

BASINS OF ATTRACTION IN NEURAL NETWORK TRAINING: A JULIA–FATOU TOPOLOGICAL VIEW OF LOSS LANDSCAPES

SAMUEL BUSAYO OLUWAFEMI^{1*} AND MOSES OBINNA FRANCIS²

ABSTRACT. We study gradient-based training in deep learning from a complex-dynamical perspective. Under a local analytic continuation assumption, the training update is modeled as iteration of a holomorphic map, which organizes the loss landscape into stable regions and sensitive boundary sets. Using fixed-point stability ideas, we derive explicit local criteria distinguishing attraction from repulsion and obtain a linear convergence rate inside attracting neighborhoods. We also extend the local stability analysis to higher-dimensional complex parameter spaces via a spectral condition. At the global level, we prove three complementary results: an escape-radius condition that forces divergence for polynomial-gradient surrogates, persistence of attracting fixed points under small learning-rate perturbations, and an instability principle for rational maps based on the density of repelling periodic points on the Julia set. These results link step-size sensitivity and initialization dependence to basin geometry and support visualization-driven diagnostics for stability in optimization.

Keywords. Deep learning, loss landscape, holomorphic iteration, Julia set, Fatou set

2020 Mathematics Subject Classification. Primary 37F10; Secondary 68T07

1. INTRODUCTION

Deep neural networks achieve strong empirical performance largely through iterative gradient-based training [5, 1]. Given data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ and parameters $\theta \in \mathbb{R}^d$, training is often posed as

$$\min_{\theta \in \mathbb{R}^d} \mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^N \ell(f_{\theta}(x_i), y_i), \quad (1.1)$$

and implemented by gradient descent or SGD,

$$\theta_{k+1} = \theta_k - \eta_k \nabla \mathcal{L}(\theta_k), \quad (1.2)$$

where $\eta_k > 0$ is the learning rate [1, 12]. Despite its simplicity, (1.2) operates on typically non-convex, high-dimensional loss landscapes whose basin structure

Date: Received: Jan 5, 2026; Revised: Feb 3, 2026; Accepted: Feb 15, 2026.

* Corresponding author: Samuel Busayo Oluwafemi

© The Author(s) 2025. This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of the licence, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

and sensitivity to initialization and step size are central to training behavior [8, 11, 6]. We adopt a dynamical-systems viewpoint: for fixed $\eta > 0$, (1.2) defines an iteration map $\Phi(\theta) = \theta - \eta \nabla \mathcal{L}(\theta)$, so training trajectories are orbits of Φ [6]. Under a local analytic continuation (complexification) assumption, we model the update as iteration of a holomorphic map. In the scalar case we extend the parameter to $z \in \mathbb{C}$ and study

$$z_{n+1} = f_\eta(z_n) = z_n - \eta L'(z_n), \quad z_n \in \Omega \subset \mathbb{C}, \quad (1.3)$$

where L is holomorphic on Ω . This connects training dynamics to one-variable complex iteration, where stability regions are organized by the Fatou set and sensitive boundaries by the Julia set [10, 3]. While practical training evolves on \mathbb{R}^d , when L and η are real on the real slice, this slice is forward invariant; thus the complexified picture serves as a rigorous envelope for basin topology, with real instability governed by proximity to basin boundaries. The holomorphic assumption is exact for losses built from analytic primitives and smooth activations (e.g. tanh, sigmoid, softplus, GELU), and applies locally on smooth regions of piecewise-smooth architectures or smooth surrogates.

Our contributions include (i) explicit local attraction/repulsion and linear-rate criteria via the fixed-point multiplier, (ii) a basin-boundary theorem placing attracting basins in the Fatou set and their boundaries in the Julia set, and (iii) a local spectral attraction condition in \mathbb{C}^d . Several local statements recover familiar real step-size restrictions, while the global gain is a topological description of basin geometry and sensitivity.

Potential applications in computer science. The basin viewpoint suggests practical tools for ML systems: stability-oriented diagnostics for learning-rate tuning and initialization, automatic detection of sensitive boundary regions during training, and divergence guardrails for surrogate polynomial-gradient models. These ideas align with stability and sensitivity needs in modern pipelines such as feature-attribution auditing and related AMCS topics [13, 14, 15].

2. PRELIMINARIES

Definition 2.1. Let $\Omega \subset \mathbb{C}$ be an open set. A function $f : \Omega \rightarrow \mathbb{C}$ is holomorphic on Ω if for every $z_0 \in \Omega$ the complex derivative

$$f'(z_0) = \lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists.

Definition 2.2. Let $g : U \rightarrow \mathbb{R}$ be a real-valued function on an open set $U \subset \mathbb{R}$. A function $G : \Omega \rightarrow \mathbb{C}$, defined on an open set $\Omega \subset \mathbb{C}$ with $U \subset \Omega \cap \mathbb{R}$, is called a local analytic continuation of g if G is holomorphic on Ω and $G(x) = g(x)$ for all $x \in U$.

Definition 2.3. The n -th iterate of f is defined inductively by

$$f^{\circ 0}(z) = z, \quad f^{\circ(n+1)}(z) = f(f^{\circ n}(z)), \quad n \geq 0.$$

Definition 2.4. Given $z_0 \in \Omega$, the forward orbit of z_0 under f is

$$\mathcal{O}^+(z_0) = \{z_0, f(z_0), f^{\circ 2}(z_0), \dots\}.$$

Definition 2.5. The ω -limit set of z_0 is the set of all accumulation points of its orbit:

$$\omega(z_0) = \left\{ w \in \Omega : \exists n_k \rightarrow \infty \text{ with } f^{\circ n_k}(z_0) \rightarrow w \right\}.$$

Definition 2.6. A point $z^* \in \Omega$ is a fixed point of f if $f(z^*) = z^*$.

Definition 2.7. A point $z^* \in \Omega$ is periodic of period $p \geq 1$ if $f^{\circ p}(z^*) = z^*$. The smallest such p is called the prime period.

Definition 2.8. If z^* is a fixed point of a holomorphic map f , the multiplier of f at z^* is

$$\lambda(z^*) = f'(z^*).$$

If z^* is periodic of period p , its multiplier is $(f^{\circ p})'(z^*)$.

Definition 2.9. Let z^* be a fixed point of a holomorphic map f with multiplier $\lambda = f'(z^*)$.

- (1) z^* is attracting if $|\lambda| < 1$.
- (2) z^* is repelling if $|\lambda| > 1$.
- (3) z^* is neutral if $|\lambda| = 1$.

Definition 2.10. If z^* is an attracting fixed point of f , its basin of attraction is

$$\mathcal{B}(z^*) = \{z \in \Omega : f^{\circ n}(z) \rightarrow z^* \text{ as } n \rightarrow \infty\}.$$

Definition 2.11. A set $A \subset \Omega$ is forward invariant if $f(A) \subseteq A$. It is backward invariant if $f^{-1}(A) \subseteq A$. It is completely invariant if $f(A) = A = f^{-1}(A)$.

Definition 2.12. Let $U \subset \mathbb{C}$ be open and let \mathcal{F} be a family of holomorphic functions on U . We say \mathcal{F} is normal on U if every sequence in \mathcal{F} has a subsequence that converges uniformly on compact subsets of U (possibly to ∞).

Definition 2.13. The Fatou set of f is

$$F(f) = \left\{ z \in \Omega : \{f^{\circ n}\}_{n \geq 0} \text{ is normal in a neighborhood of } z \right\}.$$

Definition 2.14. The Julia set of f is the complement of the Fatou set:

$$J(f) = \Omega \setminus F(f).$$

Definition 2.15. Let L be holomorphic on $\Omega \subset \mathbb{C}$ and let $\eta > 0$. The associated holomorphic training map is

$$f_\eta(z) = z - \eta L'(z), \quad z \in \Omega. \quad (2.1)$$

Definition 2.16. Let L be holomorphic on $\Omega \subset \mathbb{C}^d$ with holomorphic gradient ∇L . The associated map is

$$f_\eta(z) = z - \eta \nabla L(z), \quad z \in \Omega.$$

3. MAIN RESULTS

In this section we establish analytic and topological results for the complexified training iteration

$$z_{n+1} = f_\eta(z_n) = z_n - \eta L'(z_n), \quad z_n \in \Omega \subset \mathbb{C}, \quad (3.1)$$

where $\eta > 0$ is the learning rate and L is holomorphic on Ω . Several local statements below recover familiar real-valued step-size restrictions (e.g. $\eta < 2/L''$ in the scalar convex case), serving as a consistency check with standard numerical analysis. The main added value is the global, topological interpretation: basins of attraction are Fatou components, while sensitive basin boundaries lie in the Julia set, providing a rigorous vocabulary for initialization sensitivity and step-size fragility. Throughout, note that if $L(x) \in \mathbb{R}$ for $x \in \Omega \cap \mathbb{R}$ and $\eta \in \mathbb{R}$, then f_η maps the real slice into itself, so standard gradient descent on \mathbb{R} is the restriction of (3.1) to $\Omega \cap \mathbb{R}$.

Lemma 3.1. *Assume L is holomorphic on an open set $\Omega \subset \mathbb{C}^d$ and satisfies the conjugation symmetry*

$$L(\bar{z}) = \overline{L(z)} \quad \text{whenever } z, \bar{z} \in \Omega.$$

Let $\eta \in \mathbb{R}$. Then the associated training map

$$f_\eta(z) = z - \eta \nabla L(z)$$

leaves the real slice forward invariant:

$$f_\eta(\Omega \cap \mathbb{R}^d) \subseteq \Omega \cap \mathbb{R}^d.$$

Proof. Fix $x \in \Omega \cap \mathbb{R}^d$. For each coordinate j , define $\psi_j(t) = L(x + te_j)$ for real t near 0. By the symmetry assumption, $\psi_j(t) \in \mathbb{R}$ for real t , hence $\psi_j'(0) \in \mathbb{R}$. But $\psi_j'(0) = \partial_{z_j} L(x)$ (holomorphic directional derivative), so $\nabla L(x) \in \mathbb{R}^d$. With $\eta \in \mathbb{R}$ we obtain $f_\eta(x) = x - \eta \nabla L(x) \in \mathbb{R}^d$. \square

Lemma 3.2. *A point $z^* \in \Omega$ satisfies $L'(z^*) = 0$ if and only if z^* is a fixed point of f_η , i.e. $f_\eta(z^*) = z^*$.*

Proof. From (3.1), $f_\eta(z^*) = z^*$ holds exactly when $z^* - \eta L'(z^*) = z^*$, i.e. $L'(z^*) = 0$. \square

Lemma 3.3. *Let z^* be a fixed point of f_η . Then the multiplier at z^* is*

$$\lambda(z^*) = f'_\eta(z^*) = 1 - \eta L''(z^*).$$

Proof. Differentiate $f_\eta(z) = z - \eta L'(z)$ to obtain $f'_\eta(z) = 1 - \eta L''(z)$ and evaluate at z^* . \square

Theorem 3.4. *Let $z^* \in \Omega$ satisfy $L'(z^*) = 0$. If*

$$|1 - \eta L''(z^*)| < 1, \quad (3.2)$$

then z^ is an attracting fixed point of f_η and there exists $r > 0$ such that for every z_0 with $|z_0 - z^*| < r$,*

$$\lim_{n \rightarrow \infty} f_\eta^{\circ n}(z_0) = z^*.$$

Proof. By Lemma 3.2, z^* is a fixed point. By Lemma 3.3, the multiplier is $\lambda = f'_\eta(z^*) = 1 - \eta L''(z^*)$, and (3.2) implies $|\lambda| < 1$. By continuity of f'_η , choose $r > 0$ and $q \in (0, 1)$ such that $\sup_{|z-z^*|<r} |f'_\eta(z)| \leq q$. Then for all z, w in the disk $D(z^*, r)$,

$$|f_\eta(z) - f_\eta(w)| \leq q|z - w|,$$

so f_η is a contraction on $D(z^*, r)$. Banach's fixed-point theorem yields convergence of every orbit starting in $D(z^*, r)$ to the unique fixed point in the disk, which is z^* . \square

Remark 3.5. If $z^* \in \mathbb{R}$ and $L''(z^*) > 0$ is real, then (3.2) is equivalent to $0 < \eta < 2/L''(z^*)$. More generally, if $L''(z^*) = a + ib$ with $a = \Re L''(z^*)$ and $b = \Im L''(z^*)$, then for real $\eta > 0$,

$$|1 - \eta L''(z^*)| < 1 \iff 0 < \eta < \frac{2 \Re L''(z^*)}{|L''(z^*)|^2},$$

which requires $\Re L''(z^*) > 0$. This makes explicit how the complex multiplier condition reduces to the usual scalar restriction in the real case.

Corollary 3.6. *Let z^* be a fixed point of f_η with multiplier $\lambda = f'_\eta(z^*) = 1 - \eta L''(z^*)$.*

- (1) *If $|\lambda| > 1$, then z^* is repelling and unstable: for any neighborhood U of z^* there exist $z_0 \in U$ such that $f_\eta^{\circ n}(z_0)$ does not converge to z^* .*
- (2) *If $|\lambda| = 1$, then z^* is neutral. In particular, z^* is not a contracting fixed point, so convergence of nearby orbits is not guaranteed. Moreover, arbitrarily small perturbations of the learning rate can change the stability type.*

Proof. We argue by a first-order expansion of a holomorphic map near a fixed point. Since f_η is holomorphic and $f_\eta(z^*) = z^*$, for w near 0 we can write

$$f_\eta(z^* + w) - z^* = \lambda w + r(w), \quad \text{where } \frac{r(w)}{w} \rightarrow 0 \text{ as } w \rightarrow 0.$$

Equivalently, for every $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $0 < |w| < \delta$,

$$|r(w)| \leq \varepsilon|w|. \tag{3.3}$$

(1) *Repelling case $|\lambda| > 1$.* Choose $\varepsilon > 0$ so that $|\lambda| - \varepsilon > 1$, and take $\delta > 0$ such that (3.3) holds. For $z = z^* + w$ with $0 < |w| < \delta$,

$$|f_\eta(z) - z^*| = |\lambda w + r(w)| \geq |\lambda||w| - |r(w)| \geq (|\lambda| - \varepsilon)|w| > |w|.$$

Thus, as long as an orbit stays in $D(z^*, \delta)$, its distance to z^* strictly increases by at least the factor $|\lambda| - \varepsilon > 1$, so it cannot converge to z^* . Since every neighborhood U of z^* contains points $z_0 = z^* + w$ with $0 < |w| < \delta$, this proves instability: there exists $z_0 \in U$ whose iterates do not converge to z^* .

(2) *Neutral case $|\lambda| = 1$.* We first show that z^* is not a contraction point. Fix any $q \in (0, 1)$. If f_η were a contraction on some disk $D(z^*, \rho)$ with Lipschitz constant q , then (by differentiability) we would have $|f'_\eta(z^*)| \leq q < 1$, contradicting

$|f'_\eta(z^*)| = |\lambda| = 1$. Hence there is no neighborhood of z^* on which f_η is a contraction, and therefore convergence of nearby orbits is not guaranteed in general. For the perturbation claim, note that the multiplier depends continuously on η :

$$\lambda(\eta) = 1 - \eta L''(z^*).$$

If $|\lambda(\eta_0)| = 1$ and $L''(z^*) \neq 0$, then for η arbitrarily close to η_0 one can have $|\lambda(\eta)| \neq 1$; in particular there exist η as close as desired to η_0 with $|\lambda(\eta)| > 1$ (repelling) or $|\lambda(\eta)| < 1$ (attracting), depending on the direction of perturbation. Thus the stability type can change under arbitrarily small learning-rate perturbations at the neutral threshold. \square

Theorem 3.7. *Assume the hypotheses of Theorem 3.4. Then there exist $r > 0$ and $q \in (0, 1)$ such that for all $z_0 \in D(z^*, r)$,*

$$|z_n - z^*| \leq q^n |z_0 - z^*|, \quad z_n = f_\eta^{\circ n}(z_0).$$

Proof. As in the proof of Theorem 3.4, choose $r > 0$ and $q \in (0, 1)$ such that

$$\sup_{|z - z^*| < r} |f'_\eta(z)| \leq q.$$

We have

$$|f_\eta(z) - f_\eta(z^*)| \leq q|z - z^*| \quad \text{for all } z \in D(z^*, r).$$

In particular, if $z_n \in D(z^*, r)$ then $|z_{n+1} - z^*| \leq q|z_n - z^*| < r$, so $z_{n+1} \in D(z^*, r)$. Thus $D(z^*, r)$ is forward invariant and the inequality applies at every iterate. Iterating gives

$$|z_n - z^*| \leq q^n |z_0 - z^*|.$$

\square

Theorem 3.8. *Let z^* satisfy $L'(z^*) = 0$ and set $\lambda(\eta) = 1 - \eta L''(z^*)$.*

- (1) *If $|\lambda(\eta)| < 1$, then z^* is attracting.*
- (2) *If $|\lambda(\eta)| > 1$, then z^* is repelling.*
- (3) *The stability type can change only when $|\lambda(\eta)| = 1$.*

Proof. Items (1) and (2) follow from Theorem 3.4 and Corollary 3.6. Item (3) holds because the stability classification depends only on whether $|\lambda(\eta)|$ is less than, equal to, or greater than 1. \square

Proposition 3.9. *If z^* is an attracting fixed point of f_η , then its basin of attraction*

$$\mathcal{B}(z^*) = \{z \in \Omega : f_\eta^{\circ n}(z) \rightarrow z^*\}$$

is nonempty, open, and forward invariant: $f_\eta(\mathcal{B}(z^)) \subseteq \mathcal{B}(z^*)$.*

Proof. Since z^* is attracting, there exists $r > 0$ such that $D(z^*, r) \subset \mathcal{B}(z^*)$, hence $\mathcal{B}(z^*)$ is nonempty. Openness follows because attraction persists under small perturbations of the initial condition inside a contracting neighborhood. Forward invariance follows from $f_\eta^{\circ n}(f_\eta(z)) = f_\eta^{\circ(n+1)}(z)$. \square

Theorem 3.10. *Basins lie in the Fatou set and their boundaries lie in the Julia set. Assume f_η is holomorphic on $\Omega \subset \mathbb{C}$. If z^* is an attracting fixed point, then*

$$\mathcal{B}(z^*) \subseteq F(f_\eta) \quad \text{and} \quad \partial\mathcal{B}(z^*) \subseteq J(f_\eta).$$

Proof. On $\mathcal{B}(z^*)$ the iterates $f_\eta^{\circ n}$ converge locally uniformly to the constant map z^* , hence form a normal family. Therefore $\mathcal{B}(z^*) \subseteq F(f_\eta)$ by definition of the Fatou set. If $z \in \partial\mathcal{B}(z^*)$, then every neighborhood of z contains points that converge to z^* and points that do not, so the family $\{f_\eta^{\circ n}\}$ cannot be normal in any neighborhood of z . Thus $z \notin F(f_\eta)$ and $z \in J(f_\eta)$. \square

When $\Omega \cap \mathbb{R}$ is forward invariant, Theorem 3.10 implies that boundaries separating real basins of attraction are contained in $J(f_\eta) \cap \mathbb{R}$. In practice, real training trajectories sample this real slice of the complex basin geometry; sensitivity arises when initializations lie close to the basin boundary within that slice. In several complex variables (\mathbb{C}^d with $d \geq 2$), global Julia–Fatou theory is substantially more subtle than in one variable. Accordingly, we use the \mathbb{C}^d extension below primarily to capture local attraction near critical points via a spectral condition.

For $z \in \mathbb{C}^d$, consider the complexified update

$$z_{n+1} = f_\eta(z_n) = z_n - \eta \nabla L(z_n), \quad (3.4)$$

with Jacobian $Df_\eta(z) = I - \eta \nabla^2 L(z)$.

Theorem 3.11. *Let z^* satisfy $\nabla L(z^*) = 0$ and assume $\nabla^2 L(z^*)$ exists. If*

$$\rho(I - \eta \nabla^2 L(z^*)) < 1,$$

then z^ is locally attracting for the iteration (3.4).*

Proof. If the spectral radius is < 1 , there exists an equivalent norm on \mathbb{C}^d for which the induced operator norm satisfies $\|Df_\eta(z^*)\| < 1$. By continuity of Df_η , the contraction persists on a sufficiently small neighborhood of z^* . Banach’s fixed-point theorem then yields local attraction of z^* for the iteration (3.4). \square

If $\nabla^2 L(z^*)$ is real symmetric with eigenvalues $\{\mu_i\}$, then $\rho(I - \eta \nabla^2 L(z^*)) < 1$ is equivalent to $|1 - \eta \mu_i| < 1$ for all i , which yields the familiar restriction $0 < \eta < 2/\mu_{\max}$ when $\mu_i > 0$.

Example 3.12. Let $L(z) = \frac{1}{2}z^2$. Then $f_\eta(z) = (1 - \eta)z$. The unique fixed point is 0, attracting if and only if $0 < \eta < 2$. For $0 < \eta < 2$, $\mathcal{B}(0) = \mathbb{C}$.

Example 3.13. Let $L(z) = \frac{1}{4}(z^2 - 1)^2$, so $L'(z) = z^3 - z$ and $f_\eta(z) = (1 + \eta)z - \eta z^3$. Fixed points are $\{-1, 0, 1\}$, where 0 is repelling for all $\eta > 0$, and ± 1 are attracting if and only if $0 < \eta < 1$. Thus, for $0 < \eta < 1$ there exist at least two distinct basins $\mathcal{B}(1)$ and $\mathcal{B}(-1)$ separated by a boundary contained in $J(f_\eta)$.

Theorem 3.14. *Assume $L'(z)$ is a polynomial of degree $m \geq 2$ with leading term $a_m z^m$ where $a_m \neq 0$, and define $f_\eta(z) = z - \eta L'(z)$ for $\eta > 0$. Then for each fixed $\eta > 0$ there exists $R = R(\eta) > 0$ such that for all $z \in \mathbb{C}$ with $|z| > R$ one has*

$$|f_\eta(z)| > |z|.$$

In particular, for every z_0 with $|z_0| > R$,

$$|f_\eta^{\circ n}(z_0)| \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

Proof. Write

$$L'(z) = a_m z^m + a_{m-1} z^{m-1} + \cdots + a_0$$

and let

$$P(z) = a_{m-1} z^{m-1} + \cdots + a_0$$

. Since $P(z) = o(|z|^m)$ as $|z| \rightarrow \infty$, choose $R_1 > 0$ such that

$$|P(z)| \leq \frac{|a_m|}{2} |z|^m \quad \text{for all } |z| \geq R_1.$$

For $|z| \geq R_1$,

$$|f_\eta(z)| = |z - \eta a_m z^m - \eta P(z)| \geq \eta |a_m| |z|^m - |z| - \eta |P(z)| \geq \frac{\eta |a_m|}{2} |z|^m - |z|.$$

Choose $R \geq R_1$ large enough so that $\frac{\eta |a_m|}{2} |z|^m - |z| > |z|$ for all $|z| > R$, equivalently $\frac{\eta |a_m|}{2} |z|^{m-1} > 2$. Then $|f_\eta(z)| > |z|$ for $|z| > R$, and iteration yields $|f_\eta^{\circ n}(z_0)| \rightarrow \infty$ whenever $|z_0| > R$. \square

Theorem 3.15. *Let L be holomorphic on an open set $\Omega \subset \mathbb{C}$ and define $f_\eta(z) = z - \eta L'(z)$. Assume there exists $\eta_0 > 0$ and $z_0^* \in \Omega$ such that*

$$f_{\eta_0}(z_0^*) = z_0^* \quad \text{and} \quad f'_{\eta_0}(z_0^*) \neq 1.$$

Then there exist $\delta > 0$ and a holomorphic function $\eta \mapsto z^(\eta)$ defined for $|\eta - \eta_0| < \delta$ such that*

$$f_\eta(z^*(\eta)) = z^*(\eta), \quad z^*(\eta_0) = z_0^*.$$

Moreover, if $|f'_{\eta_0}(z_0^)| < 1$ (so z_0^* is attracting at η_0), then for δ sufficiently small the fixed point $z^*(\eta)$ remains attracting for all $|\eta - \eta_0| < \delta$.*

Proof. Define

$$G(z, \eta) = f_\eta(z) - z = -\eta L'(z).$$

Then $G(z_0^*, \eta_0) = 0$. Differentiate with respect to z :

$$\partial_z G(z, \eta) = f'_\eta(z) - 1 = -\eta L''(z).$$

At (z_0^*, η_0) we have $\partial_z G(z_0^*, \eta_0) = f'_{\eta_0}(z_0^*) - 1 \neq 0$ by assumption. Hence, by the holomorphic implicit function theorem, there exist $\delta > 0$ and a holomorphic function $z^*(\eta)$ for $|\eta - \eta_0| < \delta$ such that $G(z^*(\eta), \eta) = 0$, i.e. $f_\eta(z^*(\eta)) = z^*(\eta)$. For the stability claim, note that $(\eta, z) \mapsto f'_\eta(z)$ is continuous. Since $|f'_{\eta_0}(z_0^*)| < 1$, choose δ small enough so that $|f'_\eta(z^*(\eta))| < 1$ for all $|\eta - \eta_0| < \delta$, making $z^*(\eta)$ attracting. \square

Theorem 3.16. *Assume $f : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ is a rational map of degree at least 2. Then*

$$J(f) = \overline{\{z \in \widehat{\mathbb{C}} : z \text{ is a repelling periodic point of } f\}}.$$

In particular, $J(f)$ is the smallest closed, completely invariant subset of $\widehat{\mathbb{C}}$ containing all repelling periodic orbits.

Proof sketch. Let \mathcal{R} be the set of repelling periodic points. If $p \in \mathcal{R}$ has period k , then $(f^{\circ k})'(p)$ has modulus > 1 , so the family of iterates cannot be normal in any neighborhood of p ; hence $\mathcal{R} \subset J(f)$ and thus $\overline{\mathcal{R}} \subset J(f)$. Conversely, fix $z_0 \in J(f)$ and a neighborhood U of z_0 . Since $\{f^{\circ n}\}$ is not normal on U , Montel's theorem implies that the forward images of U cannot omit three values. This non-normality yields an iterate $f^{\circ n}$ and a disk $W \Subset U$ such that $f^{\circ n}(W) \supset \overline{W}$ and an inverse branch $g : W \rightarrow W$ mapping W strictly into itself. By the Schwarz lemma (after conformal identification with \mathbb{D}), g has an attracting fixed point $p \in W$, which is therefore a repelling fixed point of $f^{\circ n}$. Hence U contains a repelling periodic point, so $z_0 \in \overline{\mathcal{R}}$ and $J(f) \subset \overline{\mathcal{R}}$. \square

Theorem 3.17. *Let L be holomorphic on an open set $\Omega \subset \mathbb{C}$ and define*

$$f_\eta(z) = z - \eta L'(z), \quad \eta \in \mathbb{R}.$$

Assume there exist $\eta_0 > 0$ and $z^ \in \Omega$ such that $L'(z^*) = 0$ and*

$$|f'_{\eta_0}(z^*)| = |1 - \eta_0 L''(z^*)| < 1.$$

Then there exist constants $R > 0$, $r \in (0, R)$, $q \in (0, 1)$, and $\delta > 0$ such that for every η with $|\eta - \eta_0| < \delta$ the following hold:

- (1) f_η has a unique fixed point $z^*(\eta)$ in the closed disk $\overline{D(z^*, R)}$, and $z^*(\eta) \rightarrow z^*$ as $\eta \rightarrow \eta_0$.
- (2) The disk $D(z^*, r)$ is contained in the basin of attraction of $z^*(\eta)$:

$$D(z^*, r) \subset \mathcal{B}_\eta(z^*(\eta)) \quad \text{for all } |\eta - \eta_0| < \delta.$$

- (3) For every $z_0 \in D(z^*, r)$ and all $n \geq 0$,

$$|f_\eta^{\circ n}(z_0) - z^*(\eta)| \leq q^n |z_0 - z^*(\eta)|.$$

Proof. Since $|f'_{\eta_0}(z^*)| < 1$ and f'_{η_0} is continuous, choose $R > 0$ and $q \in (0, 1)$ such that

$$\sup_{|z - z^*| \leq R} |f'_{\eta_0}(z)| \leq q. \quad (3.5)$$

The map $(z, \eta) \mapsto f'_\eta(z) = 1 - \eta L''(z)$ is continuous on $\overline{D(z^*, R)} \times \mathbb{R}$. Hence there exists $\delta > 0$ such that for all $|\eta - \eta_0| < \delta$,

$$\sup_{|z - z^*| \leq R} |f'_\eta(z)| \leq q. \quad (3.6)$$

By the mean-value inequality for holomorphic maps (or the integral form of the fundamental theorem of calculus), (3.6) implies that for all $z, w \in \overline{D(z^*, R)}$,

$$|f_\eta(z) - f_\eta(w)| \leq q |z - w|.$$

Thus f_η is a strict contraction on the complete metric space $\overline{D(z^*, R)}$, provided it maps the disk into itself. To ensure invariance, note that $f_{\eta_0}(z^*) = z^*$. By continuity of $(z, \eta) \mapsto f_\eta(z)$, possibly shrinking δ , we can guarantee

$$|f_\eta(z^*) - z^*| < (1 - q) \frac{R}{2} \quad \text{whenever } |\eta - \eta_0| < \delta.$$

Then for any $z \in \overline{D(z^*, R)}$,

$$|f_\eta(z) - z^*| \leq |f_\eta(z) - f_\eta(z^*)| + |f_\eta(z^*) - z^*| \leq q|z - z^*| + (1-q)\frac{R}{2} \leq qR + (1-q)\frac{R}{2} < R.$$

Hence $f_\eta(\overline{D(z^*, R)}) \subset D(z^*, R)$, so f_η maps $\overline{D(z^*, R)}$ into itself and is a contraction there.

By Banach's fixed-point theorem, f_η has a unique fixed point $z^*(\eta) \in \overline{D(z^*, R)}$, proving (1). Moreover, the fixed point depends continuously on η because it is the limit of the Picard iterates $f_\eta^{\circ n}(z^*)$ and the map $(\eta, z) \mapsto f_\eta(z)$ is continuous; thus $z^*(\eta) \rightarrow z^*$ as $\eta \rightarrow \eta_0$.

For (2), choose $r \in (0, R/2)$. If $z_0 \in D(z^*, r)$ then

$$|z_0 - z^*(\eta)| \leq |z_0 - z^*| + |z^* - z^*(\eta)| < r + \frac{R}{2} < R$$

for η sufficiently close to η_0 (shrink δ if needed so that $|z^*(\eta) - z^*| < R/2$). Thus $z_0 \in D(z^*(\eta), R)$ and every iterate remains in $\overline{D(z^*, R)}$ by invariance above, so convergence holds and $D(z^*, r) \subset \mathcal{B}_\eta(z^*(\eta))$.

For (3), since f_η is q -Lipschitz on $\overline{D(z^*, R)}$,

$$|f_\eta(z) - f_\eta(z^*(\eta))| \leq q|z - z^*(\eta)|$$

whenever $z \in \overline{D(z^*, R)}$. Apply this with $z = f_\eta^{\circ n}(z_0)$ and iterate to obtain

$$|f_\eta^{\circ n}(z_0) - z^*(\eta)| \leq q^n |z_0 - z^*(\eta)|.$$

□

Theorem 3.18. *Let*

$$L(z) = \frac{1}{4}(z^2 - 1)^2, \quad f_\eta(z) = z - \eta L'(z) = (1 + \eta)z - \eta z^3, \quad \eta > 0.$$

Then:

- (1) *The real line is forward invariant: $f_\eta(\mathbb{R}) \subset \mathbb{R}$.*
- (2) *The fixed points are $\{-1, 0, 1\}$, with multipliers*

$$f'_\eta(0) = 1 + \eta, \quad f'_\eta(\pm 1) = 1 - 2\eta.$$

Hence for $0 < \eta < 1$, the points ± 1 are attracting and 0 is repelling.

- (3) *For every $\eta > 0$, the repelling fixed point 0 lies in the Julia set: $0 \in J(f_\eta)$. Consequently, the full backward orbit*

$$\mathcal{O}^-(0) := \bigcup_{n \geq 0} f_\eta^{-n}(\{0\})$$

is contained in $J(f_\eta)$, and therefore

$$\mathcal{O}^-(0) \cap \mathbb{R} \subset J(f_\eta) \cap \mathbb{R}.$$

In particular, the real slice \mathbb{R} contains an explicit (typically infinite) set of points at which the dynamics is unstable in the Julia–Fatou sense.

Proof. (1) Since f_η is a polynomial with real coefficients, $f_\eta(x) \in \mathbb{R}$ for all $x \in \mathbb{R}$.

(2) Solve $f_\eta(z) = z$:

$$(1 + \eta)z - \eta z^3 = z \iff \eta z - \eta z^3 = 0 \iff z(1 - z^2) = 0,$$

so $z \in \{-1, 0, 1\}$. Also

$$f'_\eta(z) = (1 + \eta) - 3\eta z^2,$$

hence $f'_\eta(0) = 1 + \eta$ and $f'_\eta(\pm 1) = (1 + \eta) - 3\eta = 1 - 2\eta$. Therefore $|f'_\eta(0)| = 1 + \eta > 1$ for all $\eta > 0$, so 0 is repelling, and $|1 - 2\eta| < 1$ iff $0 < \eta < 1$, so ± 1 are attracting exactly for $0 < \eta < 1$.

(3) We show that any repelling periodic point belongs to the Julia set by a direct normality argument. Here 0 is a repelling fixed point, so we prove $0 \in J(f_\eta)$.

Let $p = 0$. Since $f_\eta(p) = p$ and $|f'_\eta(p)| > 1$, we have $(f_\eta^{on})'(p) = (f'_\eta(p))^n$, so

$$|(f_\eta^{on})'(p)| \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

If the family $\{f_\eta^{on}\}_{n \geq 0}$ were normal in a neighborhood of p , then it would be equicontinuous there. Equicontinuity in particular prevents derivatives from becoming unbounded at a point along a subsequence. This contradiction shows that $\{f_\eta^{on}\}$ is not normal in any neighborhood of p , hence $p \notin F(f_\eta)$ and therefore $p \in J(f_\eta)$.

It remains to show $\mathcal{O}^-(0) \subset J(f_\eta)$. We use complete invariance of the Julia set for polynomials: we claim $f_\eta^{-1}(F(f_\eta)) \subset F(f_\eta)$, which implies $f_\eta^{-1}(J(f_\eta)) \subset J(f_\eta)$. Indeed, if $w \in F(f_\eta)$, then there exists a neighborhood U of w on which $\{f_\eta^{on}\}$ is normal. For any z with $f_\eta(z) = w$, choose a neighborhood V of z such that $f_\eta(V) \subset U$. Then on V the family $\{f_\eta^{on}\} = \{f_\eta^{on} \circ \text{id}\}$ is normal because $\{f_\eta^{o(n+1)}\} = \{f_\eta^{on} \circ f_\eta\}$ is normal on V as composition of a normal family on U with the holomorphic map $f_\eta|_V$. Hence $z \in F(f_\eta)$, proving backward invariance of $F(f_\eta)$. Taking complements yields $f_\eta^{-1}(J(f_\eta)) \subset J(f_\eta)$. Since $0 \in J(f_\eta)$ and $J(f_\eta)$ is backward invariant, every preimage of 0 under every iterate lies in $J(f_\eta)$, i.e., $\mathcal{O}^-(0) \subset J(f_\eta)$. Intersecting with \mathbb{R} gives $\mathcal{O}^-(0) \cap \mathbb{R} \subset J(f_\eta) \cap \mathbb{R}$. \square

4. NUMERICAL EXPERIMENTS

In this section we provide computational illustrations of the Julia–Fatou basin viewpoint for a toy loss whose complexified dynamics can be explored directly. The experiments are designed to visualize three phenomena predicted by the analysis in Section 3: the existence of multiple basins of attraction, the sensitivity of basin boundaries, and the deformation (or loss) of attraction under learning-rate variation. In addition to standard two-dimensional basin plots, we also include a three-dimensional rendering of convergence-time geometry, which helps highlight how slow convergence concentrates near basin boundaries.

Toy model and induced training map. We consider the double-well quartic loss

$$L(z) = \frac{1}{4}(z^2 - 1)^2, \quad L'(z) = z^3 - z,$$

which induces the holomorphic training map

$$z_{n+1} = f_\eta(z_n) = z_n - \eta L'(z_n) = (1 + \eta)z_n - \eta z_n^3, \quad \eta > 0. \quad (4.1)$$

The fixed points are $z^* \in \{-1, 0, 1\}$. Since

$$f'_\eta(z) = 1 - \eta(3z^2 - 1),$$

we have $f'_\eta(0) = 1 + \eta$, so 0 is repelling for every $\eta > 0$, while $f'_\eta(1) = f'_\eta(-1) = 1 - 2\eta$, so ± 1 are attracting precisely when $|1 - 2\eta| < 1$, i.e. $0 < \eta < 1$. Thus, for $0 < \eta < 1$ the complex plane decomposes into at least two stable basins $\mathcal{B}(1)$ and $\mathcal{B}(-1)$ separated by a sensitive boundary region.

Sampling window and stopping rules. We sample initial conditions $z_0 = x + iy$ on a uniform grid over the window $(x, y) \in [-2, 2] \times [-2, 2]$ and iterate (4.1) up to a maximum of N_{\max} steps. Each initialization is assigned an outcome label according to the first event detected:

- *Convergence to +1*: if $|z_n - 1| < \varepsilon$ for some $n \leq N_{\max}$;
- *Convergence to -1*: if $|z_n + 1| < \varepsilon$ for some $n \leq N_{\max}$;
- *Divergence*: if $|z_n| > R_{\text{esc}}$ for some $n \leq N_{\max}$;
- *Undecided (cap reached)*: otherwise, if none of the above occurs by N_{\max} .

In addition, we record the *hit time* $T(z_0)$, defined as the iteration index n at which convergence or divergence is first detected, with $T(z_0) = N_{\max}$ for undecided points. This produces an iteration-count field that serves as a numerical proxy for local convergence speed.

Two-dimensional basin and hit-time visualizations. Our primary visualization is the two-dimensional basin label field on $[-2, 2] \times [-2, 2]$, which shows how the plane partitions into regions converging to different attractors (or diverging). We also plot the corresponding hit-time field (displayed on a log scale in the figures) to emphasize slow convergence near basin boundaries. In the experiments below we use $n = 600$, $\varepsilon = 10^{-3}$, $R_{\text{esc}} = 10$, and $N_{\max} = 160$ unless stated otherwise. Figure 1 shows representative 2D basin decompositions for $\eta \in \{0.2, 0.5, 0.8, 1.05\}$. In the stable regime ($0 < \eta < 1$), two dominant basins are visible, separated by a complex boundary. As η increases toward the threshold $\eta = 1$, the boundary becomes more intricate and the region of slow convergence expands. For $\eta > 1$, the fixed points ± 1 lose attraction and the plots exhibit widespread instability and divergence, consistent with the multiplier-based stability switch.

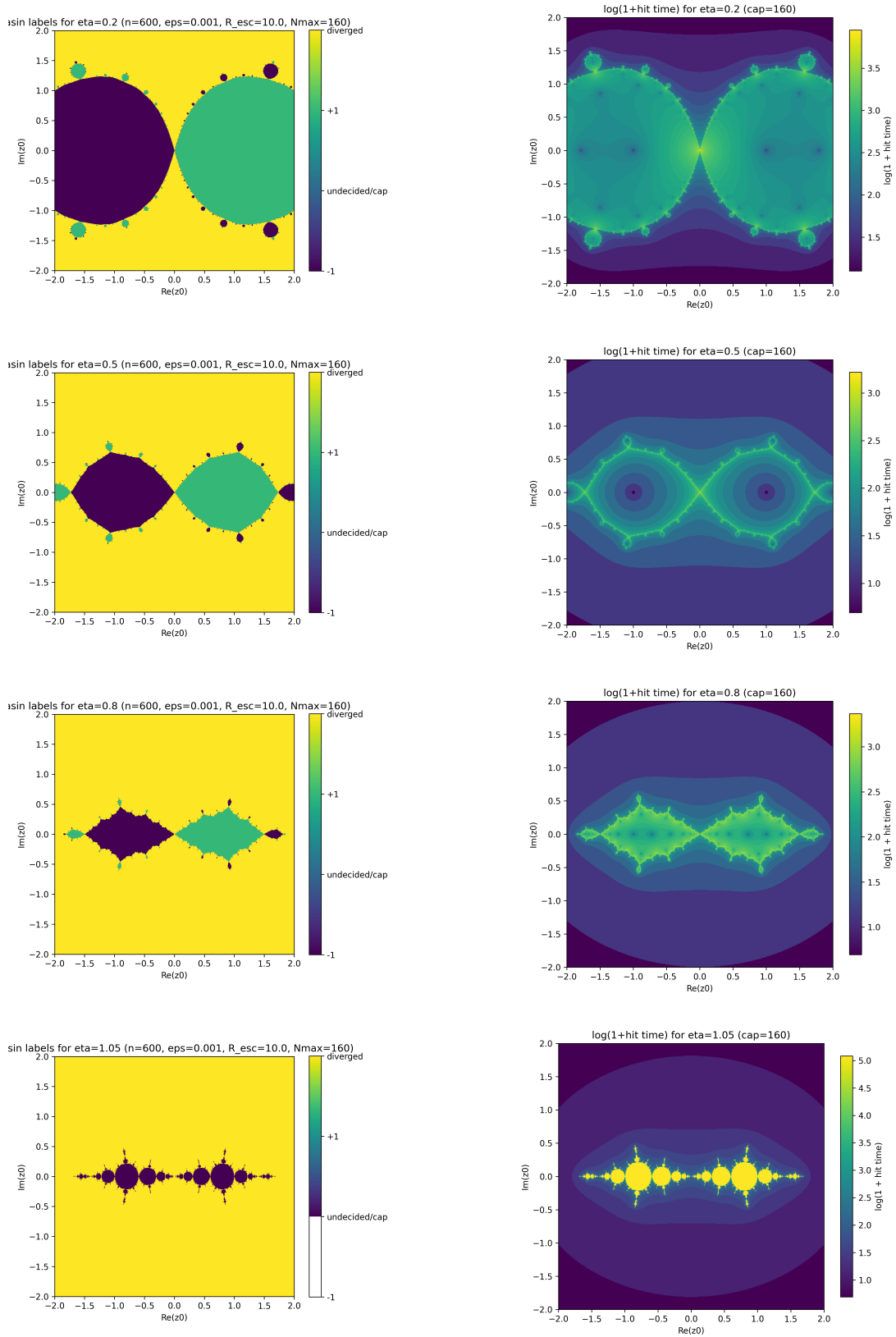


FIGURE 1. Two-dimensional basin labels (left panels) and hit-time fields (right panels, shown as $\log(1 + T)$) for the map (4.1) on $[-2, 2] \times [-2, 2]$ at $\eta \in \{0.2, 0.5, 0.8, 1.05\}$.

Labels indicate convergence to $+1$, convergence to -1 , divergence (when $|z_n| > R_{\text{esc}}$), or undecided (iteration cap N_{max} reached). The hit-time fields highlight slow convergence near basin boundaries and increased instability as η approaches and exceeds the stability threshold $\eta = 1$.

Real-slice visualization (restriction to \mathbb{R}). Although the complex plane provides a global view of basin geometry, practical gradient descent operates on the real slice $\Omega \cap \mathbb{R}$. When $L(x) \in \mathbb{R}$ for $x \in \Omega \cap \mathbb{R}$ and $\eta \in \mathbb{R}$, the real slice is forward invariant under f_η , so the real training dynamics are obtained by restricting (4.1) to $y = 0$. To highlight how basin structure impacts real initializations, we scan initial conditions $z_0 = x_0 \in [-2, 2] \subset \mathbb{R}$ and record both the observed outcome and the corresponding hit time.

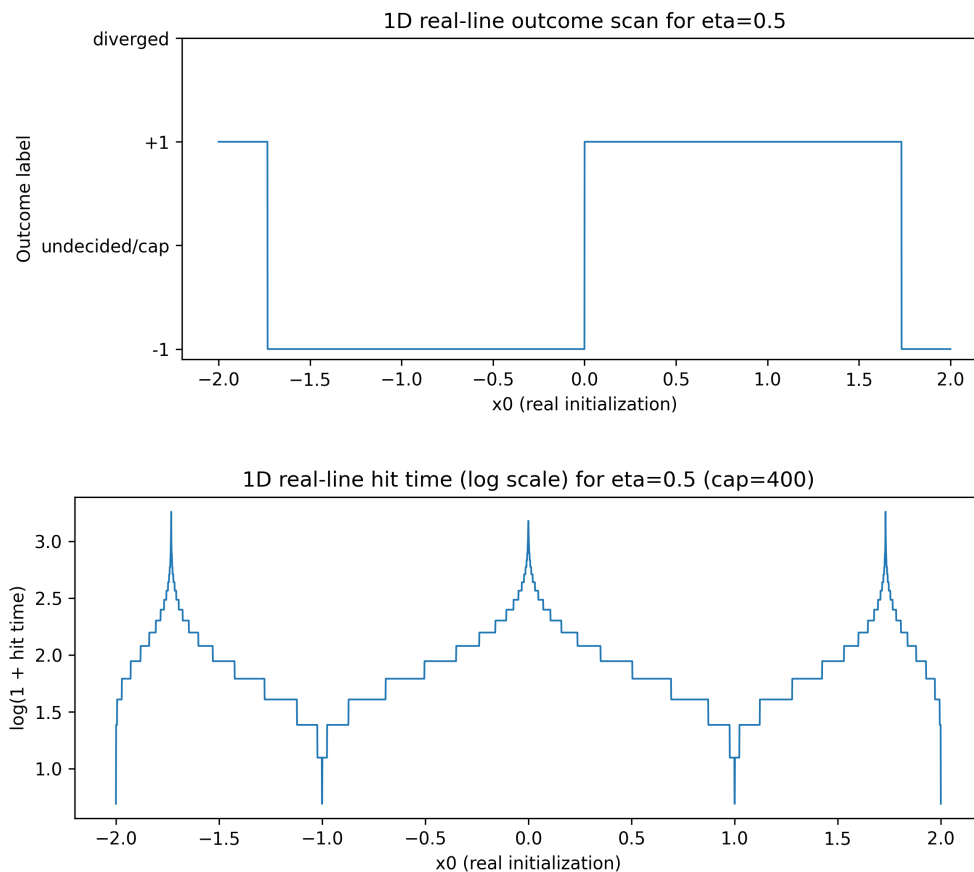


FIGURE 2. Real-slice behavior for (4.1) at $\eta = 0.5$ with initializations $z_0 = x_0 \in [-2, 2] \subset \mathbb{R}$. Top: outcome label (convergence to $+1$, convergence to -1 , divergence, or cap reached). Bottom: hit time $T(x_0)$ displayed as $\log(1 + T)$. The sharp transitions and spikes in hit time occur near the intersections of basin boundaries with the real axis, illustrating that sensitivity is present even on the real training slice.

Three-dimensional convergence-time geometry. To complement the 2D plots, we visualize the hit-time field as a three-dimensional surface over the initialization plane. Specifically, we plot

$$(x, y) \mapsto T(x + iy),$$

where T is the recorded hit time (capped at N_{\max}). This surface representation emphasizes that the basin boundary is not only a geometric separator but also a region where convergence becomes slow: the surface typically rises sharply near boundary filaments and forms broad “plateaus” at the cap value in regions that fail to converge within N_{\max} .

Figure 3 shows a representative 3D surface for $\eta = 0.5$, which lies in the stable regime. The largest hit times occur along ridges aligned with the basin boundary, matching the theoretical picture that dynamics are least contractive near Julia-set structures.

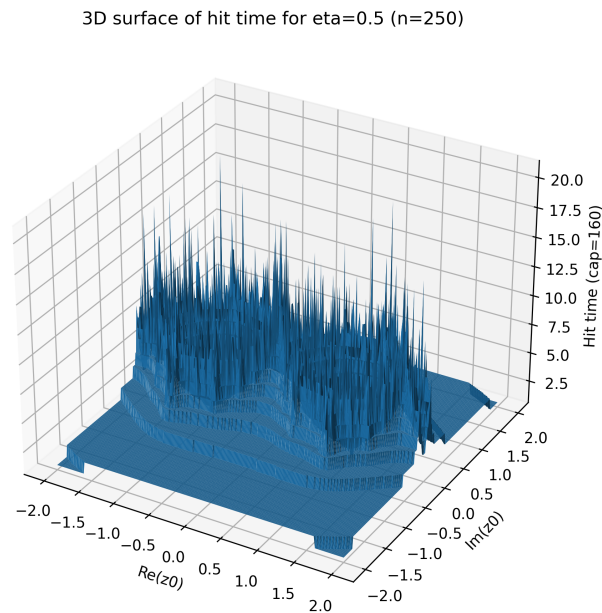


FIGURE 3. Three-dimensional visualization of the hit-time field $(x, y) \mapsto T(x + iy)$ for (4.1) at $\eta = 0.5$ on $[-2, 2] \times [-2, 2]$. Peaks and ridges in T concentrate near the basin boundary, indicating slow convergence in the sensitive region separating the two attracting basins.

Reproducibility. For completeness, we summarize the computational procedure used to generate the figures.

Algorithm. Basin labeling and hit-time computation.

- (1) Fix $\eta > 0$, window $[x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]$, grid size $n \times n$, tolerance ε , escape radius R_{esc} , and iteration cap N_{max} .
- (2) For each grid point $z_0 = x + iy$, iterate $z_{k+1} = f_\eta(z_k)$ for $k = 0, 1, \dots, N_{\text{max}} - 1$.
- (3) If $|z_k - 1| < \varepsilon$, label $z_0 \mapsto +1$ and set $T(z_0) = k$.
- (4) Else if $|z_k + 1| < \varepsilon$, label $z_0 \mapsto -1$ and set $T(z_0) = k$.
- (5) Else if $|z_k| > R_{\text{esc}}$, label $z_0 \mapsto \infty$ and set $T(z_0) = k$.
- (6) If no event occurs by N_{max} , label z_0 as undecided and set $T(z_0) = N_{\text{max}}$.
- (7) Plot the basin label field and the hit-time field (optionally using $\log(1 + T)$ for visualization).

5. CONCLUSION

We model training updates as holomorphic iterations and describe stability through basin geometry: attracting regions correspond to Fatou components, while sensitive boundaries lie in the Julia set. We derive multiplier-based local stability and linear convergence, extend local attraction to \mathbb{C}^d via a spectral criterion, and establish global results on escape-radius divergence, persistence under small learning-rate changes, and Julia-set instability in rational cases. These results clarify how step size and initialization drive stable versus unstable training behavior.

ACKNOWLEDGEMENTS

The authors thank the Editors and anonymous reviewers for their careful reading and constructive comments, which improved the clarity and presentation of this manuscript.

REFERENCES

1. Bottou, L., Curtis, F. E., & Nocedal, J. (2018). Optimization methods for large-scale machine learning. *SIAM Review*, 60(2), 223–311. <https://doi.org/10.1137/16M1080173>
2. Dauphin, Y. N., Pascanu, R., Gulcehre, Ç., Cho, K., Ganguli, S., & Bengio, Y. (2014). Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *Advances in Neural Information Processing Systems (NeurIPS)*, 27. <https://doi.org/10.48550/arXiv.1406.2572>
3. Devaney, R. L. (2018). *An introduction to chaotic dynamical systems* (2nd ed.). CRC Press. <https://doi.org/10.1201/9780429502309>
4. Garipov, T., Izmailov, P., Podoprikin, D., Vetrov, D. P., & Wilson, A. G. (2018). Loss surfaces, mode connectivity, and fast ensembling of deep neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 31. <https://doi.org/10.48550/arXiv.1802.10026>
5. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. <https://doi.org/10.1038/nature14539>
6. Hardt, M., Recht, B., & Singer, Y. (2016). Train faster, generalize better: Stability of stochastic gradient descent. *Proceedings of the 33rd International Conference on Machine Learning (ICML)*. <https://doi.org/10.48550/arXiv.1509.01240>

7. Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., & Tang, P. T. P. (2017). On large-batch training for deep learning: Generalization gap and sharp minima. *International Conference on Learning Representations (ICLR)*. <https://doi.org/10.48550/arXiv.1609.04836>
8. Li, H., Xu, Z., Taylor, G., Studer, C., & Goldstein, T. (2018). Visualizing the loss landscape of neural nets. *Advances in Neural Information Processing Systems (NeurIPS)*, 31. <https://doi.org/10.48550/arXiv.1712.09913>
9. Li, Q., Lin, T., & Shen, Z. (2023). Deep learning via dynamical systems: An approximation perspective. *Journal of the European Mathematical Society*, 25(5), 1671–1709. <https://doi.org/10.4171/JEMS/1221>
10. Milnor, J. (2011). *Dynamics in one complex variable* (3rd ed.). Princeton University Press. <https://doi.org/10.1515/9781400835539>
11. Poggio, T., Banburski, A., & Liao, Q. (2020). Theoretical issues in deep networks. *Proceedings of the National Academy of Sciences*, 117(48), 30039–30045. <https://doi.org/10.1073/pnas.1907369117>
12. Rane, N. L., Mallick, S. K., Kaya, Ö., & Rane, J. (2024). Techniques and optimization algorithms in deep learning: A review. In *Applied machine learning and deep learning: Architectures and techniques* (pp. 59–79). Deep Science Publishing. https://doi.org/10.70593/978-81-981271-4-3_3
13. Chaudhry, A., & Kristensson, K. (2025). FEM for convergence analysis of ReLU Neural Networks. *Annals of Mathematics and Computer Science*, 25, 1–16. <https://doi.org/10.56947/amcs.v25.392>
14. Ramdani, S. (2025). Feature scaling and data normalization: Choosing the right method. *Annals of Mathematics and Computer Science*, 25, 1–16. <https://doi.org/10.56947/amcs.v25.428>
15. Khan, A., & Ahmad, N. (2025). A comprehensive review of feature attribution methods in deep learning. *Annals of Mathematics and Computer Science*, 29, 1–25. <https://doi.org/10.56947/amcs.v29.635>

¹ DEPARTMENT OF MATHEMATICAL SCIENCE, FACULTY OF SCIENCE AND TECHNOLOGY, BINGHAM UNIVERSITY KARU, NASARAWA, NIGERIA

Email address: busayo-oluwafemi.samuel@binghamuni.edu.ng

² DEPARTMENT OF MATHEMATICAL SCIENCE, FACULTY OF SCIENCE AND TECHNOLOGY, BINGHAM UNIVERSITY KARU, NASARAWA, NIGERIA

Email address: francis-moses.obinna@binghamuni.edu.ng