

Positive Transverse Response in Quadratic Population Networks

Anherutowa Calvo
9D Labs

September 2026

Abstract

We study how gradient flow on a dominant quadratic target changes the response of a two-layer network to an orthogonal target. For centered quadratic neurons on S^4 with compact, moment-matched initialization, we prove that training increases the tangent-kernel capacity of the dominant mode and induces a positive transverse coupling. The residual of an infinitesimal transverse target remains negative at every finite training time and converges to zero. The proof combines characteristic pairing with computer-assisted bounds on a scalar barrier. We also obtain a finite-mixing extension on compact exposure intervals for sufficiently small mixing. The finite-mixing statement is not uniform at the limiting exposure. Independent arbitrary-precision interval checks validate the numerical obligations of the barrier and freeze bounds; ordinary simulations separately test the perturbation orders and suggest failures at large mixing.

1 Introduction

Training changes the tangent kernel of a neural network and therefore changes its response to target components that are weakly represented in the training signal. We study this effect in a quadratic model for which the parameter dynamics admit a low-dimensional characteristic representation. Our question is whether training on a dominant mode can induce a coupling that reduces the residual of an orthogonal mode. The contribution is a sign theorem for a transported response kernel and a computer-assisted residual barrier in this particular model. We do not claim that feature learning, evolving kernels, or preferential learning of stronger quadratic features are new phenomena.

We distinguish the diagonal kernel coefficient κ , which measures capacity in a target direction, from the off-diagonal coefficient χ , which couples two residuals. The ratio r of weak to dominant residuals measures their relative decay. These quantities answer different questions: positive coupling can assist a weak mode while its residual still decays more slowly than the dominant residual.

The tangent-kernel residual equation is standard (Jacot et al., 2018; Lee et al., 2019), as are the distinctions between lazy and feature-learning regimes (Chizat et al., 2019; Yang and Hu, 2021). Our analysis uses the distributional description of two-layer networks (Mei et al., 2018; Chizat and Bach, 2018; Rotskoff and Vanden-Eijnden, 2022; Mei et al., 2019). The specific results are positivity of a transported transverse kernel and a barrier estimate that prevents the linearized weak residual from crossing zero.

Section 2 defines the model and its compact initialization. The compact support is essential: Gaussian initialization has characteristics with arbitrarily short blow-up times for the exposure

flow considered here. Section 3 summarizes the results. Sections 4–6 establish capacity growth, transverse positivity, and the residual barrier. Section 7 discusses the limits of the conclusions, and Section 8 places the analysis in context. Appendix A describes the numerical certificates and their reproduction. Figure 1 illustrates the linearized trajectory.

2 Model and initialization

Let $d = 5$ unless otherwise stated. Data x is uniform on S^{d-1} . The two-layer network uses the centered quadratic

$$\sigma_w(x) = (w^\top x)^2 - \frac{\|w\|^2}{d}.$$

A particle is $(a, w) \in \mathbb{R} \times \mathbb{R}^d$. The continuum output is

$$f(x) = \int a \sigma_w(x) d\rho(a, w) = x^\top \overline{M} x, \quad \overline{M} = \int a \left(w w^\top - \frac{\|w\|^2}{d} I \right) d\rho.$$

The teacher is $y_A(x) = x^\top A x$ with

$$A = u u^\top - \frac{1}{d} I, \quad \|u\| = 1, \quad L := \text{tr}(A^2) = \frac{d-1}{d}.$$

At $d = 5$, $L = 4/5$ and $\mu := 1/d = 1/5$. The residual matrix is $R = \overline{M} - A$. Population squared-loss gradient flow on the quadratic clock is

$$\frac{da}{d\sigma} = -w^\top R w, \quad \frac{dw}{d\sigma} = -2a R w. \quad (1)$$

For the loss $\frac{1}{2} \mathbb{E}_x (f - y)^2$ with unit parameter learning rate, $\sigma = 2t/[d(d+2)]$, because $\mathbb{E}_x (x^\top R x)^2 = 2 \text{tr}(R^2)/[d(d+2)]$ for traceless R . Thus quadratic time is a fixed positive rescaling of physical training time; exposure, in contrast, has a state-dependent speed.

If ρ is axially symmetric about u , then $R = \Delta A$ and each particle reduces to (a, h, b) with $h = (w \cdot u)^2$ and $b = \|w_\perp\|^2$:

$$a' = -\Delta q, \quad h' = -4\Delta L a h, \quad b' = \frac{4\Delta}{d} a b, \quad q = L h - \mu b. \quad (2)$$

The gap obeys

$$\frac{d\Delta}{d\sigma} = -\frac{\kappa}{L} \Delta, \quad \kappa = \mathbb{E} [q^2 + 4a^2(L^2 h + \mu^2 b)].$$

Thus $\text{sign}(\Delta)$ is conserved, and $|\Delta|$ is strictly decreasing whenever $\kappa > 0$ and $\Delta \neq 0$.

Assume $\Delta(0) = -1$ and $\Delta < 0$. Exposure $\lambda(\sigma) = -\int_0^\sigma \Delta$ turns (2) into

$$\frac{da}{d\lambda} = q, \quad \frac{dh}{d\lambda} = 4L a h, \quad \frac{db}{d\lambda} = -4\mu a b. \quad (3)$$

The invariants $I := a^2 - \frac{1}{2}(h + b)$ and $h b^{d-1}$ are conserved. Freeze, when it exists, is the first λ_* with $\int_0^{\lambda_*} \kappa = L$.

Lemma 1 (Gaussian characteristics explode). *At $d = 5$, on the cone $a_0 > 0$, $R_0 := b_0/h_0 \leq r < 4$, and $I_0 \leq 0$, one has $a' \geq C_r a^2$ with $C_r = 2(L - \mu r)/(1 + r) > 0$. Hence $T_{\text{blow}} \leq 1/(C_r a_0)$. For $r = 1$, $C_r = 3/5$. Gaussian initialization has positive density on this cone, so for every $\lambda > 0$,*

$$\Pr_{\rho_0}(T_{\text{blow}} < \lambda) > 0.$$

There is no positive exposure interval on which the unbounded Gaussian characteristic flow exists ρ_0 -almost surely.

Proof. On the cone, $a' = h(L - \mu R)$. If $a_0 > 0$ and $R_0 \leq r < L/\mu$, then $a' > 0$ and $R' = -4aR \leq 0$, so $R(\lambda) \leq r$. The bound $I \leq 0$ gives $h + b \geq 2a^2$, hence $h \geq 2a^2/(1 + r)$ and $a' \geq C_r a^2$. Comparison with $x' = C_r x^2$ explodes at $1/(C_r a_0)$. Gaussian full support supplies a positive-measure set with $a_0 > 1/(C_r \lambda)$. $\square$

The unbounded characteristic field therefore requires a different initialization for the present analysis. Regularity assumptions in mean-field theory control the coefficients or initial support (Mei et al., 2018, 2019). The compact law below provides a common interval of characteristic existence while retaining the moments needed for the local calculations.

Lemma 2 (Moment-matched compact law). *Let $d = 5$. Draw $w_0 = R\omega$ with $\omega \sim \text{Unif}(S^4)$ independent of the two-point radius*

$$R^2 \in \left\{ \frac{7 - \sqrt{14}}{5}, \frac{7 + \sqrt{14}}{5} \right\}, \quad p_R^\pm = \frac{1}{2} \pm \frac{1}{\sqrt{14}}.$$

Independently draw $a_0 = \epsilon \sqrt{S_a}$ with $\epsilon \in \{\pm 1\}$ equiprobable and

$$S_a \in \{3 - \sqrt{6}, 3 + \sqrt{6}\}, \quad p_a^\pm = \frac{1}{2} \pm \frac{1}{\sqrt{6}}.$$

Then

$$\mathbb{E}R^2 = 1, \quad \mathbb{E}R^4 = \frac{7}{5}, \quad \mathbb{E}R^6 = \frac{63}{25}, \quad \mathbb{E}a_0^2 = 1, \quad \mathbb{E}a_0^4 = 3, \quad \mathbb{E}a_0^6 = 15,$$

which are the isotropic Gaussian moments of $w \sim N(0, I_5/5)$ and the even moments of $a \sim N(0, 1)$ through degree six.

Proof. Write $x_\pm = (7 \pm \sqrt{14})/5$. Then $x_+ + x_- = 14/5$ and

$$p_R^- x_- + p_R^+ x_+ = \frac{1}{2} \cdot \frac{14}{5} + \frac{1}{\sqrt{14}}(x_- - x_+) = \frac{7}{5} - \frac{2}{5} = 1.$$

The identities $\mathbb{E}R^4 = 7/5$ and $\mathbb{E}R^6 = 63/25$ are the same two-point quadrature. Likewise $s_\pm = 3 \pm \sqrt{6}$ gives $\mathbb{E}S_a = 1$, $\mathbb{E}S_a^2 = 3$, $\mathbb{E}S_a^3 = 15$. $\square$

Haar on S^4 gives $x = h_0/R^2 \sim \text{Beta}(1/2, 2)$. The substitution $x = y^2$ yields

$$dv(y) = \frac{3}{2}(1 - y^2) dy, \quad 0 \leq y \leq 1. \quad (4)$$

Lemma 3 (Crude existence). *On (3), $z^2 = a^2 + \|w\|^2$ obeys $z' \leq \sqrt{2}Lz^2$. Lemma 2 has $z_0 \leq Z_0$ with $Z_0^2 = 3 + \sqrt{6} + (7 + \sqrt{14})/5$. Every characteristic exists at least until*

$$T_0 = \frac{1}{\sqrt{2}LZ_0} \approx 0.320664.$$

Proof. $|a'| \leq L\|w\|^2 \leq Lz^2$ and $\|w'\| \leq 2L|a|\|w\| \leq Lz^2$. Then $zz' \leq Lz^2(|a| + \|w\|)$ and $|a| + \|w\| \leq \sqrt{2}z$. The comparison $x' = \sqrt{2}Lx^2$, $x(0) = Z_0$, explodes at T_0 . $\square$

Outward-rounded interval evaluation gives $T_0 \in (0.32066420257840156, 0.32066420257840267)$, hence $0.3204 < T_0$.

3 Main results

Write $\mathcal{V} = \text{Sym}_0^2(u^\perp)$. An equal-norm probe $B \in \mathcal{V}$ has $\text{tr}(AB) = 0$ and $\|B\|_F^2 = L$. Linear response means an infinitesimal B perturbation of the pure- A flow of Lemma 2: the weak residual is $z = \partial_\varepsilon \beta_2|_{\varepsilon=0}$ for teacher $T = A + \varepsilon B$, with $z(0) = -1$. Let Δ be the dominant residual, $r = z/\Delta$, and $M(s, t) = \partial_s C_{s,t}(B, B)$ the paired susceptibility kernel.

If the residual remains in $\text{span}\{A_1, A_2\}$,

$$\dot{\beta} = -\frac{1}{L}\Gamma\beta, \quad \Gamma = \begin{pmatrix} \kappa_1 & \chi \\ \chi & \kappa_2 \end{pmatrix}. \quad (5)$$

The diagonal entries measure capacity; χ describes coupling. The relative residual r also depends on the trajectory.

For the initialization of Lemma 2 in dimension $d = 5$, we establish the following results.

1. *Capacity on A .* For $d = 5$ and $\lambda > 0$, $\kappa_A(\lambda) > \kappa_A(0)$ on every common existence interval (Theorems 1–2).

2. *Useful transverse capacity.* For every equal-norm $B \in \mathcal{V}$,

$$\kappa_\perp(0) = \frac{88}{125}, \quad \kappa_\perp''(0) = \frac{40832}{15625} > 0, \quad c_{\text{cross}}''(0) = \frac{67584}{15625} > 0,$$

and $M(s, t) \geq 0$ on $0 \leq t \leq s \leq 2/5$ (Theorems 3–5).

3. *Location of the limiting exposure.* Lemma 5 gives $\lambda_* < 2/5$, hence $M \geq 0$ through freeze (Theorem 6).

4. *Linearized no-hit.* $H_1 > 0$ on $(0, \lambda_*)$ (Lemmas 7–8) and $g := -\Delta > 0$ imply $r > 1$ and $z < \Delta < 0$ (Theorems 7–8).

5. *Small finite mixing.* For the block or Gram–Schmidt pair and every $0 < T < \lambda_*$ there exists $\varepsilon_T > 0$ such that $|\beta_2|$ decreases on $[0, T]$ and $\chi_\varepsilon \gtrsim \varepsilon \lambda^2 > 0$ on $(0, T]$ (Theorem 9).

Remark. The finite-mixing threshold in (5) depends on the compact exposure interval. Section 7 gives numerical counterexamples to extensions with unrestricted mixing.

4 Capacity growth

Write $\phi = q^2 + 4a^2s$ with $s = w^\top A^2 w$, so $\kappa = \mathbb{E}[\phi]$. Along (3),

$$\phi'' = 16[q^2s + a^2(4s^2 + 7qt) + 4a^4v],$$

where $t = w^\top A^3 w$ and $v = w^\top A^4 w$. In coordinates $X = Lh, Y = \mu b$,

$$4s^2 + 7qt = 11L^2X^2 + (22L\mu - 7)XY + 11\mu^2Y^2.$$

The determinant of this quadratic form is $(7/4)(44L\mu - 7)$. Hence $4s^2 + 7qt$ is positive definite whenever $L\mu = (d-1)/d^2 > 7/44$, i.e. for integer $2 \leq d \leq 5$. Also $s \geq 0$ and $v \geq 0$. Therefore $\phi'' \geq 0$ pointwise on every characteristic existence interval.

At initialization, evenness in a gives $\kappa'(0) = 0$. Wick through degree six, using only the moments of Lemma 2, gives

$$\mathbb{E}[q^2s] = \frac{2L^2 + 8T_4}{d^3}, \quad \mathbb{E}[s^2] = \frac{L^2 + 2T_4}{d^2}, \quad \mathbb{E}[qt] = \frac{2T_4}{d^2}, \quad \mathbb{E}[v] = \frac{T_4}{d},$$

where $T_4 = \text{tr}(A^4)$. Independence and $\mathbb{E}a^2 = 1$, $\mathbb{E}a^4 = 3$ then yield

$$\kappa''(0) = \frac{32(d-1)(2d+1)(3d^3 - 4d^2 - 4d + 12)}{d^6} > 0 \quad (d \geq 2). \quad (6)$$

At $d = 5$, $\kappa(0) = 88/125$ and $\kappa''(0) = 375936/15625$.

Theorem 1 (Local capacity). *Consider the one-mode centered quadratic model with independent a, w and isotropic even initialization matching the respective Gaussian moments of $w \sim N(0, I_d/d)$ and $a \sim N(0, 1)$ through degree six. Assume a common characteristic interval near zero on which two derivatives may be passed under the expectation. Lemma 2 supplies these conditions at $d = 5$; the unbounded Gaussian law does not. Then $\kappa'(0) = 0$ and (6) holds. At $d = 5$, $\kappa(0) = 0.704$ and $\kappa''(0) = 24.059904$.*

Theorem 2 (Capacity on common existence). *For $2 \leq d \leq 5$, on every common interval of characteristic existence on which differentiation under the expectation is valid, assume the independence, isotropy, evenness and separate Gaussian moment matching of Theorem 1. Then $\kappa''(0) > 0$, $\kappa'' \geq 0$, hence $\kappa(\lambda) > \kappa(0)$ for $\lambda > 0$. Unbounded Gaussian initialization does not supply such an interval. Lemma 2 does, on $[0, T_0)$ and, by Lemma 4, on $[0, 2/5]$.*

Proof. Pointwise $\phi'' \geq 0$ implies convexity of any empirical or population κ formed from characteristics that exist up to λ_+ . Continuity of κ'' at 0 and $\kappa''(0) > 0$ give $\kappa' > 0$ thereafter. $\square$

The algebraic form $4s^2 + 7qt$ is indefinite at $d = 6$.

5 Transverse response

Let A_1, A_2 be traceless symmetric with $\text{tr}(A_i A_j) = L\delta_{ij}$. Write $q_i = w^\top A_i w$ and

$$\Gamma_{ij} = \int (q_i q_j + 4a^2 w^\top S_{ij} w) d\rho, \quad S_{ij} = \frac{1}{2}(A_i A_j + A_j A_i).$$

Under Lemma 2, with $\text{tr}(A_1 A_2) = 0$ and equal norms, $\chi(0) = \chi'(0) = 0$ and $\kappa_1(0) = \kappa_2(0) = 88/125$.

Two explicit pairs close the plane: the block pair $B = c \text{diag}(0, 1, 1, -1, -1)$ with $c = 1/\sqrt{5}$, and the Gram–Schmidt pair obtained as the $O(d-2)$ -invariant completion of a second axis $v \perp u$. A raw second rank-one quadratic A_v is *not* orthogonal to A .

5.1 The transported response

We use the unperturbed exposure λ throughout the linearization. For a prescribed scalar forcing $r(\lambda)$, the particle field is

$$a' = w^\top (A + \varepsilon r B) w, \quad w' = 2a(A + \varepsilon r B)w.$$

It follows from (1) by dividing by $g = -\Delta$ and using $\beta_2 = \varepsilon z + O(\varepsilon^2)$, so that $-z/g = z/\Delta = r$. The first variation of the dominant residual is zero. Indeed the pure- A flow is invariant under $O(u^\perp)$, whereas a first variation in a traceless transverse tensor has zero invariant scalar component. Thus no first-order change of the dominant clock enters this equation.

Here is a coordinate-free definition of the two-time kernel. Let $F_D(a, w) = (w^\top D w, 2a D w)$, let $\Psi_{s,t}$ be the tangent propagator of F_A along a base characteristic, and put $P_D(a, w) = a w^\top D w$. Define

$$C_{s,t}(B_1, B_2) = \mathbb{E}[DP_{B_1}(a_s, w_s) \Psi_{s,t} F_{B_2}(a_t, w_t)], \quad M(s, t) = \partial_s C_{s,t}(B, B). \quad (7)$$

There is no division by L in this definition. The division by L appears when projecting the output matrix onto the coefficient of B . In particular $C_{t,t}(B, B) = \kappa_{\perp}(t)$. Compact support on a common existence interval justifies differentiating and integrating these smooth functions under the expectation.

For the block probe write b_+, b_- for the squared norms in its two 2-dimensional eigenspaces, and $v = (b_+ - b_-)/b$ along the base flow. Conditional on h, b , Haar measure gives $v \sim \text{Unif}[-1, 1]$, so $\mathbb{E}_v v = 0$ and $\mathbb{E}_v v^2 = 1/3$. At $b = 0$ all transverse contributions vanish and the formulas extend continuously. Introduce η, p, u by

$$\delta a = pv, \quad \delta h = 4Lh\eta v, \quad \delta b_{\pm} = b_{\pm}(-4\mu\eta v \pm u).$$

Differentiating the particle equations, with zero initial variations, gives the closed forced system

$$\eta' = p, \quad p' = 4(L^2h + \mu^2b)\eta - \mu bu + cbr, \quad u' = 4car, \quad c = 1/\sqrt{5}. \quad (8)$$

Thus η is an integrated output-weight variation, not δa . The identity

$$W := ap - q\eta - bu/4 = 0$$

follows by differentiating W in (8); every term cancels, including the forcing terms. An impulse of unit exposure mass at t has right initial data

$$\eta(t+) = 0, \quad p(t+) = cb_t, \quad u(t+) = 4ca_t.$$

After the impulse u is constant and $W = 0$, hence

$$(\eta/a)' = ca_tb/a^2, \quad \eta(s) = ca_t a_s \mathcal{J}_{t,s}, \quad \mathcal{J}_{t,s} = \int_t^s b_{\xi}/a_{\xi}^2 d\xi.$$

Division by a is legitimate on the nonvanishing branches established below; the invariant comparison ensures existence on the full interval.

Angular averaging of δP_B gives

$$C_{s,t}^{\text{atom}} = cb_s(p_s/3 + a_s u_s - (4\mu/3)a_s \eta_s).$$

For general forcing, differentiating this expression gives $\partial_s \mathbb{E}_v \delta P_B = \kappa_{\perp}^{\text{atom}} r + \delta \Gamma_{AB}^{\text{atom}}$, with

$$\delta \Gamma_{AB}^{\text{atom}} = c \left[\frac{2}{5} D u + \frac{16}{75} b(a^2 + h + b)\eta \right], \quad D = b(2h - 2a^2 - b).$$

Substituting the impulse solution gives the explicit kernel split

$$m(s, t) = c^2 \left[\frac{16}{75} a_t a_s b_s (a_s^2 + h_s + b_s) \mathcal{J}_{t,s} + \frac{8}{5} a_t D_s \right], \quad (9)$$

where $\mathcal{J}_{t,s} = \int_t^s b_{\xi}/a_{\xi}^2 d\xi$ and $D = b(2h - 2a^2 - b)$. Although an individual particle can contribute negatively to m , we obtain a nonnegative contribution after pairing the two signs of the initial output weight.

For general forcing, Duhamel's formula gives $\delta \mathbb{E} P_B(s) = \int_0^s C_{s,t}(B, B) r(t) dt$. Differentiate and use the diagonal identity for C to obtain

$$Lz' = \kappa_{\perp} r + c, \quad c(s) = \delta \Gamma_{AB}(s) = \int_0^s M(s, t) r(t) dt.$$

This also follows by linearizing (5); the two normalizations agree. Here the expectation includes both output signs, the four (A, R) masses, and density (4). The factor $1/3$ from the remaining angular variable has already been included in the split.

Time reversal $S(a, w) = (-a, w)$ intertwines $+\lambda$ and $-\lambda$. Write $\bar{m} = \frac{1}{2}(m(s, t) - m(-s, -t))$.

On every Lemma 2 characteristic, a does not change sign on $[-2/5, 2/5]$. Put $P = 8a^4 + 14a^2b + b^2 - 6bh + 8h^2$. Write $S = h + b$, $x = h/S$, and $\rho = a^2/S$. Completing the square in

$$\frac{P}{S^2} = 15\left(x - \frac{7\rho + 4}{15}\right)^2 + \frac{71\rho^2 + 154\rho - 1}{15}$$

shows $P > 0$ for $\rho > \rho_* := (-77 + 20\sqrt{15})/71$. The $I > 0$ atoms have $\rho > 1/2$; the remaining atom has $\rho > 1/25 > \rho_*$ throughout this window. Indeed, write $I = -J$ for that atom. Its radicals give $0 < J < 9/16$ and $A^2 > J$. On the positive- a branch, $|a'| \leq 2L(a^2 + J)$, so $\theta = \arctan(a/\sqrt{J})$ satisfies $|\theta'| \leq 2L\sqrt{J} < 6/5$. Thus $\theta > \pi/4 - 12/25 > 3/10$ on $[-2/5, 2/5]$; this also precludes a first zero of a . Since $\rho = \frac{1}{2}\sin^2 \theta$, the bound $\sin(3/10) \geq 3/10 - (3/10)^3/6$ gives $\rho > 1/25$. The $I > 0$ atoms cannot reach $a = 0$ by the invariant, and the negative branches follow by time reversal. Hence $D' = \frac{4}{5}ab(2a^2 + 3b + 2h) \geq 0$ and $(aD)' = \frac{b}{5}P \geq 0$ on the $+A$ branch.

For the local statement write $c_{\text{cross}}(s) = \int_0^s M(s, t) dt$, the coupling response to constant unit forcing. It is denoted c_1 in the barrier argument.

Theorem 3 (Local opening). *On the Lemma 2 law, along the pure- A germ,*

$$\kappa_{\perp}(0) = \frac{88}{125}, \quad \kappa_{\perp}''(0) = \frac{40832}{15625} > 0, \quad c_{\text{cross}}''(0) = \frac{67584}{15625} > 0.$$

These are exact moments, of degree at most six separately in a and w .

Proof. Apply the characteristic generator $q\partial_a + 4Lah\partial_h - 4\mu ab\partial_b$ twice to $\kappa_{\perp}^{\text{atom}} = c^2(b^2/3 + 4a^2b)$. For the constant-forcing response, differentiate the forced equations (8) twice with $r = 1$ and zero initial variations. Output-sign averaging removes the first derivatives. The axial monomial rule in Appendix A, together with the displayed initialization masses, gives the stated rational second derivatives. $\square$

Theorem 4 (Paired characteristic positivity). *For every Lemma 2 atom and every $y \in [0, 1]$,*

$$\bar{m}_{A,R}(s, t; y) \geq 0 \quad (0 \leq t \leq s \leq 2/5).$$

Averaging against (4) and the four atom masses, $M(s, t) \geq 0$ on the same triangle, and $M(s, t) > 0$ for $s > 0$.

Proof. The $\mathcal{J}$ term of (9) is nonnegative for both signs. The paired second term is a positive multiple of $a(t)D(s) - a(-t)D(-s)$. Time reversal identifies $a_-(\lambda) = -a_+(-\lambda)$ and $D_-(\lambda) = D_+(-\lambda)$. Monotonicity of D and $G := aD$ gives

$$a(t)D(s) \geq a(t)D(t) = G(t) \geq G(-t) = a(-t)D(-t) \geq a(-t)D(-s)$$

on $0 \leq t \leq s \leq 2/5$. For $y < 1$, $b > 0$ and $P > 0$, so G is strictly increasing. If $t > 0$ the middle inequality is strict; if $t = 0 < s$, strict increase of D gives strictness instead. This set has positive Haar mass. At $y = 1$ the transverse contribution is zero, as required. $\square$

Lemma 4 (Invariant comparison). *Every Lemma 2 characteristic exists on $[-2/5, 2/5]$.*

Proof. The invariant I gives $h + b = 2(a^2 - I)$. The bound $|a'| \leq 2L(a^2 - I)$ controls both time directions on the positive branch; time reversal covers the negative branch. Comparison with $Y' = 2L(Y^2 - I)$, $Y(0) = A > 0$, explodes at

$$\tau(A, I) = \frac{1}{4L\sqrt{I}} \log \frac{A + \sqrt{I}}{A - \sqrt{I}} \quad (I > 0), \quad \tau(A, -I) = \frac{\arctan(\sqrt{I}/A)}{2L\sqrt{I}} \quad (I = -J < 0).$$

Writing $a = A + v$ gives $\tau = \int_0^\infty [2L(v^2 + 2Av + R^2/2)]^{-1} dv$. It is decreasing in both A and R , so the high-high atom is worst. The four atoms therefore have times strictly larger than $2/5$ once that atom is checked. The worst atom has $s^2 = I/A^2 > 4/5$, because $7 + \sqrt{14} < 2(3 + \sqrt{6})$ reduces to $14 < 81/4$. The first 20 terms give $\tau \geq (2LA_+)^{-1} \sum_{k=0}^{19} (4/5)^k / (2k+1) > 0.4318 > 2/5$; the last inequality is independently interval-checked with the freeze comparison. Every omitted term is positive. $\square$

Theorem 5 ($O(4)$ sector). *Along the pure- A flow of a Haar, even-in- a initialization, on a common existence interval with finite differentiable response moments,*

$$C_{s,t}(B_1, B_2) = \gamma(s, t) \operatorname{tr}(B_1 B_2) \quad (B_i \in \mathcal{V}).$$

For the specific initialization of Lemma 2, Theorem 4 supplies one positive direction. Hence $M(B, B) \geq 0$ for every $B \in \mathcal{V}$ on $0 \leq t \leq s \leq 2/5$, and $c''_{\text{GS}}(0) = c''_{\text{block}}(0) = 67584/15625$.

Proof. The pure- A field and Haar initialization are $O(u^\perp)$ -invariant. The representation $\mathcal{V} \cong Y_2(S^3)$ is irreducible, so the only invariant quadratic form is a multiple of the Frobenius inner product. Schur's lemma therefore extends the block-pair identity to every equal-norm probe in this sector. $\square$

Convexity $\phi'' \geq 0$ on $[-2/5, 2/5]$ gives only the baseline

$$\int_0^{2/5} \kappa_A \geq \frac{2}{5} \cdot \frac{88}{125} = \frac{176}{625} = 0.2816 < L = \frac{4}{5},$$

which is not enough by itself.

Lemma 5 (Freeze before $2/5$). *The high- A , high- R atom has weight*

$$w_{++} = \left(\frac{1}{2} - \frac{1}{\sqrt{6}}\right) \left(\frac{1}{2} - \frac{1}{\sqrt{14}}\right) > \frac{213}{10000},$$

by the integer comparison $9837749^2 > 21 \cdot 2130000^2$. Using the truncated exponential E_8 and Haar integration on $y \geq 9/20$, an elementary integrand $\mathcal{J}_8$ satisfies $\mathcal{J}_8 \geq 25.020665 > 25$ by outward interval Riemann (512 panels). Hence

$$\int_0^{2/5} \kappa_A > \frac{176}{625} + \frac{213}{10000} \cdot 25 = \frac{8141}{10000} > \frac{4}{5},$$

so $\lambda_* < 2/5$.

Proof. Write $T = 2/5$, $A_+ = \sqrt{3 + \sqrt{6}}$, $R_+^2 = (7 + \sqrt{14})/5$, $x = y^2$, and $E_8(z) = \sum_{k=0}^8 z^k/k!$. On the positive high atom, $x \geq 81/400 > \mu$ gives $q(0) > 0$. Since $q' = 4a(L^2h + \mu^2b) > 0$, we have $a \geq A_+$,

$h(t) \geq R_+^2 x e^{4LA_+ t}$, and $b(t) \leq R_+^2(1-x)$. Integrating $a' = Lh - \mu b$ once and twice, and using $e^z \geq E_8(z)$ for $z \geq 0$, gives $a(T) \geq \underline{a}(x)$ and $\int_0^T a \geq \underline{\zeta}(x)$, where, with $E = E_8(4LA_+T)$,

$$\begin{aligned}\underline{a}(x) &= A_+ + \frac{R_+^2 x}{4A_+}(E-1) - \mu R_+^2(1-x)T, \\ \underline{\zeta}(x) &= A_+T + \frac{R_+^2 x}{4A_+} \left(\frac{E-1}{4LA_+} - T \right) - \frac{\mu R_+^2(1-x)T^2}{2}, \\ \underline{q}(x) &= LR_+^2 x E_8(4L\underline{\zeta}(x)) - \mu R_+^2(1-x).\end{aligned}$$

Both affine slopes are positive. At $x = 81/400$ interval arithmetic gives $\underline{a} > 3.07$, $\underline{\zeta} > 1.004$, and $\underline{q} > 8.27$. Moreover $\underline{q}' > 0$ because $E'_8 = E_7 > 0$ on this domain. It follows that $q(T) \geq \underline{q}(x) > 0$ and products preserve order.

Since $(aq)' = \phi \geq 0$, the sign-paired integral of ϕ is bounded below by $[\underline{a} \underline{q} - A_+ R_+^2(x - \mu)]/2$. For every atom and angle, even pairing and convexity also give the baseline $T\phi(0, x)$. We improve the baseline only on the high atom and the angular cap. Explicitly,

$$\begin{aligned}\phi_0(x) &= [R_+^2(x - \mu)]^2 + 4A_+^2 R_+^2 \{L^2 x + \mu^2(1-x)\}, \\ \mathcal{J}_8 &= \int_{9/20}^1 \frac{3}{2}(1-y^2) \left\{ \frac{\underline{a}(y^2)\underline{q}(y^2) - A_+ R_+^2(y^2 - \mu)}{2} - T\phi_0(y^2) \right\} dy.\end{aligned}$$

Adding the cap correction is valid even where its integrand is negative. A separate implementation with 50-decimal-digit interval arithmetic and 512 rational panels encloses $25.0206653827 < \mathcal{J}_8 < 25.4271031699$. The displayed decimal endpoints are rounded outward. Together with the exact atom-weight comparison, this proves the claim. $\square$

Theorem 6 (Pairing through freeze). *Lemma 5 puts freeze before $2/5$, so Theorems 4 and 5 give $M \geq 0$ on $0 \leq t \leq s < \lambda_*$.*

6 Linearized residual dynamics

Let z be the linearized weak residual and Δ the dominant residual, both starting at -1 . On the exposure clock,

$$r' = \frac{Gr - c}{Lg}, \quad G = \kappa_A - \kappa_\perp, \quad g = -\Delta, \quad c(\lambda) = \int_0^\lambda M(\lambda, t) r(t) dt. \quad (10)$$

Define $c_1(\lambda) = \int_0^\lambda M(\lambda, t) dt$ and the step barrier $H_1 = G - c_1$. Write $F = Gr - c$.

Lemma 6 (Constant-forcing bootstrap). *For the initialized linear-response system above, suppose $0 < T < \lambda_*$, $H_1 > 0$ on $(0, T]$, $g > 0$ on $[0, T]$, and $M \geq 0$ on $0 \leq t \leq s \leq T$. Then $F > 0$, $r' > 0$, and $r > 1$ on $(0, T]$. In particular $z < \Delta < 0$.*

Proof. Even germs give $F(0) = F'(0) = 0$. Since $r(0) = 1$ and $r'(0) = 0$, differentiating the convolution at zero gives

$$F''(0) = H_1''(0) = \kappa_A''(0) - \kappa_\perp''(0) - c''_{\text{cross}}(0) = \frac{53504}{3125} > 0.$$

Thus $F > 0$ on a right neighbourhood of 0. If $\tau \in (0, T]$ were a later first zero of F , then $r' \geq 0$ on $[0, \tau]$, so r is nondecreasing and $c(\tau) \leq r(\tau)c_1(\tau)$. Hence $F(\tau) \geq r(\tau)H_1(\tau) > 0$, a contradiction. $\square$

The bootstrap requires the stronger condition $H_1 > 0$ in addition to $M \geq 0$. This barrier need not be positive for each atom separately; its population bound depends on angular averaging.

On $r \equiv 1$ the barrier is a functional of the base characteristics alone. With $\zeta' = a$ and $J' = b\zeta/a^2$,

$$c_1 = c^2 \mathbb{E} \left[\frac{8}{5} D\zeta + \frac{16}{75} b(a^2 + h + b)a \int_0^\lambda \frac{b(s)\zeta(s)}{a(s)^2} ds \right]. \quad (11)$$

Lemma 7 (Early step barrier). *Taylor-1 Picard tubes (Nedialkov et al., 1999; Tucker, 2011; Moore et al., 2009) with outward IEEE nextafter (IEEE, 2019), Haar y-cells, outward-rounded Lemma 2 radicals, and directed Haar sums enclose*

$$H_1'' \geq 16.485 \quad \text{on } [0, 0.045], \quad H_1 \geq 0.000648 \quad \text{on } [0.045, T_0]$$

at 140 cells and $h = 10^{-3}$. A separate 40-decimal-digit interval checker validates every proposed trajectory box and recomputes the bounds without importing the IEEE implementation. Also $g \geq 0.343$ through T_0 , $0.3204 < T_0$, and $a \neq 0$ on tubes. H_1 is evaluated on each time tube by $H_1(X_n) + [0, h]H_1'(Y)$, with H_1'' on Y .

Theorem 7 (Linearized no-hit on $(0, T_0)$). *Assume Lemma 2, $d = 5$, and any equal-norm $B \in \mathcal{V}$. Then $r > 1$ and $z < \Delta < 0$ on $(0, T_0)$.*

Proof. Lemma 6 on $(0, T_0)$, using Lemma 7 and $M \geq 0$ from Theorems 4–5. Equal-norm modes share H_1 by Theorem 5. $\square$

Lemma 8 (Late step barrier). *The same Picard method encloses*

$$H_1(\lambda) \geq 4.12036 \quad \text{for } 0.3204 \leq \lambda \leq 2/5$$

at 140 cells, verified by the same independent checker. Minima are taken over complete time tubes, not grid endpoints only.

Theorem 8 (Linearized no-hit through freeze). *Assume Lemma 2, $d = 5$, and any equal-norm $B \in \mathcal{V}$. Then $r > 1$ and $z < \Delta < 0$ on $(0, \lambda_*)$. The weak residual approaches 0 from below as quadratic time tends to infinity.*

Proof. Lemma 6 on every compact $T < \lambda_*$, using Lemmas 7–8 and Theorem 6. For the limiting statement, the linearized residual equation in the quadratic clock is $\dot{z} = -(\kappa_\perp z + c\Delta)/L$. Since $M \geq 0$ and $r > 1$, $c \geq 0$; hence $\dot{z} \geq -(\kappa_\perp/L)z > 0$ while $z < 0$. The compact characteristic family extends beyond λ_* , and $\kappa_\perp$ is continuous and strictly positive there: its nonnegative integrand contains $a^2 b$, positive on a set of positive initial mass. Thus it has a positive lower bound k_* on $[0, \lambda_*]$. The equation $\dot{g} = -(\kappa_A/L)g$, with bounded κ_A , prevents g from reaching zero at finite quadratic time. Exposure is increasing and bounded above by λ_* . If its limit were strictly smaller, continuity would bound g away from zero and $\dot{\lambda} = g$ would force further growth. Hence $\lambda \rightarrow \lambda_*$ as $\sigma \rightarrow \infty$. Comparison gives $0 < -z(\sigma) \leq e^{-k_*\sigma/L}$, proving arrival at zero from below. $\square$

Figure 1 shows a numerical trajectory truncated near T_0 . Both residual magnitudes decrease, while $r = |z|/|\Delta|$ increases from one. The rigorous bounds are supplied by Lemmas 7–8.

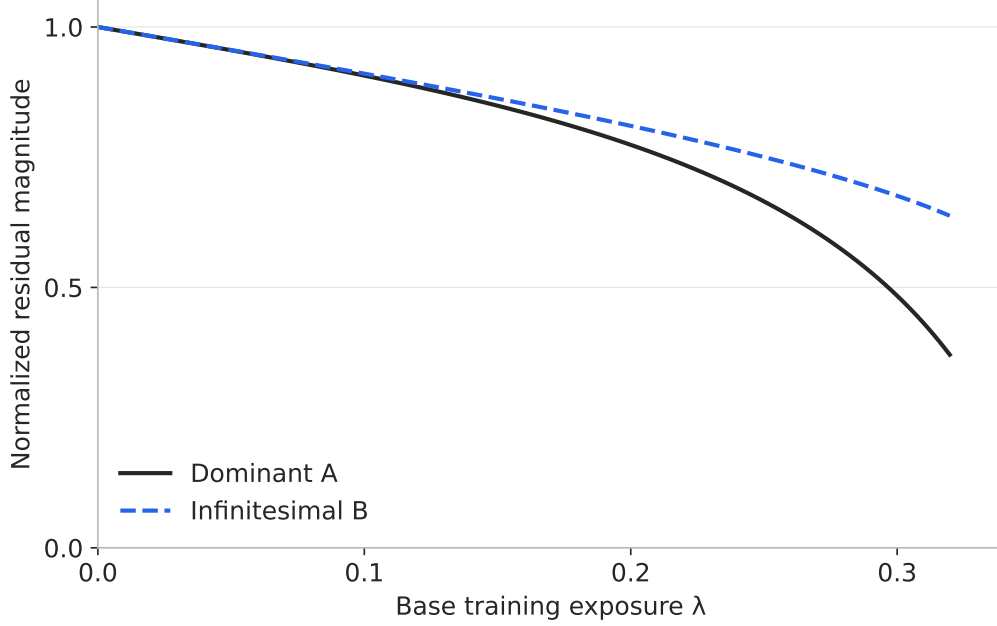


Figure 1: Ordinary RK4/quadrature illustration of the dominant residual and normalized infinitesimal transverse residual, plotted against base exposure. The weak residual decreases while remaining larger in magnitude. The figure is generated from `figure-trajectory.json`; it is not a certificate. The certified barrier is given in Lemmas 7–8.

6.1 Finite perturbations before freeze

All comparisons in this subsection use the *base* exposure $\lambda = \lambda_0(\sigma)$, also for the perturbed trajectory. This is a common coordinate on $[0, \sigma_0(T)]$ for $T < \lambda_*$, where $\sigma_0(T) = \int_0^T g^{-1} d\lambda < \infty$. It avoids comparing trajectories at different physical times.

The particle vector field and its population moments define a smooth ODE on bounded continuous functions of the compact initial labels. On a compact interval strictly before freeze, the base trajectory has a bounded continuation and $g \geq g_T > 0$. Standard parameter dependence therefore supplies a common bounded tube and bounded parameter and time derivatives for sufficiently small $|\varepsilon|$. For either closing pair, Taylor expansion gives, uniformly on $[0, T]$,

$$\begin{aligned} \beta_2^\varepsilon &= \varepsilon z + O_T(\varepsilon^2), & \beta_1^\varepsilon &= \Delta + O_T(\varepsilon^2), \\ \chi_\varepsilon &= \varepsilon c + O_T(\varepsilon^2 \lambda^2), & \kappa_2^\varepsilon &= \kappa_\perp + O_T(|\varepsilon|). \end{aligned}$$

The vanishing first variation of β_1 follows from transverse tracelessness and $O(4)$ invariance. The weighted coupling remainder requires more than ordinary Taylor expansion. For every ε , $\chi_\varepsilon(0) = \partial_\lambda \chi_\varepsilon(0) = 0$ by isotropy and output-sign symmetry. Also $\chi_0 \equiv 0$. Apply Taylor’s formula twice in ε and twice in λ to the remainder, using a uniform bound on $\partial_\varepsilon^2 \partial_\lambda^2 \chi$ on the compact tube. This gives $|\chi_\varepsilon - \varepsilon c| \leq C_T \varepsilon^2 \lambda^2$ including at zero.

The $O(|\varepsilon|)$ capacity remainder is necessary for the GS probe

$$B = \text{diag}(0, \sqrt{3/5}, -1/\sqrt{15}, -1/\sqrt{15}, -1/\sqrt{15}).$$

An exact calculation in physical quadratic time gives

$$\partial_\sigma^2 \kappa_2^\varepsilon(0) = \frac{40832}{15625} - \frac{107008\sqrt{15}}{46875} \varepsilon + \frac{40832}{1875} \varepsilon^2.$$

To reproduce it, set $x_i = w_i^2$, $D_i = A_i + \varepsilon B_i$, apply $\mathcal{D} = (\sum_i D_i x_i) \partial_a + 4a \sum_i D_i x_i \partial_{x_i}$ twice to $(\sum_i B_i x_i)^2 + 4a^2 \sum_i B_i^2 x_i$, and take the compact-law moments. The residual-feedback term is a multiple of $\mathbb{E} \mathcal{D} \phi = 0$ by output-sign symmetry. These moments are evaluated exactly in the release’s symbolic checker. The nonzero linear term rules out a quadratic capacity remainder. For the block pair, exchanging its two transverse blocks sends ε to $-\varepsilon$, so capacity is even and does admit an $O_T(\varepsilon^2)$ remainder. The common weaker bound above suffices for both pairs. No infinite-time interchange of the $\varepsilon \rightarrow 0$ and $\sigma \rightarrow \infty$ limits is made.

Theorem 9 (Small- ε usefulness). *Under the initialization of Lemma 2 and the linearized result of Theorem 8, for the $d = 5$ block pair or Gram–Schmidt pair, and every $0 < T < \lambda_*$, there exists $\varepsilon_T > 0$ such that $0 < \varepsilon < \varepsilon_T$ implies $|\beta_2|$ decreases strictly on $[0, T]$, β_2 does not change sign, and*

$$\chi_\varepsilon(\lambda) \geq \frac{1}{2} m_T \varepsilon \lambda^2 > 0 \quad (0 < \lambda \leq T).$$

The interval uses the base exposure defined above. The threshold ε_T may tend to zero as T approaches λ_* .

Proof. Theorem 8 and compactness give $\inf_{[0, T]} |z| > 0$ and $\Delta \leq -\delta_T < 0$. Strict positivity of the paired kernel gives $\mathfrak{c}(\lambda) > 0$ for $\lambda > 0$, while its germ is $\mathfrak{c}(\lambda) = \frac{1}{2} c''_{\text{cross}}(0) \lambda^2 + o(\lambda^2)$. Consequently $m_T := \inf_{0 < \lambda \leq T} \mathfrak{c}(\lambda) / \lambda^2 > 0$. Smooth parameter dependence on the compact tube, together with $\chi_\varepsilon(0) = \partial_\lambda \chi_\varepsilon(0) = 0$ for every sufficiently small ε (isotropy and evenness in a), gives the uniform remainder $|\chi_\varepsilon - \varepsilon \mathfrak{c}| \leq C_T \varepsilon^2 \lambda^2$. Choose ε_T small enough that $C_T \varepsilon_T \leq m_T/2$ and the residual expansions preserve $\beta_1 < 0$ and $\beta_2 < 0$ on $[0, T]$. Then $\chi_\varepsilon \geq m_T \varepsilon \lambda^2 / 2$. Finally, $\dot{\beta}_2 = -(\chi_\varepsilon \beta_1 + \kappa_2 \beta_2) / L > 0$, since $\kappa_2 > 0$ persists on the compact tube. This proves strict decrease of $|\beta_2|$; the positive change of clock preserves it. $\square$

7 Limitations and counterexamples

The assumptions and the order of averaging are essential to the results. Lemma 1 excludes unbounded Gaussian initialization from the classical exposure flow used here. The positivity of M requires pairing, and the positivity of H_1 requires angular averaging; neither follows from positivity of each individual contribution. Moreover, nonnegative coupling alone does not establish the no-hit conclusion, which uses the barrier in Lemma 6.

The accompanying numerical experiments also distinguish capacity from training performance. In a finite-network example, two datasets have integrated target kernel masses of 37.50 and 87.13, but final target errors of 0.00987 and 8.29, respectively. Thus the larger integrated mass need not identify the better training dataset.

For finite mixing, float64 simulations of the compact-initialization flow give $\chi_{\min} \approx -0.373$ for the Gram–Schmidt pair at $\varepsilon = 0.8$. For the block pair at $\varepsilon = 2$, they give $\max \beta_2 \approx 1.2 \times 10^{-3}$, indicating overshoot. These are numerical indications of failure, not interval-certified counterexamples to statements about the continuum flow. Their runners and parameters are included in the accompanying code.

Whether the weak residual remains negative for every $\varepsilon \in (0, 1]$ is unresolved by the present analysis. The finite-mixing result in Theorem 9 is restricted to sufficiently small ε on each compact exposure interval.

8 Related work

The residual law $\dot{r} = -Kr$ in a frozen tangent kernel is the neural tangent kernel of Jacot et al. (2018) and the infinite-width linearisation of Lee et al. (2019). Lazy training (Chizat et al., 2019) and

feature-learning parameterisations (Yang and Hu, 2021) explain when that kernel moves. Kernel alignment during training (Baratin et al., 2021; Shan and Bordelon, 2021; Atanasov et al., 2022) is a spectral statement about K_t and the target. Dynamical mean-field theory describes the evolution of kernel order parameters in wide networks (Bordelon and Pehlevan, 2022; Lauditi et al., 2026). Here we focus on a transverse mode and distinguish its diagonal capacity, its coupling to the dominant mode, and its relative residual decay.

Two-layer mean-field theory supplies a distributional PDE (Mei et al., 2018; Chizat and Bach, 2018; Rotskoff and Vanden-Eijnden, 2022) and a kernel limit under extra regularity (Mei et al., 2019). For the cubic exposure flow considered here, the Gaussian characteristic blow-up in Lemma 1 motivates the compact initialization of Lemma 2.

Martin et al. (2026) study extensive-width quadratic networks with Gaussian data, uniform output weights, and a high-dimensional finite-sample limit; their conclusions concern recovery and estimator spectra. Ben Arous et al. (2025) analyze population gradient flow and online SGD with nonnegative, power-law teacher strengths. Their Theorem 1 gives feature-alignment transitions and a limiting risk curve. The student trains its first-layer matrix rather than signed output weights. These results directly address stronger and weaker quadratic features, but their assumptions and conclusions differ from the fixed- $d = 5$, compact-population, transverse sign theorem here.

Earlier, Sarao Mannelli et al. (2020) study Gaussian-input quadratic networks with fixed positive second-layer weights, including optimization and population convergence. This is another relevant quadratic flow, not a theorem about the sign of the present transported response. In contrast, Bordelon and Pehlevan (2022) provide a general self-consistent kernel-evolution framework in an infinite-width feature-learning regime; it should not be described as the same quadratic model. The silent-alignment theorem of Atanasov et al. (2022) relates a final predictor to a final kernel under a two-phase kernel evolution. It does not assert this transverse no-hit property. Feature-learning limits (Yang and Hu, 2021) and learned spectral dynamics (Lauditi et al., 2026) provide broader context. We make no priority claim for the general premise that training changes useful features.

Pairing positivity in Theorems 4–5 uses time-reversal and monotonicity of D and aD . This argument is distinct from the Green-kernel positivity methods associated with disconjugacy (Coppel, 1971).

Computer-assisted claims use interval Picard tubes for ordinary differential equations (Nedialkov et al., 1999; Tucker, 2011) with directed rounding from IEEE 754 (IEEE, 2019) and the interval calculus of Moore et al. (2009). The computation combines these methods with Haar integration, outward-rounded initialization, and evaluation of the functional H_1 .

9 Conclusion

The result is a sign-controlled response and residual barrier for a specified quadratic population model. It separates accessible parameter directions from the direction in which training actually moves an error. The linearized transverse residual converges without crossing zero; finite perturbations are controlled only on compact intervals before the limiting exposure. Extending that finite-perturbation result uniformly to infinite time remains open.

A Numerical verification

The numerical obligations are the freeze comparison and the scalar barrier bounds. The files under `certificates/` contain all initial boxes, step endpoints, and continuous time tubes for eight signed

atoms, 140 angular cells, and 400 time steps. The time step is interpreted as its exact binary64 value; 400 steps cover $[0, 2/5]$.

The independent checker imports neither the original interval primitives nor its success flags. With 40-decimal-digit `mpmath.iv` arithmetic, it independently constructs the radical initialization and the exact polynomial Haar cell integrals. On every cell it checks

$$X_n + [0, h]f(Y_n) \subseteq Y_n, \quad X_n + hf(X_n) + \frac{1}{2}h^2(Df f)(Y_n) \subseteq X_{n+1}.$$

The intersections with $h, b \geq 0$ use the analytic invariance of those coordinates. Every a -tube excludes zero. A separate implementation of second-order automatic differentiation evaluates H'_1 , H''_1 and κ'_A on the entire tube. Positive atom and angular weights are then applied and summed. The enclosure $H_1(X_n) + [0, h]H'_1(Y_n)$ covers every intermediate time. The checker verifies the initial domain, continuous coverage, aggregation, overlap of the early and late windows, and all positivity inequalities used in the no-hit theorem. It does not rely on agreement between two grids. The common mathematical vector field and barrier formula remain shared; the reduction to these formulas is proved in the text.

Exact symbolic checks separately verify the invariants, capacity identity, Wronskian, coupling split, GS expansion and zero-time germs. For clarity, the axial monomial rule used to reproduce the latter is

$$\mathbb{E}[a^{2m} h^n b^p] = \mathbb{E}[a^{2m}] \mathbb{E}[R^{2(n+p)}] \frac{(1/2)_n (2)_p}{(5/2)_{n+p}}, \quad \mathbb{E}[a^{2m+1} h^n b^p] = 0.$$

Here $(x)_k$ is the rising factorial. Apply the polynomial characteristic generator to the displayed observables before evaluating this rule. For c_1 also use $\zeta' = a$, $J' = b\zeta/a^2$ and $\zeta(0) = J(0) = 0$. This yields respectively the second derivatives 375936/15625, 40832/15625, 67584/15625, and 53504/3125 for $\kappa_A, \kappa_\perp, c_1, H_1$. These are exact algebra, not fitted coefficients.

The independent freeze checker uses 50 decimal digits and rational panel endpoints. It verifies the comparison slopes, endpoint positivity, atom weight, common-existence lower bound and integral threshold. The remaining trusted components are the analytical reductions, Python and its standard library, SymPy for symbolic simplification, mpmath interval arithmetic, NumPy's evidence-file decoding, and the underlying hardware. This is computer-assisted mathematics, not a formal proof-assistant verification.

A.1 Reproduction and ordinary numerical tests

Code and computational evidence are available in the 9D-Labs/learning-changes-toy GitHub repository. The accompanying research package contains the experiment code, interval evidence, independent checkers and machine-readable figure data.

From a clean installation using `requirements-lock.txt`, run

```
python scripts/check_exact_identities.py
python scripts/check_freeze_independent.py
python experiments/export_barrier_evidence.py
python scripts/check_barrier_independent.py results-candidate
python experiments/validate_perturbations.py
python experiments/validate_precision.py
python experiments/check_freeze_distance.py
python experiments/validate_perturbations.py --figures
python scripts/plot_figures.py --results-dir results-candidate
```

To check the distributed tubes without regenerating them, use

```
python scripts/check_barrier_independent.py certificates
```

The lighter `python verify.py` checks hashes and selected stored inequalities only. The release documentation maps every statement to its proof and evidence and records source and artifact hashes.

The empirical suite compares full two-mode dynamics with the linearized prediction at common physical times, for both probe pairs and $\varepsilon \in \{0.04, 0.02, 0.01, 0.005, 0.0025\}$. It halves the RK4 step from 0.002 to 0.001 and increases each quadrature order from 16 to 24. A separate 50-digit arithmetic comparison uses identical quadrature nodes through physical time 0.1; it isolates rounding from quadrature error. On the tested platform `longdouble` equals `binary64` and is not counted as a precision refinement. Numerical estimates of the freeze location are refined separately; distances smaller than their resolution are not interpreted. None of these trajectories proves uniform validity as $T \uparrow \lambda_*$.

References

- Alexander Atanasov, Blake Bordelon, and Cengiz Pehlevan. Neural networks as kernel learners: The silent alignment effect. In *International Conference on Learning Representations*, 2022.
- Aristide Baratin, Thomas George, César Laurent, R. Devon Hjelm, Guillaume Lajoie, Pascal Vincent, and Simon Lacoste-Julien. Implicit regularization via neural feature alignment. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 2269–2277. PMLR, 2021.
- Gérard Ben Arous, Murat A. Erdogdu, Nuri Mert Vural, and Denny Wu. Learning quadratic neural networks in high dimensions: SGD dynamics and scaling laws. *arXiv preprint arXiv:2508.03688*, 2025. URL <https://arxiv.org/abs/2508.03688v3>. Version 3, December 31, 2025; NeurIPS 2025.
- Blake Bordelon and Cengiz Pehlevan. Self-consistent dynamical field theory of kernel evolution in wide neural networks. In *Advances in Neural Information Processing Systems*, volume 35, pages 32240–32256, 2022.
- Lénaïc Chizat and Francis Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. In *Advances in Neural Information Processing Systems*, volume 31, 2018.
- Lénaïc Chizat, Edouard Oyallon, and Francis Bach. On lazy training in differentiable programming. In *Advances in Neural Information Processing Systems*, volume 32, 2019.
- W. A. Coppel. *Disconjugacy*, volume 220 of *Lecture Notes in Mathematics*. Springer, Berlin, 1971.
- IEEE. *IEEE Standard for Floating-Point Arithmetic*. IEEE, 2019. IEEE Std 754-2019.
- Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018.
- Clarissa Lauditi, Cengiz Pehlevan, and Blake Bordelon. Spectral dynamics in deep networks: Feature learning, outlier escape, and learning rate transfer. *arXiv preprint arXiv:2605.07870*, 2026. URL <https://arxiv.org/abs/2605.07870>.
- Jaehoon Lee, Lechao Xiao, Samuel S. Schoenholz, Yasaman Bahri, Roman Novak, Jascha Sohl-Dickstein, and Jeffrey Pennington. Wide neural networks of any depth evolve as linear models under gradient descent. In *Advances in Neural Information Processing Systems*, volume 32, 2019.
- Simon Martin, Giulio Biroli, and Francis Bach. High-dimensional analysis of gradient flow for extensive-width quadratic neural networks. *Journal of Machine Learning Research*, 27(136):1–182, 2026.
- Song Mei, Andrea Montanari, and Phan-Minh Nguyen. A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences*, 115(33):E7665–E7671, 2018. doi: 10.1073/pnas.1806579115.
- Song Mei, Theodor Misiakiewicz, and Andrea Montanari. Mean-field theory of two-layers neural networks: dimension-free bounds and kernel limit. In *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 2388–2464. PMLR, 2019.
- Ramon E. Moore, R. Baker Kearfott, and Michael J. Cloud. *Introduction to Interval Analysis*. SIAM, Philadelphia, 2009.
- Nedialko S. Nedialkov, Kenneth R. Jackson, and George F. Corliss. Validated solutions of initial value problems for ordinary differential equations. *Applied Mathematics and Computation*, 105(1):21–68, 1999. doi: 10.1016/S0096-3003(98)10083-8.

- Grant Rotskoff and Eric Vanden-Eijnden. Trainability and accuracy of artificial neural networks: An interacting particle system approach. *Communications on Pure and Applied Mathematics*, 75(9):1889–1935, 2022. doi: 10.1002/cpa.22074.
- Stefano Sarao Mannelli, Eric Vanden-Eijnden, and Lenka Zdeborová. Optimization and generalization of shallow neural networks with quadratic activation functions. In *Advances in Neural Information Processing Systems*, volume 33, pages 13445–13455, 2020.
- Haozhe Shan and Blake Bordelon. A theory of neural tangent kernel alignment and its influence on training. *arXiv preprint arXiv:2105.14301*, 2021. URL <https://arxiv.org/abs/2105.14301>.
- Warwick Tucker. *Validated Numerics: A Short Introduction to Rigorous Computations*. Princeton University Press, Princeton, 2011.
- Greg Yang and Edward J. Hu. Tensor programs IV: Feature learning in infinite-width neural networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 11727–11737. PMLR, 2021.