $N_{P,1}$ is not convergent. For this target, the Newton method may fail if one choose bad initial points.

In the following, we present an example of such a bad target P. We begin with the following lemma.

Lemma 2.3. The following hold.

(i) If $P(z^*) = 0$, then z^* is a fixed point of $N_{P,\lambda}$ for every $\lambda \in \mathbb{C}$.

(ii) Moreover, the derivative satisfies $N_{P,\lambda}'(z^*) = 1 - \lambda/m$, where m is the multiplicity of the root z^* of P.

(iii) If $|\lambda - 1| < 1$, then each of the roots z^* of P is an attracting fixed point of $N_{P,\lambda}$.

In particular, if $\lambda = 1$ and $m = 1$, then the root is at the same time a critical point and an attracting fixed point of $N_{P,1}$. In the following, we assume P has no multiple roots. Then we can show $\deg N_{P,\lambda} = \deg P$, and we have the "free" critical point if $\deg P \geq 3$, which sometimes interferes with the root-finding.

Corollary 2.4. Suppose P has exactly 3 roots, all of which are simple. Then $\deg N_{P,\lambda} = 3$ and $N_{P,\lambda}$ has three common attracting fixed points if $|\lambda - 1| < 1$. Moreover, these three are critical points of the original Newton map $N_{P,1}$. In addition to these three points, the map $N_{P,1}$ can have at most one more critical point.

This motivates us to study the following example.

Example 2.5. The polynomial $P(z) = z^3 - 2z + 2$ has exactly 3 simple roots. For this P, the map $N_{P,1}(z) = (2z^3 - 2)/(3z^2 - 2)$ has a critical point $c = 0$, which satisfies $N_{P,1}(0) = 1$ and $N_{P,1}(1) = 0$.

We now visualize this in Figure 1 and Figure 2. Every initial points is colored[1] according to where it converges to (or does not converge). Namely,

- the red region is an attracting basin of the real negative root ≈ -1.77,
- the purple region is an attracting basin of the complex root $\approx 0.88 + 0.59i$,
- the yellow region is an attracting basin of the complex root $\approx 0.88 - 0.59i$, and
- the black part illustrates the set of all initial points where the orbit starting from there will NOT converge to any of these three roots.

In other words, if we start the original Newton method from non-black initial points, then we can find some root of P. The black part illustrates the initial points on which the Newton algorithm fails. Actually, the black part is the union of the Julia set and the attracting basin of the cycle $c = 0 \mapsto 1 \mapsto 0$, where the Julia set means the common boundary of the four attracting basins.

We can prove the converse of Example 2.5. Suppose that the target polynomial Q has exactly 3 simple roots and $N_{Q,1}$ has a periodic critical point of period 2. Then Q should be $P(z) = z^3 - 2z + 2$ or its affine change of variable. For more comprehensive study of the case where the target P is degree 3, see [CGS].

Thus, the polynomial $P(z) = z^3 - 2z + 2$ is the most simplest example such that $N_{P,1}$ is NOT convergent for P. Hence, we will work on this polynomial as the target of the randomized Newton method.

We end this section by showing a bifurcation diagram, Figure 3.

3 Sumi's randomized algorithm

Sumi showed a theorem which states that if large noise is inserted in the original Newton method, then we can find a numerical root of a given polynomial with probability one. See [Sumi21, Theorem 4.4. (vii)].

First, we consider the following setting.

[1] This article will be printed in black and white, but the online version will be available in color.

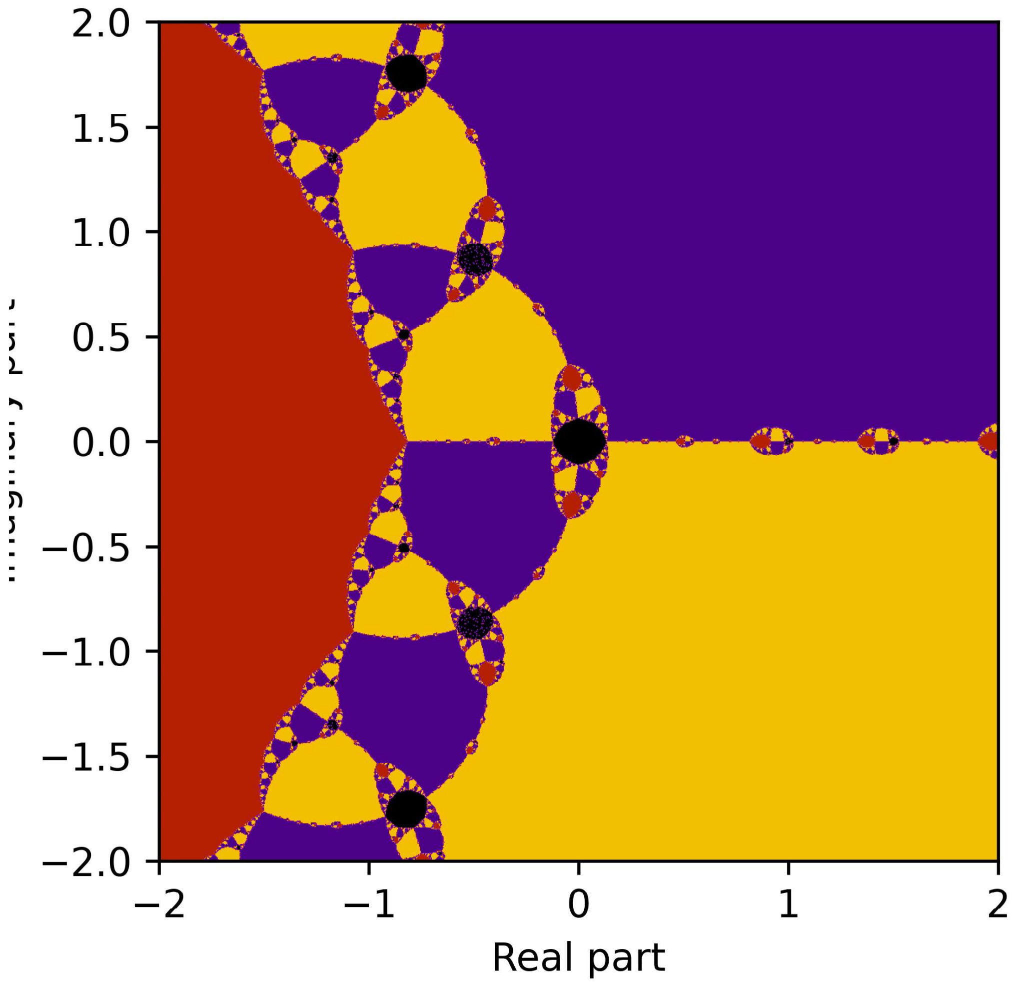

Figure 1: Attracting basins of $N_{P,1}$ for $P(z) = z^3 - 2z + 2$ of Example 2.5. The black part is the set of all initial points on which the Newton algorithm fails. Actually, the black part is the union of an attracting basin of the cycle $c = 0 \mapsto 1 \mapsto 0$ and the Julia set. We can see a black-colored open set near the origin which is too big to ignore.

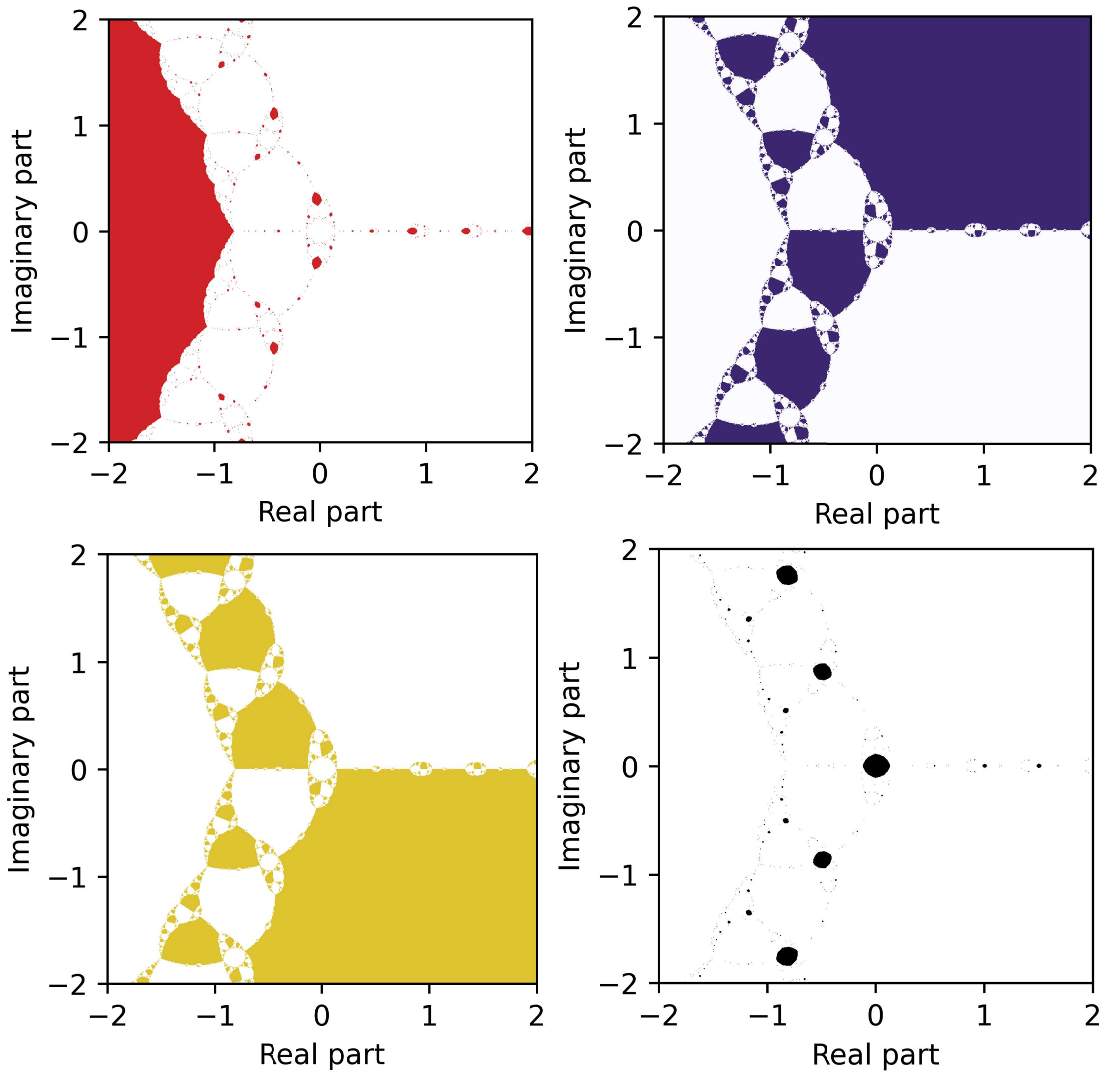

Figure 2: Decomposition of Figure 1 into four basins. The four have the same boundary, the Julia set. We can see that the upper right picture (purple) is complex conjugate of the lower left picture (yellow) since these are basins associated to the complex conjugate roots of the real polynomial.

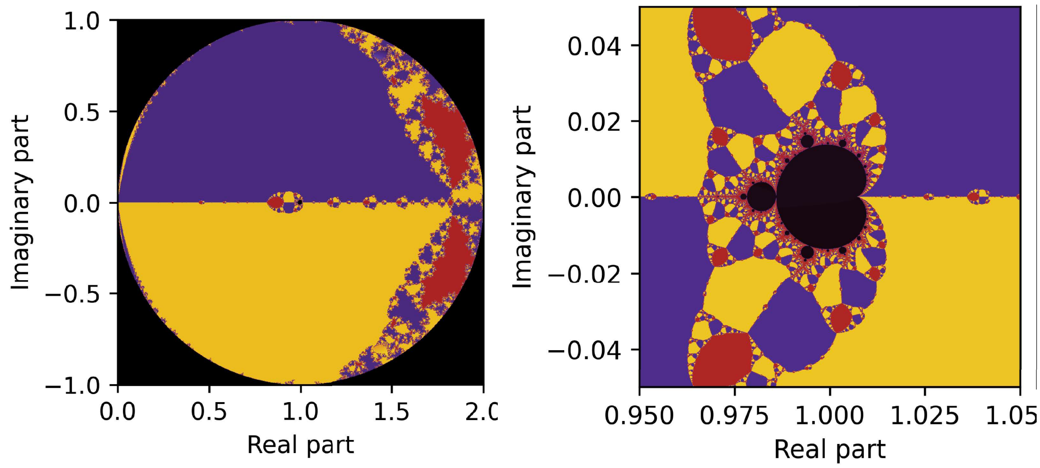

Figure 3: The bifurcation diagram of $\{N_{P,\lambda}\colon |\lambda - 1| < 1\}$ with $P(z) = z^3 - 2z + 2$. We divide the λ-plane into small pieces and pick a representative λ from each of them. For a fixed λ, we calculate the "critical" orbit $N_{P,\lambda}^{\circ n}(0)$ and we color the piece according to where it converges (or does not converge). The color corresponds to the color of attracting basins given in Example 2.5. The figure on the right is a zoomed-in view of the center $\lambda = 1$ of the left figure, on which we can see a small Mandelbrot set. The cusp of the main cardioid locates at $\lambda \approx 1.007$.

Setting 3.1. Let μ be a Borel probability measure on $\mathbb{C}$. We choose parameters $\lambda_1, \lambda_2, \dots$ independently following the measure μ and consider the random iterations $N_{P,\lambda_n} \circ \cdots \circ N_{P,\lambda_2} \circ N_{P,\lambda_1}$.

Theorem 3.2. Suppose that a probability measure μ is absolutely continuous with respect to 2-dimensional Lebesgue measure and

$$\{\lambda \in \mathbb{C}\colon |\lambda - 1| \leq \frac{1}{2}\} \subset \operatorname{int}(\operatorname{supp}\mu) \text{ and } \operatorname{supp}\mu \subset \{\lambda \in \mathbb{C}\colon |\lambda - 1| < 1\},$$

where $\operatorname{int}(S)$ denotes the interior of S with respect to the usual topology of $\mathbb{C}$.

Then for every target polynomial P of degree two or more, we have the following. Fix an initial point $z_0 \in \mathbb{C}$ with $P'(z_0) = 0$ and $P(z_0) \neq 0$. For almost every choice of a random sequence $(\lambda_1, \lambda_2, \dots)$, there exists a root z^* of P such that the orbit converges to it: $N_{P,\lambda_n} \circ \cdots \circ N_{P,\lambda_2} \circ N_{P,\lambda_1}(z_0) \to z^*$ as $n \to \infty$.

Note that the limit z^* can depend on the choice of $(\lambda_1, \lambda_2, \dots)$ even if the random orbit starts from the same initial point z_0.

The theorem gives us a random algorithm to find roots of a given polynomial.

Algorithm 3.3. Let P be a polynomial of degree two or more. Let μ_r be the normalized Lebesgue measure on the disk $\{\lambda \in \mathbb{C}\colon |\lambda - 1| \leq r\}$. Generate an i.i.d. sequence $\lambda_1, \lambda_2, \dots$ following μ_r. Then for a generic choice of initial point z_0, the orbit $N_{P,\lambda_n} \circ \cdots \circ N_{P,\lambda_2} \circ N_{P,\lambda_1}(z_0)$ approximates some root of P.

Theorem 3.2 states that the algorithm succeeds with probability one if $1/2 < r < 1$.

Example 3.4. Set $P(z) = z^3 - 2z + 2$ and $r = 0.6$. Then almost surely, the random orbit $N_{P,\lambda_n} \circ \cdots \circ N_{P,\lambda_2} \circ N_{P,\lambda_1}(0)$ converges to some root of P as $n \to \infty$.

The above example illustrates the significant difference from the deterministic case of Example 2.5. Numerical experiments show that 8 (resp. 492, 499) of the 1000 random orbits converge to the red (resp. purple, yellow) root, where the color refers to Example 2.5. See also Figure 4.

Remark 3.5. By Lemma 2.3(ii), we can deduce that the speed of convergence gets slower as r increases. Thus, it is important to find the smallest r such that Algorithm 3.3 succeeds with probability one. Theorem 3.2 implies that the smallest value is less that $1/2$ in general.

4 Results of Numerical Experiments

In this section, we verify Algorithm 3.3 for small noise amplitude r by numerical experiments. We are interested in whether or not there is an attractor near the origin. We draw pictures like Figure 4 for various r for the fixed target $P(z) = z^3 - 2z + 2$. By seeing the black part, we can determine whether the algorithm typically successes or fails.

The result is shown in Figure 5, which suggests that very small r, say 0.01, is enough to have the typical success. If $r = 0.01$, then we can expect very fast convergence since the derivative of $N_{P,\lambda}$ is at most 0.01 near the roots of P. Besides, if $r = 0.005$, then the algorithm fails when the initial point z_0 is chosen near the origin.

It seems that there exists a threshold value r_* which satisfies that the bad attractor near the origin persists until $r \leq r_*$ and it vanishes if $r > r_*$. The numerical experiments suggests that $0.005 \leq r_* < 0.01$.

The disappearance of the attractor by noise has already been observed in the author's paper [W22]. In the paper, he gives some tools to estimate the threshold (bifurcation) parameter values. The most fundamental tool is the observation that "the stochastic bifurcation occurs before the deterministic bifurcation." See [W22, Lemma 4.9].

For our situation of the random relaxed Newton method, this observation seems to be still valid. Namely, the stochastic bifurcation occurs at $r_* \in [0.005, 0.01]$, while the deterministic bifurcation occurs at $r \approx 0.007$ as one can see in Figure 3. Here, the latter value is coming from the small Mandelbrot set, in particular the cusp of the main cardioid.

Another numerical experiment suggests that the stochastic bifurcation parameter r_* satisfies $r_* < 0.006$. Namely, let the noise amplitude $r = 0.006$ and the target $P(z) = z^3 - 2z + 2$. For the initial point $z_0 = 0$, we calculate 1000 random orbits $N_{P,\lambda_n} \circ \cdots \circ N_{P,\lambda_2} \circ N_{P,\lambda_1}(0)$ with the maximum iterations $n \leq 10^6$. The result is that 2 out of 1000 orbits converge to the root $\approx 0.88 + 0.59i$, the purple one. This shows that the bad attractor near the origin vanishes when $r = 0.006$. The author think that if the number

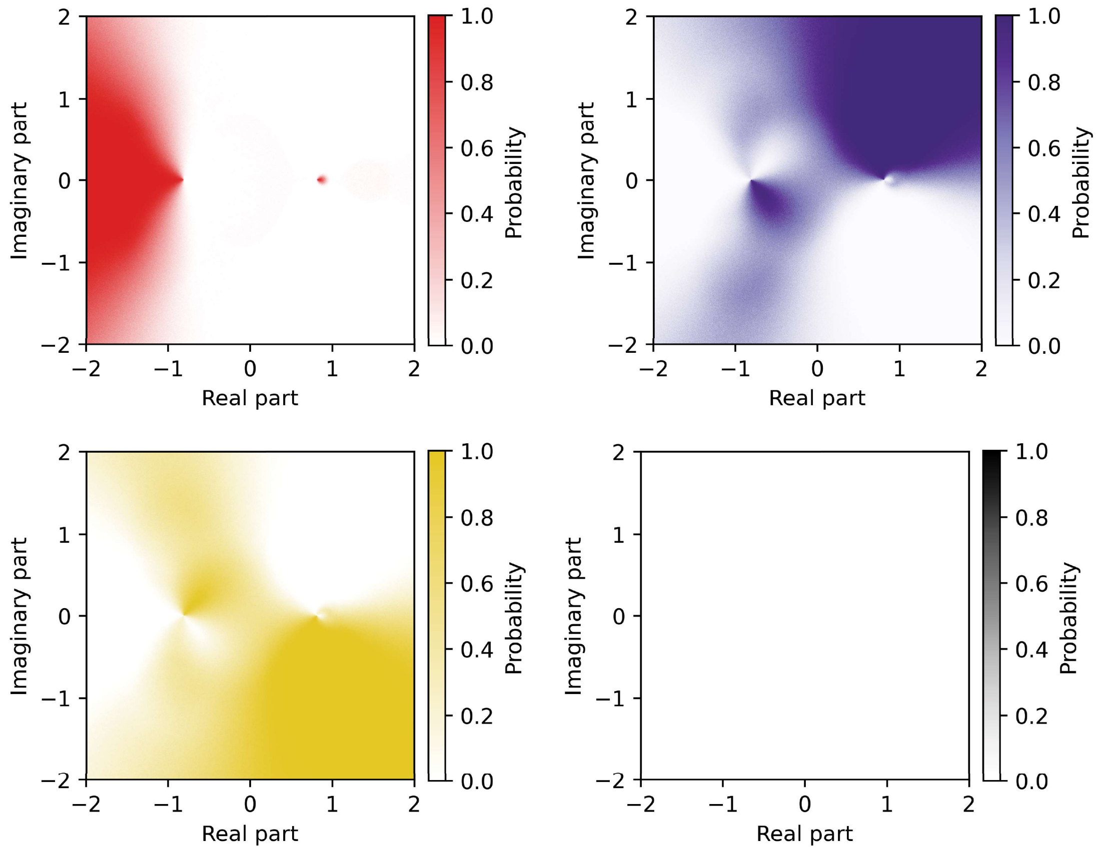

Figure 4: Each of these four figures illustrates the probability of random orbits tending to the corresponding attracting fixed point. Compare these with Figure 2. In detail, we set $P(z) = z^3 - 2z + 2$ and set $r = 0.6$ as in Example 3.4 . We divide the z-plane into small pieces and pick a representative z_0 from each of them. For each z_0, we calculate 100 random orbits $N_{P,\lambda_n} \circ \cdots \circ N_{P,\lambda_2} \circ N_{P,\lambda_1}(z_0)$ with $n \leq 10^3$, and we count the number of orbits which converge to each of the three roots (or do not converge to any of them). The color corresponds to the color of roots given in Example 2.5. In the lower right figure, the bad attractor where the root-finding algorithm fails vanishes due to noise!

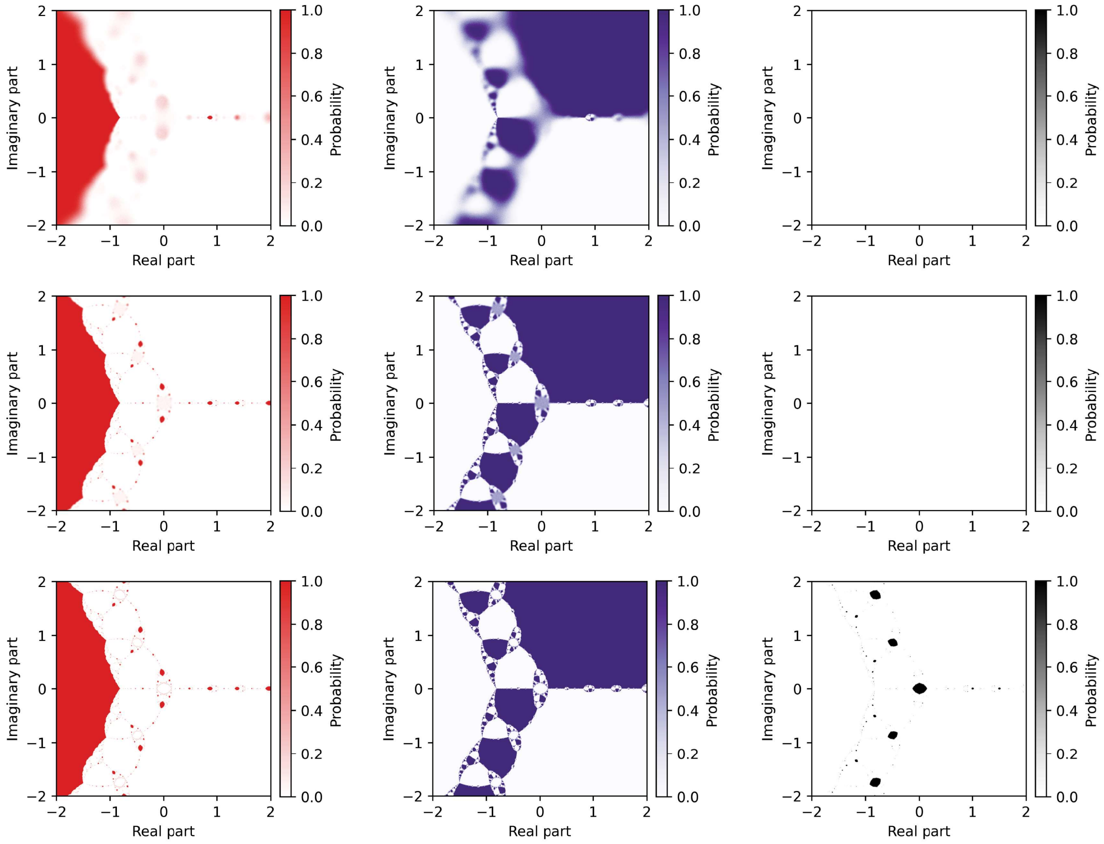

Figure 5: Three horizontal rows correspond to the cases where $r = 0.1$, 0.01 and 0.005, respectively. Here we omit the yellow basins because they are the complex conjugate of the purple ones. This experiment shows that successes and failures vary in the range of $r \in [0.005, 0.01]$; the bad attractor near the origin vanishes when $r = 0.01$, but it does not vanish when $r = 0.005$. The number of (random) iterations is at most 10^3.

of iterations were increased, then nearly half of the random orbits would be expected to converge to the purple root.

5 Another variation of the algorithm

In this section, we verify a variant of Algorithm 3.3, which is known as simulated annealing. In the following algorithm, we consider non-i.i.d. sequences $\lambda_1, \lambda_2, \ldots$.

Algorithm 5.1. Let P be a polynomial of degree two or more. Let μ_r be the normalized Lebesgue measure on the disk $\{\lambda \in \mathbb{C} \colon |\lambda - 1| \leq r\}$ and take $r_n > 0$ suitably for every $n \in \mathbb{N}$. Generate a random number λ_n following μ_{r_n} for every $n \in \mathbb{N}$. Then for a generic choice of initial point z_0, the orbit $N_{P,\lambda_n} \circ \cdots \circ N_{P,\lambda_2} \circ N_{P,\lambda_1}(z_0)$ may approximates some root of P.

Generally, we take the noise amplitudes r_n to converge to 0. This may allow us to have both the typical global convergence and the local fast convergence at the same time. We verify this in Figure 6, in which we take $r_n = r^n$ for some r. Until now, there is no rigorous proof that Algorithm 5.1 works well. We cannot give it here, so we just test the algorithm numerically.

6 Conclusion and Discussions

We tested whether Algorithm 3.3 works well for small noise amplitude r. The results show that Algorithm 3.3 successfully works for almost every initial point when $r \geq 0.01$. Also, there exists a non-empty open set B such that Algorithm 3.3 does not work if the initial point is chosen from B when $r \leq 0.005$. Another experiment suggests that the threshold value r_* is between 0.005 and 0.006 if we allow a very large number of iterations.

The value r_* is much smaller than the theoretical value given in Theorem 3.2. This is meaningful because the speed of (local) convergence is very fast if the noise amplitude r is very small. The author numerically verified that r_* is small than the distance from $\lambda = 1$ to the boundary of the small Mandelbrot set shown in Figure 3. This is closely related to the results of [W22], in which the author gives quantitative estimates of the stochastic bifurcation parameters.

These study gives us the following way to find the optimal noise amplitude. For every polynomial P, consider $C_P = \{\lambda \in \mathbb{C} \colon N_{P,\lambda} \text{ is convergent for } P\}$ and define $r_{\mathrm{d}}(P)$ as the distance from $\lambda = 1$ to C_P, which may be 0. Then $\tilde{r}_{\mathrm{d}} := \sup_P r_{\mathrm{d}}(P)$ gives a universal noise amplitude with which the random relaxed Newton method work successfully for a generic choice of initial point. That is, if $r > \tilde{r}_{\mathrm{d}}$, then Algorithm 3.3 works well with probability one.

Calculating $\tilde{r}_{\mathrm{d}}$ remains a future challenge.

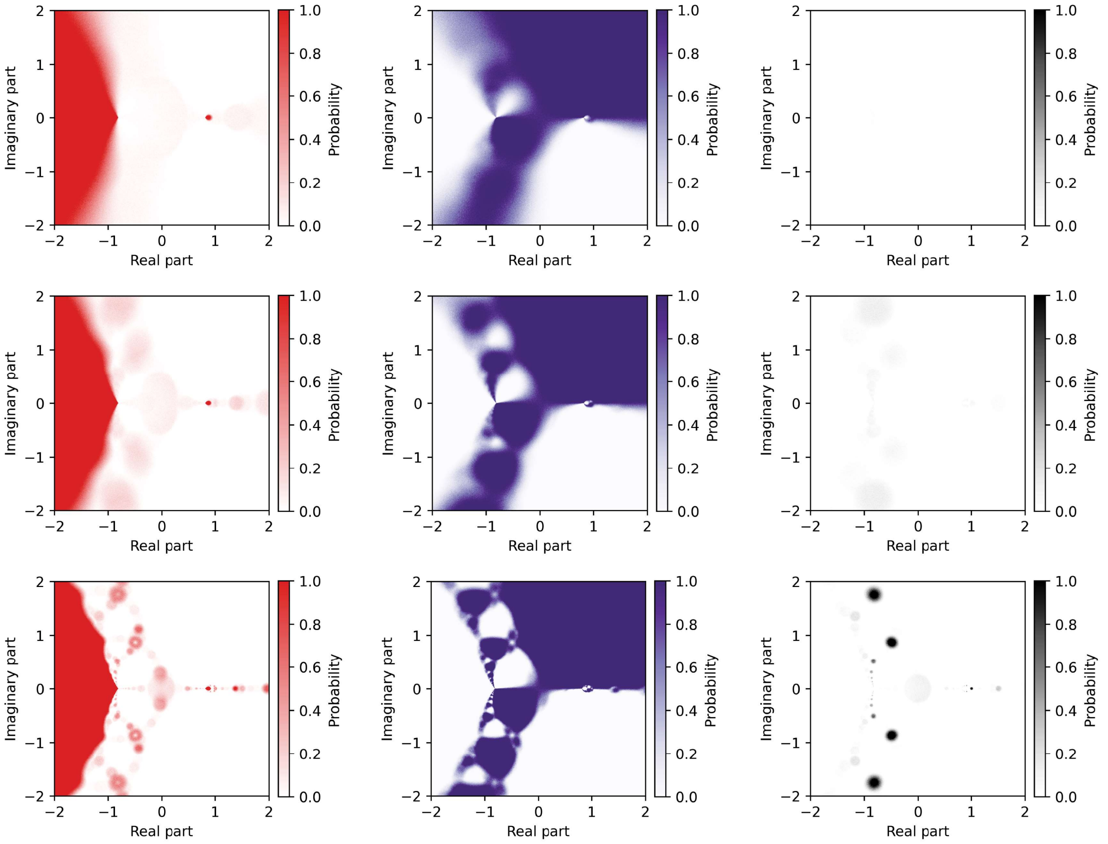

Figure 6: We test Algorithm 5.1 for $P(z) = z^3 - 2z + 2$. Here we take the noise amplitude r_n such that there exists r such that $r_n = r^n$ for every $n \in \mathbb{N}$. Three horizontal rows correspond to the cases where $r = 0.56$, 0.31 and 0.1, respectively. This experiment shows that the algorithm successfully works when $r = 0.56$. Also, the algorithm works well with a high probability when $r = 0.31$. An interesting point is that when $r = 0.1$, the computation starting from the initial point 0 works well with high probability, but the computation starting from other points does not work so well. This suggests that the noise decreases too quickly and hence global convergence does not hold when $r = 0.1$.

References

[Bishop] Bishop, Christopher M.: Pattern recognition and machine learning. Information Science and Statistics. Springer, New York, 2006.

[CGS] Curry, James H.; Garnett, Lucy; Sullivan, Dennis: On the iteration of a rational function: computer experiments with Newton's method. Comm. Math. Phys. 91 (1983), no. 2, 267–277.

[FR] Fowler, David; Robson, Eleanor: Square root approximations in old Babylonian mathematics: YBC 7289 in context. Historia Math. 25 (1998), no. 4, 366–378.

[Mc87] McMullen, Curt: Families of rational maps and iterative root-finding algorithms. Ann. of Math. (2) 125 (1987), no. 3, 467–493.

[MNTU] Morosawa, S.; Nishimura, Y.; Taniguchi, M.; Ueda, T.: Holomorphic dynamics. Cambridge Studies in Advanced Mathematics, 66. Cambridge University Press, Cambridge, 2000.

[Mi] Milnor, John: Dynamics in one complex variable, third edition. Annals of Mathematical Studies. 160, Princeton University Press, 2006.

[Sm85] Smale, Steve: On the efficiency of algorithms of analysis. Bull. Amer. Math. Soc. (N.S.) 13 (1985), no. 2, 87–121.

[Sumi21] Sumi, Hiroki: Negativity of Lyapunov exponents and convergence of generic random polynomial dynamical systems and random relaxed Newton's methods. Comm. Math. Phys. 384 (2021), no. 3, 1513–1583.

[W22] Watanabe, Takayuki: On the stochastic bifurcations regarding random iterations of polynomials of the form $z^2 + c_n$, preprint, arXiv:2206.06702 [math.DS].