BRANCH GEOMETRY AND FINITE-RADIUS SENSITIVITY OF HARD-RELU TRAINING

Anonymous authors
Paper under double-blind review

ABSTRACT

Differentiating a training run gives a local sensitivity, whereas an outer update makes a finite change to its initialization or training parameters. For hard-ReLU gradient descent (GD), we characterize the response when the perturbation radius is proportional to the GD step size. In this regime, integer changes in activation-crossing indices contribute at leading order. Smooth Euler bias shifts each crossing phase, and earlier rounding decisions move later branch boundaries. For piecewise-C^2 dynamics with finitely many separated, same-direction transverse events, we derive the discrete crossing indices and an endpoint expansion with uniform $O(\eta^2)$ error away from recursive phase boundaries. Contractive affine regions admit an explicit finite remainder that, together with complete branch verification, certifies candidate ordering. A strongly convex scalar example reverses both infinitesimal predictions on 10.55% of 9,801 tested meshes, close to the proved asymptotic frequency of 10.53%. Coupled-event experiments isolate downstream rounding feedback; a nonlinear-network study shows how prediction accuracy changes with radius and branch validity. Together with local automatic differentiation and uniform flow consistency, the theory gives three sufficient response regimes. It characterizes finite objective comparisons under stated conditions; it does not guarantee optimizer gains.

1 INTRODUCTION

An outer-learning algorithm turns a sensitivity into a finite proposal: change an initialization, reweight training examples, or adjust a continuous hyperparameter, then train again. Differentiation through finite optimization is used in hyperparameter learning, bilevel optimization and meta-learning (Maclaurin et al., 2015; Franceschi et al., 2018; Finn et al., 2017; Ren et al., 2018). The operational question is: *at a prescribed radius ϵ, which sensitivity predicts the objective difference between two proposals after finite training?* We study fixed-step hard-ReLU gradient descent (GD), holding the task, perturbation direction and training horizon fixed.

At a nonresonant input, automatic differentiation (AD) gives the exact derivative of the executed training program. Its finite-radius accuracy depends on the surrounding activation cell. The program is smooth inside that cell but can jump when the perturbation changes a discrete crossing index. Refining the GD mesh changes those cells; near resonance their radius can be much smaller than the step size η. The general distinction between an exact hypergradient and an informative finite-update prediction has precedent (Maclaurin et al., 2015, Sec. 4). Here it becomes a quantitative question about the geometry of the refining program.

Continuous training-flow sensitivity shifts event times continuously. Finite GD changes them in integer steps. At $\epsilon = \zeta\eta$ with fixed $\zeta > 0$, a one-step change contributes at the same order as the proposed perturbation. Smooth Euler error also enters at this order: it changes the incoming state and hence which integer is selected. With coupled events, the first rounding decision shifts the next discrete boundary. The response is then generally phase-dependent and need not equal either infinitesimal prediction. Figure 1 shows this resolution between a branch tangent and a continuously shifted event time.

Relation to classical sensitivity analysis. Saltation matrices, event-time sensitivity, finite perturbation propagation and its downstream event changes are classical (Kong et al., 2024; Wardi

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

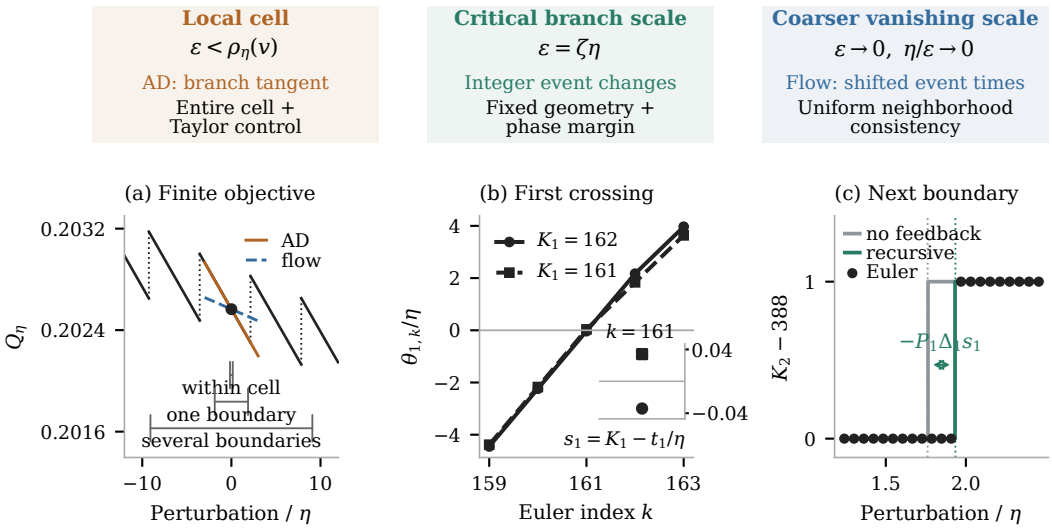

Figure 1: **One objective, three perturbation resolutions.** Top: sufficient regimes, with their separate hypotheses; these are not a universal partition of an ϵ/η axis. (A) Exact scalar finite-GD branches at $N = 1000$, with the same base point and intervals crossing zero, one or several boundaries. Tangents use that base objective; the flow slope differentiates the continuous objective. (B) Two nearby inputs cross the first surface at different integer indices. (C) The transported rounding contribution moves the second discrete boundary; full recursion, ablation and independently rerun Euler indices are compared. Windows in (B)–(C) are fixed by geometry, without using the objective. Curves and points are computed; colored cards and arrows explain which mechanism each resolution retains.

et al., 2017; Cassandras et al., 2010). So are discretized sensitivity failure and event-aware recovery schemes (Stewart & Anitescu, 2010; Nurkanović et al., 2024). Our result concerns the joint limit of the changing Euler program and a perturbation ηz. Theorem 2 derives its actual integer crossings, including smooth Euler bias and ordered upstream rounding, and proves an endpoint expansion uniform away from the recursively computed boundaries. A separate affine bound controls a finite comparison. The contribution is this uniform characterization of the response as the program and perturbation scale change together.

Three contributions.

1. **Finite-radius response theory.** For a fixed finite number of separated transverse events, Theorem 2 derives a recursive phase-resolved response for piecewise-C^2 dynamics at $\epsilon = \Theta(\eta)$. Its uniform expansion supplies the critical part of a radius-resolution principle connecting branch-preserving AD and flow sensitivity.

2. **Finite comparison control.** Theorem 3 gives explicit remainder bounds in contractive affine regions. Complete word verification and a strict error margin certify two-candidate comparisons; failures remain unresolved.

3. **Mechanism tests and nonlinear response evidence.** Scalar phase frequencies and a coupled feedback ablation test the theory. A fixed width-4 network cohort tests prediction across four radii, retaining uncovered cases, incorrect branch predictions and the full error tails.

The main theorem varies initialization. Appendix H.1 treats continuous outer-parameter derivatives, and Appendix F.4 states the conditions for extending the critical law by state augmentation. Discrete architecture changes and optimizer improvement guarantees remain outside this scope.

2 THE FINITE OBJECTIVE AND ITS SENSITIVITY OBJECTS

Fix a horizon $T > 0$, an integer N and step size $\eta = T/N$. Write $\theta_{k+1} = \theta_k + \eta f_{\alpha_k}(\theta_k)$, with regional field $f_\alpha = -\nabla L_\alpha$ and endpoint $\Theta_{\eta,T}(\theta_0) = \theta_N$. The *activation word* $(\alpha_0, \dots, \alpha_{N-1})$ records every executed training mask and fixes the regional formulas. A *nonresonant* point is one where every executed preactivation has a strict sign. There,

$$J_{k+1} = (I + \eta D f_{\alpha_k}(\theta_k))J_k, \qquad J_0 = I, \qquad J_{\eta,T}^{\mathrm{AD}} = D\Theta_{\eta,T}. \tag{1}$$

At an exact landing the selected AD product may exist even when the classical finite-program derivative does not.

The foundational limit assumes bounded C^1 regional extensions, regular C^2 interfaces, finitely many separated events, isolating tubes and positive non-event margins. Each interface is locally oriented from incoming to outgoing, with both one-sided normal speeds bounded below by $\nu > 0$. At a discrete exact hit, the state convention selects an adjacent field with that positive speed. Appendix A states the full geometric conditions.

Theorem 1 (Regional limit of finite-program derivatives)**.** *For this fixed itinerary, the GD interpolation has uniform state error $O(\eta)$, and the selected derivative products converge to the regional limit J_{reg}. Writing Φ_i for ordinary variational propagation between events,*

$$\begin{aligned} J_{\mathrm{reg}} &= \Phi_m \cdots \Phi_0, \qquad \Xi_j = I + \frac{\Delta_j n_j^\top}{n_j^\top f_j^-}, \\ D\varphi_T &= \Phi_m \Xi_m \Phi_{m-1} \Xi_{m-1} \cdots \Phi_1 \Xi_1 \Phi_0. \end{aligned} \tag{2}$$

Here $\Delta_j = f_j^+ - f_j^-$ and n_j is the oriented normal; for $m = 0$, both products are Φ_0. On nonresonant meshes the selected product is the exact finite-program derivative. Its regional-limit error is $o(1)$, and $O(\eta)$ when the regional Jacobians are locally Lipschitz.

The state/tangent distinction supplies the reference objects; its mechanism is classical (Stewart & Anitescu, 2010). Appendix C.1 gives the proof under the stated geometry.

For a terminal objective q, write $Q_\eta(\lambda) = q(\Theta_{\eta,T}(\lambda), \lambda)$ and $Q_0(\lambda) = q(\varphi_T(\lambda), \lambda)$. Distinguish $g_\eta = \nabla Q_\eta$, its regional limit g_{reg}, and $g_0 = \nabla Q_0$. The latter two gradients use terminal derivatives of q at the flow endpoint; g_η uses the finite endpoint. Terminal regularity is required in addition to strict training signs. Our target is

$$D_{\eta,\epsilon}(v) = \frac{Q_\eta(\lambda + \epsilon v) - Q_\eta(\lambda - \epsilon v)}{2\epsilon}, \qquad \|v\| = 1. \tag{3}$$

Its sign compares the two actual proposals: $D_{\eta,\epsilon} > 0$ means the negative proposal has lower objective. The critical theorem treats initialization perturbations, $\lambda = \theta_0$, and $q = q(\theta_T)$. Parameter sensitivities have separate supporting formulas. A matrix gap, an objective-gradient gap and a finite proposal reversal are different evidence levels.

3 THE RADIUS-RESOLUTION PRINCIPLE

Fix the initialization λ, unit direction v and objective Q_η; write $x_T = \varphi_T(\lambda)$. Define $\rho_\eta(v)$ as the supremum of radii whose entire open perturbation interval preserves every training and terminal-validation branch. A phase condition on a few evaluated endpoints is a different test.

Response at three sufficient resolutions. The following synthesis combines Taylor's formula, Theorem 2, and uniform objective consistency.

1. **Branch-preserving: exact finite-program AD.** If $0 < \epsilon < \rho_\eta(v)$ and $|\partial_s^3 Q_\eta(\lambda + sv)| \le M_{3,\eta}$ on the interval, then $|D_{\eta,\epsilon} - g_\eta^\top v| \le M_{3,\eta}\epsilon^2/6$. The remainder must vanish along a refining-grid sequence.

2. **Critical: recursively rounded branch response.** Fix $\zeta > 0$ and $\epsilon = \zeta\eta$. Under Theorem 2's geometry, with $q \in C^2$ near x_T, on sufficiently fine grids with a common positive recursive phase margin on $K = \{0, -\zeta v, \zeta v\}$,

$$D_{\eta,\zeta\eta} = \underbrace{\frac{\nabla q(x_T)^\top [e_{p_N}(\zeta v) - e_{p_N}(-\zeta v)]}{2\zeta}}_{\widehat{D}_{\eta,\zeta}(v)} + O_K(\eta/\zeta).$$

   The interval between these points may cross branches.

3. **Coarser vanishing: flow sensitivity.** If $\sup_U |Q_\eta - Q_0| \le C_0\eta$ on a fixed neighborhood and $Q_0 \in C^1(U)$, with $\lambda + [-\epsilon, \epsilon]v \subset U$, then $|D_{\eta,\epsilon} - g_0^\top v| \le C_0\eta/\epsilon + \omega_{\nabla Q_0}(\epsilon)$, where $\omega_{\nabla Q_0}$ is the gradient's local continuity modulus. Thus $\epsilon \to 0$ and $\eta/\epsilon \to 0$ suffice.

These are sufficient regimes, not an exhaustive partition. Resonance can make ρ_η/η arbitrarily small; $\epsilon \ll \eta$ alone is insufficient. The first regime differentiates one finite program, whose regional limit need not be the flow gradient in the third regime. Each proposed displacement must be evaluated by the recursion; coordinate responses do not define a vector whose inner product predicts every other finite proposal. Appendix F.4 gives the synthesis proof, the C^2 local remainder and an explicit counterexample.

4 Critical response through finitely many events

The incoming error combines smooth Euler bias with earlier event rounding.

Fix m events at $0 < t_1 < \cdots < t_m < T$, separated from each other and the endpoints by $\Delta > 0$. On segment i, let f_i have a bounded C^2 extension and let $x(t)$ be the base path. In each isolating event tube exactly one C^2 surface $g_j = 0$ can vanish; all others have a positive margin. Away from these tubes all surfaces have a positive margin on a fixed spatial neighborhood of the base path. Orient each g_j so that both adjacent speeds $\nabla g_j^\top f_{j-1}$ and $\nabla g_j^\top f_j$ exceed $\nu > 0$. These conditions concern continuous geometry; they assume no discrete indices.

Set $t_0 = 0, t_{m+1} = T$, $x_i = x(t_i)$ and $L_i = t_{i+1} - t_i$. Let P_i be regional variational propagation and compute the smooth bias

$$\dot{w} = Df_i(x(t))w - \tfrac{1}{2}Df_i(x(t))f_i(x(t)), \quad w(t_i) = 0, \quad w_i = w(t_{i+1}). \tag{4}$$

Define $n_j = \nabla g_j(x_j)$, $u_j = n_j^\top f_{j-1}(x_j)$ and $\Delta_j = f_j(x_j) - f_{j-1}(x_j)$. At phase $p \in \mathbb{T}^m$ and scaled displacement z, put $b_0 = z$ and recurse:

$$\begin{aligned} a_j &= P_{j-1}b_{j-1} + w_{j-1}, & s_j &= \left\lceil p_j - \frac{n_j^\top a_j}{u_j} \right\rceil - p_j, \\ b_j &= a_j - \Delta_j s_j, \quad 1 \le j \le m, & e_p(z) &= P_m b_m + w_m. \end{aligned} \tag{5}$$

The phase margin is the distance of each recursively computed ceiling argument to the integers. Raw event phases alone omit the incoming error.

Theorem 2 (Finite-event critical-scale branch response)**.** *Under the stated geometry, let $K \subset \mathbb{R}^d$ be compact, $0 \in K$, and $\mu > 0$. For every sufficiently fine canonical grid $\eta = T/N$ whose phase vector $p_N = (\{t_j/\eta\})_{j=1}^m$ has recursive phase margin at least μ uniformly on K, every Euler path from $x_0 + \eta z$, $z \in K$, has exactly the base event sequence. Its first positive indices are*

$$k_j(z) = \left\lceil t_j/\eta - n_j^\top a_j(p_N, z)/u_j \right\rceil.$$

All executed signs are strict, and uniformly on K,

$$\Theta_{\eta,T}(x_0 + \eta z) = x_{m+1} + \eta e_{p_N}(z) + O_K(\eta^2). \tag{6}$$

For $q \in C^2$ near the endpoint, the rescaled objective change equals $\nabla q(x_{m+1})^\top [e_{p_N}(z) - e_{p_N}(0)] + O_K(\eta)$. On phase-convergent subsequences this becomes the corresponding e_p law with error $O_K(\eta + \mathrm{dist}(p_N, p))$. Constants depend on fixed m, T, K, field and surface derivative bounds, transversality, event separation and tube margins. The mesh threshold also depends on μ; the endpoint remainder constant need not.

From endpoint error to response. Here a_j is the incoming error at the nominal event clock, s_j is the crossing delay in units of η, and b_j is the outgoing error returned to that same clock. Applying the endpoint expansion to $z = \pm\zeta v$ gives the critical response in Section 3.

Why coupling matters. With $P_{j-1:r} = P_{j-1} \cdots P_r$, the jth scaled boundary contains

$$p_j - \frac{n_j^\top A_j^{\mathrm{free}}(z)}{u_j} + \sum_{r<j} \frac{n_j^\top P_{j-1:r} \Delta_r}{u_j} s_r \in \mathbb{Z}, \tag{7}$$

where A_j^{free} contains initial displacement and smooth bias. Changing one upstream index can therefore change several later integers. The no-upstream-feedback ablation deletes these threshold terms while retaining the transported endpoint contributions.

Scope and proof. Appendix D derives the smooth bias from the one-step truncation error, identifies each first positive index by two strict signs and normal monotonicity, and propagates the restart error $a_j - \Delta_j s_j$. C^2 regularity suffices: surface curvature enters only $O(\eta^2)$. The theorem requires no phase independence, contractivity or affine structure. Exact scaled boundaries, grazing and simultaneous events are excluded. Section 6 gives a finite bound under stronger affine assumptions.

A concrete coupled training risk. For $L(\theta) = \frac{1}{2}\|A\theta - b\|^2 + \frac{1}{2}\sum_{j=1}^{2}[\mathrm{ReLU}(n_j^\top \theta) - c_j]^2$, $A^\top A \succ 0$ and $c_j \le 0$ give global strong convexity. Regional fields are affine, with $M_S = -A^\top A - \sum_{j\in S} n_j n_j^\top$, and (4) reduces to $w_i = -L_i P_i M_i f_i(x_i)/2$. The fixed example $A^\top A = \left(\begin{smallmatrix} 1 & .4 \\ .4 & 1.3 \end{smallmatrix}\right)$, $A^\top b = (2,2)$, $n_j = e_j$, $c = (-.4, -.6)$, $\theta_0 = (-.4, -1)$ and $T = 1$ has two provably separated events. Its regional commutator has norm .4, and its downstream coefficient $n_2^\top P_1 \Delta_1 / u_2$ is strictly positive. Appendix D.6 gives the geometry and exact finite-cell jumps.

5 An exact scalar response classification

Let $L_{a,c}(x) = \frac{1}{2}(x-a)^2 + \frac{1}{2}(\mathrm{ReLU}(x) - c)^2$, with $a > 0, c > -a$ and $x \in (a(1-e^T), 0)$. For $\eta = T/N < 1/2$, set $t = \log((a-x)/a)$, $P = (a+c)/2$, $A = 1-\eta$, $B_\eta = 1-2\eta$ and $z_k = a(1-A^{-k})$. For $x \in (z_K, z_{K-1})$, $1 \le K < N$, the endpoint on this cell and the right-minus-left jump at an interior boundary $1 \le k < N$ are

$$F_\eta(y) = P + [a + (y-a)A^K - P]B_\eta^{N-K}, \qquad [F_\eta]_{z_k} = \eta c B_\eta^{N-k-1}. \tag{8}$$

The training-cell radius is $\rho_\eta = \min(x - z_K, z_{K-1} - x)$; a smooth terminal q adds no branch constraint. Writing $B = a - x$, $E = e^{-2(T-t)}$, $J = e^{-t}E$ and $F_0 = P(1-E)$, the critical response at phase $p = \lim\{t/[-\log(1-\eta)]\}$ is

$$R_\zeta(p) = q'(F_0)\left[J + \frac{cE}{2\zeta} H(p,\zeta)\right], \quad H(p,\zeta) = \#\{j \in \mathbb{Z} : B|p-j| < \zeta\}. \tag{9}$$

The phase law excludes $\zeta = B|p-j|$ for every integer j. The corrected scalar phase includes the Euler bias in (4). Its cell radius can satisfy $\rho_\eta/\eta \to 0$, so $\epsilon \ll \eta$ alone does not guarantee a branch-preserving perturbation.

For $\ell = 2\zeta/B$, the count is $\lfloor \ell \rfloor$ or $\lceil \ell \rceil$, with uniform-phase masses $1-r, r$, where $r = \ell - \lfloor \ell \rfloor$. Consequently

$$\mathbb{E}_p R_\zeta = Q_0'(x), \qquad \mathrm{Var}_p\, R_\zeta = [q'(F_0)cE/(2\zeta)]^2 r(1-r). \tag{10}$$

Positive integer ℓ cancels phase dependence on the allowed phases. For irrational t/T, an equidistribution proof with uniform error control away from boundaries yields canonical-mesh response frequencies. Corollary 22 in Appendix F.3 treats zero levels and exact-hit exceptions. This scalar classification is a consequence of branch counting; its phase average is not an expectation over training seeds or data.

6 A FINITE COMPARISON CERTIFICATE

Contractive affine regions admit an explicit finite error bound, conditional on verifying the candidate word.

Theorem 3 (Contractive finite response)**.** *Let m be fixed and $f_i(x) = M_i x + d_i$, with $M_i = M_i^\top \prec 0$. Choose supplied ordered knots $0 = t_0 < \cdots < t_{m+1} = T$, let $L_i = t_{i+1} - t_i$, $P_i = e^{L_i M_i}$, and obtain scheduled states x_i by the corresponding exact regional flows from x_0. For a candidate word require $0 = k_0 < \cdots < k_{m+1} = N$ and $\eta\|M_i\|_2 < 1$. Put $h_i = -M_i^{-1} d_i$, $\delta_i = k_{i+1} - k_i - L_i/\eta$, $A_i = P_i(\delta_i M_i - L_i M_i^2/2)$, and $B_i = P_i[L_i M_i^3/3 - \delta_i M_i^2/2 + (\delta_i M_i - L_i M_i^2/2)^2/2]$. Starting from $e_0 = z, v_0 = 0$, propagate*

$$e_{i+1} = P_i e_i + A_i(x_i - h_i), \qquad v_{i+1} = P_i v_i + A_i e_i + B_i(x_i - h_i).$$

If the entire candidate word passes its strict sign test, the endpoint $\widetilde{x} = x_{m+1} + \eta e_{m+1} + \eta^2 v_{m+1}$ has error at most the explicit spectral remainder R_{m+1} in Eq. (54) in Appendix E.2. For bounded displacements, offsets and regional data, with $\eta\|M_i\| \le r_* < 1$, the bound is uniformly $O(\eta^3)$.

Let $\widetilde{x}_\pm$ and $R_\pm$ correspond to $z = \pm\zeta v$. If q is C^2 on an open neighborhood of both endpoint-error balls and $\|\nabla^2 q\| \le M_q$ on the balls, then

$$\widehat{D}^{(2)} = \frac{q(\widetilde{x}_+) - q(\widetilde{x}_-)}{2\zeta\eta},$$
$$|D_{\eta,\zeta\eta} - \widehat{D}^{(2)}| \le \frac{\sum_{\sigma=\pm}[\|\nabla q(\widetilde{x}_\sigma)\| R_\sigma + M_q R_\sigma^2/2]}{2\zeta\eta} =: \mathcal{E}^{(2)}. \tag{11}$$

A strict margin $|\widehat{D}^{(2)}| > \mathcal{E}^{(2)}$ certifies the candidate ordering and a gap of at least $2\zeta\eta(|\widehat{D}^{(2)}| - \mathcal{E}^{(2)})$. Failed words, violated domain conditions and insufficient margins are unresolved.

At fixed positive ζ, these bounds give $O(\eta^2)$ response error. Ordered mixed terms such as $A_i e_i$ retain prior event rounding.

Scope and cost. Appendix E.2 proves the spectral tails and contractive defect recurrence. For affine switching surfaces in the plane, each surface along a block is a constant plus two geometric sequences. Its forward difference has at most one zero, so at most four integer extrema verify all its executed signs. For S surfaces, two candidates need at most $8(m+1)S$ scalar sign evaluations, plus regional algebra. The spectral remainder bound holds in any dimension; the constant-size word test is planar. The experiments numerically evaluate this real-arithmetic certificate with independent high-precision checks, not formal interval arithmetic. Base reference data are disclosed. Exact affine word evaluation is also cheap; the certificate diagnoses responses without claiming a speed advantage.

7 EXPERIMENTS: RESPONSE AT A SPECIFIED RADIUS

We compare predicted central responses with the actual finite-GD response in (3), using absolute error. Each experiment fixes the task, direction and radius rule before evaluation. E1–E2 use preserved scalar and affine arrays; E3 reanalyzes stored network candidate pairs. Protocols and full populations are in Appendix J.1.

Evidence and comparison targets. Response error measures prediction accuracy; a sign reversal changes the ordering of two candidates, and their objective gap measures its size. Neither establishes a multi-step optimization gain. E1–E2 test response and ordering; E3 tests nonlinear response accuracy. The separate historical gradient-ranking and outer-update studies retain their own cohorts.

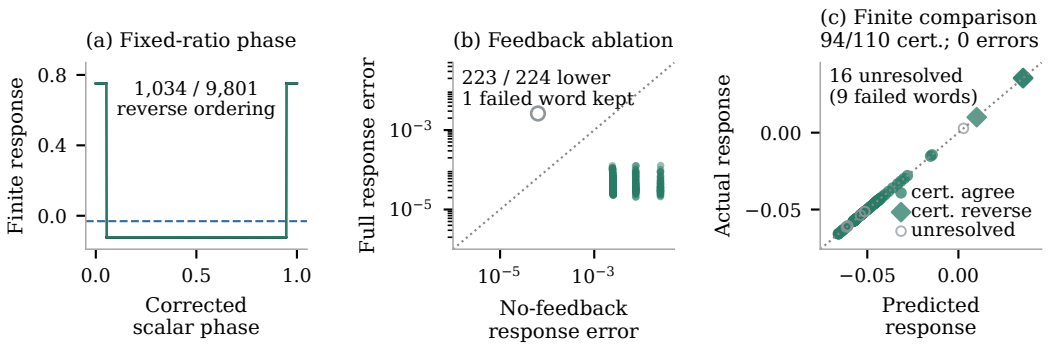

Figure 2: **Mechanism and finite comparison evidence.** (A) All 9,801 scalar responses, the critical phase law and flow response. (B) The 224 feedback-active coupled rows; the hollow failed-word exception is retained. The diagonal denotes equal prediction error. (C) All 101 valid coarse words compared with GD reruns; the finite test certifies 94, including three reversals across both couplings. Seven valid words and nine failed words are unresolved. Counts refer to a numerically evaluated real-arithmetic bound, not interval certification.

E1: a fixed-ratio response distribution. For the scalar task $a = 4, c = -3, T = 1, t = \sqrt{2}/4$, $q(y) = (y-1)^2/2$ and $\epsilon = 0.3/N$, the limiting responses are -0.122817 and 0.751714. Both infinitesimal signs are negative. The proposal ordering reverses on $1{,}034/9{,}801$ frozen meshes (10.5499%), versus the proved asymptotic frequency 10.5328%. Independent 60-digit Euler checks of the 20 closest critical endpoints differ by at most 2.27×10^{-12}. Integer phase lengths $\ell = 1, 2, 3$ approach the flow response. Figure 2(A) shows the complete fixed-ratio population.

E2: upstream feedback and a finite decision. The coupled risk in Section 4 uses $v = e_1$ and $q(\theta) = (\theta_1 - 1)^2/2$, with zero coupling as a control. Across $N = 1000, \ldots, 5000$ and three fixed ratios, the full and ablated recursions predict different signed second-index changes on 224 of 12,003 coupled rows. This selection uses no rerun objective. The full law has lower error on $223/224$, with mean 5.94×10^{-5} versus 8.17×10^{-3}, and zero sign errors versus eight. The retained exception has an invalid predicted word. The second-order test certifies $23{,}997/24{,}006$ dense comparisons across both couplings, with zero incorrect decisions; nine words fail. Of 110 separately fixed coarse comparisons, 94 are certified and 16 unresolved. Two coupled certified reversals have actual objective gaps 7.08×10^{-5} and 2.01×10^{-5}, respectively 0.097% and 0.027% of base loss. A third coupled reversal remains unresolved. These are modest two-candidate effects, not optimization gains.

E3: accuracy changes with proposal radius. We use a width-4 ReLU network with 17 parameters on a fixed two-moons split (64 training and 64 validation samples), trained by full-batch GD in float64 to $T = 1$. All 20 fixed seeds, $N = 500, 2000$, radii $10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}$ and direction e_1 are retained. Figure 3 compares errors on the frozen validation half-MSE. One seed has no admissible reference: every predictor comparison therefore uses the same 38 seed–mesh observations per radius, 152 of 160 rows in total. The smooth *branch-response predictor* ("Branch") uses the critical recursion with base reference events and regional variational/bias solves. "Smooth" in stored analyses names this regional expansion, not smoothed ReLU. Independently rerun perturbed GD supplies the evaluation target.

At the smallest radius, AD has median error 9.56×10^{-12} and the smallest error among the four stored predictors on $25/38$ rows. The branch-response predictor has the smallest error on $25/38$ and $30/38$ at 10^{-4} and 10^{-3}. At 10^{-2}, flow improves substantially over AD: median errors 1.87×10^{-4} versus 1.72×10^{-3}. The branch-response predictor is best on $22/38$ rows and flow on $11/38$. The best predictor therefore changes with radius; the largest tested radius does not establish a flow-dominated regime.

Conditioning exposes the failure tail. Of 152 reference rows, 81 match both complete training words, nine have strictly ordered proposed events but an incorrect word, and 62 have unordered proposed events. The branch-response predictor's median/maximum errors in these strata are respec-

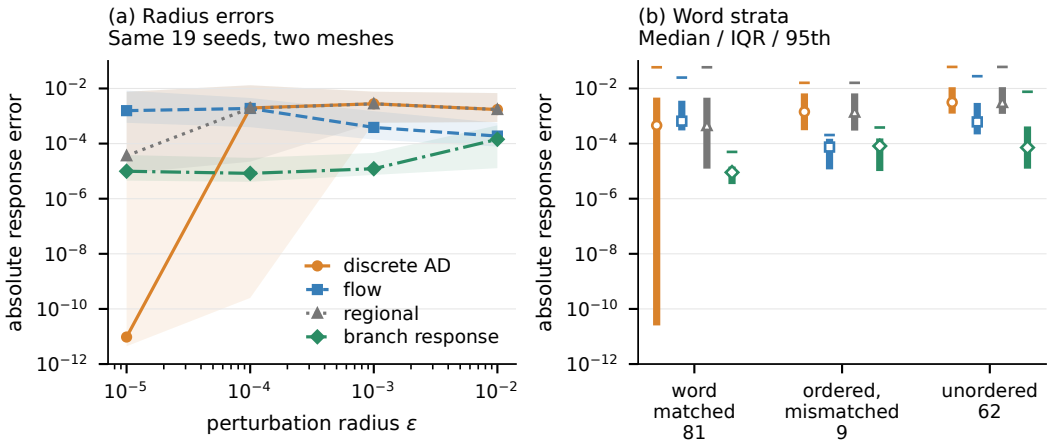

Figure 3: **Nonlinear response fidelity and its branch hypothesis.** (A) The same 19 seeds and two meshes at every radius; lines are medians and bands interquartile ranges (IQRs). All 152 reference rows enter. (B) Complete-word matched (81), strict ordered but mismatched (9), and unordered (62) rows. Marks show medians, IQR and 95th percentiles; the strata diagnose validity after rerunning the candidates. Eight uncovered rows remain in the released 160-row population. These descriptive summaries are not sampling confidence intervals or a finite nonlinear guarantee. Phase-margin diagnostics appear in Appendix J.2.

tively $9.03 \times 10^{-6}/7.77 \times 10^{-4}$, $8.16 \times 10^{-5}/5.27 \times 10^{-4}$, and $7.17 \times 10^{-5}/4.87 \times 10^{-2}$. The strata differ in radius and event count, so their contrast is descriptive. Even among the 81 matched rows, two change the terminal validation branch. Recursive phase margin only modestly discriminates word correctness (area under the receiver operating characteristic curve 0.647); it is not a validity certificate. No network sign reversal occurs. Appendix J.2 reports subgroup diagnostics, full tails and absolute objective gaps.

Decision information beyond a scalar sign. On the same 38 seed–mesh pairs, the stored full initialization gradients of flow and finite GD select different largest absolute coordinates on $6/38$ and different ordered top-two coordinates on $16/38$; their minimum cosine is 0.858. These are derivative-ranking diagnostics, not evaluated vector proposals. The critical scan has only direction e_1, so it cannot establish a critical-predictor ranking or cosine. The separate frozen four-weight outer audit has no top-one disagreement, one top-two disagreement among 15 seeds, and small finite-proposal regrets. Its five-step flow paths have lower final objective than AD on $9/11$ covered seeds, with median flow-minus-AD difference -2.41×10^{-7}.

8 Related Work

Differentiating finite training. Hyperparameter optimization differentiates validation performance through a finite training run (Maclaurin et al., 2015); bilevel formulations make this finite inner objective explicit (Franceschi et al., 2018). Initialization meta-learning updates parameters through one or more adaptation steps (Finn et al., 2017), while example reweighting uses a validation derivative through a one-step training surrogate (Ren et al., 2018). These applications motivate a finite outer change, and the need to quantify its prediction error at a specified radius. Maclaurin et al. (2015, Sec. 4, Fig. 10) already observed that hypergradients can be uninformative about a broader objective trend in unstable training. We quantify a different, transverse-switching mechanism that also occurs in our strongly convex examples. The cited stochastic, momentum and meta-learning algorithms are motivation, not applications of our fixed-horizon Euler theorem without checking its assumptions.

Event and finite perturbation analysis. Saltation, endogenous event sensitivity and ordered flow selections are classical (Hiskens & Pai, 2000; Cassandras et al., 2010; Burden et al., 2016; Kong et al., 2024). Finite perturbation analysis already reconstructs changed sample paths and propagates

downstream branch decisions (Ho et al., 1983; Ho & Cao, 1991; Ho, 1992; Cassandras & Ho, 1985; Wardi et al., 2017). Stewart–Anitescu demonstrate discretized sensitivity failure and artificial optimization structure (Stewart & Anitescu, 2010); FESD recovers flow sensitivity using event-aligned discretization (Nurkanović et al., 2024). Our fixed-grid program can be encoded in finite perturbation analysis (FPA). We derive its integer crossings and a uniform endpoint expansion when perturbations scale with the refining mesh, then gives a finite comparison bound in the contractive affine subclass.

Derivative targets and limiting objects. Fixed-program AD correctness (Lee et al., 2020; 2023; Bolte & Pauwels, 2020; Park et al., 2024), contractive fixed-point differentiation (Bolte et al., 2022; Grazzi et al., 2024), and training-flow approximations (Elkabetz & Cohen, 2021) concern distinct objects. Architectural Euler limits may share the numerical structure (Xu et al., 2023; Gao et al., 2025; Brändle et al., 2026). Our radius-dependent response complements these derivative targets. Its three sufficient regimes combine classical local/consistency bounds with the critical-scale theorem. Appendix K records exact source locators and the remaining primary-text access gaps.

9 Discussion and Limitations

The accuracy of a training sensitivity depends on the proposal radius.

The critical theorem requires finitely many separated, same-direction transverse events, bounded derivatives, isolating tubes and a common recursive phase margin. Its constants and mesh threshold depend on these conditions. Grazing, sliding, simultaneous events and event counts growing with refinement require further analysis. The finite certificate also requires contractive affine regions and complete word verification. The nonlinear predictor uses reference trajectories and has no finite certificate or established computational advantage.

The network study supports radius-dependent response accuracy but shows no sign reversal or universal ordering of predictors. Its failed words and small historical outer-update effects limit the practical conclusion. The next questions are how often consequential proposals encounter these scales in larger networks, and whether branch validity can be checked cheaply enough to guide an outer update.

Reproducibility Statement

Appendices A–E state the assumptions and prove the main results; Appendix F.4 proves the resolution synthesis. Appendices J.1–J.5 specify experiments, retained failures and independent-check coverage. The supplementary code and preserved arrays reproduce numerical summaries and vector figures; analytic curves follow the stated formulas.

Statement on the Use of AI Assistants

Generative AI tools assisted with literature search; formulating, checking, drafting, and editing mathematical claims and proofs; experimental design and implementation; result interpretation; source-code editing; figure preparation; and reproducibility audits. The tools did not serve as sources of empirical measurements. Every reported value was checked against preserved machine-readable output, selected finite-program and flow derivatives were independently cross-checked by centered finite differences with coverage detailed in Appendix J.3, and event-aware forward and reverse calculations were cross-checked in the analytic examples. The human authors retain responsibility for the final manuscript, including every claim, proof, citation, numerical result, and artifact.

References

Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Oxford University Press, Oxford, 2000. doi: 10.1093/oso/9780198502456.001.0001. URL https://academic.oup.com/book/53762.

Jérôme Bolte and Edouard Pauwels. A mathematical model for automatic differentiation in machine learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 10809–10819. Curran Associates, Inc., 2020. URL https://arxiv.org/abs/2006.02080.

Jérôme Bolte, Edouard Pauwels, and Samuel Vaiter. Automatic differentiation of non-smooth iterative algorithms. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26404–26417, 2022. doi: 10.52202/068431-1915. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/a9077da44185792cb63599cc9e0357bc-Abstract-Conference.html.

Alena Brändle, Lukas Eisenmann, Florian Götz, and Daniel Durstewitz. Continuous-time piecewise-linear recurrent neural networks, 2026. URL https://arxiv.org/abs/2602.15649. arXiv preprint, version 2, August 6, 2026.

Samuel A. Burden, S. Shankar Sastry, Daniel E. Koditschek, and Shai Revzen. Event-selected vector field discontinuities yield piecewise-differentiable flows. *SIAM Journal on Applied Dynamical Systems*, 15(2):1227–1267, 2016. doi: 10.1137/15M1016588. URL https://epubs.siam.org/doi/10.1137/15M1016588.

Christos G. Cassandras and Yu-Chi Ho. An event domain formalism for sample path perturbation analysis of discrete event dynamic systems. *IEEE Transactions on Automatic Control*, 30(12): 1217–1221, 1985. doi: 10.1109/TAC.1985.1103875.

Christos G. Cassandras, Yorai Wardi, Christos G. Panayiotou, and Chenyang Yao. Perturbation analysis and optimization of stochastic hybrid systems. *European Journal of Control*, 16(6):642–661, 2010. doi: 10.3166/ejc.16.642-661. URL https://people.bu.edu/cgc/Published/EJC_12_10.pdf.

Françoise Demengel. Fonctions à hessien borné. *Annales de l'Institut Fourier*, 34(2):155–190, 1984. doi: 10.5802/aif.969. URL https://www.numdam.org/item/AIF_1984__34_2_155_0/.

Omer Elkabetz and Nadav Cohen. Continuous vs. discrete optimization of deep neural networks. In *Advances in Neural Information Processing Systems*, volume 34, pp. 4947–4960, 2021. URL https://papers.nips.cc/paper/2021/hash/274ad4786c3abca69fa097b85867d9a4-Abstract.html.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1126–1135. PMLR, 2017. URL https://proceedings.mlr.press/v70/finn17a.html.

Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazzi, and Massimiliano Pontil. Bilevel programming for hyperparameter optimization and meta-learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1568–1577. PMLR, 2018. URL https://proceedings.mlr.press/v80/franceschi18a.html.

Tianxiang Gao, Siyuan Sun, Hailiang Liu, and Hongyang Gao. Global convergence in neural ODEs: Impact of activation functions. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/2fd37205c429a30cc1bb0afdbbf96857-Abstract-Conference.html.

Riccardo Grazzi, Massimiliano Pontil, and Saverio Salzo. Nonsmooth implicit differentiation: Deterministic and stochastic convergence rates. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 16250–16274. PMLR, 2024. URL https://proceedings.mlr.press/v235/grazzi24a.html.

Ian A. Hiskens and M. A. Pai. Trajectory sensitivity analysis of hybrid systems. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 47(2):204–220, 2000. doi: 10.1109/81.828574. URL https://ieeexplore.ieee.org/document/828574.

Yu-Chi Ho. Perturbation analysis: Concepts and algorithms. In *Proceedings of the 1992 Winter Simulation Conference*, pp. 231–240, 1992. doi: 10.1145/167293.167340. URL https://www.informs-sim.org/wsc92papers/1992_0024.pdf.

Yu-Chi Ho and Xi-Ren Cao. *Perturbation Analysis of Discrete Event Dynamic Systems*, volume 145 of *The Kluwer International Series in Engineering and Computer Science*. Kluwer Academic Publishers, 1991. doi: 10.1007/978-1-4615-4024-3.

Yu-Chi Ho, Xi-Ren Cao, and Christos G. Cassandras. Infinitesimal and finite perturbation analysis for queueing networks. *Automatica*, 19(4):439–445, 1983. doi: 10.1016/0005-1098(83)90060-2.

Nathan J. Kong, J. Joe Payne, James Zhu, and Aaron M. Johnson. Saltation matrices: The essential tool for linearizing hybrid dynamical systems. *Proceedings of the IEEE*, 112(6):585–608, 2024. doi: 10.1109/JPROC.2024.3440211. URL https://ieeexplore.ieee.org/document/10638633.

Wonyeol Lee, Hangyeol Yu, Xavier Rival, and Hongseok Yang. On correctness of automatic differentiation for non-differentiable functions. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6719–6730. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4aaa76178f8567e05c8e8295c96171d8-Abstract.html.

Wonyeol Lee, Sejun Park, and Alex Aiken. On the correctness of automatic differentiation for neural networks with machine-representable parameters. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 19094–19140. PMLR, 2023. URL https://proceedings.mlr.press/v202/lee23p.html.

Dougal Maclaurin, David Duvenaud, and Ryan P. Adams. Gradient-based hyperparameter optimization through reversible learning. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2113–2122. PMLR, 2015. URL https://proceedings.mlr.press/v37/maclaurin15.html.

Armin Nurkanović, Mario Sperl, Sebastian Albrecht, and Moritz Diehl. Finite elements with switch detection for direct optimal control of nonsmooth systems. *Numerische Mathematik*, 156(3):1115–1162, 2024. doi: 10.1007/s00211-024-01412-z. URL https://link.springer.com/article/10.1007/s00211-024-01412-z.

Sejun Park, Sanghyuk Chun, and Wonyeol Lee. What does automatic differentiation compute for neural networks? In *The Twelfth International Conference on Learning Representations*, pp. 52559–52585, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/e8711daef520be07cb9852390c673de8-Abstract-Conference.html.

Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4334–4343. PMLR, 2018. URL https://proceedings.mlr.press/v80/ren18a.html.

David E. Stewart and Mihai Anitescu. Optimal control of systems with discontinuous differential equations. *Numerische Mathematik*, 114(4):653–695, 2010. doi: 10.1007/s00211-009-0262-2. URL https://link.springer.com/article/10.1007/s00211-009-0262-2.

Yorai Wardi, Christos G. Cassandras, and Xi-Ren Cao. Perturbation analysis: A framework for data-driven control and optimization of discrete event and hybrid systems. *IFAC-PapersOnLine*, 50 (1):3028–3038, 2017. doi: 10.1016/j.ifacol.2017.08.671. URL https://people.bu.edu/cgc/Published/IFAC17_PAhistory.pdf. Author-hosted congress preprint has printed pages 3083–3093.

Yewei Xu, Shi Chen, and Qin Li. Correcting auto-differentiation in neural-ODE training, 2023. URL https://arxiv.org/abs/2306.02192. Version 3, revised 2026; accepted for publication in SIAM Journal on Applied Mathematics.

He Zhang and Wai Kin Victor Chan. Mathematical programming-based perturbation analysis for GI/G/1 queues. In *Proceedings of the 2007 Winter Simulation Conference*, pp. 553–559, 2007. doi: 10.1109/WSC.2007.4419647. URL https://www.informs-sim.org/wsc07papers/066.pdf.

GUIDE TO THE APPENDIX

The main argument is proved in Appendices D and E. The following guide separates those proofs from classical sensitivity background and supporting experiments.

Question	Where to find it
Which assumptions are required?	Appendix A; the stronger critical-scale geometry is in Appendix D.
Why does finite-program AD select the regional limit?	Appendix C.1.
How are discrete crossings and their errors derived?	Appendix D; the explicit two-event risk is in Appendix D.6.
When is a finite comparison certified?	Appendix E.2, including the complete planar word test.
How do the three resolutions connect?	Appendices F.2 and F.4; phase frequencies are in Appendix F.3.
What changes for outer parameters or smoothing?	Appendices H.1 and I.1.
What evidence supports each experiment?	Appendix J.1; network tails and coverage are in Appendices J.2 and J.5.
What is added to prior work?	Appendix K.

A ASSUMPTIONS, NOTATION, AND EXCLUDED REGIMES

Let $\theta \in \mathbb{R}^d$. In a neighborhood of a compact tube K around the reference trajectory, a finite family of open regions $\{\mathcal{R}_\alpha\}_\alpha$ forms a piecewise-smooth partition. On $\mathcal{R}_\alpha$, write f_α for the regional vector field and $A_\alpha = Df_\alpha$. The state and derivative-selection theorem is stated for this piecewise-C^1 field. In the hard-ReLU gradient specialization,

$$\mathcal{L}(\theta) = \mathcal{L}_\alpha(\theta), \qquad f_\alpha(\theta) = -\nabla\mathcal{L}_\alpha(\theta), \qquad A_\alpha(\theta) = -H_\alpha(\theta).$$

The following are the common base assumptions for the sharp-interface results; individual theorems state any additions.

(A1) Each f_α has a C^1 extension to a neighborhood of $\overline{\mathcal{R}_\alpha} \cap K$, and f_α and $A_\alpha = Df_\alpha$ are uniformly bounded there. When a result invokes gradient, Hessian, convexity, or ReLU structure, each $\mathcal{L}_\alpha$ additionally has a C^2 extension, its one-sided loss traces agree on common interfaces, and $f_\alpha = -\nabla\mathcal{L}_\alpha$.

(A2) Every interface met by the reference trajectory is locally a regular C^2 hypersurface $\Sigma_j = \{g_j = 0\}$, with $n_j(\theta) := \nabla g_j(\theta) \neq 0$. The defining function is oriented so that the trajectory moves from $g_j < 0$ to $g_j > 0$. At the event state z_j, the associated unit normal is $\nu_j = n_j(z_j)/\|n_j(z_j)\|$. We use n_j in formulas invariant to rescaling of g_j, and ν_j whenever a coefficient or spectral statement requires a unit normal.

(A3) The patched Carathéodory solution $\theta(t) = \varphi_t(\theta_0)$ exists on $[0, T]$, remains in K, and has exactly $m < \infty$ events

$$0 < t_1 < \cdots < t_m < T, \qquad z_j = \theta(t_j).$$

Consecutive events and the endpoints of the time interval are separated by at least $\Delta_t > 0$. Each event has an isolating neighborhood containing no other interface met by the trajectory. On a neighborhood of θ_0, this event count, order, separation, and itinerary persist.

(A4) The crossings are uniformly same-direction transverse. For some $c_0 > 0$, after shrinking the isolating neighborhoods if necessary,

$$n_j(\theta)^\top f_j^-(\theta) \geq c_0, \qquad n_j(\theta)^\top f_j^+(\theta) \geq c_0$$

on the respective one-sided neighborhoods.

Here "same-direction" is local to the oriented event: the normal points from the incoming region to the outgoing region. It does not require all ReLU units to switch in one common global direction; it excludes grazing, sliding, and opposing one-sided normal velocities at that crossing.

(A5) There is no grazing, sliding, chattering, simultaneous multi-surface hit, event creation, Zeno accumulation, or event at the terminal time T.

The assumptions in this section govern the derivative-selection foundation. The critical-scale theorem in Appendix D states stronger C^2 field bounds and quantitative tubes explicitly, and derives its discrete itinerary from those data.

These assumptions give a unique piecewise-classical trajectory and a locally stable event order. At an exact discrete boundary landing, the state update selects one of the adjacent one-sided fields whose oriented normal component remains at least c_0. This condition is needed for the state estimate. Whenever ordinary AD is identified with a classical derivative of the discrete endpoint map, we additionally require that no iterate land exactly on an interface. At an exact hit, software AD still returns a selected branch product, but that product need not be a classical derivative.

We use the canonical meshes $\eta_N = T/N$. Let $\theta_k^{\eta_N}$ denote the Euler/GD iterates and define the endpoint map and its piecewise-linear interpolation by

$$\Theta_{\eta_N,T}(\theta_0) := \theta_N^{\eta_N}, \qquad \Theta_{\eta_N}(t;\theta_0) := \text{the interpolation of } \{\theta_k^{\eta_N}\}_{k=0}^N.$$

We use $J_{\eta,T}^{\mathrm{sel}}$ for the product returned by the executed branch convention, including at exact hits. On a nonresonant mesh it is both ordinary branchwise AD and the classical derivative $D\Theta_{\eta,T}$, and we write it as $J_{\eta,T}^{\mathrm{AD}}$. We use J_{reg} for the event-free regional continuum product, $D\varphi_T$ for the true flow derivative, and $J_{\eta,T}^{\mathrm{SC}}$ for the event-aware corrected sensitivity. The matrix Φ_j propagates the regional variational equation between t_j^+ and t_{j+1}^-, with $t_0 = 0$ and $t_{m+1} = T$.

Rate assumption. Under (A1), the regional Jacobians are uniformly continuous on the relevant compact sets. We write

$$\omega_A(r) = \max_{\alpha} \sup \left\{ \|A_\alpha(x) - A_\alpha(y)\| : x, y \in K \cap \overline{\mathcal{R}_\alpha},\ \|x - y\| \le r \right\}.$$

Thus $\omega_A(r) \to 0$ as $r \downarrow 0$. An $O(\eta)$ sensitivity rate additionally requires locally Lipschitz regional Jacobians; in the gradient setting, $\mathcal{L}_\alpha \in C^{2,1}$ suffices.

A.1 Additional scope and excluded regimes

Multidimensional scope. The sharp-interface results are fully multidimensional. Tangential modes pass through an interface of a continuous piecewise-smooth loss unchanged, while the normal mode is multiplied by $(\nu^\top f^+)/(\nu^\top f^-)$. Regional dynamics before and after an event may rotate and couple these modes, which is why the exact multiple-event criterion uses transported factors. The smoothing theorem is narrower: it extends directly only when the system admits an autonomous scalar normal coordinate with separately controlled tangential propagation. A general curved layer introduces tangential drift, a varying normal, noncommuting matrix cocycles, and flattening-curvature terms. No general curved-interface $O(\eta/\tau)$ theorem is claimed.

Excluded regimes and rate qualifications. The sharp-interface and smoothing results retain the following restrictions.

1. A piecewise C^2 loss, hence a C^1 regional field, gives the modulus estimate for ordinary AD; an $O(\eta)$ sensitivity rate needs Lipschitz regional Hessians or an equivalent time-regularity assumption.
2. At an exact discrete interface landing, $D\Theta_{\eta,T}$ may not exist even though software AD returns a selected branch product.
3. Constants are nonuniform near grazing because event-time and saltation formulas contain $(n^\top f^-)^{-1}$.
4. Opposite-sign normal speeds can produce sliding or nonunique patched solutions. Simultaneous hits require a separate multi-surface analysis and cannot be assigned an arbitrary ordered saltation product.
5. Infinite or nonseparated event sequences invalidate the finite induction and finite-product perturbation arguments.

6. If T is an event time, the fixed-time derivative may be one-sided or convention-dependent.
7. Multiple nonidentity event factors can cancel after transport; a switch count alone is not a mismatch criterion.
8. The identity $\|\Xi\|_2 = \max\{1, |r|\}$ requires $d \geq 2$; in one dimension the norm is $|r|$.
9. Positive-definite regional Hessians do not imply global convexity of a continuous nonsmooth loss.
10. The resolved smoothing theorem includes a tail term and does not make η/τ universal across profile families. Its sharp-first identity limit is directed through nonresonant step sizes.

B Regional smoothness and classical hybrid sensitivity

B.1 Piecewise-smooth ReLU objectives

Proposition 4 (Finite-sample ReLU objectives)**.** *Consider a finite ReLU network on a finite data set, with a loss smooth in the network outputs. On every fixed activation-pattern cell, the empirical risk is smooth (and polynomial for squared loss). The risk is continuous across activation boundaries. At a point where exactly one preactivation vanishes and its parameter gradient is nonzero, the boundary is locally a regular hypersurface and the one-sided parameter gradients have well-defined traces.*

Proof. Fixing every ReLU mask replaces each activation by either zero or its preactivation. The network output is then a finite composition of affine and multilinear operations in the parameters. It is polynomial on that mask cell, and so is a squared empirical risk. ReLU itself is continuous, hence the network and risk traces agree across a mask boundary. If a single preactivation $s(\theta)$ vanishes with $\nabla s \neq 0$, the implicit-function theorem makes $\{s = 0\}$ a regular hypersurface. Smooth regional formulas provide the gradient traces. □

Proposition 5 (Local codimension of exceptional activation events)**.** *Work inside a fixed surrounding activation pattern, where a preactivation $s : \mathbb{R}^d \to \mathbb{R}$ is C^2 and the incoming regional field f^- is C^1. The following statements are local.*

1. *If $\nabla s(z) \neq 0$, then $\Sigma = \{s = 0\}$ is a regular hypersurface near z. The transverse subset $\{x \in \Sigma : \nabla s(x)^\top f^-(x) \neq 0\}$ is open in Σ.*
2. *Let $h = (\nabla s^\top f^-)|_\Sigma$. If zero is a regular value of h, then the grazing set $\{x \in \Sigma : h(x) = 0\}$ has codimension one in Σ, hence codimension two in $\mathbb{R}^d$.*
3. *If two preactivations vanish and their gradients are linearly independent, their simultaneous-hit set is a codimension-two submanifold.*
4. *Fix an executed-branch cell of an N-step GD program and one evaluated state $\theta_k^\eta(\theta_0)$. If $D_{\theta_0}(s \circ \theta_k^\eta) \neq 0$ at an exact landing, then the initializations producing that landing form a local hypersurface. For fixed N and finitely many activation surfaces, the union of all such regular exact-landing sets has Lebesgue measure zero inside the branch cell.*

Proof. The first statement is the implicit-function theorem, and continuity of $\nabla s^\top f^-$ makes its nonzero set open. The regular-value theorem applied to $h : \Sigma \to \mathbb{R}$ gives the grazing codimension. Applying the implicit-function theorem to the map (s_1, s_2) proves the simultaneous hit statement when its two gradient rows are independent. On a fixed branch cell, every executed regional update is C^1, so $s \circ \theta_k^\eta$ is C^1. The final hypersurface claim is again the implicit-function theorem; a finite union of regular hypersurfaces has Lebesgue measure zero. □

For canonical meshes $\eta_N = T/N$, a useful conditional consequence is available. If every exact-landing map in every fixed branch cell has zero as a regular value, then for each N the resonant initialization set is a finite union of measure-zero hypersurfaces. Taking the countable union over $N \in \mathbb{N}$ shows that almost every initialization is nonresonant for all canonical meshes. This statement is about initialization under the displayed regular-value condition; it does not assert that nonresonant step sizes are generic or dense for one fixed initialization.

These statements explain only local geometry under the displayed nondegeneracy conditions. They do not prove that a modern-network trajectory has a globally stable finite itinerary, that grazing or simultaneous events never occur, or that the sensitivity mismatch is prevalent.

The output is generally not piecewise affine in all network parameters. For a two-layer unit $a\ \mathrm{ReLU}(w^\top x + b)$, it is bilinear on an active cell, and the squared risk may have degree four. Only piecewise smoothness is used here. For a shallow squared-loss sample, orient $s(\theta) = q^\top \theta + b$ from inactive to active. At $s = 0$, with output coefficient c and residual $r = F_\theta(x) - y$, direct differentiation gives

$$[\nabla \mathcal{L}] = rcq$$

up to the empirical averaging factor. The jump is normal to the activation surface. A simultaneous change of several samples or units is outside the single-interface result.

B.2 Event-time derivative and saltation

For one event, let z be the nominal impact at time t_*, let $n = \nabla g(z)$, and let ζ^- be the incoming perturbation evaluated at the nominal time. The perturbed event time $t_* + \delta t$ obeys

$$0 = n^\top(\zeta^- + f^-\delta t), \qquad \delta t = -\frac{n^\top \zeta^-}{n^\top f^-}.$$

The perturbed impact point changes by $\zeta^- + f^-\delta t$, while the outgoing segment loses time δt. Returning to the nominal clock gives

$$\zeta^+ = \zeta^- + (f^- - f^+)\delta t = \left(I + \frac{(f^+ - f^-)n^\top}{n^\top f^-}\right)\zeta^-.$$

This is the saltation matrix without a state reset. The event time is C^1 by the implicit-function theorem because $n^\top f^- \neq 0$. Iterating this calculation over the stable itinerary gives the hybrid product in Theorem 6.

Normal jump for a continuous loss. At a transverse crossing, an incoming perturbation ζ^- shifts the event time and fixed-clock state as follows. For a continuous piecewise-smooth loss, tangential derivatives agree; with $\nu = n/\|n\|$ and $[\nabla \mathcal{L}] = \beta\nu$,

$$\delta t_* = -\frac{n^\top \zeta^-}{n^\top f^-}, \qquad \zeta^+ = \Xi\zeta^-, \qquad \Xi = I + \frac{(f^+ - f^-)n^\top}{n^\top f^-}, \tag{12}$$

$$\Xi = I - \frac{\beta\nu\nu^\top}{\nu^\top f^-}, \qquad \lambda_\perp(\Xi) = \frac{\nu^\top f^+}{\nu^\top f^-}. \tag{13}$$

C State consistency and derivative selection

C.1 General state and first-variation theorem

Put $\eta = \eta_N = T/N$. The executed-branch Euler/GD recursion and its selected branch product are

$$\theta^\eta_{k+1} = \theta^\eta_k + \eta f_{\alpha_k}(\theta^\eta_k), \tag{14}$$

$$J^{\mathrm{sel}}_{\eta,0} = I, \qquad J^{\mathrm{sel}}_{\eta,(k+1)\eta} = \big(I + \eta A_{\alpha_k}(\theta^\eta_k)\big)J^{\mathrm{sel}}_{\eta,k\eta}. \tag{15}$$

At an exact hit, α_k is the admissible adjacent branch selected by the convention in Appendix A. On the jth smooth segment, let $\Phi_j(t,s)$ solve

$$\partial_t \Phi_j(t,s) = A_{\alpha_j}(\theta(t))\Phi_j(t,s), \qquad \Phi_j(s,s) = I, \tag{16}$$

and abbreviate $\Phi_j = \Phi_j(t^-_{j+1}, t^+_j)$, with $t_0 = 0$ and $t_{m+1} = T$.

Let M_f and M_A bound the regional fields and Jacobians on the compact tube. Constants below may depend on $T, m, c_0^{-1}, \Delta_t^{-1}, M_f, M_A$, the C^2 geometry and radii of the isolating tubes, and the positive distance from the reference path to all other interfaces outside those tubes. They do not depend on N. This dependence is necessarily nonuniform as $c_0 \downarrow 0$.

Theorem 6 (State consistency and first-variation selection)**.** *Under (A1)–(A5), there are $C < \infty$ and N_0 such that, for every $N \geq N_0$,*

$$\sup_{0 \leq t \leq T} \|\Theta_{\eta_N}(t;\theta_0) - \varphi_t(\theta_0)\| \leq C\eta_N \tag{17}$$

and

$$\|J^{\mathrm{sel}}_{\eta_N,T} - J_{\mathrm{reg}}\| \leq C\big(\eta_N + \omega_A(C\eta_N)\big) = o(1), \qquad J_{\mathrm{reg}} := \Phi_m \cdots \Phi_0. \tag{18}$$

If the regional Jacobians are locally Lipschitz, the right-hand side is $O(\eta_N)$.

The terminal flow map is C^1 on a neighborhood with the same itinerary, and

$$D\varphi_T(\theta_0) = \Phi_m \Xi_m \Phi_{m-1} \Xi_{m-1} \cdots \Phi_1 \Xi_1 \Phi_0, \qquad \Xi_j = I + \frac{(f_j^+ - f_j^-)n_j^\top}{n_j^\top f_j^-}. \tag{19}$$

For every nonresonant subsequence $N_\ell \to \infty$, the endpoint map is classically differentiable near θ_0 and

$$D_{\theta_0}\Theta_{\eta_{N_\ell},T}(\theta_0) = J^{\mathrm{AD}}_{\eta_{N_\ell},T} = J^{\mathrm{sel}}_{\eta_{N_\ell},T} \longrightarrow J_{\mathrm{reg}}. \tag{20}$$

At a resonant mesh, the selected product still satisfies (18), but need not be a classical derivative.

The proof is split into six lemmas to expose the event-tube induction and the matrix limit.

Lemma 7 (Smooth-segment Euler stability)**.** *Suppose the exact and numerical paths use the same regional extension on grid indices $k_0, \ldots, k_1$. With $e_k = \|\theta_k^\eta - \theta(k\eta)\|$,*

$$\max_{k_0 \leq k \leq k_1} e_k \leq e^{M_A(k_1-k_0)\eta}\left[e_{k_0} + \tfrac{1}{2}M_A M_f (k_1 - k_0)\eta^2\right]. \tag{21}$$

The corresponding piecewise-linear interpolation has an additional $O(M_f\eta)$ error.

Proof. The exact regional solution has local defect at most $M_A M_f \eta^2/2$. Lipschitz continuity of the field gives

$$e_{k+1} \leq (1 + M_A\eta)e_k + \tfrac{1}{2}M_A M_f \eta^2.$$

Discrete Gronwall and $1 + x \leq e^x$ prove (21). Both the exact path and the interpolant move at uniformly bounded speed within a step. □

Lemma 8 (Single-event localization)**.** *Consider the isolating tube of event j, and suppose the numerical path reaches its incoming part with error e_{in}. For sufficiently small η, let $k^\eta_{j,\mathrm{in}}$ be its first grid index in that incoming part. The first nonnegative interface index*

$$\kappa_j^\eta := \min\{k \geq k^\eta_{j,\mathrm{in}} : g_j(\theta_k^\eta) \geq 0\}$$

exists and is unique inside the tube, the numerical path does not recross there, and

$$|\kappa_j^\eta \eta - t_j| + \|\theta^\eta_{\kappa_j^\eta} - z_j\| \leq C(e_{\mathrm{in}} + \eta). \tag{22}$$

If an exact landing executes the incoming branch once at the zero iterate, the executed branch-switch index is $\kappa_j^\eta + 1$; otherwise it is κ_j^η. Thus its clock obeys the same estimate after increasing the constant by one mesh unit. In particular, an $O(\eta)$ incoming error gives first-order event localization.

Proof. Uniformly on either one-sided part of the tube, Taylor expansion gives

$$g_j(x + \eta f_j^\pm(x)) = g_j(x) + \eta \nabla g_j(x)^\top f_j^\pm(x) + O(\eta^2). \tag{23}$$

The remainder is uniform because g_j is C^2 and $f_j^\pm$ is bounded. Shrink the isolating tube so that its compact incoming and outgoing cross-sections satisfy $g_j \leq -\gamma_j$ and $g_j \geq \gamma_j$, respectively, for some $\gamma_j > 0$. By (A4), the numerical increment of g_j is at least $c_0\eta/2$ for all small η, including an exact hit under the admissible convention. Starting on the negative cross-section, the discrete normal coordinate therefore reaches zero before the positive cross-section, does so at most once, and cannot recross before leaving the tube.

Up to the earlier of the exact and numerical crossing clocks, both paths use the incoming regional extension. Lemma 7 and the smoothness of g_j give an $O(e_{\mathrm{in}} + \eta)$ normal-coordinate error there. Between the two clocks, either one-sided field has bounded speed, while both one-sided normal velocities are at least c_0. Comparing with the corresponding incoming or outgoing regional continuation therefore bounds the clock difference by $C(e_{\mathrm{in}} + \eta)/c_0$. Bounded state speed gives (22). The possible extra incoming update at an exact zero delays the executed switch by at most one step. □

Lemma 9 (Finite-itinerary induction)**.** *For all sufficiently small η, the numerical path visits the same regions and isolating tubes as the reference path, in the same order, and has exactly one branch transition in each tube. Its error on leaving every tube is $O(\eta)$, with a constant uniform over the finite itinerary.*

Proof. Before the first tube, Lemma 7 gives $O(\eta)$ error. Lemma 8 localizes the transition. Substituting this incoming $O(\eta)$ error into (22) shows that the switch clocks differ by $O(\eta)$, hence by only $O(1)$ grid steps. Each such step contributes at most $2M_f\eta$ between the two one-sided updates; thus the outgoing restart error remains $O(\eta)$. The positive distance from all other interfaces fixes the next region. Restart Lemma 7 on the outgoing segment. The Δ_t-separated tubes and finite event count allow this argument to be iterated; a finite composition of its Gronwall constants is still independent of N. The final segment and the interpolation estimate prove (17). □

Lemma 10 (Branch-mismatch set)**.** *Let $\alpha(t)$ be the exact branch and let $\alpha_\eta(t) = \alpha_k$ on $[k\eta, (k+1)\eta)$. Define*

$$E_\eta = \{t \in [0,T] : \alpha_\eta(t) \neq \alpha(t)\}.$$

Then $|E_\eta| \le Cm\eta$. For

$$A(t) = A_{\alpha(t)}(\theta(t)), \qquad A^\eta(t) = A_{\alpha_k}(\theta_k^\eta), \quad t \in [k\eta, (k+1)\eta),$$

we have

$$\|A^\eta - A\|_{L^1(0,T)} \le C\eta + T\omega_A(C\eta). \tag{24}$$

Proof. Lemma 8 confines a branch disagreement at event j to an $O(\eta)$ interval around t_j; the intervals are disjoint for small η. Their total length is at most $Cm\eta$. Outside them, both paths use the same regional extension, and Lemma 9 plus within-step motion gives $\|A^\eta(t) - A(t)\| \le \omega_A(C\eta)$. Inside them, both coefficients have norm at most M_A. Integration proves (24). □

Lemma 11 (Fundamental-matrix stability)**.** *Let U_η and U be the fundamental matrices for the stepwise coefficient A^η and the exact coefficient A, respectively. Then*

$$\|U_\eta(T) - U(T)\| \le e^{2M_A T}\|A^\eta - A\|_{L^1(0,T)}, \tag{25}$$

$$\|J_{\eta,T}^{\mathrm{sel}} - U_\eta(T)\| \le C\eta. \tag{26}$$

Proof. Variation of constants gives

$$U_\eta(T) - U(T) = \int_0^T U_\eta(T,s)(A^\eta(s) - A(s))U(s)\,\mathrm{d}s.$$

Both propagators have norm at most $e^{M_A T}$, which proves (25). On one step, the exact cocycle of A^η is $e^{\eta A_k}$, and

$$\|e^{\eta A_k} - (I + \eta A_k)\| \le \tfrac{1}{2}\eta^2 M_A^2 e^{\eta M_A}.$$

A telescoping product estimate over $N = T/\eta$ uniformly bounded factors proves (26). □

Lemma 12 (Flow-map differentiability and saltation)**.** *On a neighborhood with the stable itinerary, every event time and event state is C^1. If ψ_t^- denotes the incoming regional flow and $\tau_j(x)$ its hitting time, then*

$$D\tau_j(x) = -\frac{n_j^\top D\psi_{\tau_j(x)}^-(x)}{n_j^\top f_j^-}. \tag{27}$$

The fixed-clock derivative across the event is Ξ_j, and finite chronological composition gives (19).

Proof. The hitting equation is $g_j(\psi_{\tau_j(x)}^-(x)) = 0$. Its derivative with respect to time is $n_j^\top f_j^- \neq 0$, so the implicit-function theorem proves the first claim and (27). An incoming perturbation ζ^- changes the event time by

$$\delta t_j = -\frac{n_j^\top \zeta^-}{n_j^\top f_j^-}.$$

Returning the outgoing perturbation to the nominal clock gives

$$\zeta^+ = \zeta^- + (f_j^- - f_j^+)\delta t_j = \left(I + \frac{(f_j^+ - f_j^-)n_j^\top}{n_j^\top f_j^-} \right) \zeta^-.$$

Regional flow maps and event maps are C^1; composing the finitely many maps in chronological order proves (19). □

Proof of Theorem 6. Lemma 9 proves the state estimate. Lemmas 10 and 11 give

$$\|J_{\eta,T}^{\mathrm{sel}} - J_{\mathrm{reg}}\| \le C\{\eta + \|A^\eta - A\|_{L^1}\} \le C\big(\eta + \omega_A(C\eta)\big).$$

Lemma 12 proves the flow derivative. If the mesh is nonresonant, every evaluated regional sign is strict and persists in a neighborhood of θ_0. The endpoint is then a finite composition of C^1 regional Euler maps, and the chain rule identifies its classical derivative with the selected branch product. At an exact landing, this neighborhood argument can fail, while the already proved product convergence remains valid. □

For completeness, applying the reverse triangle inequality to $J_{\eta,T}^{\mathrm{sel}} - D\varphi_T = (J_{\mathrm{reg}} - D\varphi_T) + (J_{\eta,T}^{\mathrm{sel}} - J_{\mathrm{reg}})$ and then using (18) gives (30). This statement concerns the selected product on every canonical mesh. It becomes a statement about the classical derivative $D\Theta_{\eta,T}$ only along nonresonant meshes, by (20).

In the continuous-loss gradient specialization, $[\nabla\mathcal{L}]_j = \beta_j \nu_j$ implies

$$\Xi_j = I - \frac{\beta_j \nu_j \nu_j^\top}{\nu_j^\top f_j^-}, \qquad r_j = \frac{\nu_j^\top f_j^+}{\nu_j^\top f_j^-}. \tag{28}$$

The matrix $\Xi_j - I$ is symmetric and has rank at most one. For Ξ_j, tangential directions have eigenvalue one, and the normal direction has eigenvalue r_j. For $d \ge 2$, $\|\Xi_j\|_2 = \max\{1, |r_j|\}$; for $d = 1$, $\|\Xi_j\|_2 = |r_j|$.

C.2 Supporting derivative identities

The finite-mesh comparison separates into an asymptotic event gap and a discretization remainder:

$$J_{\eta,T}^{\mathrm{sel}} - D\varphi_T = \big(J_{\mathrm{reg}} - D\varphi_T\big) + \big(J_{\eta,T}^{\mathrm{sel}} - J_{\mathrm{reg}}\big). \tag{29}$$

Consequently, the reverse triangle inequality gives the two-sided spectral-norm estimate

$$\big|\|J_{\eta,T}^{\mathrm{sel}} - D\varphi_T\|_2 - \|J_{\mathrm{reg}} - D\varphi_T\|_2\big| \le C\{\eta + \omega_A(C\eta)\}. \tag{30}$$

In particular, $\|J_{\eta,T}^{\mathrm{sel}} - D\varphi_T\|_2 \to \|J_{\mathrm{reg}} - D\varphi_T\|_2$. Along a nonresonant subsequence the first term is exactly $\|D\Theta_{\eta,T} - D\varphi_T\|_2$. Thus any fixed nonzero event gap remains visible on sufficiently fine meshes; the constants deteriorate near grazing through their dependence on c_0^{-1}. This is a trajectory-local visibility statement, not a uniform lower bound over changing event geometries.

The selection is not specific to Euler. Theorem 13 proves the same event-free derivative limit for a class of fixed-step, event-blind one-step schemes. It assumes regional consistency, mixed-event Jacobians $I + O(\eta)$, a uniformly bounded number of mixed steps per event, and oriented progress through each event tube. Higher regional order alone therefore does not recover saltation; the claim does not cover a Runge–Kutta, adaptive, implicit, or event-localizing method unless these hypotheses are verified for that method.

C.3 A restricted event-blind one-step extension

The Euler calculation is not the only way to obtain the event-free derivative limit. Regional consistency alone, however, is insufficient: it places no restriction on the derivative of a step that samples more than one region. The following result states the additional hypotheses needed for a safe extension.

Theorem 13 (Conditional event-blind one-step derivative criterion)**.** *Assume (A1)–(A5). Consider a fixed-step method whose executed update is a piecewise-C^1 map $x_{k+1} = \Psi_{\eta,\sigma_k}(x_k)$, where σ_k records a locally fixed finite pattern of regional evaluations. Suppose the following bounds hold uniformly on the relevant compact tube.*

1. *On a step whose evaluations remain in one region α,*

$$\Psi_{\eta,\alpha}(x) = x + \eta f_\alpha(x) + O(\eta^2), \tag{31}$$

$$D\Psi_{\eta,\alpha}(x) = I + \eta A_\alpha(x) + O(\eta^2). \tag{32}$$

2. *Every mixed step in an isolating event tube satisfies*

$$\|\Psi_{\eta,\sigma}(x) - x\| \le C\eta, \qquad \|D\Psi_{\eta,\sigma}(x) - I\| \le C\eta. \tag{33}$$

3. *The method makes oriented interface progress at least $c\eta$ in each event tube and executes at most M mixed steps per event, with $c > 0$ and $M < \infty$ independent of η.*

4. *No evaluation used by the executed program lies exactly on an interface.*

Then the numerical state converges uniformly on $[0, T]$, and the exact Jacobian of the executed program satisfies

$$D\Theta^{\Psi}_{\eta,T} \longrightarrow J_{\mathrm{reg}}. \tag{34}$$

The derivative of the limiting flow remains the saltation-interleaved product in Theorem 1. Thus the two derivatives differ whenever the transported event product is nonidentity.

Proof. On pure regional steps, (31) gives the standard first-order consistency and stability recursion. In each isolated event tube, positive progress and the mixed-step bound imply an $O(\eta)$ timing error and only $O(1)$ local defects of size $O(\eta)$. Restarting the regional recursion after every separated event gives uniform state convergence by the same finite induction as Theorem 6.

For the derivative, pure regional factors have the expansion $I + \eta A_\alpha + O(\eta^2)$. Their ordered products converge to the corresponding regional fundamental matrices. Every mixed factor is $I + O(\eta)$, and only finitely many such factors occur, so their combined product converges to I, not to a nonidentity saltation matrix. A finite product perturbation argument across the separated itinerary proves (34). Strict branch evaluations make this product the classical derivative of the executed finite program. □

The theorem includes Euler/GD. It also includes any event-blind regional one-step construction that satisfies the mixed-step hypotheses, irrespective of its higher regional order. It does not cover an event-localizing method, whose mixed-step derivative is designed to retain the event-time transfer, or an adaptive, implicit, or stage-switching scheme unless those hypotheses are verified separately. Smooth preconditioned Euler is covered relative to its own limiting field $f = -P(\theta)\nabla\mathcal{L}(\theta)$, provided the same event and regularity assumptions hold; this is not a statement about the unpreconditioned gradient flow. Fixed-coefficient momentum and heavy-ball limits are not covered.

D FINITE-EVENT CRITICAL-SCALE BRANCH LAW

This result concerns a joint limit of the step size and the initial perturbation. Event-wise propagation itself is standard in perturbation analysis. The statement below additionally derives the integer indices selected by explicit Euler, including accumulated Euler bias, and proves a uniform endpoint expansion away from the resulting scaled boundaries. No independence, density, or equidistribution of event phases is used.

D.1 GEOMETRIC AND REGULARITY HYPOTHESES

Fix $T > 0$ and a finite number $m \ge 0$ of events. A finite collection of C^2 functions g_s, $1 \le s \le S$, defines the switching surfaces. The executed field is selected by their signs. Every regional field used below has a C^2 extension to an open neighborhood of the relevant path and its event tubes. The Euler program evaluates the field at the current grid state and then takes one explicit Euler step. At a zero

sign it may use any fixed convention; our conclusion concerns grids with strict signs. These are C^2 assumptions on the regional vector fields, not just the losses: when $f_i = -\nabla L_i$, regional C^3 loss extensions suffice.

The base continuous trajectory $x(t)$ is continuous and follows the successive regional fields $f_0, \dots, f_m$ on the intervals separated by

$$0 = t_0 < t_1 < \cdots < t_m < t_{m+1} = T, \qquad L_i = t_{i+1} - t_i, \qquad x_i = x(t_i).$$

Its events are isolated single-surface crossings. More precisely, assume the following quantitative neighborhood geometry.

1. $L_i \geq \Delta > 0$. There is $a \in (0, \Delta/4)$ such that the time windows $[t_j - a, t_j + a]$ are disjoint. Within a neighborhood U_j of each such path arc, just one surface can vanish, and its two sign regions select exactly f_{j-1} and f_j. Orient its local defining function as $g_j^* = \sigma_j g_{\ell_j}$, $\sigma_j \in \{-1, 1\}$, so that the crossing is from negative to positive. Thus the same G_1, G_2 bounds apply to the oriented functions. The other defining functions have absolute values at least $\gamma > 0$ on U_j.
2. Both adjacent smooth extensions satisfy

   $$\nabla g_j^*(y)^\top f_{j-1}(y) \geq \nu > 0, \qquad \nabla g_j^*(y)^\top f_j(y) \geq \nu > 0 \quad (y \in U_j).$$

   Outside these time windows, every surface has absolute value at least γ on the base path. These base arcs and event arcs have a common positive spatial tube radius ρ within the indicated neighborhoods. The same physical surface may occur at different, separated events; its orientation is chosen separately at each occurrence.
3. On a closed neighborhood containing these tubes, the short smooth extensions at their endpoints, and straight segments used in Taylor estimates, there are bounds

   $$\|f_i\| \leq F_0, \quad \|Df_i\| \leq F_1, \quad \|D^2 f_i\| \leq F_2, \qquad \|\nabla g_s\| \leq G_1, \quad \|\nabla^2 g_s\| \leq G_2.$$

   The indicated neighborhood is fixed independently of the mesh.

For example, isolated transverse events of a fixed finite base path give these constants after shrinking the windows and tubes, provided the outgoing normal speed has the same strictly positive sign. Merely having a nonzero incoming speed is not sufficient: the outgoing condition excludes immediate return, sticking, or sliding. No discrete crossing index, discrete event order, or perturbed itinerary is assumed.

On segment i, write $\Phi_i(t, s)$ for the ordinary smooth variational propagator along its base path, and define $P_i = \Phi_i(t_{i+1}, t_i)$. Let $w_i(t)$ solve

$$\dot{w}_i = Df_i(x(t))w_i - \tfrac{1}{2} Df_i(x(t))f_i(x(t)), \qquad w_i(t_i) = 0, \qquad w_i = w_i(t_{i+1}). \tag{35}$$

Also set

$$n_j = \nabla g_j^*(x_j), \qquad u_j = n_j^\top f_{j-1}(x_j) \geq \nu, \qquad \Delta_j = f_j(x_j) - f_{j-1}(x_j).$$

These are base-path quantities. In particular, P_i contains no saltation matrix, and w_i is the smooth-region Euler bias.

D.2 Recursive branch law and theorem

For $p \in \mathbb{T}^m$ and $z \in \mathbb{R}^d$, put $b_0 = z$ and recurse chronologically:

$$\begin{aligned} a_j &= P_{j-1} b_{j-1} + w_{j-1}, \\ c_j(p, z) &= p_j - n_j^\top a_j / u_j, \\ s_j &= \lceil c_j(p, z) \rceil - p_j, \qquad b_j = a_j - \Delta_j s_j, \qquad 1 \leq j \leq m, \\ e_p(z) &= P_m b_m + w_m. \end{aligned} \tag{36}$$

Integer changes in a lift of p_j leave s_j unchanged. At an integer c_j the formula provides a candidate value but the following theorem does not identify the actual branch. For $m = 0$ the recursion is simply $e(z) = P_0 z + w_0$.

Detailed statement of Theorem 2. Under the preceding hypotheses, let $K \subset \mathbb{R}^d$ be compact, $Z = \sup_{z\in K} \|z\| < \infty$, and $0 < \mu < 1/2$. There are finite constants C, B, depending only on

$$(m, T, Z, F_0, F_1, F_2, G_1, G_2, \nu^{-1}, \Delta^{-1}, a^{-1}, \rho^{-1}, \gamma^{-1}),$$

and a sufficient mesh threshold $\eta_* > 0$, also depending on μ, such that the following holds. For every integer N with $\eta = T/N \le \eta_*$, define $p_N = (\{t_1/\eta\}, \dots, \{t_m/\eta\})$. If

$$\inf_{z\in K} \min_{1\le j\le m} \operatorname{dist}(c_j(p_N, z), \mathbb{Z}) \ge \mu, \tag{37}$$

then every perturbed Euler path starting at $x_0 + \eta z$, $z \in K$, has exactly the base sequence of m events, with first positive indices

$$k_j(z) = \left\lceil \frac{t_j}{\eta} - \frac{n_j^\top a_j(p_N, z)}{u_j} \right\rceil. \tag{38}$$

All its executed signs are strict. Uniformly on K,

$$\|\Theta_{\eta,T}(x_0 + \eta z) - x_{m+1} - \eta e_{p_N}(z)\| \le C\eta^2, \qquad \sup_{z\in K} \|e_{p_N}(z)\| \le B. \tag{39}$$

For $m = 0$ the phase condition is vacuous and there are no event indices. The compact set K need not be connected: finite candidate sets may occupy different ceiling cells while satisfying (37) at every evaluated point. A connected set spanning a ceiling transition cannot satisfy a positive uniform margin. The arguments c_j are affine within preceding ceiling cells and may jump across upstream boundaries. The proof below imposes $\eta J_j < \nu\mu/2$ for every event, among other sufficient restrictions; it asserts no universal scaling of η_* in μ.

If $0 \in K$, $q \in C^2$ near x_{m+1}, and $\mathcal{R}_\eta(z) = [q(\Theta_{\eta,T}(x_0 + \eta z)) - q(\Theta_{\eta,T}(x_0))]/\eta$, then uniformly on K,

$$\mathcal{R}_\eta(z) = \nabla q(x_{m+1})^\top [e_{p_N}(z) - e_{p_N}(0)] + O_K(\eta). \tag{40}$$

If $p_N \to p$ along a subsequence and every $c_j(p, z)$ on K has distance at least μ from the integers, the same conclusion holds on that subsequence for all sufficiently large N, with e_{p_N} replaced by e_p and error $O_K(\eta + \operatorname{dist}_{\mathbb{T}^m}(p_N, p))$.

D.3 A RESTART-AWARE SMOOTH EULER ESTIMATE

We first prove the regional estimate, including its bias sign. Let $x' = f(x)$ on a smooth reference segment. For an initial fixed-clock error b, let

$$E' = Df(x)E - \tfrac{1}{2}Df(x)f(x), \qquad E(t_i) = b.$$

Thus $E(t_{i+1}) = P_i b + w_i$. Define the comparison curve $y(t) = x(t) + \eta E(t)$. Since $x'' = Df(x)f(x)$, its one-step residual is

$$\begin{aligned} y(t+\eta) - y(t) - \eta f(y(t)) &= \eta^2[\tfrac{1}{2}Df(x)f(x) + E' - Df(x)E] + O(\eta^3) \\ &= O(\eta^3). \end{aligned}$$

This proves the negative sign in (35). Only C^2 regularity is needed: $x''' = D^2 f(x)[f, f] + Df(x)Df(x)f$ is bounded, and differentiating the equation for E once uses only $D^2 f$. At an early grid restart, $t_i + \eta\sigma < t_i$, both x and E mean their hypothetical smooth continuation under the outgoing regional extension; no claim is made that the switched base flow follows that extension before t_i. The fixed neighborhoods permit these short backward continuations. The bounds below use absolute elapsed time and therefore cover them as well as forward continuations. For explicit bounds, on a time interval of length at most $D = T + 2$ put

$$\begin{aligned} \Gamma &= e^{F_1 D}, & X_3 &= F_2 F_0^2 + F_1^2 F_0, \\ V(U) &= \Gamma(U + \tfrac{1}{2}F_1 F_0 D), & V'(U) &= F_1 V(U) + \tfrac{1}{2}F_1 F_0, \\ V''(U) &= F_2 F_0 V(U) + F_1 V'(U) + \tfrac{1}{2}X_3, & A(U) &= X_3/6 + V''(U)/2 + F_2 V(U)^2/2, \\ H_U(s) &= \tfrac{1}{2}F_1 F_0 s^2 + V'(U)|s|. \end{aligned}$$

For $\|b\| \le U$, the one-step residual is at most $A(U)\eta^3$. This follows from Taylor's integral remainders for x, E, and f; the last contributes at most $F_2\eta^3 V(U)^2/2$.

Suppose a regional Euler block starts at the grid clock $t_i + \eta\sigma$ with state

$$x_i + \eta\{b + f_i(x_i)\sigma\} + r, \qquad \|r\| \le C_i\eta^2.$$

Taylor's formula gives an initial discrepancy from $y(t_i + \eta\sigma)$ of at most $\eta^2[C_i + H_U(\sigma)]$. A discrete Gronwall estimate for the one-step residual, in the actual order of Euler steps, then gives at the grid clock $t_{i+1} + \eta s$ the expansion

$$x_{i+1} + \eta\{P_i b + w_i + f_i(x_{i+1})s\} + r', \quad \|r'\| \le \eta^2\{\Gamma[C_i + H_U(\sigma) + DA(U)] + H_U(s)\}. \tag{41}$$

Indeed, each discrepancy grows by at most $1 + \eta F_1$ and there are at most D/η residuals. This argument also holds at every intermediate grid clock, not just at the endpoint. Short extensions of length $\eta|\sigma|$ or $\eta|s|$ are allowed inside the fixed neighborhoods. A stopped-neighborhood argument closes the estimate: derive it up to the first possible exit, and choose the mesh so its error is smaller than the available tube radius; that exit is then impossible.

D.4 Proof of the theorem and explicit dependency recursion

Proof. First bound the proposed offsets without choosing any mesh or branch. Set $U_0 = Z$. For $j = 1, \dots, m$ recursively define

$$A_j = e^{F_1 L_{j-1}} U_{j-1} + \tfrac{1}{2} F_1 F_0 L_{j-1} e^{F_1 L_{j-1}},$$
$$S_j = 1 + G_1 A_j/\nu, \qquad U_j = A_j + 2F_0 S_j.$$

Then $\|a_j\| \le A_j$, $|s_j| \le S_j$, and $\|b_j\| \le U_j$ for every phase and $z \in K$. These follow from $|\lceil p - \xi \rceil - p| \le 1 + |\xi|$, the variational bound $\|P_i\| \le e^{F_1 L_i}$, the corresponding integral bound for w_i, and $\|\Delta_j\| \le 2F_0$. Thus $B = e^{F_1 L_m}(U_m + F_1 F_0 L_m/2)$ suffices for the endpoint coefficient. Put $S_0 = S_{m+1} = 0$, $C_0 = 0$, and set

$$C_{i+1} = \Gamma[C_i + H_{U_i}(S_i) + DA(U_i)] + H_{U_i}(S_{i+1} + \mathbf{1}_{i<m}), \qquad 0 \le i \le m. \tag{42}$$

The extra one covers the grid point just before a proposed crossing. This recursion uses only the displayed data and bounds every remainder needed in (41). In particular $C = C_{m+1}$ is an admissible endpoint constant.

We now identify the actual first event. The smooth Euler extension of f_0 at a grid point $k\eta = t_1 + \eta s$ has the expansion

$$x_1 + \eta[a_1 + f_0(x_1)s] + r, \qquad \|r\| \le C_1\eta^2.$$

Applying the nonlinear surface and using $g_1^*(x_1) = 0$ gives

$$g_1^*(\theta_k) = \eta u_1(s + n_1^\top a_1/u_1) + O(\eta^2). \tag{43}$$

For $|s| \le S_1 + 1$ and $\eta C_1 \le 1$, the error in this equation is at most $J_1\eta^2$, where generally

$$J_j = G_1 C_j + \tfrac{1}{2} G_2[A_j + F_0(S_j + 1) + 1]^2.$$

Thus curvature enters only the $O(\eta^2)$ remainder. It produces no additional term in the first-order ceiling argument.

At the integer k_1 proposed by (38) the leading surface value is at least $\eta u_1\mu$; at $k_1 - 1$ it is at most $-\eta u_1\mu$. For $\eta J_1 < \nu\mu/2$ these are strictly positive and negative, respectively. Outside the first event window, the smooth regional estimate and surface margin exclude any crossing. Inside its tube, for either adjacent regional extension,

$$g_1^*(y + \eta f(y)) - g_1^*(y) \ge \eta\nu - \tfrac{1}{2}\eta^2 G_2 F_0^2 \ge \eta\nu/2$$

when $\eta G_2 F_0^2 \le \nu$. Consequently no earlier zero or positive grid state can precede the negative value at $k_1 - 1$: k_1 is exactly the first positive index. The other surfaces cannot change in this tube. The outgoing monotonicity excludes a return across the same surface.

At this actual crossing state the expansion can be rewritten exactly at first order as

$$x_1 + \eta[a_1 + f_0(x_1)s_1] + r = x_1 + \eta[b_1 + f_1(x_1)s_1] + r, \qquad b_1 = a_1 - \Delta_1 s_1.$$

This is the restart error at the fixed reference clock; it is not an independent continuously shifted event time. Apply (41) in region 1, starting at the actual grid clock $k_1\eta$. It gives the incoming error $a_2 = P_1 b_1 + w_1$, including the first index's rounding error. The same two-sign and monotonicity argument at the second surface derives k_2. Induction derives every k_j, its restart b_j, and the expansion on the next smooth block. No commutation of variational propagators is used.

Here are explicit sufficient types of mesh restrictions completing that induction. In addition to $\eta \le 1$, require

$$\eta \max_j (S_j + 1) < \min\{a/4, 1\}, \quad \eta \max_j C_j \le 1, \quad \eta \max_j J_j < \nu\mu/2, \quad \eta G_2 F_0^2 \le \nu.$$

The first puts all candidate crossings inside their separated windows, so $k_1 < \cdots < k_m$ and $0 < k_1$, $k_m < N$. The regional lemma bounds the discrepancy from the base path throughout all blocks by ηB_{tube} for a finite constant depending on the same recursion; for example, after decreasing the mesh to make the second-order errors at most η, one may take

$$B_{\mathrm{tube}} = 1 + \max_i V(U_i) + 2F_0 \max_j (S_j + 1).$$

Choose additionally $\eta B_{\mathrm{tube}} < \rho/4$, $\eta F_0 < \rho/4$, and $\eta G_1 B_{\mathrm{tube}} < \gamma/2$. The tube bounds keep every relevant state and Euler step in the stipulated neighborhoods. The surface bounds exclude extra events in the intervening corridors and changes of other surfaces in event tubes. The intermediate regional estimate invoked here is proved on the candidate smooth extension before assuming its signs, so this is not circular. There are only finitely many restrictions, all computable from the displayed data, and they give a positive η_*. The final block has no further ceiling and gives (39) with C_{m+1}.

For the objective statement, take a ball of radius $r_q > 0$ about x_{m+1} where q is C^2 and $\|\nabla^2 q\| \le M_q$. Decrease the objective's mesh threshold, if needed, so that $\eta B + \eta^2 C < r_q$. This additional threshold does not change the state constants or its geometric hypotheses. If $\|\nabla q(x_{m+1})\| \le G_q$, Taylor's formula gives an error in (40) bounded by

$$2\eta\{G_q C + \tfrac{1}{2} M_q (B + \eta C)^2\}.$$

Finally, the ceiling arguments are affine in (p, z) on each preceding ceiling cell. A uniform positive margin preserves all their integer values under a sufficiently small change of a consistent local phase lift, by induction over j. The coefficients are bounded by the same propagator and jump bounds. Therefore e_p is uniformly Lipschitz in phase on these neighborhoods, proving the subsequence conclusion. □

D.5 Downstream boundaries, reductions, and excluded limits

For $j > r$, write $P_{j-1:r} = P_{j-1} P_{j-2} \cdots P_r$ in chronological order. Unrolling the recursion gives

$$a_j = A_j^{\mathrm{free}}(z) - \sum_{r<j} P_{j-1:r} \Delta_r s_r,$$

where A_j^{free} includes the initial displacement and all smooth Euler biases but no event-rounding updates. Hence the jth scaled boundary is

$$p_j - \frac{n_j^\top A_j^{\mathrm{free}}(z)}{u_j} + \sum_{r<j} \frac{n_j^\top P_{j-1:r} \Delta_r}{u_j} s_r \in \mathbb{Z}. \tag{44}$$

This expression states the contribution of every upstream index to a later threshold. Deleting these terms defines the *no-upstream-feedback ablation*; it is not a published competing algorithm. Inside a fixed complete ceiling cell, $D_z e_p = P_m \cdots P_0$. Across a cell change with integer offset differences h_j, the terminal coefficient changes by $-\sum_j P_{m:j} \Delta_j h_j$, with the downstream integers themselves determined recursively. Nonlinear surface curvature affects the finite location of the boundaries at the next order, so their leading scaled hyperplanes need not be the exact finite-grid boundaries.

For $f_i(x) = M_i x + d_i$, $\Phi_i(t, s) = e^{(t-s)M_i}$ and $f_i(x(t)) = e^{(t-t_i)M_i} f_i(x_i)$. Variation of constants in (35) therefore gives exactly

$$w_i = -\tfrac{1}{2} L_i e^{L_i M_i} M_i f_i(x_i),$$

recovering the affine recursion. No symmetry or contractivity is needed for the smooth first-order theorem. The sharper finite spectral certificate below retains those stronger affine hypotheses.

With one event this is the single ceiling law. With zero events it is the ordinary first-order Euler-error expansion. If an event has $\Delta_j = 0$, its rounding update vanishes at first order, although the variational field may still change. With all jumps zero the first-order response is phase-independent and equals the continuous variational response. Independent coordinate fields remove the relevant cross-coordinate coefficients in (44); commutativity alone does not imply that every such coefficient is zero.

The theorem makes no uniform assertion when m grows, ν or event separation tends to zero, a new surface enters a tube, or a derived phase margin collapses. At an exact scaled boundary, the second-order surface terms can choose the integer; the leading ceiling law alone does not determine it. Coincident events, sliding, grazing, and reset maps are outside the stated class. A numerical nonlinear network trace does not verify these neighborhood hypotheses merely by having a plausible event list.

D.6 Two coupled events as an explicit example

The general theorem specializes to the following exact affine geometry. Its two-index asymptotic proof is the $m = 2$ case above; the finite boundaries and invariant-region checks below supply the example-specific information.

D.6.1 A strongly convex affine realization

Let $H = A^\top A \succ 0$, $d = A^\top b$, and consider

$$L(\theta) = \tfrac{1}{2}\|A\theta - b\|^2 + \tfrac{1}{2}\sum_{j=1}^{2}[\mathrm{ReLU}(n_j^\top \theta - \tau_j) - c_j]^2, \qquad c_j \le 0. \tag{45}$$

Each scalar residual term is convex: its derivative jumps from zero to $-c_j \ge 0$ at zero and has nonnegative derivative slope on each side. Thus L is globally $\lambda_{\min}(H)$-strongly convex. In active set S, the negative gradient is affine,

$$f_S(\theta) = M_S\theta + d_S, \quad M_S = -H - \sum_{j\in S} n_j n_j^\top, \quad d_S = d + \sum_{j\in S}(\tau_j + c_j)n_j.$$

The regional matrices are symmetric negative definite; this does not make them commute. The jump on activating surface j is $\Delta_j = c_j n_j$.

Assume the base flow from θ_0 has precisely two activations, in order $1, 2$, at $0 < t_1 < t_2 < T$, and no deactivations. There are disjoint event tubes, positive distance from every other surface outside its tube, and a positive minimum separation Δ of events and endpoints. In each tube both adjacent normal speeds are at least $\nu > 0$. These are hypotheses about the continuous base path and its neighborhood, not about a desired discrete word. They exclude grazing, recrossing, simultaneity and new events. Fixed nonzero coupling is allowed throughout. Write $t_0 = 0, t_3 = T$, $x_i = \varphi_{t_i}(\theta_0)$, and let M_i, d_i be the fields on the three successive regions, indexed $i = 0, 1, 2$. Define

$$L_i = t_{i+1} - t_i, \qquad P_i = e^{L_i M_i}, \qquad w_i = -\tfrac{1}{2}L_i P_i M_i f_i(x_i), \tag{46}$$

$$u_j = n_j^\top f_{j-1}(x_j) > 0, \qquad \Delta_j = f_j(x_j) - f_{j-1}(x_j), \qquad j = 1, 2. \tag{47}$$

Here w_i is the first-order Euler drift at a fixed clock, not an event transfer. Assume $q \in C^2$ on a neighborhood of the terminal states.

D.6.2 The two-event recursion

For $p = (p_1, p_2)$ on the two-dimensional torus and a scaled displacement $z \in \mathbb{R}^d$, construct $e_p(z)$ as follows:

$$\begin{aligned} a_1 &= P_0 z + w_0, & s_1 &= \lceil p_1 - n_1^\top a_1/u_1 \rceil - p_1, & b_1 &= a_1 - \Delta_1 s_1, \\ a_2 &= P_1 b_1 + w_1, & s_2 &= \lceil p_2 - n_2^\top a_2/u_2 \rceil - p_2, & b_2 &= a_2 - \Delta_2 s_2, \\ e_p(z) &= P_2 b_2 + w_2. \end{aligned} \tag{48}$$

Ceilings at integer arguments are excluded in the local-limit statement. The phases p_j alone are not corrected event phases: the drift and incoming error inside each ceiling are essential.

The coupling and the branch jumps. There are exact finite-grid boundaries behind this expansion. Write $\mathcal{E}_i(y) = (I + \eta M_i)y + \eta d_i$. Before the first event, the boundary for landing at index k is the affine hyperplane $n_1^\top \mathcal{E}_0^k(\theta_0 + \eta z) = \tau_1$, restricted by all earlier strict signs. On a fixed first-index cell k_1, the analogous second boundary is $n_2^\top \mathcal{E}_1^{k_2-k_1}\mathcal{E}_0^{k_1}(\theta_0 + \eta z) = \tau_2$. These equations follow by solving the executed affine recurrences; their expansion gives the boundaries below. If a first boundary changes the first positive index from $k+1$ to k while the second index stays k_2, the right (earlier-crossing) minus left endpoint jump is exactly

$$\eta(I + \eta M_2)^{N-k_2}(I + \eta M_1)^{k_2-k-1}\Delta_1.$$

Indeed the incoming state at index k is on the surface, and the two one-step continuations differ by $\eta\Delta_1$; all subsequent affine differences propagate in the displayed order. If the second index also changes, this expression alone is insufficient; the recursive second ceiling determines its additional leading jump.

Inside a fixed first ceiling cell, the first scaled boundary is $p_1 - n_1^\top(P_0 z + w_0)/u_1 \in \mathbb{Z}$. The second boundary is

$$p_2 - \frac{n_2^\top[P_1(P_0 z + w_0) + w_1]}{u_2} + \frac{n_2^\top P_1 \Delta_1}{u_2} s_1 \in \mathbb{Z}. \tag{49}$$

The last term is the first event's rounding displacement of the second boundary. Deleting it is the no-upstream-feedback ablation used in E2. When a first-event boundary changes s_1 by an integer h, the direct terminal error changes by $-P_2 P_1 \Delta_1 h$; the second ceiling can also change by an integer h_2, adding $-P_2\Delta_2 h_2$. These formulas give the endpoint jumps of the scaled map and preserve multiplication order. On a fixed complete word the slope is $P_2 P_1 P_0$, as required by the regional derivative limit. Saltation instead continuously changes the event time; it contains neither ceiling nor downstream boundary shift.

Nonseparability and limiting checks. In E2, $H = \left(\begin{smallmatrix} 1 & 0.4 \\ 0.4 & 1.3 \end{smallmatrix}\right)$, $d = (2, 2)$, $n_1 = e_1, n_2 = e_2, \tau_j = 0$, $c = (-0.4, -0.6)$ and $\theta_0 = (-0.4, -1)$, $T = 1$. Its H is positive definite, and $\|[M_1, M_0]\|_2 = 0.4$. A fixed linear change producing independent scalar regional equations would simultaneously diagonalize these matrices, contradicting their nonzero commutator. The second threshold coefficient $n_2^\top P_1\Delta_1/u_2$ is nonzero (numerically 0.01336). The argument uses the specified fixed coupling, not its continuity at zero. Its two-event geometry can be checked without selecting a trajectory by its observed response. While both coordinates are negative, they increase, with $2 \le \dot{x} \le 2.8$ and $2 \le \dot{y} \le 3.46$. Hence $1/7 \le t_1 \le 1/5$ and $-5/7 \le y(t_1) \le -0.308$; the second coordinate cannot cross first. In the next region the rectangle $[0, 1] \times [-1, 0]$ is invariant until the second crossing, and $1.6 \le \dot{y} \le 3.3$. Its duration is between $0.308/3.3$ and $(5/7)/1.6$, so $t_2 < 0.647 < T$. Afterward $[0, 1] \times [0, 14/23]$ is invariant: all four boundary fields point inward. Outgoing normal speeds at the two crossings are at least 1.6 and 1, respectively. Compact subsegments and continuity supply the stated isolating tubes and margins. Finally the off-diagonal entry of $e^{L_1 M_1}$ is strictly negative for $L_1 > 0$ (diagonalizing this symmetric two-by-two matrix gives a negative multiple of a hyperbolic sine). Since $\Delta_1 = -0.4e_1$, the coupling coefficient is strictly positive, not merely nonzero to numerical precision. For H diagonal and these coordinate normals, this coefficient is zero and each event depends only on its own coordinate. Deleting the second event recovers the single-event ceiling law; in the scalar residual case $a_1 = e^{-t}z + mt/2$, so the effective phase is $\{t/\eta - t/2\}$, agreeing with $\{t/[-\log(1-\eta)]\}$ to $O(\eta)$. Its jumps reproduce Theorem 20. If $\Delta_j = 0$, that event's rounding displacement vanishes at first order, though the regional Hessian can still change. With both jumps zero the response is the ordinary continuous variational response; no spurious phase effect survives. No independence or equidistribution of the two event phases is assumed or needed.

E FINITE-GRID COMPARISON CERTIFICATES IN THE AFFINE CLASS

The asymptotic branch law applies to smooth regional fields. The explicit spectral certificates in this section retain symmetric negative-definite affine regional matrices; the nonlinear theorem does not inherit these bounds. All sign guarantees below are real-arithmetic statements, evaluated numerically in the experiments.

E.1 A FIRST-ORDER COMPARISON BOUND FOR THE TWO-EVENT EXAMPLE

The following finite bound is conservative but does not hide an unspecified asymptotic constant. It also handles inexact event localization. Choose any strictly ordered *supplied* knots $0 < \widehat{t}_1 < \widehat{t}_2 < T$. Construct scheduled states $\widehat{x}_i$ by exact regional flows for these durations, whether or not the knots are exact event times. Compute the recursion with this scheduled data, including the actual field differences $f_j(\widehat{x}_j) - f_{j-1}(\widehat{x}_j)$ and positive incoming speeds. Its candidate integers need not be right. Verify the candidate word by evaluating all its regional affine grid extensions and checking the required strict signs at every executed state. If this fails, return *unresolved*; do not repair the prediction with observed endpoint objectives. A validated root/tube method can replace this exhaustive check, but is not assumed here.

Use homogeneous coordinates

$$G_i = \begin{pmatrix} M_i & d_i \\ 0 & 0 \end{pmatrix}, \qquad Y_0 = \begin{pmatrix} \theta_0 \\ 1 \end{pmatrix}, \qquad Z = \begin{pmatrix} z \\ 0 \end{pmatrix}.$$

Here $G_i \in \mathbb{R}^{(d+1)\times(d+1)}$ and $Y_0, Z \in \mathbb{R}^{d+1}$. Thus $I + \eta G_i$ represents one affine Euler update and $\|Z\|_2 = \|z\|_2$. All norms in this bound are Euclidean or induced operator norms. Set $k_0 = 0, k_3 = N$, $s_j = k_j - \widehat{t}_j/\eta$, $s_0 = s_3 = 0$, $\delta_i = s_{i+1} - s_i$ and $\widehat{L}_i = \widehat{t}_{i+1} - \widehat{t}_i$. Let $g_i = \|G_i\|$, $r_i = \eta g_i < 1$, and define

$$\begin{aligned} B_i &= \frac{|\delta_i| g_i}{1 - r_i} + \frac{(\widehat{L}_i + \eta|\delta_i|) g_i^2}{2(1 - r_i)^2}, \\ C_i &= \frac{|\delta_i| g_i^2}{(1 - r_i)^2} + \frac{2(\widehat{L}_i + \eta|\delta_i|) g_i^3}{3(1 - r_i)^3}, \\ K_i &= 1 + 2\|M_i^{-1} d_i\|, \qquad B_* = \textstyle\sum_i B_i, \quad C_* = \textstyle\sum_i C_i, \\ r_\eta(z) &= \frac{\eta^2}{2} \prod_{i=0}^{2} K_i \left[(B_*^2 + C_*)(\|Y_0\| + \eta\|z\|) + 2B_*\|z\| \right]. \end{aligned} \tag{50}$$

Require $\widehat{L}_i + \eta\delta_i > 0$. Let $\widehat{e}(z)$ denote the scheduled-data recursion and $\widehat{g} = \nabla q(\widehat{x}_3)$.

Proposition 14 (Finite comparison certificate)**.** *If the preceding word checks pass, then*

$$\|\Theta_{\eta,T}(\theta_0 + \eta z) - \widehat{x}_3 - \eta\widehat{e}(z)\| \le r_\eta(z).$$

For $z_\pm = \pm\zeta v$, $\|v\| = 1$, define, using only the supplied base data and the recursion,

$$\widehat{D} = \frac{\widehat{g}^\top [\widehat{e}(z_+) - \widehat{e}(z_-)]}{2\zeta}.$$

If $\|\nabla^2 q\| \le M_q$ on the endpoint Taylor segments, then

$$|D_{\eta,\zeta\eta}(v) - \widehat{D}| \le \mathcal{E} = \frac{\|\widehat{g}\|(r_+ + r_-) + \frac{1}{2}M_q \sum_{\sigma=\pm}(\eta\|\widehat{e}(z_\sigma)\| + r_\sigma)^2}{2\zeta\eta}, \qquad r_\sigma = r_\eta(z_\sigma). \tag{51}$$

Consequently $|\widehat{D}| > \mathcal{E}$ certifies the comparison sign; otherwise the certificate returns unresolved. For a fixed positive margin, bounded offsets and a nonzero limiting response this test succeeds on all sufficiently fine meshes. It is not necessarily effective at coarse meshes.

Proof. The word check and induction over its grid states identify the actual GD endpoint with the stated ordered affine product. For $0 \le s \le \eta$ define the analytic family

$$Y(s) = U_2(s)U_1(s)U_0(s)(Y_0 + sZ), \qquad U_i(s) = \exp[(\widehat{L}_i + s\delta_i)G_{i,s}].$$

$$G_{i,s} = s^{-1}\log(I + sG_i), \qquad G_{i,0} = G_i.$$

At $s = \eta$ this is exactly the candidate Euler word, because $\widehat{L}_i/\eta + \delta_i = k_{i+1} - k_i$. At zero it is the scheduled flow endpoint. Differentiating at zero, including the duration derivatives, gives exactly $\widehat{e}(z)$ in the first d coordinates: the terms from $G'_{i,s}|_0 = -G_i^2/2$ give w_i, and adjacent duration terms

combine as $-\Delta_j s_j$ at each restart. This calculation does not require the supplied knots to be exact surface roots.

The absolutely convergent logarithm series gives

$$\|G_{i,s}\| \le \frac{g_i}{1-r_i}, \quad \|G'_{i,s}\| \le \frac{g_i^2}{2(1-r_i)^2}, \quad \|G''_{i,s}\| \le \frac{2g_i^3}{3(1-r_i)^3}.$$

For example the first derivative has coefficients $k/(k+1) \le k/2$; the second has $k(k-1)/(k+1) \le k(k-1)/3$. Summing the geometric-series derivatives proves the bounds. The exponent and its derivatives commute, being functions of the same G_i. Its first two derivatives are bounded by B_i, C_i. The top-left block of $U_i(s)$ is a contraction: the eigenvalues of M_i are negative, $1+s\lambda > 0$, and the duration is positive on $[0,\eta]$. The affine equilibrium is $-M_i^{-1}d_i$, so $\|U_i(s)\| \le 1+2\|M_i^{-1}d_i\| = K_i$. Hence

$$\|U_i'\| \le K_i B_i, \qquad \|U_i''\| \le K_i(B_i^2 + C_i).$$

Apply the product rule in its actual order, and the norm triangle inequality, to obtain

$$\sup_{0\le s\le\eta} \|Y''(s)\| \le \prod_{i=0}^{2} K_i[(B_*^2 + C_*)(\|Y_0\| + \eta\|z\|) + 2B_*\|z\|].$$

Taylor's integral remainder proves the state bound. Taylor expansion of q about $\widehat{x}_3$ bounds each objective error by $\|\widehat{g}\|r_\sigma + \frac{1}{2}M_q(\eta\|\widehat{e}(z_\sigma)\| + r_\sigma)^2$. Subtract and divide by $2\zeta\eta$ to obtain the certificate. □

Localization, numerical arithmetic, and cost. No root error is silently dropped in the finite bound: supplied knots define the expansion center, and their entire timing displacement occurs in δ_i and the computed scheduled states. Poor localization can make the word check fail or the bound large. In the asymptotic theorem the knots are the exact base events; if approximate knots are used for that conclusion their errors must vanish relative to η, or be retained as phase errors. The finite certificate is a real-arithmetic statement. Experiments evaluate its bounds and sign tests numerically, with independent Euler and higher-precision checks; they are not interval-arithmetic machine certificates. Reference trajectories are base-point information, explicitly disclosed as controlled theory inputs. Three regional propagators, two ceilings, and a terminal gradient suffice for the prediction. The exhaustive word check costs $O(N)$ sign work and removes a claim of computational advantage over evaluating two candidates. This describes the original exhaustive test. The planar spectral test in Appendix E.2 replaces it with a bounded number of sign evaluations; exact affine candidate evaluation is also cheap, so no general training speedup follows.

E.2 Ordered spectral jets for a finite number of events

The preceding homogenized bound pays for affine translations repeatedly. The regional symmetric negative definiteness permits a sharper bound in physical coordinates. We state a second-order version; dropping its quadratic term gives the first-order bound used in the ablation. Neither version assumes that a predicted word is correct.

Data and finite-grid domain. For a fixed finite number m of events, use any supplied ordered knots and their scheduled exact regional states x_i as in the preceding construction, now indexed by $i = 0, \dots, m$; hats are suppressed in this subsection. For each proposed displacement z, compute candidate indices by the finite-event recursion. Put $m_i = k_{i+1} - k_i > 0$, $L_i = t_{i+1} - t_i > 0$, $\delta_i = m_i - L_i/\eta$, and $h_i = -M_i^{-1}d_i$. Require $\eta\|M_i\|_2 < 1$ for each region. All norms below are Euclidean/operator norms. For each eigenvalue $\lambda < 0$ of M_i, define

$$\begin{aligned}
&\alpha_i(\lambda) = \delta_i\lambda - \tfrac{1}{2}L_i\lambda^2, && \beta_i(\lambda) = \tfrac{1}{3}L_i\lambda^3 - \tfrac{1}{2}\delta_i\lambda^2,\\
&t_{2i}(\lambda) = \eta^2\left[\tfrac{1}{2}|\delta_i|\lambda^2 + \frac{m_i\eta|\lambda|^3}{3(1-\eta|\lambda|)}\right], && a_{2i}(\lambda) = \eta|\alpha_i(\lambda)| + t_{2i}(\lambda),\\
&t_{3i}(\lambda) = \eta^3\left[\tfrac{1}{3}|\delta_i||\lambda|^3 + \frac{m_i\eta\lambda^4}{4(1-\eta|\lambda|)}\right], && b_i(\lambda) = \eta^2|\beta_i(\lambda)| + t_{3i}(\lambda),\\
&a_{3i}(\lambda) = \eta|\alpha_i(\lambda)| + b_i(\lambda). &&
\end{aligned} \tag{52}$$

The following maxima are over that finite spectrum:

$$\begin{aligned} d_{1i} &= \max_{\lambda} e^{L_i\lambda}(e^{a_{2i}} - 1), \\ d_{2i} &= \max_{\lambda} e^{L_i\lambda}(t_{2i} + \tfrac{1}{2}a_{2i}^2 e^{a_{2i}}), \\ d_{3i} &= \max_{\lambda} e^{L_i\lambda}[t_{3i} + \tfrac{1}{2}b_i(2\eta|\alpha_i| + b_i) + \tfrac{1}{6}a_{3i}^3 e^{a_{3i}}], \\ \chi_i &= \max_{\lambda}(1+\eta\lambda)^{m_i} \le 1. \end{aligned} \tag{53}$$

Functional calculus for the symmetric M_i defines $A_i = P_i\alpha_i(M_i)$ and $B_i = P_i[\beta_i(M_i) + \alpha_i(M_i)^2/2]$. Set $e_0 = z$, $v_0 = 0$, $R_0 = 0$, and apply the ordered recursion for $i = 0, \dots, m$

$$\begin{aligned} e_{i+1} &= P_i e_i + A_i(x_i - h_i), \\ v_{i+1} &= P_i v_i + A_i e_i + B_i(x_i - h_i), \\ R_{i+1} &= \chi_i R_i + d_{3i}\|x_i - h_i\| + \eta d_{2i}\|e_i\| + \eta^2 d_{1i}\|v_i\|. \end{aligned} \tag{54}$$

These e_i are errors at the moving grid endpoints, unlike the intermediate fixed-clock errors in the ceiling recursion. At the final clock $e_{m+1} = e_{p_N}(z)$ with the supplied scheduled data. Matrix order is never interchanged. In particular, $A_1 e_1$ includes the first rounding offset inside the second region's second-order error.

Detailed finite statement of Theorem 3. For every finite η, displacement z and candidate word satisfying the preceding domain conditions, if its strict signs are verified, then

$$\|\Theta_{\eta,T}(\theta_0 + \eta z) - (x_{m+1} + \eta e_{m+1} + \eta^2 v_{m+1})\| \le R_{m+1}. \tag{55}$$

This statement does not require a limiting phase or an unreported "sufficiently small" mesh. For fixed m, bounded $z, \delta_i, L_i, M_i, h_i, x_i$ and a common bound $\eta\|M_i\| \le r_* < 1$, it has $R_{m+1} = O(\eta^3)$ uniformly. The explicit formulas state every constant; there is no inverse phase margin in the bound. Near-boundary failure is handled by the word test, not by declaring the candidate itinerary true. Unbounded scaled displacements retain the finite inequality but not this uniform asymptotic rate.

Proof. The exact regional matrix is $E_i = (I + \eta M_i)^{m_i}$. On an eigenvalue, write $a = m_i \log(1 + \eta\lambda) - L_i\lambda$. The absolutely convergent logarithm series yields both

$$|a - \eta\alpha_i| \le t_{2i}, \qquad |a - \eta\alpha_i - \eta^2\beta_i| \le t_{3i}.$$

For the second inequality, the cubic logarithm term contributes $\eta^3\delta_i\lambda^3/3$, and the remaining series is bounded by $m_i\eta^4|\lambda|^4/[4(1 - \eta|\lambda|)]$. The quadratic truncation gives the first inequality in the same way. Taylor's integral remainder for the scalar exponential gives $|e^a - 1 - a| \le a^2 e^{|a|}/2$ and $|e^a - 1 - a - a^2/2| \le |a|^3 e^{|a|}/6$. Moreover, $|a^2 - \eta^2\alpha_i^2| \le b_i(2\eta|\alpha_i| + b_i)$. Multiplication by $e^{L_i\lambda}$ and the spectral theorem therefore give

$$\|E_i - P_i\| \le d_{1i}, \quad \|E_i - P_i - \eta A_i\| \le d_{2i}, \quad \|E_i - P_i - \eta A_i - \eta^2 B_i\| \le d_{3i}.$$

The exact regional affine step is $h_i + E_i(y - h_i)$, whereas $x_{i+1} = h_i + P_i(x_i - h_i)$. Substituting $y = x_i + \eta e_i + \eta^2 v_i + \rho_i$ and subtracting the displayed jet leaves exactly

$$E_i\rho_i + (E_i - P_i - \eta A_i - \eta^2 B_i)(x_i - h_i) + \eta(E_i - P_i - \eta A_i)e_i + \eta^2(E_i - P_i)v_i.$$

Its norm is at most R_{i+1} because $\|E_i\| = \chi_i$. Induction from the exact initial state proves the result for the candidate affine product. The strict word test identifies that product with actual GD by induction over executed states. It does not use the two rerun objective values. The rate follows directly from $d_{ji} = O(\eta^j)$ and bounded ordered recursions. □

First-order ablation and objective control. For the unchanged first-order predictor, set $r_0 = 0$ and use

$$r_{i+1} = \chi_i r_i + d_{2i}\|x_i - h_i\| + \eta d_{1i}\|e_i\|.$$

The same subtraction proves $\|\Theta_{\eta,T} - (x_{m+1} + \eta e_{m+1})\| \le r_{m+1}$; inserting it into (51) isolates the effect of the tighter bound. For second order let $\widetilde{x}_\pm = x_{m+1} + \eta e_{m+1}(\pm\zeta v) + \eta^2 v_{m+1}(\pm\zeta v)$ and

$R_\pm = R_{m+1}(\pm\zeta v)$. If q is C^2 on an open neighborhood of the two closed balls of radii $R_\pm$ about these points and $\|\nabla^2 q\| \le M_q$ on the balls, then

$$\widehat{D}^{(2)} = \frac{q(\widetilde{x}_+) - q(\widetilde{x}_-)}{2\zeta\eta},$$
$$\mathcal{E}^{(2)} = \frac{\sum_{\sigma=\pm}[\|\nabla q(\widetilde{x}_\sigma)\| R_\sigma + M_q R_\sigma^2/2]}{2\zeta\eta}. \tag{56}$$

Taylor's integral formula on each ball proves $|D_{\eta,\zeta\eta}(v) - \widehat{D}^{(2)}| \le \mathcal{E}^{(2)}$. At fixed positive ζ and bounded offsets this response error is $O(\eta^2)$, compared with the first-order $O(\eta)$ bound. If both words pass and $|\widehat{D}^{(2)}| > \mathcal{E}^{(2)}$, the smaller predicted objective is the smaller actual objective, and the actual gap is at least $2\zeta\eta(|\widehat{D}^{(2)}| - \mathcal{E}^{(2)})$. Thus a contrary AD/flow ordering has at least this two-candidate regret. This last decision statement is an elementary corollary, not an optimization convergence theorem. Equality is unresolved.

Constant-size planar word verification. For a two-dimensional symmetric negative definite regional matrix, diagonalize $M = V\,\mathrm{diag}(\lambda_1, \lambda_2)V^\top$. For a proposed block of n_b executed states and any affine surface, its signed value at integer $k \in \{0, \dots, n_b - 1\}$ is exactly

$$g(k) = a + b_1 r_1^k + b_2 r_2^k, \qquad r_l = 1 + \eta\lambda_l \in (0, 1).$$

The coefficients are obtained from the known block initial state and equilibrium; no continuous event root within this block is assumed. Its forward difference is $g(k+1) - g(k) = b_1(r_1 - 1)r_1^k + b_2(r_2 - 1)r_2^k$. If the two coefficients have the same sign, or one vanishes, this sequence has no sign change. If the rates coincide, combine coefficients and reach the same conclusion. Otherwise its unique possible zero is

$$k_* = \frac{\log|b_2(r_2 - 1)| - \log|b_1(r_1 - 1)|}{\log r_1 - \log r_2}.$$

Checking the endpoints and the clipped integers $\lfloor k_* \rfloor + 1, \lceil k_* \rceil + 1$ gives the exact minimum and maximum on all integer grid states: the forward difference has constant sign on either side of its sole zero, so no other integer extremum is possible. If k_* is an integer, $g(k_*) = g(k_* + 1)$; testing $k_* + 1$ therefore captures their common extremal value. For $n_b = 1$ only that state is needed. Check the required strict sign of every surface at these at most four indices, then propagate the block initial state by its exact spectral power to start the next block. Each restart belongs to the next region; the terminal state has no executed mask. These tests are necessary and sufficient for the entire proposed strict word. Zero signs return unresolved regardless of the convention used for ReLU at zero. Let S be the total number of surfaces selecting a region, including those absent from the event list. For two candidates and $m + 1$ regions, at most $8(m + 1)S$ scalar sign evaluations suffice, independently of N. The number 48 applies only when $m = 2$ and $S = 2$. This is a real-arithmetic operation count; bit precision near a zero is not uniformly bounded.

Endpoint masks alone cannot replace this test, even for a contracting symmetric regional matrix. Take $\eta = 1$, $M = \mathrm{diag}(-0.1, -0.19)$, zero affine forcing, initial state $(1, 1)$ and surface $g(x) = 0.24 - x_1 + x_2$. Then $g(k) = 0.24 - 0.9^k + 0.81^k$ is positive at $k = 0$ and $k = 20$ but negative at $k = 6$. Thus a proposed word that maintains the positive side throughout is false despite its matching endpoint signs. The spectral-extremum test rejects it. This example concerns verification of a regional extension, not a claim that the true switched path follows that invalid extension after its first crossing.

Scope, cost, and limiting cases. The spectral remainder is valid in any dimension with symmetric negative definite regional matrices; the constant-size test just proved is planar. The experiment uses base-point event roots, scheduled states and regional spectra. It does not use perturbed rerun states in its prediction or word test. Checking a later block requires its exact initial affine state; therefore the check is not a claim to avoid all candidate-state work. It avoids an N-step replay, but exact word evaluation in this solvable class is equally inexpensive. No speed advantage over that stronger baseline is asserted. Matrix-log expansions, contractivity and the two-exponential sign calculation are standard tools; their role here is to turn the coupled response into a nonvacuous finite decision guarantee.

With no event the recursion is the ordinary second-order Euler-error expansion. With commuting independent coordinate fields the coordinates decouple. Repeated eigenvalues are covered by the

spectral formulas and the merged-coefficient word test. Vanishing jumps remove the first-order rounding effect, but may leave a second-order effect when regional Hessians change; setting both fields equal removes that artificial event at every order. At a failed word, no error guarantee is issued. None of these finite results asserts uniformity as $\eta\|M_i\|$ approaches one, as a phase margin vanishes in the local-limit theorem, or as the number of events grows. Floating-point and high-precision validations below are not formal interval-arithmetic certificates.

General-dimensional work and arithmetic. The spectral remainder and jet recursion hold in any dimension d. Regional eigendecompositions cost $O((m+1)d^3)$; after these and the base data are available, ordered matrix-vector jets cost $O((m+1)d^2)$ per candidate, excluding evaluation of q and its gradient. The planar extremum formula does not establish a dimension-independent word test. A general safe alternative is to check all N executed states against all S surfaces, costing $O(N(d^2+Sd))$ per candidate for dense affine fields. A scalar surface sequence in dimension d contains up to d exponential modes; any faster general root isolation method requires its own proof and precision analysis. Neither cheap word evaluation in this solvable affine class nor the certificate proves a speedup for nonlinear network training.

The theorem is exact in real arithmetic. Its numerical implementation is a numerically evaluated deterministic certificate, with floating-point and independent high-precision checks. Those checks do not constitute formal interval-arithmetic certification. The piecewise-smooth endpoint theorem does not inherit this affine finite-grid certificate.

F Scalar branch cells, response levels, and frequencies

F.1 Finite-sample ReLU empirical-risk realization

Proposition 15 (Exact residual-ReLU flow and discrete derivatives)**.** *Fix $m > 0$, $c > -m$, and $T > 0$, and consider*

$$\mathcal{L}_{m,c}(\theta) = \tfrac{1}{2}(\theta - m)^2 + \tfrac{1}{2}(\mathrm{ReLU}(\theta) - c)^2. \tag{57}$$

For $\theta_0 \in (m(1-e^T), 0)$, let $t_* = \log[(m-\theta_0)/m]$ and $r = (m+c)/m > 0$. The flow has exactly one stable transverse crossing before T, and

$$J_{\mathrm{reg}} = e^{-t_*}e^{-2(T-t_*)}, \qquad D\varphi_T = rJ_{\mathrm{reg}}.$$

The regional Hessians are $H^- = 1$ and $H^+ = 2$, while the interface coefficient is $\beta = -c$. If $c \le 0$, the objective is globally 1-strongly convex; in the crossing range its unique minimizer is $(m+c)/2 > 0$.

For $\eta = T/N < 1/2$, define

$$\kappa_\eta = \frac{t_*}{-\log(1-\eta)}.$$

If $\kappa_\eta \notin \mathbb{N}$, the step size is nonresonant, the first positive iterate is $K_\eta = \lceil \kappa_\eta \rceil$, and

$$D\Theta_{\eta,T} = J^{\mathrm{AD}}_{\eta,T} = (1-\eta)^{K_\eta}(1-2\eta)^{N-K_\eta} \longrightarrow J_{\mathrm{reg}}. \tag{58}$$

Moreover, the piecewise-linear GD interpolation converges uniformly to the flow at $O(\eta)$ for each fixed pair (m, c).

Proof. For the constant input $x = 1$, define the standard two-output residual-ReLU network and target

$$z_\theta(1) = \theta, \qquad F_\theta(1) = \begin{bmatrix} \theta \\ \mathrm{ReLU}(\theta) \end{bmatrix}, \qquad y = \begin{bmatrix} m \\ c \end{bmatrix}.$$

Then $\frac{1}{2}\|F_\theta(1) - y\|_2^2$ is exactly (57). Its regional vector fields and Hessians are

$$f^-(\theta) = m - \theta, \quad H^- = 1, \qquad f^+(\theta) = m + c - 2\theta, \quad H^+ = 2.$$

Before the event, $\theta(t) = m + (\theta_0 - m)e^{-t}$, which reaches zero at t_*. The initialization interval is exactly the condition $0 < t_* < T$. The one-sided speeds at zero are m and $m + c$, so the crossing is stable, same-direction transverse, and has saltation factor $r = (m+c)/m$. After the event,

$$\theta(t) = \frac{m+c}{2}\big(1 - e^{-2(t-t_*)}\big) > 0,$$

so there is no recrossing. Multiplication of the regional propagators and the event factor gives the stated formulas for J_{reg} and $D\varphi_T$.

The normal derivative jumps from $-m$ to $-(m+c)$, hence $\beta = -c$. If $c \le 0$, the function $h(\theta) = \frac{1}{2}(\mathrm{ReLU}(\theta) - c)^2$ is convex: its left derivative at zero is 0, its right derivative has trace $-c \ge 0$, and the latter then increases with slope one. Since $\frac{1}{2}(\theta - m)^2$ is 1-strongly convex, their sum is globally 1-strongly convex. For $c > -m$, its positive-region critical point $(m+c)/2$ is positive and is therefore the unique minimizer. For $c > 0$, the downward derivative jump $-c < 0$ instead proves global nonconvexity.

For the discrete dynamics, while $\theta_k^\eta \le 0$,

$$\theta_k^\eta = m + (\theta_0 - m)(1-\eta)^k.$$

Thus an exact landing occurs precisely when κ_η is an integer, and otherwise the first positive iterate is $K_\eta = \lceil \kappa_\eta \rceil$. The incoming update gives $0 < \theta_{K_\eta}^\eta < \eta m$. Since $0 < 1 - 2\eta < 1$, the outgoing recurrence

$$\theta_k^\eta = \frac{m+c}{2} + \left(\theta_{K_\eta}^\eta - \frac{m+c}{2}\right)(1-2\eta)^{k-K_\eta}$$

is a convex combination of positive quantities at each step and therefore stays positive. Hence the discrete itinerary also has one crossing. Its branch sequence is locally fixed at nonresonant sizes, and differentiating the two affine recurrences gives (58). Because $-\log(1-\eta) = \eta + O(\eta^2)$, $K_\eta \eta = t_* + O(\eta)$. Taking logarithms in (58) therefore gives the claimed Jacobian limit. The standard uniform comparisons $(1-\eta)^k = e^{-k\eta}[1 + O(\eta)]$ and $(1-2\eta)^k = e^{-2k\eta}[1 + O(\eta)]$, together with the $O(\eta)$ crossing-time and overshoot errors, prove the uniform state bound. Its constant may depend on (m, c). □

Corollary 16 (Uniform-margin amplification and convex attenuation)**.** *Fix $a > 0$, $t_* \in (0, T)$, and a prescribed event ratio $r > 0$. Choose*

$$(m, c) = \begin{cases} (a/r, (r-1)a/r), & 0 < r \le 1, \\ (a, (r-1)a), & r \ge 1, \end{cases} \qquad \theta_0 = m(1 - e^{t_*}).$$

Then the event occurs at t_*, $\min\{f^-(0), f^+(0)\} = a$, and $D\varphi_T = r J_{\mathrm{reg}}$. If $r \le 1$, the objective is globally 1-strongly convex. In particular, for $R > 1$, the choice $m = Ra$, $c = -(R-1)a$ gives

$$D\varphi_T = R^{-1} J_{\mathrm{reg}}, \qquad \frac{|J_{\mathrm{reg}}|}{|D\varphi_T|} = R, \tag{59}$$

while a common interface tube $|\theta| < a/4$ has normal-speed lower bound $a/2$ for every R. The limit is taken for each fixed R and then $\eta \downarrow 0$; the Euler state-error constant is not asserted to be uniform in R.

Proof. The parameter choice gives $m + c = rm$ and makes the smaller of m and $m + c$ equal to a. The event-time formula and Proposition 15 prove the ratio. When $r \le 1$, $c \le 0$, so global 1-strong convexity follows from that proposition. In the reciprocal family, the incoming speed is Ra. On the incoming side of $|\theta| < a/4$ it is at least Ra, while on the outgoing side the speed is $a - 2\theta \ge a/2$. This proves the uniform transversality claim. □

Proposition 17 (Exact-hit set for the residual-ReLU family)**.** *For the pre-event recurrence in Proposition 15, the positive step sizes that place an evaluated iterate exactly at the interface are*

$$\eta_k = 1 - e^{-t_*/k}, \qquad k \in \mathbb{N}. \tag{60}$$

This set is countable, has Lebesgue measure zero, and accumulates at zero. Requiring $T = N\eta$ selects only its intersection with the mesh family $\{T/N : N \in \mathbb{N}\}$. For every $0 < t_* < T$, infinitely many canonical meshes are nonresonant.

Proof. An exact hit at iterate k means $(1-\eta)^k = e^{-t_*}$, which is equivalent to (60). The set is countable and $1 - e^{-t_*/k} \sim t_*/k$, proving accumulation at zero. On a resonant canonical mesh, the corresponding index is

$$k_N = \frac{t_*}{-\log(1 - T/N)} = \frac{t_*}{T} N - \frac{t_*}{2} + O(N^{-1}).$$

If every sufficiently large canonical mesh were resonant, both k_N and k_{N+1} would be integers, while their integer difference would converge to $t_*/T \in (0,1)$, which is impossible. Hence an infinite nonresonant subsequence exists. □

F.2 Perturbation radius, discrete branches, and derivative selection

This section separates an elementary consequence of uniform state accuracy from an exact, phase-dependent calculation for the residual-ReLU construction. The latter resolves a finite perturbation question that the limiting Jacobian alone cannot answer. Neither result is a claim of universal priority for two-scale numerical differentiation.

F.2.1 One target, three derivatives, and a directional branch radius

Let λ include the initialization and/or parameters of the regional field, and fix a scalar terminal objective

$$Q_\eta(\lambda) = q(\Theta_{\eta,T}(\lambda), \lambda), \qquad Q_0(\lambda) = q(\varphi_T(\lambda), \lambda).$$

$$D_{\eta,\epsilon}(v) = \frac{Q_\eta(\lambda + \epsilon v) - Q_\eta(\lambda - \epsilon v)}{2\epsilon}, \quad \|v\| = 1.$$

Write $g_\eta = \nabla Q_\eta(\lambda)$ for the exact finite-program gradient, $g_{\mathrm{reg}} = Z_{\mathrm{reg}}^\top q_\theta + q_\lambda$ for its regional limit, and $g_0 = \nabla Q_0(\lambda) = Z_{\mathrm{flow}}^\top q_\theta + q_\lambda$, where the last two expressions evaluate q at the flow endpoint. These identities require smooth terminal traces; a terminal ReLU boundary needs a separate analysis even if every training evaluation is nonresonant.

For a fixed mesh with strict training and terminal-validation masks, define $\rho_\eta(v)$ as the supremum of $r > 0$ for which every $|s| < r$ retains *all* executed training branches and terminal-validation branches at $\lambda + sv$, within the parameter domain and regional smooth extensions. This is a connected directional cell radius, not an endpoint mask test. If $\psi_i(s)$ are all signed branch tests of the base-branch extensions, and $|\psi_i'(s)| \le L_i$ on $[-r_0, r_0]$, then

$$\rho_\eta(v) \ge \min\{r_0, \min_{i:L_i>0} |\psi_i(0)|/L_i\}.$$

Constant nonzero tests impose no restriction. The bound follows from the mean-value theorem and induction over the finite executed program: before the first changed test the base extension is the actual program. The L_i must be bounds on an interval; a derivative sampled only at zero does not certify them. They and the margins can depend on η.

Proposition 18 (Elementary response bounds)**.** *If $\epsilon < \rho_\eta(v)$ and $s \mapsto Q_\eta(\lambda + sv)$ has third derivative bounded by $M_{3,\eta}$ on $[-\epsilon, \epsilon]$, then*

$$|D_{\eta,\epsilon}(v) - g_\eta^\top v| \le M_{3,\eta}\epsilon^2/6.$$

With only a second derivative bound $M_{2,\eta}$, the bound is $M_{2,\eta}\epsilon/2$. Thus a varying-mesh conclusion requires the corresponding remainder to vanish, as well as branch preservation.

Suppose on a fixed neighborhood U that $\sup_{\mu\in U} \|\Theta_{\eta,T}(\mu) - \varphi_T(\mu)\| \le C_s\eta$, and q is uniformly L_q-Lipschitz in its state argument on segments joining these endpoints. If $\lambda + [-\epsilon, \epsilon]v \subset U$ and Q_0 is C^1 there, then

$$|D_{\eta,\epsilon}(v) - g_0^\top v| \le L_q C_s \eta/\epsilon + \omega_{\nabla Q_0}(\epsilon). \tag{61}$$

If the third directional derivative of Q_0 is bounded by $M_{3,0}$, replace the modulus term by $M_{3,0}\epsilon^2/6$.

Proof. Taylor expansion at $s = 0$ cancels the even quadratic term in the central difference. Each cubic remainder is bounded by $M_3\epsilon^3/6$, giving the first claim after division by 2ϵ. Alternatively integrate the directional first derivative over $[-\epsilon, \epsilon]$; its difference from its value at zero is bounded by $M_2|s|$, whose average is $M_2\epsilon/2$. Uniform state accuracy and the Lipschitz bound give $|Q_\eta(\mu) - Q_0(\mu)| \le L_q C_s \eta$. Apply this at the two perturbed endpoints, subtract, and divide by 2ϵ. The integral formula for the central difference of Q_0 supplies the remaining modulus or Taylor bound. No branch preservation of Q_η is needed for this last step. □

When the neighborhood bound is available. For initialization perturbations, restrict (A1)–(A5) to a compact closure of a sufficiently small neighborhood with common isolating tubes, a common finite itinerary, uniformly positive endpoint/event separation Δ_t, normal speed c_0, and distance from other interfaces outside the tubes. Require uniform regional field/Jacobian bounds and C^2 interface geometry on a common compact tube. Then the proof of Theorem 6 is uniform on that neighborhood. Indeed each smooth-segment recurrence has the same Gronwall constant; the interface coordinate error divided by c_0 bounds each clock error; bounded speed converts it to a restart error; a finite induction over the same number of tubes bounds every restart by $C_j\eta$. Taking their maximum and the common mesh threshold proves the asserted uniform bound. The constants depend on T, m, M_f, M_A, c_0^{-1}, the tube sizes and geometry, and the inverse separation and non-event margins, not the grid or perturbed initialization. For field/initialization/surface parameters, require the same assumptions uniformly in λ with jointly smooth regional extensions and smooth initialization; apply the same argument to $(\theta, \lambda)' = (f(\theta, \lambda), 0)$. The state estimate alone needs no differentiability across discrete cells. Thus $\epsilon \to 0$ and $\eta/\epsilon \to 0$ suffice *with* these uniform hypotheses and terminal regularity. Pointwise accuracy at a single base point, or a collapsing event separation or crossing speed, does not supply this conclusion. This is a corollary of numerical consistency and the triangle inequality, not a new sensitivity mechanism.

F.2.2 Exact cells and jumps of the residual-ReLU endpoint

Fix $m > 0$, $c > -m$, $T > 0$, and $x \in U_T = (m(1 - e^T), 0)$. Put

$$t = \log((m - x)/m), \quad P = (m + c)/2, \quad B = m - x,$$
$$E = e^{-2(T-t)}, \quad F_0(x) = P(1 - E), \quad J = e^{-t}E.$$

Here x is the initialization and the direction is the fixed scalar $v = 1$. For $\eta = T/N < 1/2$, let $a = 1 - \eta$, $b = 1 - 2\eta$, $h = -\log a$, $\kappa = t/h$, and

$$z_k = m(1 - a^{-k}), \qquad C_K = (z_K, z_{K-1}).$$

Only cells with $1 \le K < N$ are used below; compact subsets of U_T satisfy this condition for all sufficiently fine meshes. Boundary initializations are excluded when a classical derivative is asserted.

Theorem 19 (Exact branch and jump response)**.** *On C_K, the Euler endpoint and its derivative are*

$$F_\eta(y) = F_K(y) := P + \big[m + (y - m)a^K - P\big]b^{N-K}, \qquad s_K = F_K'(y) = a^K b^{N-K}. \tag{62}$$

For a smooth terminal objective the training-cell radius at $x \in C_K$ is

$$\rho_\eta = \min\{x - z_K, z_{K-1} - x\}, \qquad |C_K| = m\eta a^{-K}. \tag{63}$$

If the objective has additional branches this radius must be intersected with their inverse images. At an interior boundary z_k, $1 \le k < N$, the right-minus-left endpoint jump is

$$[F_\eta]_{z_k} := F_k(z_k) - F_{k+1}(z_k) = \eta c\, b^{N-k-1}. \tag{64}$$

For $q \in C^1$, perturbation endpoints away from these boundaries, and an interval contained in the above cells, the exact identity is

$$D_{\eta,\epsilon}(1) = \frac{1}{2\epsilon}\left[\int_{x-\epsilon}^{x+\epsilon} q'(F_\eta(y))s_{K(y)}\, dy + \sum_{z_k \in (x-\epsilon, x+\epsilon)} \big(q(F_k(z_k)) - q(F_{k+1}(z_k))\big)\right]. \tag{65}$$

For $q(z) = \frac{1}{2}(z - y_*)^2$ and $\epsilon < \rho_\eta$, the central response equals the exact discrete derivative without a remainder.

Proof. The incoming recurrence solves to $m + (y - m)a^k$. Its first positive state has index K exactly when $z_K < y < z_{K-1}$. That state lies in $(0, m\eta)$. Every subsequent update is $b\theta + \eta(m + c)$, a convex combination of positive values because $0 < b < 1$ and $P > 0$. There is no recrossing. Solving this recurrence proves (62) and the cell endpoints; subtracting z_K from z_{K-1} gives the width. At $y = z_k$, the limit from the right uses k incoming steps and then starts the outgoing recurrence from zero; the limit from the left uses one extra incoming step and starts it from $m\eta$. Consequently

$$F_k(z_k) = P(1 - b^{N-k}), \qquad F_{k+1}(z_k) = P + (m\eta - P)b^{N-k-1}.$$

Their difference is $[P(1-b)-m\eta]b^{N-k-1} = \eta c b^{N-k-1}$. Apply the fundamental theorem of calculus separately on each smooth subinterval and add the intervening right-minus-left jumps; telescoping gives (65). Inside one cell a quadratic q composed with affine F_K is quadratic, whose central difference is exact. The value assigned at an interior jump does not affect this identity; if a perturbation endpoint itself hits a jump, its selected value must be used instead of an unspecified one-sided limit. □

Uniform constants in this construction. For $t(y) \in [\delta, T-\delta]$, $0 < \delta < T/2$, and $\eta < \min\{1/4, \delta\}$, the preceding formulas give

$$\sup_y |F_\eta(y) - F_0(y)| \le C_s \eta, \qquad C_s = m + P(2+6T). \tag{66}$$

To verify this bound, $h - \eta \le \eta^2/[2(1-\eta)]$ and $K = \lceil t/h \rceil$ give $|K\eta - t| \le (1+T)\eta$ (at a boundary either adjacent choice obeys the same bound). Also $|\log(1-2\eta) + 2\eta| \le 4\eta^2$. Thus the exponents of b^{N-K} and $e^{-2(T-t)}$ differ by at most $(2+6T)\eta$. They are nonpositive, where the exponential is 1-Lipschitz. The incoming overshoot is at most $m\eta$, proving the bound. All endpoints lie between zero and $\max\{P, m/4\}$, so for the squared target y_* one may take $L_q = |y_*| + \max\{P, m/4\}$. No event separation other than the two endpoint distances is needed for this one-event result. Since $F_0(y) = P[1 - e^{-2T}(1 - y/m)^2]$, $Q_0 = q \circ F_0$ is explicitly smooth for a smooth q. The constants cease to be uniform if the instance scales without bound; $c = -m$ is excluded because outgoing transversality fails.

F.2.3 Critical scales and a fixed-initialization counterexample

Theorem 20 (Phase-resolved critical limit)**.** *For the fixed instance and initialization above, consider nonresonant canonical meshes with fractional phases $p_N = \{t/h\} \to p \in [0,1]$ and $\epsilon_N/\eta_N \to \zeta \in (0,\infty)$. Suppose $\zeta \ne B|p-j|$ for every $j \in \mathbb{Z}$. For $q \in C^1$ near $F_0(x)$,*

$$D_{\eta_N,\epsilon_N}(1) \longrightarrow q'(F_0(x))\left[J + \frac{cE}{2\zeta} H(p,\zeta)\right], \qquad H(p,\zeta) = \#\{j \in \mathbb{Z} : B|p-j| < \zeta\}, \tag{67}$$

$$\rho_{\eta_N}/\eta_N \longrightarrow B \min\{p, 1-p\}. \tag{68}$$

Thus ϵ/η alone does not determine the limit. At the excluded endpoint-equality phases, the higher-order endpoint approach and the value selected at a hit matter, and the formula is not asserted.

Proof. Write $\kappa = \lfloor \kappa \rfloor + p_N$. For each fixed integer j,

$$\frac{z_{\lfloor \kappa \rfloor + j} - x}{\eta} = \frac{B[1 - \exp(h(j - p_N))]}{\eta} \longrightarrow B(p-j).$$

The two nearest boundaries prove the radius formula. On an $O(\eta)$ interval the cell widths divided by η are uniformly bounded below, so only a bounded number of boundaries can be present. The strict endpoint condition makes their count eventually $H(p,\zeta)$. On every encountered cell $K\eta \to t$, hence $s_K \to J$, uniformly over this bounded set; also $F_\eta(y) \to F_0(x)$ uniformly. The integral term in (65) therefore converges to $q'(F_0(x))J$. For each encountered boundary, (64) divided by η converges to cE. The mean-value theorem and continuity of q' give the corresponding objective jump divided by η as $q'(F_0(x))cE + o(1)$. Divide their finite sum by 2ϵ to finish the proof. □

There is no unspecified universal constant in (67): B, E, J, c and the integer count are given explicitly. For a compact family of strictly separated scaled endpoints, the proof is uniform when B is bounded away from zero, m, c, T, t stay in compact admissible sets, and q' has a common continuity modulus. Without a positive gap from $\zeta = B|p-j|$, stabilization of the count is not uniform. At $c = 0$ the jumps vanish and the leading phase dependence disappears. For large ζ, $H(p,\zeta)/(2\zeta) \to 1/B$ uniformly in p, and $J + cE/B = (m+c)E/B = DF_0(x)$. This checks agreement with the coarser-radius bound, not an interchange of unproved joint limits.

Corollary 21 (Small relative radius need not select the discrete gradient)**.** *If t/T is irrational and $cq'(F_0(x)) \ne 0$, there is a nonresonant canonical subsequence and radii ϵ_N such that*

$$\epsilon_N/\eta_N \to 0, \qquad \epsilon_N > \rho_{\eta_N}, \qquad |D_{\eta_N,\epsilon_N}(1)| \to \infty,$$

although $Q'_{\eta_N}(x) \to q'(F_0(x))J$ is finite. In particular, with $m = 4, c = -3, T = 1, t = \sqrt{2}/4$ and the fixed objective $q(z) = \frac{1}{2}(z-1)^2$, the exact discrete and flow derivatives are both negative, but along that subsequence $D_{\eta_N,\epsilon_N}(1) > 0$. Consequently the ordering of the two predefined proposals $x \pm \epsilon_N$ is opposite to the ordering predicted by either infinitesimal derivative.

Proof. The expansion $\kappa_N = (t/T)N - t/2 + O(N^{-1})$ and the equidistribution proof in Corollary 22 give arbitrarily large indices with fractional phase in every open subinterval of $(0, 1)$. A diagonal selection therefore gives strictly positive $p_N \to 0$ without hitting zero. Equation (68) gives $\rho_N/\eta_N \to 0$. Set $\epsilon_N = \sqrt{\rho_N \eta_N}$. For large N this interval contains the nearest boundary and no other boundary, since neighboring widths are asymptotic to $B\eta_N$. Its regular response is bounded, whereas its single objective jump divided by $2\epsilon_N$ is

$$\frac{\eta_N}{2\epsilon_N}\big[cEq'(F_0(x)) + o(1)\big],$$

which diverges with the sign of $cq'(F_0(x))$. For the stated squared objective $0 < F_0(x) < 1/2$, hence $q'(F_0(x)) < 0$; $c < 0$ makes the jump contribution positive, while $J > 0$ and $DF_0(x) > 0$ make both derivatives negative. The central-response sign is precisely the sign of $Q_\eta(x + \epsilon) - Q_\eta(x - \epsilon)$, proving the proposal comparison. □

The radii in this counterexample are phase-adaptive mechanism probes. The task and direction, and the base initialization, are fixed. The result does not assert that such phases are typical at an externally prescribed radius, nor that a normalized flow gradient differs in this scalar example. Absolute objective differences still vanish: a large response here divides a small $O(\eta)$ jump by an even smaller radius. This explains why infinitesimal correctness supplies no mesh-independent decision radius. Theorem 2 handles fixed finite separated event sequences; changing itineraries, colliding events and certified network cell radii remain open.

F.3 Critical-scale response classification

The following consolidates consequences of Theorem 20; it is not a separate sensitivity mechanism. Phase measure below is Lebesgue measure on $[0, 1)$, not a distribution over data, seeds, or trained models.

Corollary 22 (Classification and canonical-mesh frequencies)**.** *In Theorem 20, fix $\zeta > 0$ and put*

$$\ell = 2\zeta/B, \quad n = \lfloor \ell \rfloor, \quad r = \ell - n, \quad a = q'(F_0)J, \quad b = q'(F_0)cE/(2\zeta).$$

Except at the finitely many endpoint phases $B|p - j| = \zeta$, the count H takes n on a set of measure $1 - r$ and $n + 1$ on a set of measure r (the second set is empty when $r = 0$). Consequently the critical response has values $a + bn$ and $a + b(n + 1)$ with those masses, and

$$\int_0^1 R_\zeta(p)\, dp = Q'_0(x), \qquad \mathrm{Var}_p\, R_\zeta = b^2 r(1 - r). \tag{69}$$

The variance vanishes exactly when $cq'(F_0) = 0$ or ℓ is an integer. For positive integer ℓ, every allowed phase has response $Q'_0(x)$.

If t/T is irrational and $\epsilon_N = \zeta T/N$, then the empirical distribution of the actual responses $D_{T/N,\epsilon_N}$ converges to this two-point distribution. In particular, if every level with positive mass is nonzero, the frequency of positive responses is

$$(1 - r)\mathbf{1}_{a+bn>0} + r\mathbf{1}_{a+b(n+1)>0}.$$

This is also the frequency with which the predefined positive proposal has higher objective than the negative proposal. For comparison to any fixed nonzero derivative sign s, replace each indicator by $\mathbf{1}_{s(a+bk)<0}$. If a positive-mass level is zero, its limiting sign is not determined: the lower and upper frequency bounds for positive responses are the masses of $R_\zeta > 0$ and $R_\zeta \geq 0$, respectively.

Proof. Write $u = p - \zeta/B$. The number of integers in $(u, u + \ell)$, away from integer endpoints, is $\lfloor u + \ell \rfloor - \lfloor u \rfloor$. It equals $n + \mathbf{1}_{\{u\}>1-r}$ outside the endpoint phases. Translation on the circle preserves measure, proving the two masses. The mean is $a + b\ell = q'(F_0)(J + cE/B) = Q'_0(x)$; the

variance is that of a Bernoulli variable of mass r, multiplied by b^2. Since E, B, ζ are strictly positive, these are exactly the stated degeneracies.

For the frequency statement one needs equidistribution, not density. Expanding $-\log(1 - T/N)$ gives

$$\frac{t}{-\log(1 - T/N)} = \alpha N - t/2 + O(N^{-1}), \qquad \alpha = t/T.$$

For each nonzero integer h, the geometric sum of $\exp(2\pi i h \alpha N)$ is bounded by $2/|1 - e^{2\pi i h \alpha}|$. Changing its argument by $O(N^{-1})$ changes the normalized sum through M by $O_h((1+\log M)/M)$. Thus every nonconstant Fourier average of the actual phases tends to zero. Trigonometric-polynomial approximation gives convergence for continuous periodic functions; bounding interval indicators above and below by continuous functions gives equidistribution.

It remains to transfer phase measure to actual finite-grid responses. Remove a neighborhood of width δ around each of the finitely many endpoint phases. On its compact complement, the boundary expansion in Theorem 20 is uniform, the integer counts stabilize, and $D_{T/N,\zeta T/N} - R_\zeta(p_N) \to 0$ uniformly. Inside the removed set the responses remain uniformly bounded: the cell widths are bounded below by a constant times η, there are only boundedly many jumps, and each jump is $O(\eta)$. Equidistribution bounds the upper density of removed indices by the set's measure. First let $N \to \infty$, then $\delta \downarrow 0$. This proves the distribution and moment limits and, when the occupied levels are nonzero, the sign frequencies. The same argument bounds zero-level sign frequencies without assigning their signs. An exact finite-grid endpoint hit can only occur in the removed sets eventually, so its convention cannot change these frequencies. □

At $q'(F_0) = 0$ both limiting responses vanish and higher-order terms decide finite comparisons. At $c = 0$ the phase term disappears; a nonzero $q'(F_0)$ still gives the common nonzero derivative sign. Rational t/T is not covered by the equidistribution statement. Positive integer ℓ is a special critical-scale cancellation, not a phase-independent law for arbitrary radii. No joint or independent phase distribution is asserted for the coupled system.

F.4 Proof and scope of the resolution synthesis

The three sufficient regimes in Section 3 combine existing statements; this section supplies the complete deduction and its scope. Fix $Q_\eta = q \circ \Theta_{\eta,T}$, $Q_0 = q \circ \varphi_T$, a base initialization x, and a unit direction v. Write $g_\eta = \nabla Q_\eta(x)$ at a nonresonant program, g_{reg} for its regional limit, and $g_0 = \nabla Q_0(x)$. Terminal covectors in the latter two objects are evaluated at the flow endpoint. The central response is $D_{\eta,\epsilon}(v) = [Q_\eta(x+\epsilon v) - Q_\eta(x-\epsilon v)]/(2\epsilon)$. The critical clause assumes the fixed finite-event geometry of Theorem 2, $q \in C^2$ near its terminal point, and a common positive recursive phase margin on the stated compact candidate set. The coarse clause assumes a fixed neighborhood containing the whole proposed perturbation interval.

The directional cell is a whole-interval property. At a nonresonant grid, list all signed tests $\psi_i(s)$ of the base-branch program, including terminal validation and any domain tests. Assume their smooth extensions exist on $[-r_0, r_0]$, their values at zero are nonzero, and $|\psi_i'(s)| \le L_i$ throughout that interval. Then

$$\rho_\eta(v) \ge \min\{r_0, \min_{i:L_i>0} |\psi_i(0)|/L_i\}.$$

Tests with $L_i = 0$ impose no restriction. The mean-value theorem keeps each extended test strictly on its original side inside the displayed radius. Induction through the program then identifies the extended execution with the actual one. Bounds computed only at the base point, or agreement of two endpoint masks, do not establish this claim. All tests, margins, derivative bounds and the resulting radius may depend on η.

Within-cell remainder with only C^2 regularity. Let $h(s) = Q_\eta(x + sv)$. Two integral Taylor formulas give exactly

$$\frac{h(\epsilon) - h(-\epsilon)}{2\epsilon} - h'(0) = \frac{1}{2\epsilon} \int_0^\epsilon (\epsilon - u)[h''(u) - h''(-u)]\, du.$$

Define the centered modulus $\omega_{h''}(r) = \sup_{|s|\le r} |h''(s) - h''(0)|$. The absolute value is at most

$$\frac{1}{\epsilon}\int_0^\epsilon (\epsilon - u)\omega_{h''}(u)\,du \le \frac{\epsilon}{2}\omega_{h''}(\epsilon).$$

For fixed η, continuity of h'' makes this $o(\epsilon)$. Across varying grids, pointwise C^2 regularity gives no uniform modulus: the displayed product must tend to zero. A simpler bound $\sup|h''| \le M_{2,\eta}$ gives $M_{2,\eta}\epsilon/2$ directly from the same identity. If $|h'''| \le M_{3,\eta}$, then $|h''(u) - h''(-u)| \le 2M_{3,\eta}u$, and integration gives $M_{3,\eta}\epsilon^2/6$. The critical theorem's C^2 regional-field hypotheses do not by themselves supply this additional third-derivative bound. Under the derivative-selection theorem, write $\alpha_\eta = \|g_\eta - g_{\mathrm{reg}}\| \to 0$ at nonresonant meshes. The within-cell bound plus α_η also bounds $|D_{\eta,\epsilon}(v) - g_{\mathrm{reg}}^\top v|$. This extra limit concerns the regional object, not ∇Q_0.

Critical response and an explicit objective constant. The endpoint theorem provides, uniformly on the specified compact K,

$$\Theta_{\eta,T}(x + \eta z) = x_T + \eta e_{p_N}(z) + r_\eta(z), \quad \|e_{p_N}(z)\| \le B, \quad \|r_\eta(z)\| \le C\eta^2.$$

Let $G = \|\nabla q(x_T)\|$ and bound the Hessian of q by M_q on a ball around x_T. Add to the mesh threshold the requirement that $\eta B + C\eta^2$ lies inside this ball, and fix an upper mesh $\bar\eta$. Taylor's formula yields

$$\left|Q_\eta(x + \eta z) - q(x_T) - \eta\nabla q(x_T)^\top e_{p_N}(z)\right| \le C_Q\eta^2, \qquad C_Q = GC + \tfrac{1}{2}M_q(B + C\bar\eta)^2.$$

Apply this at $z = \pm\zeta v$ and divide their difference by $2\zeta\eta$. This proves the critical bound. The detailed endpoint theorem does not need $0 \in K$ unless a base-subtracted response is requested, so a two-point $K = \{-\zeta v, \zeta v\}$ is sufficient for this central comparison. The three-point version in the main box matches the main theorem's presentation and is slightly stronger than necessary.

The constant depends on the fixed event count, time, scaled displacement bound, field and surface derivative bounds, normal speeds, event separation and isolating-tube margins. The mesh threshold also depends on the positive recursive phase margin μ; this controls the actual integer choices against the $O(\eta^2)$ surface residuals. It is not enough to require only raw distances $\mathrm{dist}(t_j/\eta, \mathbb{Z})$. For a phase limit $p_N \to p$ whose recursive margins are positive on K, induction through the finitely many ceilings fixes their integers for nearby phases and gives $\sup_K \|e_{p_N} - e_p\| \le C_p\,\mathrm{dist}(p_N, p)$. This supplies the additional term $GC_p\,\mathrm{dist}(p_N, p)/\zeta$. The induction does not assume independent, dense or uniform event phases.

The compact set can be finite or disconnected. Requiring a positive margin on a connected interval crossing a recursive branch boundary would exclude the very cross-branch comparisons the theorem describes. By contrast, the within-cell radius deliberately requires every point of its interval to preserve the branch. These are different conditions. Fixed $\zeta > 0$ is essential to the stated $O(\eta)$ response bound. For a family ζ_η in a fixed compact subset of $(0, \infty)$, the same bound follows if one common compact K and one common phase margin contain all requested endpoints. Letting $\zeta_\eta \to 0$ or $\zeta_\eta \to \infty$ requires separate uniform remainder and geometry control; it is not an unqualified consequence.

Coarser perturbations. At the two endpoints, uniform objective consistency contributes at most $2C_0\eta$ to the numerator. Also

$$\frac{Q_0(x + \epsilon v) - Q_0(x - \epsilon v)}{2\epsilon} = \frac{1}{2\epsilon}\int_{-\epsilon}^{\epsilon} \nabla Q_0(x + sv)^\top v\,ds.$$

Subtract $\nabla Q_0(x)^\top v$, use $\|v\| = 1$, and add the two bounds. Uniform state accuracy $C_s\eta$ and a uniform state-Lipschitz constant L_q imply $C_0 = L_qC_s$. Such a state constant follows from the existing separated-crossing state proof on a fixed neighborhood with common finite itinerary, event/end-point separation, transverse speeds, field bounds, and tube/non-event margins. A state estimate at the base point alone is insufficient. Smoothness of the flow objective must include the terminal validation trace; training transversality alone does not imply $Q_0 \in C^1$. This triangle-inequality argument is a direct corollary, not a new approximation theorem.

A critical response need not be a gradient. This failure already occurs for a planar single event. Let $T = 1$, $x = (-1/2, 0)$, guard $x_1 = 0$, incoming field $(1, 0)$, outgoing field $(2, 0)$, and $q(y) = y_1$. On odd grids N, the event phase is $p = 1/2$, the smooth bias vanishes, and the endpoint response is exactly

$$e_p(z) = \big(z_1 - \lceil 1/2 - z_1 \rceil + 1/2,\ z_2\big).$$

All fields and the guard satisfy the finite-event theorem's smoothness, separation and same-direction speed conditions. Fix $\zeta = 3/4$. Coordinate comparisons give $\widehat{D}_{\eta,\zeta}(e_1) = 7/3$ and $\widehat{D}_{\eta,\zeta}(e_2) = 0$. For the unit vector $v = (1/3, 2\sqrt{2}/3)$, however,

$$\widehat{D}_{\eta,\zeta}(v) = 1/3 \neq (7/3, 0)^\top v = 7/9.$$

Every ceiling argument used here is at least $1/4$ from an integer. Thus the counterexample is strictly away from boundary ambiguity, and the critical prediction equals the exact finite response on these grids. One may define a *finite coordinate-response vector* $h_i^{(\zeta)} = \widehat{D}_{\eta,\zeta}(e_i)$ and report its cosine with another explicitly defined vector, but that cosine does not measure agreement of directional derivatives and $h^{(\zeta)\top} v$ does not generally predict another direction's finite response.

Decision and parameter scope. Any verified error bound $|D - \widehat{D}| \le E$ certifies the sign only when $|\widehat{D}| > E$; otherwise the bound returns unresolved. It also certifies an absolute candidate-objective gap of at least $2\epsilon(|\widehat{D}| - E)$. An asymptotic $O(\eta)$ with an uncomputed constant is not by itself a numerical certificate. The affine finite-word certificate supplies a separately evaluated bound; the network experiment currently supplies observed response errors. No Taylor bound guarantees that AD is the most accurate predictor at every cell-preserving radius, and no coarse-radius consistency bound guarantees that the flow proposal has lower loss on every finite grid. For a fixed-dimensional smooth outer parameter, the critical theorem can be applied to the augmented state (θ, λ) with $\dot{\lambda} = 0$ and jointly C^2 fields/guards, provided its same geometric hypotheses are checked. A parameter changing architecture, program length, or the time discretization is not covered by this constant-state augmentation.

G Supporting convexity, event geometry, and mismatch results

G.1 Atomic curvature and one-event geometry

Spatial versus pathwise curvature. The distributional Hessian has regional density and interface mass (Appendix G.3); along a trajectory the event mass is normalized by crossing speed.

Proposition 23 (Speed-normalized pathwise curvature)**.** *Along the reference itinerary, define*

$$\mathrm{d}K_\theta(t) = H_{\alpha(t)}(\theta(t))\,\mathrm{d}t + \sum_{j=1}^{m} \frac{\beta_j}{\nu_j^\top f_j^-} \nu_j \nu_j^\top\, \delta_{t_j}(\mathrm{d}t). \tag{70}$$

The left/prepoint Stieltjes system $Z(t) = I - \int_{(0,t]} \mathrm{d}K_\theta(s) Z(s^-)$ satisfies $Z(t_j^+) = \Xi_j Z(t_j^-)$ and $Z(T) = D\varphi_T$. Deleting its atoms gives the discrete-AD limit J_{reg}.

The atoms act by prepoint multiplication, not exponentiation; the spatial and pathwise distinction is detailed in Appendix G.3. BV calculus and saltation are classical (Demengel, 1984; Ambrosio et al., 2000; Hiskens & Pai, 2000; Kong et al., 2024); our result identifies which part survives in the exact discrete derivative sequence.

One-event discrepancy.

Proposition 24 (One-event rank-one geometry)**.** *For one transverse event,*

$$\Delta_T := D\varphi_T - J_{\mathrm{reg}} = -\frac{\beta}{\nu^\top f^-} (\Phi_1 \nu)(\nu^\top \Phi_0), \tag{71}$$

$$\|\Delta_T\|_2 = \|\Delta_T\|_F = \frac{|\beta|}{|\nu^\top f^-|} \|\Phi_1 \nu\|_2 \|\Phi_0^\top \nu\|_2. \tag{72}$$

If $\beta \neq 0$, the gap has rank one. For a terminal scalar objective, its outer-gradient gap then vanishes exactly when the terminal gradient is orthogonal to $\Phi_1 \nu$. Appendix G.3 gives the kernel, image, adjoint, and random-direction formulas.

G.2 Distributional Hessian and singular interface curvature

Proposition 25 (Distributional Hessian)**.** *Let a C^2 hypersurface Σ split a neighborhood into minus and plus regions. Suppose $\mathcal{L}$ is continuous and has C^2 extensions $\mathcal{L}^{\pm}$ with one-sided gradients. For the unit normal ν from minus to plus, there is a scalar trace β such that*

$$[\nabla \mathcal{L}] = \beta \nu,$$

and, as a matrix-valued distribution,

$$D^2\mathcal{L} = H^- \mathbf{1}_- \, \mathrm{d}x + H^+ \mathbf{1}_+ \, \mathrm{d}x + \beta \nu\nu^\top \, \mathcal{H}^{d-1} \lfloor_\Sigma \, .$$

Proof. The traces of $\mathcal{L}^+$ and $\mathcal{L}^-$ agree on Σ. Their tangential derivatives therefore agree, so their gradient difference is normal: $[\nabla \mathcal{L}] = \beta\nu$. Integrating each component of $\nabla\mathcal{L}$ by parts on the two regions gives the regional Hessians as volume terms. The outward normal of the minus region is ν, while that of the plus region is $-\nu$; the boundary terms combine as $[\nabla\mathcal{L}] \otimes \nu = \beta\nu\nu^\top$, which is the stated measure. □

The decomposition is standard BV calculus. Its role here is to locate the curvature omitted by regional Hessian products. Pulling the interface term through a transverse trajectory must account for the changing normal speed. In the resolved layer of Appendix I.1, the scalar identity $\mathrm{d}\log Z = \mathrm{d}\log f$ converts the concentrated curvature into f^+/f^-, exactly the normal saltation factor. A naive exponential of the bare hypersurface mass would be incorrect.

G.3 Atomic curvature and event geometry

This section proves the pathwise-curvature, one-event, convexity, and ReLU-event statements used in Sections G.1 and G.4. The Stieltjes convention below is part of the definition; it is important because the vector field itself jumps at an event.

G.3.1 Prepoint Stieltjes formulation

Along the reference trajectory, write

$$H(t) = H_{\alpha(t)}(\theta(t)), \qquad a_j = \frac{\beta_j}{\nu_j^\top f_j^-},$$

and define the finite matrix-valued measure

$$K_\theta(B) = \int_B H(t)\,\mathrm{d}t + \sum_{j:t_j \in B} a_j \nu_j \nu_j^\top. \tag{73}$$

Here and below, event data are evaluated at the nominal event state. A right-continuous matrix function Z, with left limits, solves the *prepoint* Stieltjes system when

$$Z(t) = I - \int_{(0,t]} \mathrm{d}K_\theta(s)\, Z(s^-). \tag{74}$$

Thus an atom acts linearly on the left limit. Equation (74) is not an exponential convention for matrix atoms.

Proof of Proposition 23. Off the event times, the absolutely continuous part of (74) gives

$$\dot{Z}(t) = -H(t)Z(t) = A_{\alpha(t)}(\theta(t))Z(t).$$

At t_j, subtracting the left limit from the right value gives

$$\begin{aligned} Z(t_j^+) - Z(t_j^-) &= -K_\theta(\{t_j\})Z(t_j^-) \\ &= -\frac{\beta_j}{\nu_j^\top f_j^-}\nu_j\nu_j^\top Z(t_j^-). \end{aligned}$$

Hence

$$Z(t_j^+) = \left(I - \frac{\beta_j \nu_j \nu_j^\top}{\nu_j^\top f_j^-} \right) Z(t_j^-) = \Xi_j Z(t_j^-).$$

Solving the regional linear equations and applying these finitely many jumps in chronological order yields

$$Z(T) = \Phi_m \Xi_m \Phi_{m-1} \cdots \Phi_1 \Xi_1 \Phi_0 = D\varphi_T.$$

Existence and uniqueness follow from the same finite construction. If the atoms are removed from K_θ, the solution remains continuous at each event and its terminal value is $\Phi_m \cdots \Phi_0 = J_{\mathrm{reg}}$. Theorem 1 identifies this atom-free solution with the vanishing-step branchwise-AD limit. □

The spatial measure in Proposition 25 and the pathwise measure in (73) are related but not identical objects. Since the trajectory velocity has two one-sided values at the interface, a bare distributional pullback does not select a crossing speed. The incoming-speed normalization and the prepoint convention in (74) encode the fixed-clock event derivative. In particular, $\exp(-a_j \nu_j \nu_j^\top)$ is generally different from $I - a_j \nu_j \nu_j^\top = \Xi_j$.

G.3.2 One-event rank-one geometry

Proof of Proposition 24. For one event, Theorem 1 and (13) give

$$\begin{aligned} \Delta_T &:= D\varphi_T - J_{\mathrm{reg}} = \Phi_1(\Xi - I)\Phi_0 \\ &= -\frac{\beta}{\nu^\top f^-}(\Phi_1 \nu)(\nu^\top \Phi_0). \end{aligned} \tag{75}$$

Both regional fundamental matrices are invertible, so $\Phi_1 \nu \neq 0$ and $\Phi_0^\top \nu \neq 0$. If $\beta \neq 0$, (75) is therefore a nonzero outer product and has rank one. A rank-one matrix $uv^\top$ has one nonzero singular value $\|u\|_2 \|v\|_2$. This proves

$$\|\Delta_T\|_2 = \|\Delta_T\|_F = \frac{|\beta|}{|\nu^\top f^-|} \|\Phi_1 \nu\|_2 \|\Phi_0^\top \nu\|_2.$$

The same outer-product formula gives

$$\ker \Delta_T = (\Phi_0^\top \nu)^\perp, \qquad \operatorname{im} \Delta_T = \operatorname{span}(\Phi_1 \nu),$$

and, for $g_T = \nabla q(\varphi_T(\theta_0))$,

$$\Delta_T^\top g_T = -\frac{\beta}{\nu^\top f^-}(\nu^\top \Phi_1^\top g_T) \Phi_0^\top \nu.$$

Since the last vector factor is nonzero, the hypergradient discrepancy vanishes exactly when $\beta = 0$ or $g_T \perp \Phi_1 \nu$.

Finally, if V is uniform on the unit sphere in $\mathbb{R}^d$, then $\mathbb{E}[VV^\top] = I/d$, and hence

$$\mathbb{E}\|\Delta_T V\|_2^2 = \operatorname{tr}\big(\Delta_T^\top \Delta_T \, \mathbb{E}[VV^\top]\big) = \frac{\|\Delta_T\|_F^2}{d}.$$

This is a directional statement at a fixed event, not a prevalence claim for network trajectories. □

G.4 Convex and strongly convex consequences (supporting results)

The attenuation and determinant statements below are direct consequences of classical normal-jump geometry and hybrid composition; the squared-loss and coupled examples supply explicit realizations. See Appendix K for ownership.

Does convexity restore tangent consistency? No. Convexity constrains direct event geometry, but it does not remove event timing. In fact, a strict convex event prevents complete cancellation.

Theorem 26 (Convex attenuation and no complete cancellation)**.** *Suppose, in addition to Theorem 1, that $\mathcal{L}$ is globally convex. At every traversed interface,*

$$[\nabla \mathcal{L}]_j = \beta_j \nu_j \quad \Longrightarrow \quad \beta_j \geq 0, \qquad r_j = 1 - \frac{\beta_j}{\nu_j^\top f_j^-} \in (0, 1]. \tag{76}$$

Moreover,

$$\frac{\det D\varphi_T}{\det J_{\mathrm{reg}}} = \prod_{j=1}^{m} r_j. \tag{77}$$

Thus, if any event is strict ($\beta_j > 0$, equivalently $r_j < 1$), then the product is below one and transported event factors cannot cancel completely.

Corollary 27 (Quantitative event-volume certificate)**.** *Under Theorem 26, using the transported factors of Eq. (91) in Appendix G.5, define the cumulative relative event map and its volume ratio by*

$$\mathcal{C}_T := J_{\mathrm{reg}}^{-1} D\varphi_T = \widetilde{\Xi}_m \cdots \widetilde{\Xi}_1, \qquad \delta_T := \det \mathcal{C}_T = \prod_{j=1}^{m} r_j \in (0,1].$$

Then

$$\|\mathcal{C}_T - I\|_2 \geq 1 - \delta_T^{1/d}, \qquad \|D\varphi_T - J_{\mathrm{reg}}\|_2 \geq \sigma_{\min}(J_{\mathrm{reg}})\big(1 - \delta_T^{1/d}\big). \tag{78}$$

The certificate is strict when any event is strict. It is dimension-diluted and does not identify the most affected direction.

This is a volumetric statement under same-direction transversality, not a claim that every direction contracts. Conversely, $\beta < 0$ gives $r > 1$: negative singular interface curvature amplifies the direct normal mode and is incompatible with global convexity.

ReLU event law. For one switching preactivation $s(\theta)$, let $q = \nabla_\theta s$, write $F_{\mathrm{act}} - F_{\mathrm{inact}} = a\, s + o(\|\theta - z\|)$, and let g_y be the loss gradient at the common boundary output. Corollary 32 shows that, with empirical averaging factor γ, $\chi = \langle g_y, a \rangle$, and $n = q$ for activation,

$$[\nabla_\theta \mathcal{L}] = \gamma \chi n, \qquad \lambda_\perp(\Xi) = 1 - \frac{\gamma \chi \|q\|^2}{q^\top f^-}$$

for inactive-to-active crossing. Thus the projected residual determines attenuation, invisibility, or amplification. This is a single-simple-event statement, not a simultaneous-switch or genericity result.

A minimal strongly convex residual-ReLU risk. For one constant input, consider the empirical squared loss

$$F_\theta(1) = (\theta, \mathrm{ReLU}(\theta)), \qquad \mathcal{L}_{m,c}(\theta) = \tfrac{1}{2}(\theta - m)^2 + \tfrac{1}{2}(\mathrm{ReLU}(\theta) - c)^2. \tag{79}$$

Theorem 28 (Residual-ReLU risks with prescribed sensitivity ratios)**.** *Fix $m > 0, c > -m, T > 0$. Every $\theta_0 \in U_T = (m(1 - e^T), 0)$ crosses zero exactly once, at $t_* = \log((m - \theta_0)/m)$, with*

$$(f^-, f^+, H^-, H^+, \beta, r) = \left(m, m + c, 1, 2, -c, \frac{m + c}{m}\right).$$

Along every nonresonant canonical-mesh sequence,

$$D\Theta_{\eta,T} = J_{\eta,T}^{\mathrm{AD}} \to J_{\mathrm{reg}} = e^{-t_*} e^{-2(T - t_*)}, \qquad D\varphi_T = r J_{\mathrm{reg}}. \tag{80}$$

For $c \leq 0$, the objective is globally 1-strongly convex; for $c > 0$, it is globally nonconvex. Given $R > 1$, $a > 0$, and $t_* \in (0, T)$, the choice

$$m = Ra, \qquad c = -(R - 1)a, \qquad \theta_0 = m(1 - e^{t_*})$$

is globally 1-strongly convex, has minimum one-sided speed a, and satisfies

$$D\varphi_T = R^{-1} J_{\mathrm{reg}}, \qquad |J_{\mathrm{reg}}|/|D\varphi_T| = R. \tag{81}$$

The ratio holds on the open set U_T for each fixed instance; it is not a uniform joint R, η limit. It also does not arise from unstable regional curvature: the regional Hessians are 1 and 2. In this scalar result $\|\Xi\|_2 = |r|$; after embedding the same normal transfer in $d \geq 2$, $\|\Xi\|_2 = \max\{1, |r|\}$.

Lemma 29 (Ratio–conditioning tradeoff)**.** *At a same-direction event, let $r = (\nu^\top f^+)/(\nu^\top f^-)$. If $\nu^\top f^+ \geq c_0 > 0$ and $\|f^-\| \leq M$, then*

$$\frac{1}{r} \leq \frac{M}{c_0}. \tag{82}$$

The bounded-data family

$$\mathcal{L}_R(\theta) = \tfrac{1}{2}(\theta - 1)^2 + \tfrac{1}{2}\big(\mathrm{ReLU}(\theta) + 1 - R^{-1}\big)^2 \tag{83}$$

is globally 1-strongly convex. For any $\theta_0 \in (1 - e^T, 0)$, its trajectory crosses zero before T and realizes $D\varphi_T = R^{-1} J_{\mathrm{reg}}$, but its outgoing speed is R^{-1}. The uniform-margin construction in Theorem 28 instead scales the incoming field, targets, and initialization as $\Theta(R)$. Hence an unbounded reciprocal ratio cannot coexist with both a uniform outgoing transverse margin and a uniformly bounded incoming field scale.

A qualitative outer decision.

Corollary 30 (Strongly convex initialization-gradient ranking reversal)**.** *For positive event-free sensitivities a_1, a_2 and transfer ratios r_1, r_2, the strict ranking $a_1 > a_2$ reverses after event transfer exactly when*

$$1 < \frac{a_1}{a_2} < \frac{r_2}{r_1}. \tag{84}$$

This open set is nonempty exactly when $r_2 > r_1$. A globally 1-strongly convex realization at $T = 1$ is

$$\mathcal{L}(\theta) = \sum_{i=1}^{2} \left[\tfrac{1}{2}(\theta_i - m_i)^2 + \tfrac{1}{2}(\mathrm{ReLU}(\theta_i) - c_i)^2\right], \quad \begin{matrix}(m_1, c_1) = (4, -3),\\ (m_2, c_2) = (4/3, -1/3),\end{matrix} \tag{85}$$

initialized at

$$\theta_0 = \left(4(1 - e^{0.6}), \tfrac{4}{3}(1 - e^{0.4})\right). \tag{86}$$

Its separated events have minimum speed one, and for $q(z) = z_1 + z_2$,

$$g_{\mathrm{reg}} = (e^{-1.4}, e^{-1.6}), \qquad g_{\mathrm{flow}} = (\tfrac{1}{4}e^{-1.4}, \tfrac{3}{4}e^{-1.6}). \tag{87}$$

Thus the largest-magnitude coordinate is 1 for g_{reg} and 2 for g_{flow}. Exact discrete AD has the first ranking on all sufficiently fine nonresonant meshes. The itinerary and strict reversal persist on an open neighborhood relative to the displayed separable parameter family.

Corollary 31 (Coupled strongly convex ranking reversal)**.** *Let $\mathcal{L}_0$ be the preceding loss and set*

$$\mathcal{L}_{\kappa,\varepsilon}(\theta) = \mathcal{L}_0(\theta) + \frac{\kappa}{2}\|\theta\|^2 - \varepsilon\theta_1\theta_2. \tag{88}$$

There are arbitrarily small $0 < |\varepsilon_0| < \kappa_0$ for which this finite-sample affine/residual-ReLU squared loss is genuinely coupled, globally more than 1-strongly convex, and retains the separated itinerary and opposite rankings. These properties persist on an open neighborhood of that nonzero coupling point.

A local Taylor analysis relates directional disagreement to matched-radius regret: Proposition 40 and its proof are in Appendix H.

The constructions prove low-dimensional magnitude and directional effects on explicit open sets, while Proposition 40 states only the corresponding local decision consequence. It guarantees an ordering within a sufficiently small radius controlled by its smoothness and curvature assumptions, but no ordering at an arbitrary prescribed radius outside that regime and no multi-step performance ordering. It makes no typicality claim. The proofs of the convexity and ranking statements follow below; the scalar squared-loss realization is established in Appendix F.1.

G.4.1 Convexity and the sign of an event

Proof of Theorem 26. Let z be an event point and let ν point from the minus to the plus region. For sufficiently small $\varepsilon > 0$, the points $x_\varepsilon = z - \varepsilon\nu$ and $y_\varepsilon = z + \varepsilon\nu$ lie in the adjacent smooth regions. Monotonicity of the gradient of a differentiable convex function gives

$$\big(\nabla\mathcal{L}(y_\varepsilon) - \nabla\mathcal{L}(x_\varepsilon)\big)^\top (y_\varepsilon - x_\varepsilon) \geq 0.$$

Divide by 2ε and pass to the one-sided traces. Since $[\nabla\mathcal{L}] = \beta\nu$, the limit is $\beta \geq 0$.

Gradient structure gives

$$\nu^\top f^+ = \nu^\top f^- - \beta.$$

Same-direction transversality makes both sides positive. Therefore

$$r = \frac{\nu^\top f^+}{\nu^\top f^-} = 1 - \frac{\beta}{\nu^\top f^-} \in (0, 1],$$

with $r < 1$ exactly when $\beta > 0$. This attenuates the direct normal mode; tangential modes remain eigenvectors with eigenvalue one.

Every regional fundamental matrix is invertible and has positive determinant by Liouville's formula. The matrix determinant lemma gives $\det \Xi_j = r_j$. Taking determinants in the saltation-interleaved product and in J_{reg} yields

$$\frac{\det D\varphi_T}{\det J_{\mathrm{reg}}} = \prod_{j=1}^{m} r_j, \qquad \log|\det D\varphi_T| - \log|\det J_{\mathrm{reg}}| = \sum_{j=1}^{m} \log r_j. \tag{89}$$

If any convex event is strict, one $r_j < 1$ while all other factors are at most one and positive. The product is then less than one, so $D\varphi_T \neq J_{\mathrm{reg}}$; exact transported cancellation is impossible.

If $\beta < 0$, the same speed identity gives $r > 1$ under a same-direction crossing. The trace monotonicity argument above also shows that such a downward normal-gradient jump is incompatible with global convexity. This is negative singular interface curvature. The converse statement is deliberately local: $\beta > 0$ fixes the sign of the atomic part but does not, without convex regional Hessians, prove global convexity. □

Proof of Corollary 27. Set $\mathcal{C}_T = J_{\mathrm{reg}}^{-1} D\varphi_T$ and $\delta_T = \det \mathcal{C}_T = \prod_j r_j$. All regional fundamental matrices are invertible, and every Ξ_j is invertible because $\det \Xi_j = r_j > 0$; hence J_{reg}, $D\varphi_T$, and $\mathcal{C}_T$ are invertible. If $s_{\min} = \sigma_{\min}(\mathcal{C}_T)$, then

$$s_{\min}^d \leq \prod_{i=1}^{d} \sigma_i(\mathcal{C}_T) = |\det \mathcal{C}_T| = \delta_T,$$

so $s_{\min} \leq \delta_T^{1/d} \leq 1$. For a unit right singular vector v associated with $s_{\min}$,

$$\|\mathcal{C}_T - I\|_2 \geq \|(\mathcal{C}_T - I)v\|_2 \geq 1 - \|\mathcal{C}_T v\|_2 = 1 - s_{\min} \geq 1 - \delta_T^{1/d}.$$

The multiplication order is fixed by the definition: $D\varphi_T = J_{\mathrm{reg}} \mathcal{C}_T$. Therefore

$$\|D\varphi_T - J_{\mathrm{reg}}\|_2 = \|J_{\mathrm{reg}}(\mathcal{C}_T - I)\|_2 \geq \sigma_{\min}(J_{\mathrm{reg}}) \|\mathcal{C}_T - I\|_2,$$

which proves (78). □

G.4.2 A ReLU event in residual/downstream coordinates

Corollary 32 (Residual/downstream ReLU event law)**.** *Suppose one preactivation $s(\theta)$ switches while all other masks remain fixed. Let $q = \nabla_\theta s \neq 0$, write $F_{\mathrm{act}} - F_{\mathrm{inact}} = a\, s + o(\|\theta - z\|)$, let g_y be the loss gradient at the common boundary output, and set $\chi = \langle g_y, a \rangle$. With empirical averaging factor γ and oriented normal $n = \sigma q$,*

$$[\nabla_\theta \mathcal{L}] = \gamma \chi n, \qquad \beta = \gamma \chi \|q\|.$$

For an inactive-to-active crossing,

$$\lambda_\perp(\Xi) = 1 - \frac{\gamma \chi \|q\|^2}{q^\top f^-}.$$

Under same-direction transversality, the sign of χ gives attenuation, invisibility, or amplification. For squared loss, χ is the residual projected onto the switching unit's downstream output direction.

Proof of Corollary 32. Let $s(\theta)$ be the changing preactivation and $q = \nabla_\theta s(z) \neq 0$. With every other mask fixed, write the active and inactive network outputs near z as

$$F_{\mathrm{act}}(\theta) - F_{\mathrm{inact}}(\theta) = a\, s(\theta) + o(\|\theta - z\|).$$

The two outputs agree at z, while differentiation gives

$$DF_{\mathrm{act}}(z) - DF_{\mathrm{inact}}(z) = aq^\top. \tag{90}$$

Let g_y be the loss gradient at the common output and let γ be the empirical averaging factor. The parameter-gradient jump from inactive to active is

$$\gamma\big(DF_{\mathrm{act}} - DF_{\mathrm{inact}}\big)^\top g_y = \gamma\langle g_y, a\rangle q.$$

Set $\chi = \langle g_y, a\rangle$. For activation, the oriented normal is $n = q$. For deactivation, incoming and outgoing sides are exchanged and the oriented normal is $n = -q$; both the gradient jump and n change sign. Thus, in either orientation,

$$[\nabla_\theta \mathcal{L}] = \gamma\chi n, \qquad \beta = \gamma\chi\|q\|.$$

For an inactive-to-active event, $n = q$, and substitution into (13) yields

$$\lambda_\perp(\Xi) = 1 - \frac{\gamma\chi\|q\|^2}{q^\top f^-}.$$

Under same-direction transversality, the signs of χ and β therefore give attenuation, invisibility, and amplification as stated. For squared loss, g_y is the residual, so χ is its inner product with the downstream output direction a. □

Proof of Corollary 30. For positive a_i, r_i, the two desired strict inequalities are $a_1/a_2 > 1$ and $a_1/a_2 < r_2/r_1$, proving (84). This interval is nonempty exactly when $r_2 > r_1$. Strict inequalities define an open subset of $\mathbb{R}^4_{>0}$.

Equation (85) is the half-squared risk of the one-sample, four-output network

$$F_\theta(1) = (\theta_1, \mathrm{ReLU}(\theta_1), \theta_2, \mathrm{ReLU}(\theta_2))$$

with target $(4, -3, 4/3, -1/3)$. Each coordinate loss has $c_i < 0$, so Proposition 15 makes it globally 1-strongly convex. Their sum is globally 1-strongly convex on $\mathbb{R}^2$.

The initialization formula in Proposition 15 gives coordinate event times $s_1 = 0.6$ and $s_2 = 0.4$. The corresponding ratios are

$$r_1 = \frac{m_1 + c_1}{m_1} = \frac{1}{4}, \qquad r_2 = \frac{m_2 + c_2}{m_2} = \frac{3}{4}.$$

Both outgoing speeds equal one; the incoming speeds are 4 and $4/3$. The events are therefore separated by 0.2 and have one-sided speed margin one. Since the system is separable, the event-free and flow initialization Jacobians are diagonal, with entries

$$(J_{\mathrm{reg}})_{ii} = e^{-2T+s_i}, \qquad (D\varphi_T)_{ii} = r_i e^{-2T+s_i}.$$

For $q(z) = z_1 + z_2$, its terminal gradient is $(1, 1)$, which proves (87). The regional ranking follows from $e^{-1.4} > e^{-1.6}$. The flow ranking is reversed because $e^{0.2} < 3$, so $\frac{1}{4}e^{-1.4} < \frac{3}{4}e^{-1.6}$.

At nonresonant meshes, the two scalar executed programs are locally smooth and their exact derivatives converge coordinatewise to the regional values. Because the regional inequality is strict, exact discrete AD has the same ranking for all sufficiently small step sizes. Finally, the event-time and ratio formulas are continuous where $m_i > 0$, $-m_i < c_i < 0$, and the event times lie strictly inside $(0, T)$. The two ranking margins, event separation, and normal speeds are also strict. They therefore persist on an open neighborhood relative to the displayed separable target–initialization parameter family. □

Proof of Lemma 29. Since ν is a unit vector and both normal speeds are positive,

$$\frac{1}{r} = \frac{\nu^\top f^-}{\nu^\top f^+} \le \frac{\|f^-\|}{c_0} \le \frac{M}{c_0}.$$

For (83), the parameters in (79) are $m = 1$ and $c = -1 + R^{-1}$. Proposition 15 gives global 1-strong convexity, incoming speed one, outgoing speed R^{-1}, and $D\varphi_T = R^{-1}J_{\mathrm{reg}}$. The target and initialization remain bounded for fixed t_*, but the transverse margin tends to zero. In the uniform-margin parameterization, $m = Ra$, $|c| = (R-1)a$, and $|\theta_0| = Ra(e^{t_*} - 1)$, proving the asserted linear scale growth. □

Proof of Corollary 31. On a fixed activation region, the smooth quadratic part of $\mathcal{L}_{\kappa,\varepsilon}$ has Hessian

$$Q_{\kappa,\varepsilon} = \begin{bmatrix} 1+\kappa & -\varepsilon \\ -\varepsilon & 1+\kappa \end{bmatrix}, \qquad \lambda_{\min}(Q_{\kappa,\varepsilon}) = 1 + \kappa - |\varepsilon|.$$

The two residual-ReLU terms are convex because $c_i < 0$. Hence $\mathcal{L}_{\kappa,\varepsilon}$ is globally $(1+\kappa-|\varepsilon|)$-strongly convex and is more than 1-strongly convex whenever $\kappa > |\varepsilon|$.

At $(\kappa, \varepsilon) = (0, 0)$, the event times, normal speeds, event separation, and both ranking margins computed above are strict. Regional ODE solutions depend continuously on (κ, ε). In each isolating tube, the event equation has nonzero time derivative, so the implicit-function theorem gives a unique nearby event time and state. Induction over the two separated events shows that the itinerary, regional fundamental matrices, saltation matrices, and both terminal gradients vary continuously. All strict inequalities therefore persist near the origin. Choose a sufficiently small point with $0 < |\varepsilon_0| < \kappa_0$, and then a ball around it small enough that $\varepsilon \neq 0$, $\kappa > |\varepsilon|$, and all event and ranking margins persist. This is an open set in the explicit coupled parameter family.

Finally, let $A^\top A = Q_{\kappa,\varepsilon}$ and $b = A^{-\top} m$, where $m = (m_1, m_2)$. Up to an additive constant,

$$\mathcal{L}_{\kappa,\varepsilon}(\theta) = \tfrac{1}{2}\|A\theta - b\|^2 + \tfrac{1}{2}\sum_{i=1}^{2}(\mathrm{ReLU}(\theta_i) - c_i)^2.$$

Thus the coupled objective is a finite-dimensional affine/residual-ReLU squared loss. This completes the proof. □

G.5 Multi-event cancellation criterion

Several nontrivial saltation matrices do not by themselves imply a matrix mismatch: transported event effects may cancel.

Proposition 33 (Transported saltation criterion)**.** *For $j = 1, \ldots, m$, define*

$$Q_j = \Phi_{j-1}\Phi_{j-2}\cdots\Phi_0, \qquad \widetilde{\Xi}_j = Q_j^{-1}\Xi_j Q_j. \tag{91}$$

Then

$$J_{\mathrm{reg}}^{-1} D\varphi_T = \widetilde{\Xi}_m \widetilde{\Xi}_{m-1}\cdots\widetilde{\Xi}_1. \tag{92}$$

Consequently,

$$D\varphi_T = J_{\mathrm{reg}} \quad \Longleftrightarrow \quad \widetilde{\Xi}_m \cdots \widetilde{\Xi}_1 = I. \tag{93}$$

A sufficient, but not necessary, condition for mismatch is

$$\prod_{j=1}^{m} \frac{\nu_j^\top f_j^+}{\nu_j^\top f_j^-} \neq 1. \tag{94}$$

Equality of the scalar product to one only makes the determinant of the transported matrix product equal to one; it does not imply that the matrix product is I. Regional propagation can rotate and couple the affected directions before subsequent events. For any submultiplicative matrix norm normalized by $\|I\| = 1$, set

$$S_j = \Xi_j - I, \qquad \widetilde{S}_j = Q_j^{-1} S_j Q_j, \qquad E_T = \sum_{j=1}^{m} \|\widetilde{S}_j\|. \tag{95}$$

Then the exact log-volume correction is

$$\log|\det D\varphi_T| - \log|\det J_{\mathrm{reg}}| = \sum_{j=1}^{m} \log r_j, \qquad r_j = \frac{\nu_j^\top f_j^+}{\nu_j^\top f_j^-} > 0, \tag{96}$$

and

$$\|J_{\mathrm{reg}}^{-1} D\varphi_T - I\| \leq \prod_{j=1}^{m}(1 + \|\widetilde{S}_j\|) - 1 \leq e^{E_T} - 1, \tag{97}$$

$$J_{\mathrm{reg}}^{-1} D\varphi_T = I + \sum_{j=1}^{m} \widetilde{S}_j + R_2, \tag{98}$$

$$\|R_2\| \leq e^{E_T} - 1 - E_T \leq \tfrac{1}{2} e^{E_T} E_T^2. \tag{99}$$

The first-order sum contains the transported event contributions; R_2 contains their ordered interactions. These are upper bounds only. No lower bound follows from E_T, because transported factors may cancel. For one event,

$$D\varphi_T - J_{\mathrm{reg}} = \Phi_1(\Xi_1 - I)\Phi_0, \tag{100}$$

so equality holds if and only if $\Xi_1 = I$. In the continuous-loss gradient setting this is equivalent to equality of the one-sided gradients. A single genuine gradient jump therefore forces a matrix mismatch, although perturbation directions satisfying $\nu_1^\top \Phi_0 v = 0$ are unaffected by that event.

Proof. Every regional fundamental matrix is invertible. Expanding $J_{\mathrm{reg}}^{-1} D\varphi_T$ and cancelling regional factors from the outside inward gives

$$J_{\mathrm{reg}}^{-1} D\varphi_T = (Q_m^{-1}\Xi_m Q_m)\cdots(Q_1^{-1}\Xi_1 Q_1),$$

which proves the factorization and exact cancellation criterion. Moreover,

$$\det \Xi_j = 1 + \frac{\nu_j^\top (f_j^+ - f_j^-)}{\nu_j^\top f_j^-} = \frac{\nu_j^\top f_j^+}{\nu_j^\top f_j^-},$$

so the determinant condition follows. Under (A4), every ratio is positive; taking logarithms proves (96).

Expanding the ordered product in (92) gives

$$(I + \widetilde{S}_m)\cdots(I + \widetilde{S}_1) = I + \sum_j \widetilde{S}_j + \sum_{k=2}^{m} \sum_{j_k > \cdots > j_1} \widetilde{S}_{j_k}\cdots\widetilde{S}_{j_1}.$$

Submultiplicativity therefore yields

$$\left\|(I + \widetilde{S}_m)\cdots(I + \widetilde{S}_1) - I\right\| \le \prod_j (1 + \|\widetilde{S}_j\|) - 1,$$

$$\|R_2\| \le \prod_j (1 + \|\widetilde{S}_j\|) - 1 - \sum_j \|\widetilde{S}_j\|.$$

Using $1 + x \le e^x$ proves (97) and the first inequality in (99). The final inequality follows from $e^x - 1 - x \le \frac{1}{2} e^x x^2$ for $x \ge 0$.

For one event, invertibility of Φ_0 and Φ_1 makes (100) zero exactly when $\Xi_1 - I = 0$. Since the denominator is nonzero, the latter is equivalent to $f_1^+ = f_1^-$, or $\beta_1 = 0$. □

G.6 Arbitrary-gap construction

Corollary 34 (Arbitrarily large discrete-to-flow derivative gap)**.** *Let $0 < a_- < a_+$ and, for $\theta = (x, y) \in \mathbb{R} \times \mathbb{R}^{d-1}$, define*

$$\mathcal{L}_{a_-,a_+}(x, y) = \frac{1}{2}\|y\|^2 + \begin{cases} \frac{1}{2}(x - a_-)^2, & x \le 0, \\ \frac{1}{2}(x - a_+)^2 + \frac{1}{2}(a_-^2 - a_+^2), & x \ge 0. \end{cases} \tag{101}$$

For every $T > 0$, the open strip

$$U_T = \left\{(x_0, y_0) : a_-(1 - e^T) < x_0 < 0\right\} \tag{102}$$

has a stable one-event itinerary before time T. On U_T, both regional Hessians equal I_d, and

$$J_{\mathrm{reg}} = e^{-T} I_d, \tag{103}$$

$$D\varphi_T = e^{-T}\,\mathrm{diag}\left(\frac{a_+}{a_-}, I_{d-1}\right). \tag{104}$$

Consequently,

$$\|D\varphi_T - J_{\mathrm{reg}}\|_2 = e^{-T}\left(\frac{a_+}{a_-} - 1\right), \tag{105}$$

$$\frac{\|D\varphi_T\|_2}{\|J_{\mathrm{reg}}\|_2} = \frac{a_+}{a_-}. \tag{106}$$

Thus any multiplicative gap $R > 1$ is obtained by taking $a_+/a_- = R$, and any additive gap $G > 0$ by taking $a_+/a_- = 1 + Ge^T$. If $a_+/a_- > e^T$, the true flow is expansive although ordinary branchwise AD converges to the strict contraction $e^{-T} I_d$.

For $\eta = T/N$ and any step size with no exact interface landing,

$$D\Theta_{\eta,T} = J^{\mathrm{AD}}_{\eta,T} = (1-\eta)^N I_d \longrightarrow e^{-T} I_d. \tag{107}$$

Inserting the exact event matrix once gives

$$J^{\mathrm{SC}}_{\eta,T} = (1-\eta)^N \operatorname{diag}\left(\frac{a_+}{a_-}, I_{d-1}\right) = D\varphi_T + O(\eta). \tag{108}$$

Here the inserted saltation matrix is exact. Since $T = N\eta$, $(1-\eta)^N = e^{-T} + O(\eta)$, so the entire $O(\eta)$ difference in (108) is the ordinary Euler error of the regional dynamics. There is no event-localization or one-sided-data error in this construction; those additional errors are represented by ε_η and δ_η in Theorem 38. The loss is regionwise strongly convex but not globally convex.

Proof. The additive constant makes the two traces agree at $x = 0$. On the left, $\dot{x} = a_- - x$, so a point with $x_0 < 0$ reaches the interface at

$$t_*(x_0) = \log \frac{a_- - x_0}{a_-}. \tag{109}$$

The strict strip inequalities are exactly $0 < t_*(x_0) < T$. Both one-sided speeds at impact are positive, namely a_- and a_+, so the crossing is stable and same-direction transverse. The regional fundamental matrices are scalar contractions whose total duration is T, while

$$\Xi = \operatorname{diag}\left(\frac{a_+}{a_-}, I_{d-1}\right).$$

This proves the regional and true derivatives and both gap formulas. Since both regional Hessians are I_d, every executed Euler Jacobian is $(1-\eta)I_d$, irrespective of the branch count. The regional factors commute with Ξ, proving the corrected formula. Finally, the normal derivative of $\mathcal{L}$ jumps downward from $-a_-$ to $-a_+$, so the continuous loss has negative singular interface curvature and is not globally convex. □

H Parameters, adjoints, and terminal-objective consequences

H.1 Hyperparameter sensitivity and reverse mode

This section proves Theorem 36 and Corollary 37. The formulas are classical hybrid sensitivity identities; their role here is to compare the derivative of a hard-ReLU training flow with the limit of exact derivatives of its nonresonant discrete programs.

H.1.1 Parameter-dependent assumptions

Let $\lambda \in \mathbb{R}^p$. In addition to (A1)–(A5), suppose that each regional field $f_\alpha(\theta, \lambda)$ is C^1 jointly in state and parameter on the relevant compact tube. Write

$$A_\alpha = D_\theta f_\alpha \in \mathbb{R}^{d\times d}, \qquad B_\alpha = D_\lambda f_\alpha \in \mathbb{R}^{d\times p}.$$

The matrices A_α and B_α are uniformly continuous on that tube. Each event function $g_j(\theta, \lambda)$ is C^2, and the initialization is $h(\lambda)$ with $h \in C^1$. The event count, order, separation, and same-direction transversality persist on a parameter neighborhood of the reference value. The mesh, horizon, state dimension, and program depth are held fixed while differentiating with respect to λ.

Let

$$Z(t) = D_\lambda \theta(t;\lambda) \in \mathbb{R}^{d\times p}, \quad n_j = \nabla_\theta g_j \in \mathbb{R}^d, \quad g_{\lambda,j} = \nabla_\lambda g_j \in \mathbb{R}^p.$$

Lemma 35 (Joint branch constancy for a fixed program)**.** *Fix a base parameter λ_0 and a finite Euler/GD program. If every activation function evaluated by the executed program is nonzero at λ_0, then all executed branch signs are constant on a neighborhood of λ_0. On that neighborhood the endpoint is C^1, and its classical parameter derivative is the branchwise AD recursion (113).*

Proof. The initialization and all regional update maps are continuous in λ. A nonzero finite list of activation values retains its signs on a sufficiently small neighborhood. Induction over the finite program depth therefore fixes the branch word. The endpoint is then a composition of jointly C^1 regional maps, and the chain rule gives the stated recursion. □

Theorem 36 (Parameter derivatives and the vanishing-step limit)**.** *Fix a base parameter λ_0, and assume the conditions above, including persistence of the stable separated itinerary on a neighborhood of λ_0. The horizon, mesh, and program depth are held fixed while taking D_λ. Regionally,*

$$\dot{Z} = A_\alpha Z + B_\alpha, \qquad Z(0) = Dh(\lambda).$$

At event j,

$$Z_j^+ = \Xi_j Z_j^- + C_j, \qquad C_j = \frac{(f_j^+ - f_j^-)g_{\lambda,j}^\top}{n_j^\top f_j^-}.$$

At each mesh that is nonresonant at λ_0, branchwise AD is the classical parameter derivative. Along any such vanishing-mesh sequence it converges to the regional inhomogeneous sensitivity continued without event jumps, whereas the limiting flow derivative obeys the displayed affine event transfer.

H.1.2 Forward jump and the discrete learning limit

Proof of Theorem 36. Inside a smooth region, ordinary differentiation gives

$$\dot{Z} = A_\alpha Z + B_\alpha, \qquad Z(0) = Dh(\lambda). \tag{110}$$

At event j, perturb the parameter by $\delta\lambda$ and denote the resulting event-time change by δt_j. Differentiating $g_j(\theta(t_j + \delta t_j; \lambda + \delta\lambda), \lambda + \delta\lambda) = 0$ on the incoming side gives

$$\delta t_j = -\frac{(n_j^\top Z_j^- + g_{\lambda,j}^\top)\delta\lambda}{n_j^\top f_j^-}.$$

Returning the outgoing perturbation to the nominal clock yields

$$\begin{aligned} Z_j^+\delta\lambda &= Z_j^-\delta\lambda + (f_j^- - f_j^+)\delta t_j \\ &= \left(I + \frac{(f_j^+ - f_j^-)n_j^\top}{n_j^\top f_j^-}\right) Z_j^-\delta\lambda + \frac{(f_j^+ - f_j^-)g_{\lambda,j}^\top}{n_j^\top f_j^-}\delta\lambda. \end{aligned}$$

Since this holds for every parameter perturbation,

$$Z_j^+ = \Xi_j Z_j^- + C_j, \qquad C_j = \frac{(f_j^+ - f_j^-)g_{\lambda,j}^\top}{n_j^\top f_j^-}. \tag{111}$$

The dimensions are $Z_j^\pm, C_j \in \mathbb{R}^{d\times p}$.

Equivalently, augment the state by the constant parameter:

$$y = (\theta, \lambda), \qquad \dot{y} = (f_\alpha(\theta, \lambda), 0), \qquad G_j(y) = g_j(\theta, \lambda).$$

The ordinary saltation matrix of the augmented system is

$$\mathcal{S}_j = \begin{bmatrix} \Xi_j & C_j \\ 0 & I_p \end{bmatrix}, \tag{112}$$

which gives a second derivation of both the sign and the block dimensions.

At a fixed Euler/GD program that is nonresonant at λ_0, with T, η, N held fixed while differentiating, Lemma 35 gives the exact derivative recursion

$$Z_{k+1}^\eta = \big(I + \eta A_{\alpha_k}(\theta_k^\eta, \lambda)\big) Z_k^\eta + \eta B_{\alpha_k}(\theta_k^\eta, \lambda), \qquad Z_0^\eta = Dh(\lambda). \tag{113}$$

Introduce the augmented regional coefficient and sensitivity

$$\mathcal{A}(t) = \begin{bmatrix} A(t) & B(t) \\ 0 & 0 \end{bmatrix}, \qquad W(t) = \begin{bmatrix} Z(t) \\ I_p \end{bmatrix}.$$

The discrete factor corresponding to (113) is

$$I + \eta \mathcal{A}_k = \begin{bmatrix} I + \eta A_k & \eta B_k \\ 0 & I_p \end{bmatrix}.$$

Outside event windows, the state estimate and uniform continuity imply that both A_k and B_k approach their reference regional coefficients. Inside event windows their norms are bounded and the total window length is $O(\eta)$. Consequently,

$$\|\mathcal{A}^\eta - \mathcal{A}\|_{L^1(0,T)} \le C\big[\eta + \omega_A(C\eta) + \omega_B(C\eta)\big].$$

The same fundamental-matrix comparison used in Theorem 6 gives

$$\|Z^{\mathrm{AD}}_{\eta,T} - Z_{\mathrm{reg}}(T)\| \le C\big[\eta + \omega_A(C\eta) + \omega_B(C\eta)\big], \tag{114}$$

where Z_{reg} solves (110) and is continued without a jump at each nominal event time. If A_α, B_α are locally Lipschitz, the right-hand side is $O(\eta)$.

The true flow derivative instead obeys (111). For one event, subtracting the event-free continuation and propagating to T gives the useful identity

$$Z_{\mathrm{flow}}(T) - Z_{\mathrm{reg}}(T) = \Phi_1 \frac{(f^+ - f^-)(n^\top Z^- + g_\lambda^\top)}{n^\top f^-}. \tag{115}$$

If $f^+ \neq f^-$, a parameter direction has zero event-induced gap exactly when its first-order event-time change vanishes. If $f^+ = f^-$, the gap vanishes identically. □

The formula separates three cases. If a parameter enters only the initialization, then $B = 0$, $g_\lambda = 0$, and events act through ΞZ^-. If it enters the vector field but not the interface, then regional forcing B is accumulated and subsequently transferred by Ξ, while $C = 0$. If it enters the interface explicitly, the affine term C is present. These effects coexist when a parameter enters all three objects.

H.1.3 Reverse mode

Corollary 37 (Event-aware adjoint)**.** *For a terminal scalar objective $q(\theta(T), \lambda)$, let $p(T) = \nabla_\theta q$ and solve $-\dot{p} = A_\alpha^\top p$ regionally. Then*

$$p_j^- = \Xi_j^\top p_j^+,$$

and

$$\nabla_\lambda q = q_\lambda + Dh^\top p(0) + \int_0^T B^\top p \, \mathrm{d}t + \sum_{j=1}^m g_{\lambda,j} \frac{(f_j^+ - f_j^-)^\top p_j^+}{n_j^\top f_j^-}.$$

Omitting these event terms gives the adjoint of the regional sensitivity limit, not generally the derivative of the limiting flow.

Proof of Corollary 37. For a terminal scalar objective $q(\theta(T), \lambda)$, set

$$p(T) = \nabla_\theta q(\theta(T), \lambda), \qquad -\dot{p} = A_\alpha^\top p$$

on each regional segment. The fixed-clock forward event map in (111) pulls covectors back by

$$p_j^- = \Xi_j^\top p_j^+. \tag{116}$$

Indeed,

$$\begin{aligned}(p_j^+)^\top Z_j^+ &= (p_j^+)^\top (\Xi_j Z_j^- + C_j) \\ &= (\Xi_j^\top p_j^+)^\top Z_j^- + (p_j^+)^\top C_j.\end{aligned}$$

This proves that the transpose, rather than the inverse transpose, is the correct jump for the fixed-clock derivative map.

Between events,

$$\frac{\mathrm{d}}{\mathrm{d}t}(Z^\top p) = B^\top p.$$

At an event, (116) gives

$$(Z_j^+)^\top p_j^+ - (Z_j^-)^\top p_j^- = C_j^\top p_j^+.$$

Integrating segment by segment and adding the jumps yields

$$\nabla_\lambda q = q_\lambda + Dh(\lambda)^\top p(0) + \int_0^T B(t)^\top p(t)\,\mathrm{d}t + \sum_{j=1}^m C_j^\top p_j^+, \tag{117}$$

where

$$C_j^\top p_j^+ = g_{\lambda,j}\frac{(f_j^+ - f_j^-)^\top p_j^+}{n_j^\top f_j^-}. \tag{118}$$

If $g_{\lambda,j} = 0$, the direct event accumulator vanishes, but the event can still change the gradient through the jump $p_j^- = \Xi_j^\top p_j^+$. Omitting all event terms gives the regional continuous adjoint and hence the transpose of the event-free sensitivity limit. Reverse-mode AD through every fixed nonresonant discrete program remains its exact classical derivative. □

H.2 Event-aware first-variation consistency

Let $\widehat{t}_j^\eta$ and $\widehat{z}_j^\eta$ be a posteriori event times and states with the correct event count and order. Suppose

$$\max_j \left(|\widehat{t}_j^\eta - t_j| + \|\widehat{z}_j^\eta - z_j\|\right) \le \delta_\eta. \tag{119}$$

Evaluate one-sided data at the localized event and define

$$\widehat{\Xi}_j^\eta = I + \frac{(\widehat{f}_j^+ - \widehat{f}_j^-)(\widehat{n}_j)^\top}{\widehat{n}_j^\top \widehat{f}_j^-}. \tag{120}$$

Let $\widehat{\Phi}_j^\eta$ be approximations of the regional propagators and assume

$$\max_j \|\widehat{\Xi}_j^\eta - \Xi_j\| \le \varepsilon_\eta, \qquad \max_j \|\widehat{\Phi}_j^\eta - \Phi_j\| \le \gamma_\eta. \tag{121}$$

Define the a posteriori corrected product

$$J_{\eta,T}^{\mathrm{SC}} = \widehat{\Phi}_m^\eta \widehat{\Xi}_m^\eta \widehat{\Phi}_{m-1}^\eta \widehat{\Xi}_{m-1}^\eta \cdots \widehat{\Phi}_1^\eta \widehat{\Xi}_1^\eta \widehat{\Phi}_0^\eta. \tag{122}$$

Theorem 38 (Consistency of an a posteriori event-corrected product)**.** *Under (A1)–(A5), if the supplied event data have the correct stable itinerary and $\delta_\eta + \varepsilon_\eta + \gamma_\eta \to 0$, then*

$$J_{\eta,T}^{\mathrm{SC}} \longrightarrow D\varphi_T(\theta_0). \tag{123}$$

More quantitatively,

$$\|J_{\eta,T}^{\mathrm{SC}} - D\varphi_T(\theta_0)\| \le C\left[\delta_\eta + \varepsilon_\eta + \gamma_\eta\right]. \tag{124}$$

The corrected matrix is not, in general, the derivative of unmodified hard-branch GD. It is the derivative of an event-aware solver only when that solver differentiates the localized event time and split steps.

Proof. All exact and numerical factors are uniformly bounded, and their number is fixed. A telescoping perturbation bound for the finite product, together with (19), proves (124) and the stated conclusions. □

For the intended split hard-GD construction, use the $O(\eta)$-accurate state path from Theorem 6, localize each event with error δ_η, and evaluate the one-sided fields and normals consistently. Transversality then gives $\varepsilon_\eta \le C(\delta_\eta + \text{one-sided-data error})$, while Lemmas 10–11 give

$$\gamma_\eta \le C\big[\eta + \omega_A(C\eta) + \delta_\eta\big].$$

This specialization recovers the earlier convergence rate without treating the theorem as a complete event-detection algorithm.

H.3 Scalar outer gradients and local candidate regret

Corollary 39 (Scalar outer-gradient consequence)**.** *Under Theorem 28, along any nonresonant mesh sequence, let $q \in C^1(\mathbb{R})$ satisfy $q'(\varphi_T(\theta_0)) \neq 0$, let $G_\eta = q \circ \Theta_{\eta,T}$, and let $G = q \circ \varphi_T$. Then*

$$DG_\eta \to q'(\varphi_T) J_{\mathrm{reg}}, \qquad DG = r\, q'(\varphi_T) J_{\mathrm{reg}}. \tag{125}$$

Proof of Corollary 39. At every nonresonant step size, the ordinary chain rule gives

$$DG_\eta(\theta_0) = q'(\Theta_{\eta,T}(\theta_0)) D\Theta_{\eta,T}(\theta_0).$$

Proposition 15, continuity of q', and state convergence therefore yield $DG_\eta(\theta_0) \to q'(\varphi_T(\theta_0)) J_{\mathrm{reg}}$. The stable-itinerary flow map is classically differentiable, so

$$DG(\theta_0) = q'(\varphi_T(\theta_0)) D\varphi_T(\theta_0) = r q'(\varphi_T(\theta_0)) J_{\mathrm{reg}}.$$

Both limits are nonzero because $q'(\varphi_T(\theta_0)) \neq 0$ and $J_{\mathrm{reg}} > 0$. This proves the stated factor-r relation.

At resonance, an evaluated iterate equals zero. If another step remains, the incoming and outgoing updates from that state differ by ηc, which is propagated through the remaining outgoing recurrence. The endpoint map therefore need not have a classical derivative; software AD returns a selected branch product. The state estimate still holds under the admissible positive-speed selection. This is why every derivative statement above, including Theorem 28 and Corollary 39, follows a nonresonant sequence.

Proposition 40 (Matched-radius local regret)**.** *Fix a nonresonant mesh and let the actual finite-program outer objective Q_η be C^2 on an open neighborhood of $\overline{B}(\lambda, \epsilon_0)$. Write $g_\eta = \nabla Q_\eta(\lambda) \neq 0$, let $h \neq 0$ be a surrogate gradient, and set $u = g_\eta/\|g_\eta\|_2$, $v = h/\|h\|_2$. If $\sup_{x \in \overline{B}(\lambda,\epsilon_0)} \|\nabla^2 Q_\eta(x)\|_2 \leq M$, then, for $0 \leq \epsilon \leq \epsilon_0$,*

$$\left| Q_\eta(\lambda - \epsilon v) - Q_\eta(\lambda - \epsilon u) - \epsilon \|g_\eta\|_2 (1 - u^\top v) \right| \leq M\epsilon^2. \tag{126}$$

If $u^\top v < 1$, the exact discrete-gradient proposal has strictly lower Q_η for all sufficiently small positive radii (in particular, for $0 < \epsilon < \min\{\epsilon_0, \|g_\eta\|_2(1 - u^\top v)/M\}$ when $M > 0$). If $v = u$, the two matched-radius proposals coincide, regardless of their gradient-norm ratio.

Proof of Proposition 40. For any unit vector w and $0 \leq \epsilon \leq \epsilon_0$, Taylor's formula with integral remainder gives

$$Q_\eta(\lambda - \epsilon w) = Q_\eta(\lambda) - \epsilon g_\eta^\top w + R_w(\epsilon), \qquad R_w(\epsilon) = \epsilon^2 \int_0^1 (1 - t) w^\top \nabla^2 Q_\eta(\lambda - t\epsilon w) w \, \mathrm{d}t.$$

The Hessian bound therefore gives $|R_w(\epsilon)| \leq \frac{1}{2} M\epsilon^2$. Apply this identity with $w = v$ and $w = u$, subtract, and use $g_\eta^\top u = \|g_\eta\|_2$ and $g_\eta^\top v = \|g_\eta\|_2 u^\top v$. The combined remainder satisfies

$$|R_v(\epsilon) - R_u(\epsilon)| \leq |R_v(\epsilon)| + |R_u(\epsilon)| \leq M\epsilon^2,$$

which proves (126). If $u^\top v < 1$, its right-hand first-order coefficient is positive; the displayed threshold makes the lower Taylor bound strict when $M > 0$, while the conclusion is immediate when $M = 0$. If $u = v$, the two proposal points are identical, so their objective values agree exactly. Nonresonance is used only to make the finite program's branch word locally constant and hence g_η its classical outer gradient; the proposition separately assumes the stated C^2 regularity. □

I Scalar smoothing and discretization

I.1 Smoothing and discretization theory

This section gives a self-contained scalar result that separates three effects: truncating a smoothing profile, discretizing its inner dynamics, and taking the sharp-interface limit. It also records the phase qualification needed for the sharp-first limit. Throughout, the update is written with a plus sign because the vector field already includes the negative gradient.

I.1.1 Layer assumptions, rescaling, and entry–exit convention

Let $F\colon \mathbb{R} \to \mathbb{R}$ satisfy the following standing assumptions:

(S1) $F \in C^2(\mathbb{R})$ and there are constants $0 < m \le M < \infty$ such that $m \le F(u) \le M$ for every u;

(S2) the positive traces

$$a_- := \lim_{u \to -\infty} F(u), \qquad a_+ := \lim_{u \to +\infty} F(u)$$

exist;

(S3) $L_F := \|F'\|_\infty < \infty$ and $F'' \in L^1(\mathbb{R})$.

For the sharp-first result we additionally assume

$$|uF'(u)| \longrightarrow 0 \quad \text{as } u \to \pm\infty. \tag{127}$$

Assumption (S1) is the scalar same-direction transversality condition. In particular, neither grazing nor reversal of the continuous trajectory occurs inside the layer. The scalar saltation factor is

$$\Xi := \frac{a_+}{a_-} > 0. \tag{128}$$

For a smoothing width $\tau > 0$, define

$$f_\tau(x) := F(x/\tau), \qquad x_{k+1} = x_k + \eta f_\tau(x_k). \tag{129}$$

The inner variables

$$u_k := x_k/\tau, \qquad \rho := \eta/\tau \tag{130}$$

give the exact map and fixed-itinerary cocycle

$$u_{k+1} = u_k + \rho F(u_k), \qquad \delta u_{k+1} = \big(1 + \rho F'(u_k)\big)\delta u_k. \tag{131}$$

We use the following entry–exit convention for the canonical layer. Given a normalized half-window $R > 0$, set $u_0 = -R$ and

$$N_{\rho,R} := \min\{n \ge 1 : u_n \ge R\}, \qquad J_{\rho,R} := \prod_{k=0}^{N_{\rho,R}-1} \big(1 + \rho F'(u_k)\big). \tag{132}$$

The index is finite because every increment is at least ρm. The product in (132) is the derivative of the $N_{\rho,R}$-fold iterate on a phase cell on which $N_{\rho,R}$ is locally constant. At a phase where the stopping index changes, the stopped map need not be differentiable; no derivative across such a phase boundary is claimed. The physical tube is $[-\tau R, \tau R]$. A localized interface limit requires both $R \to \infty$ and $\tau R \to 0$.

I.1.2 Exact flow transfer and resolved Euler transfer

Lemma 41 (Exact scalar flow transfer)**.** *For the inner flow and its variational equation,*

$$\dot u = F(u), \qquad \dot v = F'(u)v,$$

the transfer from $u = -R$ to $u = R$ is

$$J_R^{\mathrm{flow}} = \frac{F(R)}{F(-R)}. \tag{133}$$

Consequently, $J_R^{\mathrm{flow}} \to \Xi$ as $R \to \infty$.

Proof. Along the trajectory,

$$\frac{d}{dt} \log v = F'(u) = \frac{d}{dt} \log F(u).$$

Integrating between the entry and exit points proves (133). The trace limit follows from (S2). □

Define the logarithmic tail truncation

$$\delta_F(R) := \left|\log \frac{F(R)}{a_+}\right| + \left|\log \frac{F(-R)}{a_-}\right|. \tag{134}$$

Theorem 42 (Resolved discrete layer)**.** *Under* (S1)–(S3)*, suppose that*

$$\rho L_F \le \frac{1}{2}. \tag{135}$$

Then

$$|\log J_{\rho,R} - \log \Xi| \le C_F \rho + \delta_F(R), \tag{136}$$

where one admissible constant is

$$C_F := \frac{\|F''\|_{L^1}}{\min\{1/2, m/M\}} + \frac{ML_F}{m}. \tag{137}$$

In particular, if $\rho_j \to 0$ and $R_j \to \infty$, then $J_{\rho_j,R_j} \to \Xi$.

Proof. Write $N = N_{\rho,R}$ and, for $0 \le k < N$, set

$$A_k := 1 + \rho F'(u_k), \qquad B_k := \frac{F(u_{k+1})}{F(u_k)}.$$

Because $u_{k+1} - u_k = \rho F(u_k)$, Taylor's formula with integral remainder gives

$$\begin{aligned} |B_k - A_k| &\le \frac{1}{F(u_k)} \int_{u_k}^{u_{k+1}} (u_{k+1} - s)|F''(s)|\, ds \\ &\le \rho \int_{u_k}^{u_{k+1}} |F''(s)|\, ds. \end{aligned}$$

Condition (135) gives $A_k \ge 1/2$, whereas (S1) gives $B_k \ge m/M$. With $c := \min\{1/2, m/M\}$, the mean-value theorem for the logarithm yields

$$|\log A_k - \log B_k| \le c^{-1}|A_k - B_k|.$$

The intervals $[u_k, u_{k+1}]$ are disjoint. Summing and using the telescoping product of the B_k therefore gives

$$\left|\log J_{\rho,R} - \log \frac{F(u_N)}{F(-R)}\right| \le \frac{\rho \|F''\|_{L^1}}{c}. \tag{138}$$

The exit overshoot obeys

$$0 \le u_N - R \le \rho M.$$

Since $|(\log F)'| \le L_F/m$,

$$|\log F(u_N) - \log F(R)| \le \frac{ML_F}{m}\rho. \tag{139}$$

Combining (138), (139), and (134) proves the claim. □

Remark 43 (Rate, entry phase, and physical localization)**.** If $E_{\rho,R} := C_F\rho + \delta_F(R)$, then

$$|J_{\rho,R} - \Xi| \le \Xi e^{E_{\rho,R}} E_{\rho,R}.$$

An entry displacement $|u_0 + R| = O(\rho)$ adds only another $O(\rho)$ endpoint term to the proof. If $\delta_F(R) \le Ce^{-\gamma R}$, the choice $R(\rho) \ge \gamma^{-1}\log(1/\rho)$ gives an $O(\rho)$ bound, provided an embedded interface also satisfies $\tau R(\rho) \to 0$. For fixed R, however, the resolved limit is $F(R)/F(-R)$, not exactly Ξ; a fixed-window error must not be reported as Euler discretization error.

I.1.3 THE TWO ORDERS OF LIMITS

Iterated limits are meaningful only after fixing a common state map. Fix $x_0 < 0$ and $T > -x_0/a_-$, take $\eta = T/n$, and define

$$x_{k+1}^{\eta,\tau} = x_k^{\eta,\tau} + \eta F(x_k^{\eta,\tau}/\tau), \qquad x_0^{\eta,\tau} = x_0, \tag{140}$$

$$J_{\eta,\tau} = \prod_{k=0}^{n-1} \left(1 + \frac{\eta}{\tau} F'(x_k^{\eta,\tau}/\tau)\right). \tag{141}$$

The horizon assumption places the limiting sharp trajectory strictly on the positive side at time T.

Theorem 44 (Smooth-first iterated limit)**.** *Under* (S1)–(S3)*,*

$$\lim_{\tau\downarrow 0} \left[\lim_{n\to\infty} J_{T/n,\tau}\right] = \frac{a_+}{a_-} = \Xi. \tag{142}$$

Proof. For fixed τ, the vector field f_τ is smooth and globally Lipschitz. Standard Euler convergence applied jointly to the state and variational equations gives

$$\lim_{n\to\infty} J_{T/n,\tau} = D\varphi_T^\tau(x_0),$$

where φ_t^τ is the flow of $\dot{x} = F(x/\tau)$. The scalar autonomous identity from Lemma 41 gives

$$D\varphi_T^\tau(x_0) = \frac{F(\varphi_T^\tau(x_0)/\tau)}{F(x_0/\tau)}. \tag{143}$$

For any reached point x, elapsed time is characterized by

$$t = \int_{x_0}^{x} \frac{dz}{F(z/\tau)}.$$

The integrand is bounded by $1/m$ and converges pointwise away from zero to a_-^{-1} on the negative side and a_+^{-1} on the positive side. Dominated convergence, followed by monotonic inversion of the time-of-flight map, shows that $\varphi_T^\tau(x_0)$ converges to the sharp trajectory endpoint

$$a_+\left(T + \frac{x_0}{a_-}\right) > 0.$$

Thus the denominator and numerator in (143) converge to a_- and a_+, respectively. □

For the opposite order, define the sharp Euler orbit

$$y_{k+1} = y_k + \eta f_0(y_k), \qquad f_0(y) = \begin{cases} a_-, & y < 0, \\ a_+, & y > 0, \end{cases} \qquad y_0 = x_0. \tag{144}$$

The value of $f_0(0)$ is irrelevant when the orbit is nonresonant.

Theorem 45 (Nonresonant sharp-first limit)**.** *Assume* (S1)–(S3) *and* (127)*. Fix $\eta = T/n$ and suppose that*

$$y_k \neq 0, \qquad 0 \le k \le n-1. \tag{145}$$

Then

$$\lim_{\tau\downarrow 0} J_{\eta,\tau} = 1. \tag{146}$$

Consequently, for the set $\mathcal{E}_{\mathrm{nr}}$ of step sizes satisfying (145)*,*

$$\lim_{\substack{\eta\downarrow 0 \\ \eta\in\mathcal{E}_{\mathrm{nr}}}} \left[\lim_{\tau\downarrow 0} J_{\eta,\tau}\right] = 1. \tag{147}$$

Before the first crossing, condition (145) *is equivalent to $-x_0/(\eta a_-) \notin \mathbb{N}$.*

Proof. For fixed η, the sharp orbit has finitely many states and

$$d_\eta := \min_{0\le k<n} |y_k| > 0.$$

Induction in (140), using the trace convergence in (S2) at each state away from zero, gives $x_k^{\eta,\tau} \to y_k$ for every k. Hence $|x_k^{\eta,\tau}| \ge d_\eta/2$ for all sufficiently small τ. For each of the finitely many Jacobian factors,

$$\left|\frac{\eta}{\tau}F'(x_k^{\eta,\tau}/\tau)\right| = \frac{\eta}{|x_k^{\eta,\tau}|}\left|\frac{x_k^{\eta,\tau}}{\tau}F'(x_k^{\eta,\tau}/\tau)\right| \longrightarrow 0$$

by (127). Their product therefore converges to one. The outer limit in (147) is then the limit of the constant inner-limit value one over nonresonant step sizes. □

Remark 46 (Why the phase qualification is necessary)**.** For the tanh profile below, $F'(0) \neq 0$ and the tails are exponential. If a sharp iterate satisfies $y_j = 0$, all preceding smooth-versus-sharp state errors are exponentially small in $1/\tau$, so $x_j^{\eta,\tau}/\tau \to 0$. The corresponding factor behaves as

$$1 + \frac{\eta}{\tau}F'(0),$$

whose magnitude diverges. Resonant step sizes can accumulate at zero. Thus an unrestricted statement $\lim_{\eta\to 0}\lim_{\tau\to 0} J_{\eta,\tau} = 1$ is generally false. The proved replacement is the directed nonresonant limit (147).

I.1.4 Finite-ratio transfer depends on phase and profile

At fixed ρ, equation (131), rather than the trace ratio alone, is the limiting inner problem.

Proposition 47 (Phase and profile are genuine data)**.** *At finite ρ, the fixed-itinerary cocycle need not be determined by (a_-, a_+, ρ).*

Proof. Phase dependence follows already from the explicit tanh calculations in Subsection I.1.5: the same profile and ρ give transfers near 46 and 1 for two different incoming phases.

For profile dependence, fix a finite itinerary $u_0, \dots, u_N$ away from a stopping-index boundary. Choose a smooth bump h supported in a small neighborhood of u_0, disjoint from all later sampled states, such that $h(u_0) = 0$ and $h'(u_0) \neq 0$. For sufficiently small ε, the function $\widetilde{F} = F + \varepsilon h$ remains positive and has the same traces. Every sampled state is unchanged because $\widetilde{F}(u_k) = F(u_k)$, but the first cocycle factor changes because $\widetilde{F}'(u_0) \neq F'(u_0)$. Choose the base itinerary so that the product of the remaining factors is nonzero—for example, for the increasing tanh profile below all factors are positive. The total transfer then changes while the traces do not. □

I.1.5 Tanh-profile constructions

Consider

$$F(u) = a_- + (a_+ - a_-)\frac{1+\tanh(u/2)}{2}, \qquad F'(u) = \frac{a_+ - a_-}{4}\operatorname{sech}^2(u/2). \tag{148}$$

This profile satisfies all assumptions above. In the expanding example $a_- = 0.2$ and $a_+ = 2$, so $\Xi = 10$ and $\|F'\|_\infty = 0.45$.

Skip. With $u_0 = -20$, exit threshold 20, and $\rho = 205$,

$$u_1 = -20 + 205F(-20) = 21.00000076056569 > 20.$$

There is one factor and

$$J = 1 + 205F'(-20) = 1.0000007605656835. \tag{149}$$

This is a constructive skip phase, not a universal large-ρ limit.

Tuned hit. Choose

$$\rho_* := \frac{20}{F(-20)} = 99.99999814496176. \tag{150}$$

Then $u_1 = 0$ and $u_2 = \rho_* F(0) = 109.99999795946 > 20$. Thus

$$J = \big(1 + \rho_* F'(-20)\big)\big(1 + \rho_* F'(0)\big) = 46.0000162315841. \tag{151}$$

More generally, taking $u_0 = -R$ and $\rho_R = R/F(-R)$ lands exactly at the maximizer $u = 0$. For the increasing profile the following step exits $+R$ for all sufficiently large R, and

$$J_{\rho_R, R} \sim \rho_R F'(0) \longrightarrow \infty.$$

The phase is retuned with R (and hence with ρ_R); this is not a fixed-phase divergence statement.

Same ratio, different phases. At $\rho = 100$ with exit threshold 20, direct iteration gives

$$u_0 = -20: \quad J = 46.00001706635, \qquad u_0 = -10: \quad J = 1.01634222248. \tag{152}$$

This supplies the phase comparison used in Proposition 47.

Orientation reversal of the discrete cocycle. For the contracting twin $a_- = 2$, $a_+ = 0.2$, take $u_0 = -20$, $R = 20$, and $\rho = 5$. The orbit contains

$$u_1 = -10.00000001855, \qquad u_2 = -0.00040859936, \qquad u_3 = 5.50051074919.$$

At the near-central sample,

$$1 + 5F'(u_2) = -1.24999990609 < 0. \tag{153}$$

The two preceding factors are positive. For $u \geq u_3$, the magnitude of the negative derivative in (148) decreases, and all remaining factors are at least 0.9635. Therefore exactly one sampled factor is negative and

$$J = -1.17941975827. \tag{154}$$

The continuous same-direction saltation factor is $a_+/a_- = 0.1 > 0$; the sign reversal is a finite-step critical-regime effect. The inequality $\rho \max(-F') > 1$ merely makes a negative factor possible. It does not force a given phase to sample that factor, nor does it determine the parity of all negative factors in a longer itinerary.

I.1.6 Effective resolution and scope beyond the scalar flat layer

The local dimensionless stiffness is

$$\rho_{\text{eff}} := \eta \|f_\tau'\|_\infty = \rho \|F'\|_\infty. \tag{155}$$

For (148) with traces 0.2 and 2, $\rho_{\text{eff}} = 0.45\rho$. The condition $\rho_{\text{eff}} \leq 1/2$ is a sufficient condition for Theorem 42; it is not a necessary condition for an accidental finite-ρ value to lie near Ξ. Across a family of profiles, ρ_{eff} also does not replace the tail error, positivity bounds, or $\|F''\|_{L^1}$ appearing in (136). The contrast quantity $\rho|a_+ - a_-|/a_-$ measures net trace gain, not local spike resolution.

The preceding proofs are scalar. They extend without new ideas only when a multidimensional system genuinely reduces to an autonomous normal coordinate, for example at a flat interface with fixed normal, a rank-one normal jump, and separately controlled tangential dynamics. In that setting the scalar normal factor is embedded in the usual matrix

$$I + \frac{(f^+ - f^-)\nu^\top}{\nu^\top f^-}.$$

The scalar proof does not establish a general matrix theorem: matrix cocycles do not telescope through a scalar speed ratio, normal–tangential blocks need not commute, and tangential drift can change both interface location and incoming phase. For a curved interface, flattening changes the normal across the physical tube of radius $\varepsilon = \tau R$ and introduces curvature terms. Any claimed $O(\varepsilon) + O(\eta/\tau)$ bound therefore requires a separate geometric argument with bounded curvature, transverse speed, tangential regularity, and $\varepsilon \to 0$. Grazing, sliding, chattering, and simultaneous interface hits are outside the scope of this section.

Logical scope. The results above do not assert an unrestricted sharp-first iterated limit, an exact saltation value at fixed R, an $O(\rho)$ error without a tail term, or universality of η/τ over arbitrary smoothing families. They also do not assert that every unresolved phase skips the layer, that the tuned-hit divergence holds at fixed phase, or that $\rho \max(-F') > 1$ forces orientation reversal. Finally, the scalar argument does not by itself prove a curved-interface or general multidimensional extension.

J Experimental protocols and evidence coverage

J.1 Finite-radius experiments: configuration, inputs and populations

All experiments compare the same terminal objective at two specified initializations. The datasets, seeds, objectives and directions are fixed. Deterministic mesh populations and seed–mesh quartiles are descriptive summaries, not independent statistical samples or confidence intervals. Results, scripts, full error distributions and file manifests accompany the paper. Scalar and affine analyses reuse preserved arrays. The smooth branch-response predictor (Branch) was computed from the same reference trajectories and previously evaluated candidate pairs. Here "smooth" refers to the regional expansion, not to smoothing the activation.

E1: exact scalar response. The scalar risk has target $a = 4$, residual target $c = -3$, $T = 1$, initialization $x = 4(1 - e^{\sqrt{2}/4})$, terminal objective $q(y) = (y-1)^2/2$ and direction one. All $N = 200, \dots, 10000$ are used at $\epsilon = 0.3/N$: 9,801 total rows, 9,801 finite responses, and 1,034 proposal reversals. The limiting negative and positive levels are -0.1228170958 and 0.7517137144; their predicted positive frequency is 0.1053282752, versus $1034/9801$ observed. The complete phase-count classification, rather than near-resonant quotient growth, determines this comparison. The smallest critical-endpoint margins select 20 precision checks; independent 60-digit Euler differs from the recorded response by at most 2.27×10^{-12}. Integer-length tests use $\ell = 1, 2, 3$ at $N = 200, 1000, 10000, 100000$. At the finest grid the response errors are below 10^{-5}. Absolute objective differences remain in every row.

E2: fixed coupled risk and feedback-active subset. The two-event risk and its proven geometry are in Appendix D.6. The coupling values are 0 and 0.4, direction e_1, horizon one and $q(\theta) = (\theta_1 - 1)^2/2$. The original 72 comparisons cross $N = 200, 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000$ with $\zeta = 0.1, 0.3, 1, 3$. The dense evaluation crosses all $N = 1000, \dots, 5000$ with $\zeta = 0.3, 1, 3$: 24,006 rows. The coarse evaluation crosses $N = 20, 30, 50, 75, 100, 150, 200, 300, 500, 750, 1000$ with $\epsilon = 0.001, 0.003, 0.01, 0.03, 0.1$: 110 rows. These populations were fixed before their respective evaluations. The second-order refinement was developed after the first-order results, using the whole population and the unchanged task; it is a method development study rather than an independent confirmatory replication.

The no-upstream-feedback ablation keeps the regional propagators, Euler bias and transported endpoint contributions but removes the first-event rounding contribution from the second threshold. The informative subset is defined by a difference in the predicted signed second-index count $k_2(z_+) - k_2(z_-)$ between full and ablated recursions. This test uses no rerun objective. It selects 224 of 12,003 coupled rows. The full recursion has lower error on 223, the ablation on one; there are no ties at tolerance 10^{-12}. Full mean/median absolute error is $5.93829 \times 10^{-5}/3.87803 \times 10^{-5}$, versus $8.16781 \times 10^{-3}/2.52606 \times 10^{-3}$ for the ablation. The full and ablated sign-error counts are zero and eight. The exception favoring the ablation has a failed predicted word and is retained. On the 223 valid informative words, every row favors the full recursion.

Every row is eligible for evaluation. Unresolved includes failed words; the categories are not additive. The coarse unresolved set contains nine failed words and seven valid words with insufficient sign margins. Certificate columns refer to the second-order deterministic test, whereas the feedback-error comparison isolates the first-order recursive law. The initial homogenized certificate gave 7/72 and 0/24006. Contractive first-order bounds give 70/72, 23997/24006 and 71/110 without changing that predictor; second order gives the displayed counts.

What the finite decisions change. All 19 coupled dense comparisons opposing AD and flow are certified; their actual gaps are about 1.3×10^{-6}–5.1×10^{-6}. The coarse coupled cases at $N = 150, 200$, $\epsilon = 0.001$, have actual positive responses 0.035408 and 0.010055, while both

Table 1: Finite affine comparison tests. Incorrect counts refer to certified comparisons; unresolved includes both invalid words and insufficient margins.

Population	Total	Valid word	Certified	Incorrect	Unresolved
Original E2	72	71	71	0	1
Dense E2, both couplings	24006	23997	23997	0	9
Dense E2, coupled	12003	11999	11999	0	4
Feedback-active coupled	224	223	223	0	1
Coarse E2	110	101	94	0	16

infinitesimal predictors are negative. The second-order response errors are bounded by 0.001118 and 0.000371. Guaranteed loss gaps rounded down are 6.85×10^{-5} and 1.93×10^{-5}, below the actual gaps 7.08×10^{-5} and 2.01×10^{-5}. The coupled disagreement at $N = 75$, $\epsilon = 0.003$ remains unresolved. The largest relative candidate gaps in this coarse scan agree with the infinitesimal predictions; larger objective effects are not automatically an advantage for the new predictor.

Controlled smooth-field and curved-surface check. Apply $h(x) = (x_1, x_2 + x_1^2/4)$ to the fixed coupled flow and set $f_y(h(x)) = Dh(x)f_x(x)$. Its guards are $g_1(y) = y_1$ and $g_2(y) = y_2 - y_1^2/4$. The diffeomorphism carries the proven isolating tubes and invariant region to bounded tubes with the same event order and one-sided normal speeds. This checks the abstract piecewise-C^2 theorem; it does not preserve Euclidean gradient structure. Euler is stepped directly in y, since $h(x + \eta f_x) - h(x) - \eta Dh(x)f_x = (0, \eta^2 f_{x,1}^2/4)$. We fix $y_0 = (-.4, -.96)$, direction e_1, $q(y) = (y_1 - 1)^2/2$, $N = 200, 500, 1000, 2000, 5000, 10000$ and $z = -1, 0, 1$ before execution. All 18 rows retain the predicted word; 11 meet the preset recursive phase margin 0.05. Across all rows the observed endpoint error divided by η^2 ranges from 0.0561 to 0.298. Three of six central pairs have both endpoints margin-eligible; two meshes also include the base point under that margin. These are phase checks, not a numerical certification of the theorem's sufficient mesh threshold. The six response errors divided by η are at most 0.141; none reverses the prediction's sign. These are observed ratios, not certified constants. Three 60-digit direct-y Euler checks at the geometrically selected closest-margin mesh $N = 10000$ agree in event indices and differ from float64 endpoints by at most 1.24×10^{-15}. No finite nonlinear certificate is applied.

E3: nonlinear training and branch-response inputs. The original width-4 network has 17 parameters. All seeds $0, \ldots, 19$, $N = 500, 2000$, and radii $10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}$ are kept. The direction is the first initialization coordinate, and the objective is the frozen validation half-MSE. Seed 16 has no admissible flow reference, leaving 152 comparable reference rows out of 160. Every plotted radius uses the same 19 seeds and 38 seed–mesh observations.

The Branch predictor takes the frozen base event times and directed surface sequence, reconstructs regional masks, and integrates the state, regional variational matrix and (4) with DOP853 at relative/absolute tolerances 2×10^{-11} and 2×10^{-13}. The reconstructed endpoint, regional product and saltation-bearing derivative differ from the preserved reference by at most 1.42×10^{-14}, 3.20×10^{-13} and 3.29×10^{-13}. Independent Euler step-halving on each seed's longest regional segment checks the smooth bias, with observed error ratios below 0.500. Only this base information constructs the response. Actual perturbed endpoints, masks and objectives are used afterward to evaluate it.

For both perturbed candidates, 90/152 proposed index sequences are strict and ordered; 81/152 match the complete rerun training word. Terminal validation masks match the reference branch on 138/152 rows. These are computed diagnostics, not a proof of all uniform-neighborhood hypotheses. All 152 predictions, including failures, enter the main comparison. There is no finite nonlinear certificate: certified/incorrect-certificate counts are inapplicable, rather than zero passed tests.

Entries are absolute response errors. No network sign reversal occurs relative to AD, flow or the Branch prediction. The Branch predictor's median at 10^{-4} conceals a failed-word maximum error 0.0486928; its mean is 1.29792×10^{-3}. At 10^{-2} the Branch and flow means are 1.02079×10^{-3} and 1.10035×10^{-3}. All quartiles, tail errors and the regional limit are included in the supplementary arrays. Small medians do not validate incorrect words.

Table 2: E3 absolute response errors and complete training-word matches. Every radius uses the same 38 seed–mesh observations. Branch denotes the branch-response predictor.

Radius	AD median	Flow median	Branch median	Matched training word
10^{-5}	$9.56 \cdot 10^{-12}$	$1.57 \cdot 10^{-3}$	$9.92 \cdot 10^{-6}$	29/38
10^{-4}	$1.96 \cdot 10^{-3}$	$1.88 \cdot 10^{-3}$	$8.32 \cdot 10^{-6}$	29/38
10^{-3}	$2.78 \cdot 10^{-3}$	$3.86 \cdot 10^{-4}$	$1.22 \cdot 10^{-5}$	17/38
10^{-2}	$1.72 \cdot 10^{-3}$	$1.87 \cdot 10^{-4}$	$1.43 \cdot 10^{-4}$	6/38

Sign policy, validation and reproducibility. Descriptive sign comparisons classify values of absolute magnitude at most 10^{-12} as zero, never as a reversal. Error differences within 10^{-12} are ties. The deterministic certificate uses its strict unrounded sign margin, independently of this descriptive tolerance. Its inequalities and strict word tests are evaluated in floating point. Independent checks include 24 distinct 60-digit Euler comparisons, 353 high-precision matrix-power state bounds, and 1,200 planar extrema tests including interior extrema and repeated eigenvalues. They validate the implementation numerically; formal interval arithmetic is not used. A separate algebra check covers 24 ordered affine-block examples in dimensions two and three, with zero, one, three and five event slots. All 65-digit exact matrix-power errors satisfy the finite second-order remainder (maximum error/bound 0.10441). These prescribed blocks check the finite-event algebra; they do not verify event geometry. Hash checks preserve all 951 earlier experiment files and 13 certificate-stage files.

The supplied scripts reproduce the calculations and vector figures. From the paper directory, run

```
python scripts/network_smooth_response.py
python scripts/validate_curved_response.py
python scripts/final_figures_analysis.py
python scripts/validate_final_finite_events.py
```

The evidence package also includes the original scalar, coupled and network generators, their protocols and raw arrays. Detailed hashes and build manifests are supplementary artifacts rather than mathematical assumptions.

J.2 Frozen network outputs: resolution, tails, and decision evidence

This appendix reanalyzes saved outputs without running training, perturbed trajectories, or a reference solver. The descriptive protocol was fixed before aggregation. E3 retains all 20 seeds at both meshes $N \in \{500, 2000\}$ and four radii $\epsilon \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$. Eight seed-16 rows lack a reference; four-method comparisons use the same 19 seeds and 38 seed–mesh observations at each radius. These repeated observations are not independent trials.

Definitions and raw-output checks. For both candidates, the complete predicted training word is compared with the saved training-word hash. The three disjoint reference strata are M: both complete words match (81 rows); O: both predicted words are strictly ordered, but at least one complete word differs (9); and U: at least one predicted word is not strictly ordered (62). Terminal validation-mask agreement is recorded separately and does not remove rows. The M/O distinction requires actual word comparisons. Unordered indices (U) can be detected at prediction time and left unresolved, but all 62 U rows remain in the reported comparison. The uncovered eight rows remain explicit in the full-population table. Only 79/81 complete-training-word-matched rows also retain both reference terminal validation masks. Matching the training word alone therefore does not establish the smooth-terminal hypothesis. Across the increasing radii, the matched counts are 29, 29, 17, 6. The pooled strata differ in radius and event-count composition; their error contrast is descriptive and cannot isolate a causal effect of word correctness. The stored flow and regional objective gradients use the validation gradient at the frozen continuous endpoint: $D\phi_T^\top \nabla q(\phi_T)$ and $J_{\mathrm{reg}}^\top \nabla q(\phi_T)$. Discrete AD uses the finite endpoint. Recomputing all three full vectors for 19 seeds and four historical meshes reproduces every stored component exactly in this environment; the suspected use of a finite terminal gradient for the flow object is therefore rejected. Saved Q_+ and Q_- reconstruct every central response; saved endpoint-error vectors reconstruct every branch prediction;

Stratum	Method	n	Median	Mean	90th	95th	Maximum
M	AD	81	0.000457	0.00958	0.024	0.0588	0.157
M	Flow	81	0.000652	0.00642	0.00832	0.0249	0.135
M	Regional	81	0.000453	0.00959	0.024	0.0588	0.157
M	Branch	81	9.03e-06	2.85e-05	4.12e-05	4.99e-05	0.000777
O	AD	9	0.00143	0.0044	0.00991	0.016	0.0222
O	Flow	9	7.49e-05	8.79e-05	0.000179	0.000204	0.00023
O	Regional	9	0.00139	0.00438	0.00991	0.016	0.0222
O	Branch	9	8.16e-05	0.000123	0.000241	0.000384	0.000527
U	AD	62	0.00314	0.0146	0.0549	0.0606	0.155
U	Flow	62	0.000613	0.00769	0.0112	0.0279	0.133
U	Regional	62	0.00311	0.0146	0.0549	0.0605	0.155
U	Branch	62	7.17e-05	0.00171	0.00459	0.0076	0.0487

Table 3: Absolute finite-response errors, pooled across the frozen radii and meshes, in the three exact strata defined above. All rows are retained. Scientific notation uses e for powers of ten. These quantiles describe this finite, correlated cohort.

ϵ	AD wins	Flow	Regional	Branch	Branch mean	95th	Maximum
1e-05	25/38	0/38	0/38	13/38	2.76e-05	0.000101	0.000242
0.0001	12/38	1/38	0/38	25/38	0.0013	6.19e-05	0.0487
0.001	2/38	6/38	0/38	30/38	0.000527	0.00447	0.00767
0.01	1/38	11/38	4/38	22/38	0.00102	0.00751	0.00777

Table 4: Lowest absolute-error counts among the four methods on the same 38 observations per radius. The prespecified absolute tie tolerance is 10^{-12}; no ties occur. Means and upper tails accompany the radius plot rather than reporting medians alone.

all stored word, phase-margin and terminal flags agree with their raw records. This checks data lineage and arithmetic, not an independent trajectory solve.

Radius comparisons do not identify a universal winner. At the largest radius, branch-response mean error is 0.00102 versus flow 0.0011; its 95th percentile and maximum are slightly larger. Thus neither the means nor the winner counts support a claim that flow uniformly dominates at the largest tested radius. The largest branch error, 0.0487, occurs at seed 9, $N = 500$, $\epsilon = 0.0001$, in stratum U; its recursive margin is 0.00246 and the actual absolute two-candidate objective difference is 2.74e-05. These quantities have different units: using the rounded values and $\epsilon = 10^{-4}$ gives a gap-prediction error of about $2\epsilon(0.0487) = 9.74 \times 10^{-6}$ and $|D| \approx 0.137$. This is an illustrative conversion, not a new experiment. All four predictors preserve the sign on the fixed E3 direction in all 152 reference-covered rows. This establishes no optimization benefit or direction reversal.

Phase margin, event count, and mesh resolution. The recursive margin is the minimum distance to an integer of every event ceiling argument, over both candidates. It includes upstream feedback in the prediction, but it is not a measured distance to the exact nonlinear discrete branch boundary. Strict order also requires $0 < K_1 < \cdots < K_m < N$. The reported shortest reference segment includes the initial-to-first and last-to-terminal intervals; it is not silently replaced by an inter-event-only minimum. Event counts are grouped by their exact integer value, and mesh summaries use the same seeds. Reference event counts range from 1 to 46 (median 23). The pooled margin AUC is 0.647 and the pooled Spearman correlation with branch-response error is -0.156. The prespecified $N = 2000$, $\epsilon = 10^{-4}$ subgroup has AUC 0.324; hence even the direction of association is not uniform across the tested groups. The fixed bins are $[0, .001), [.001, .005), [.005, .01), [.01, .05), [.05, .1), [.1, .25), [.25, .5]$. Every bin count, error distribution, exact-event-count group, and paired seed–mesh radius transition is retained. In this shallow model the activation guards are affine in the parameters ($D^2 g = 0$), so this subgroup result does not identify guard curvature as its cause. The computed phase margin alone does not control accumulated regional approximation errors or establish the theorem's finite-mesh

N	Rows	Word matched	Unordered	Margin median	IQR	AUC
500	76	34	38	0.0077	[0.00299, 0.0265]	0.676
2000	76	47	24	0.0111	[0.00341, 0.0321]	0.597

Table 5: Descriptive discrimination of complete word matching by a larger recursive margin. The remaining ordered-mismatched rows number 4 at $N = 500$ and 5 at $N = 2000$, completing each 76-row partition. AUC is the rank probability with half credit for ties; no threshold is fitted, and no independence-based uncertainty interval is claimed.

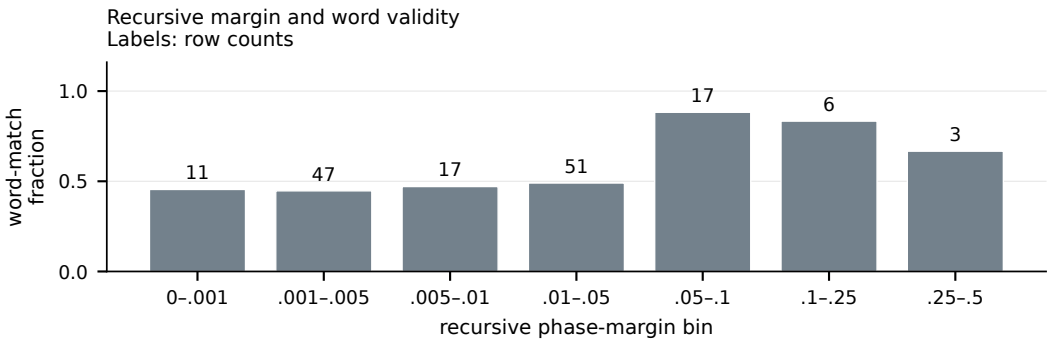

Figure 4: Exact complete training-word match fractions in the prespecified recursive-margin bins for all 152 reference-covered width-4 rows. Labels give dependent seed–mesh–radius row counts, not independent trials. The eight uncovered rows have no margin. No bin is chosen or removed after observing the errors. The margin is a descriptive diagnostic and is not a nonlinear certificate.

conditions. Multiple crossings within one GD step can indicate insufficient grid resolution; they do not establish simultaneous continuous-time events.

Full initialization gradients versus finite vector decisions. On the same 38 E3 seed–mesh pairs, the full flow gradient differs from AD in at least one component sign on 16/38 (8/19 unique seeds), in the largest absolute coordinate on 6/38 (3/19 seeds), and in the ordered two largest absolute coordinates on 16/38 (8/19 seeds). The minimum cosine is 0.858397, the maximum normalized-direction distance is 0.532172, and its median is 0.300401. Regional gradients change none of these signs or ranks and have maximum normalized-direction distance 0.001488. These are objective-gradient and normalized-direction diagnostics, distinct from matrix discrepancies. They do not evaluate finite proposals along these full-vector directions. The 76-row, four-mesh historical diagnostic is reported separately, rather than doubling the E3 denominator.

Existing outer-variable proposals and preserved negative results. The separate frozen four-dimensional training-group-weight audit starts with 20 seeds, retains 15 eligible seeds and 14 fixed common-itinerary seeds across all primary radii, and stores 660 proposals including boundary probes. No proposals are rerun here. Historical method names are retained: these are not finite proposals generated by the new branch-response predictor. Flow outer gradients differ from AD in some component sign on 2/15, in ordered top-two absolute-coordinate ranks on 1/15, and in top-one ranks on 0/15. Event-aware outer gradients differ from AD in some component sign on 2/15, in ordered top-two absolute-coordinate ranks on 1/15, and in top-one ranks on 0/15. At the three smallest primary radii, flow-minus-AD proposal-objective median differences are 9.96e-10, 1.99e-09, 3.99e-09, all positive in all 14 common seeds. At radius .005 the signs split 7 lower/7 higher; the median is 8.39e-10 and the largest absolute difference is 4.5e-06. After the stored five matched outer steps, flow achieves a lower final objective on 13/15 completed seeds and 9/11 seeds satisfying every stored itinerary flag; on the latter fixed subset the median flow-minus-AD final-objective difference is -2.41e-07. Thus gradient/rank differences do not entail uniformly better finite proposals, and the negative result for a universal AD advantage is preserved.

Reproduction and unavailable extensions. From the paper directory, run:

```
python -X utf8 scripts/resolution_network_analysis.py
```

The frozen analysis protocol, row IDs, source hashes, exact numeric CSV tables, and figure inputs are under `experiments/resolution_revision_20260906/`. The script reads saved arrays and raw JSON only. The current E3 candidate-response and recursive-margin outputs exist only for width 4. Widths 8 and 16 retain historical base-path resonance and matrix checks, which are distinct quantities and are not presented as wider-network recursive-margin validation. Finite vector proposals driven by the critical response, a nonlinear network certificate, and computational savings over two direct candidate evaluations are not established by these outputs.

J.3 Frozen network population and reference construction

The supporting matrix audit uses the frozen two-moons split, with 64 training and 64 validation samples, noise 0.08, standardized training inputs and targets in $\{-1, +1\}$. Full-batch hard-ReLU squared-loss training uses float64, $T = 1$, widths $4, 8, 16$ (respectively $17, 33, 65$ parameters) and seeds $0, \ldots, 19$ at every width. The reference-admissible counts are $19/20, 20/20, 18/20$: 57 of the fixed 60 configurations, evaluated at $N = 200, 500, 1000, 2000$ for 228 mesh rows. Width-4 seed 16 has a numerically grazing one-sided speed; width-16 seeds 4 and 15 have opposing one-sided normal velocities. Zero cases are event-admissible but fail the subsequent reference comparison. No sensitivity-gap threshold selects this population.

Within each activation region, DOP853 integrates the state, regional variational equation and four group-weight sensitivity columns. Each sample–unit preactivation is projected onto the accepted step's dense polynomial. A Bernstein same-sign test rejects root-free polynomials; otherwise derivative roots partition the step into monotonicity cells. Companion roots and dense evaluations cross-check candidates, ambiguities fail closed, and continuation stops at the earliest root before restarting under the outgoing field. The 8-, 32- and 128-point scans check this inventory afterward; they are not independent root locators. Coarse/medium/tight $(\mathrm{rtol}, \mathrm{atol}, h_{\max})$ are $(3 \cdot 10^{-8}, 3 \cdot 10^{-10}, .005)$, $(10^{-9}, 10^{-11}, .003)$ and $(3 \cdot 10^{-11}, 3 \cdot 10^{-13}, .002)$.

Admissibility requires unit-normal speeds $\geq 10^{-5}$, event separation $\geq 2 \cdot 10^{-7}$, normalized non-event margin $\geq 10^{-9}$, and absolute/relative root residuals $\leq 2 \cdot 10^{-9}$. Initial/terminal events, near-simultaneity, opposing speeds and numerical grazing fail closed. The normalized margin is $|s_i| / \max(1, \|n_i\|_2)$. All three solves must agree on directed event signatures; medium/tight endpoints, event times, regional and flow Jacobians, parameter sensitivities and transported products are also compared. These are floating-point checks on numerical interpolants, not an interval proof of the exact ODE event inventory.

The 57 references contain 2,357 events. Maximum root residual and polynomial/evaluator discrepancy are $3.33 \cdot 10^{-16}$ and $1.33 \cdot 10^{-15}$; minimum unit-normal speed, event separation and normalized non-event margin are $5.70 \cdot 10^{-4}$, $1.18 \cdot 10^{-6}$ and $3.04 \cdot 10^{-8}$. All signatures agree across tolerances and no inventoried root is missed by the auxiliary scans. Across 228 GD rows no evaluated zero activation is detected; the minimum dimensionless resonance distance is

$$\min_{k,i} \frac{|s_i(\theta_k)|}{\eta \max(|n_i^\top f(\theta_k)|, 10^{-14})} = 2.09 \cdot 10^{-5}.$$

This supports numerical nonresonance of those evaluated programs.

J.4 Correction sources and independent evidence

The final driver passes initialization, weights, mesh, dataset, width and configuration to `gd_with_derivatives`, but no reference events. Six replays (seed zero at each width, $N = 200, 2000$) reproduce saved J^{AD} and J^{SC} arrays exactly in the recorded environment. Thus labeling this network correction reference-assisted reconstruction would be inaccurate. Reference events are used in the separate transported-product identities and exact scalar correction controls.

Multiple events within one step. For this shallow network each preactivation is affine in all parameters, hence affine in fractional time on a straight Euler segment. With strict endpoints it has at most one zero, exactly characterized by opposite signs. The code processes every changed sample–unit pair, sorts $-s_i(\theta_k)/(s_i(\theta_{k+1}) - s_i(\theta_k))$, and applies all transfers in that order. The

Table 6: Sources in the frozen multi-width implementation. Reference and coarse-GD columns describe separate computations.

Quantity	Reference computation	Coarse-GD correction computation
Event count/order	Dense DOP853 roots; earliest root followed by restart	All changed affine preactivations on each Euler segment, sorted by fractional crossing location
Event time/state	Root-localized dense state	Linear interpolation of the actual Euler segment
Normals, one-sided fields	Analytic affine normals; fields at reference root under adjacent masks	Same formulas at interpolated coarse state, toggling masks in estimated order
Regional propagation	DOP853 variational solve between reference events	Entire-step factors $I + \eta A(\theta_k)$; no within-step regional subintegration
J_{reg}	Reference regional matrices without saltation	Not used to construct J^{SC}
$D\varphi_T$	Reference regional matrices and reference transfers	Comparison target only
J^{SC}	Not supplied by reference inventory	Full coarse factor followed by all estimated transfers at that step

width-4 $N = 200$ replay has one two-event step; width 16 has two multi-event steps, with up to three events in a step. Completeness holds for these straight segments, not for the true flow, which may recross between grid times. It does not extend to general deep-network preactivations along parameter segments.

The reference locator audits each regional dense polynomial, advances only to its earliest root, switches the field, and restarts. Subsequent candidates are recomputed under outgoing dynamics. Its solver step need not be below the smallest reference event separation 1.18×10^{-6}. Coarse endpoint masks alone do not supply that inventory.

Retained itinerary mismatches. All 228 historical rows pass the correction's local speed and near-simultaneity checks. Only 165/228 stored event-identity sequences match the reference: $31, 38, 46, 50$ of 57 at $N = 200, 500, 1000, 2000$. All 63 mismatched rows remain in the frozen summaries. The historical flag `correction_covered` does not establish the correct-itinerary hypothesis of Theorem 38. The seven unmatched finest-grid cases remain in the full-cohort summary; the separate matched-cohort comparison uses a fixed intersection across all four meshes (Appendix J.5).

For matched seed-zero replays, maximum event-time errors at $N = 200$ are 7.77×10^{-3} (width 4) and 2.59×10^{-3} (width 8); at $N = 2000$ they are $5.18 \times 10^{-4}, 3.32 \times 10^{-4}, 2.25 \times 10^{-4}$ by width. Width 16 at $N = 200$ has unequal order despite equal counts (43); no paired localization error is assigned. A small corrected matrix error is a controlled consistency diagnostic, not certified event recovery or the derivative of the unmodified finite GD program.

Three validation claims. The 57 transported-product and determinant tests are algebraic self-consistency. Cross-tolerance root, state and variational comparisons test reference stability. Perturb-and-rerun checks in Table 7 validate selected responses, not full Jacobians. Leading singular-gap directions are deliberately chosen worst-direction checks, not typical learning evidence. The new coordinate scan is reported separately and changes no historical coverage denominator.

J.5 Derivative coverage and fixed-cohort matrix summaries

Relative errors divide by $\max(\|\text{reference}\|, 10^{-14})$, using Euclidean vector and spectral matrix norms. State errors use endpoints, not the path supremum. Table 7 summarizes the independent perturbation coverage. The first two rows of Table 7 use eight normalized Rademacher directions and overlap only width-4 seed 0 at $T = 1$ with four meshes. The additional audit fixes three cases per width by minimum normal speed, maximum limiting gap, and upper-median limiting gap (ties by seed), then tests one leading right singular-gap direction per case. This is a worst-direction diagnostic. Discrete perturbations shrink from 10^{-6} to 10^{-10} to preserve the full training word; flow perturbations use $2 \cdot 10^{-6}$ and require the same directed itinerary. The positive width-8 seed-3 flow perturbation

Table 7: Independent checks retain their eligible denominators. A passed directional test is not a full-Jacobian check; uncovered perturbations remain in the checked count.

Check	Eligible	Checked	Passed	Maximum relative error
Discrete Rademacher JVP	228	4	4	$4.42 \cdot 10^{-8}$
Flow Rademacher JVP	57	1	1	$3.69 \cdot 10^{-8}$
Discrete singular-gap JVP	228	36	36	$1.96 \cdot 10^{-7}$
Flow singular-gap JVP	57	9	8	$7.60 \cdot 10^{-8}$
Initialization full Jacobian	57	0	0	–
Discrete group hypergradient	15	15	15	$5.79 \cdot 10^{-9}$
Network forward–adjoint	57	0	0	–

changes event order and remains uncovered. In total, 40/228 distinct meshes and 10/57 references have independent initialization checks attempted; nine references have covered passing checks. No full-Jacobian or network adjoint coverage is inferred.

At $N = 2000$, pooled median endpoint, AD-to-regional and corrected-to-flow relative errors are $3.58 \cdot 10^{-5}$, $1.31 \cdot 10^{-4}$ and $1.51 \cdot 10^{-4}$. Median AD-to-flow and regional-to-flow gaps are 0.384737 and 0.384733; their absolute difference has median $3.45 \cdot 10^{-6}$. Median per-case slopes on $N = 500, 1000, 2000$ are $0.998, 1.001, 0.997$ for the first three errors and $3.50 \cdot 10^{-6}$ for AD-to-flow error. The maximum transported-product, flow-reconstruction and determinant/speed-ratio residuals are $2.00 \cdot 10^{-14}$, $7.35 \cdot 10^{-15}$ and $2.67 \cdot 10^{-14}$. These last checks are algebraic identities on reference quantities.

All 57 configurations and 228 rows remain in the full summaries. The fixed intersection matching the reference itinerary at every mesh contains 31 configurations. Slopes of the final-three-mesh *pooled median* corrected error are 1.0094 for all 57 and 0.9986 for the fixed 31. These are distinct from the preceding median-of-case-slopes statistic. Each curve uses the same IDs at every mesh; `correction_fixed_cohort.json` records them. The per-width matrix summaries are retained in Figure 6.

Table 8: Fixed-width summaries at $N = 2000$. Event counts and relative matrix errors are medians over the stated reference population.

Width	Parameters	References	Events	G_∞	e_{reg}	e_{corr}
4	17	19/20	23.0	.346	$1.12 \cdot 10^{-4}$	$1.42 \cdot 10^{-4}$
8	33	20/20	36.0	.397	$1.29 \cdot 10^{-4}$	$1.27 \cdot 10^{-4}$
16	65	18/20	53.5	.413	$1.39 \cdot 10^{-4}$	$3.13 \cdot 10^{-4}$

Gate sensitivity retains the original 60-case denominator. Normal-speed thresholds $10^{-6}, 10^{-5}, 10^{-4}$ retain 57 references, and 10^{-3} retains 53. Separation thresholds $2 \cdot 10^{-8}, 2 \cdot 10^{-7}, 10^{-6}$ retain 57, while $2 \cdot 10^{-6}, 10^{-5}$ retain 56. Normalized near-touch thresholds through 10^{-8} retain 57, and 10^{-7} retains 54; root-residual thresholds 10^{-10} through 10^{-7} leave the count unchanged. Conditional median limiting gaps range from 0.376 to 0.386. These are deterministic screening checks, not prevalence estimates.

J.6 Outer-weight provenance and negative optimization result

With $\ell_i(\theta) = \frac{1}{2}(f_\theta(x_i) - y_i)^2$, four groups defined by class and the sign of the first standardized coordinate have counts $24, 8, 9, 23$. The implemented objectives are

$$L_{\mathrm{train}}(\theta, \lambda) = \frac{1}{64} \sum_{i \in \mathcal{T}} \lambda_{g(i)} \ell_i(\theta), \tag{156}$$

$$Q_{\mathrm{disc}}(\lambda) = \frac{1}{64} \sum_{i \in \mathcal{V}} \ell_i(\Theta_{\eta,T}(\theta_0; \lambda)). \tag{157}$$

The denominator is the sample count, not the sum of weights. Gradients are with respect to λ, with base point $(1, 1, 1, 1)$. Proposals are $\lambda - \epsilon g/\|g\|_2$ without projection. Nonpositive weights reject a proposal or stop a path. Every evaluation reruns all $N = 2000$, $T = 1$ inner steps from the same initialization; there is no warm start.

The width-4 population contains 19 final references, but its outer cohort contains 15 because earlier gates excluded seeds 5 and 18 by separation, 9 by reference validation and 12 by the flow finite-difference gate. Seed 16 also fails the final grazing audit. This 19-to-15 reduction is not a new application of the final gates. The 15 retained references agree with the final pipeline in event signature and in all compared endpoint, initialization and group-sensitivity arrays to relative error below $8.66 \cdot 10^{-13}$. All 660 one-step proposals and every five-step path were rerun with those final references.

Table 9: All 20 width-4 seeds. Calibration checks the executed training word at one radius, not a complete smooth neighborhood.

Seeds	Reference	Outer cohort	Calibration	Restriction
2, 4, 7, 10, 11, 13, 15	yes	yes	yes	None
0, 1, 3, 6, 8, 17, 19	yes	yes	no	Training word changes
14	yes	yes	no	Proposal itinerary changes
5,18	yes	no	–	Earlier separation gate
9	yes	no	–	Earlier reference gate
12	yes	no	–	Earlier flow FD gate
16	no	no	–	Grazing; earlier separation gate

Every outer-cohort seed passes the reexecuted four-coordinate discrete hypergradient finite differences, with maximum relative error $5.79 \cdot 10^{-9}$. Four proposals (AD, regional limit, hybrid flow, and corrected flow) use eight primary radii from $5 \cdot 10^{-5}$ to $5 \cdot 10^{-3}$ and three boundary probes. Fourteen of 15 seeds retain the common continuous and discrete event identities at all primary radii; seed 14 fails this condition. The stricter calibration has seven seeds at $\epsilon = 5 \cdot 10^{-5}$. Its hashes cover training masks only, not terminal validation masks or every point between evaluated endpoints. Thus the full provenance is 57 references across widths, 19 at width 4, 15 outer-cohort seeds, 14 common-itinerary seeds, and seven calibration seeds; these are different denominators and tests.

For the three smallest radii, hybrid-minus-AD median loss differences are $9.96 \cdot 10^{-10}$, $1.99 \cdot 10^{-9}$ and $3.99 \cdot 10^{-9}$, positive on all 14 seeds; signs mix at larger radii. Base hybrid/discrete gradients differ in a component sign on 2/15 seeds, an ordered top-two choice on 1/15, and no top-one choice. The five-step audit recomputes each derivative and uses radius 0.005. Eleven of 15 seeds retain a common initial itinerary along all four complete paths. Median final loss differences relative to AD are $1.69 \cdot 10^{-9}$ for the regional path, $-2.41 \cdot 10^{-7}$ for hybrid flow and $-2.52 \cdot 10^{-7}$ for correction. Hybrid and corrected paths each attain lower loss on 9/11 seeds. This negative optimization result is retained in Figure 7; a derivative identity does not rank nonlinear multi-step optimizers.

J.7 Supporting scalar, parameter, and resolution checks

The curvature-phase control fixes $T = .8$, $t_* = .3$ and event ratios $r \in \{1/40, 1/20, 1/10, 1/5, 1/2, 1, 2, 5, 10, 20, 40\}$, with $m = 1/\min(1, r)$, $c = (r - 1)m$ and $\theta_0 = m(1 - e^{\cdot 3})$. Eight step sizes $.04, .02, .01, .005, .002, .001, .0005, .0002$ yield 88 nonresonant one-crossing rows. The objective is $q(z) = z^2/2$ and the scalar correction inserts the exact event ratio. Minimum speed, dimensionless resonance distance and branch radius are 1, 0.150005 and $4.05 \cdot 10^{-5}$. Maximum endpoint/outer AD discrepancies are $3.82 \cdot 10^{-14}$/$4.11 \cdot 10^{-13}$; corresponding finite-difference discrepancies are $3.66 \cdot 10^{-9}$/$3.37 \cdot 10^{-8}$.

The earlier scalar radius grid contains 78 rows and six record-low near-resonance probes, in addition to the complete phase population in E1. The probe $N = 9801$, $\rho_\eta = 8.74 \cdot 10^{-10}$, $\epsilon = \sqrt{\rho_\eta \eta} = 2.99 \cdot 10^{-7}$ has response 89.53 but absolute objective difference only $5.35 \cdot 10^{-5}$. It deliberately probes resonance, not typical impact. Within-cell response agrees with AD to $6.15 \cdot 10^{-13}$; the integral-plus-jump residual is at most $1.69 \cdot 10^{-9}$. An independent Euler loop checks 18 endpoints

and derivatives. The earlier width-4 seed-0 coordinate scan retains all 24 mesh–radius comparisons and 12 perturbed reference flows. Every response is negative. Training and terminal validation branches can change separately; no interval-certified branch radius is claimed.

The two-parameter control uses $F_{(u,v),\lambda}(1) = (u, v, \mathrm{ReLU}(u-\lambda) + .5v)$, target $(1, .2, 1)$, $T = .8$, $\lambda = 0$ and $(u_0, v_0) = (1 - e^{\cdot 3}, 0)$. Its one event is at .3, state $(0, .1751180039)$, with speeds $1, 1.9124409981$ and $\beta = -.9124409981$. On the same eight step sizes, maximum PyTorch, discrete finite-difference and flow finite-difference discrepancies are $4.87 \cdot 10^{-14}$, $2.59 \cdot 10^{-8}$ and $1.62 \cdot 10^{-9}$. Continuous and independently implemented discrete forward–adjoint pairings agree within $1.39 \cdot 10^{-17}$ and $1.43 \cdot 10^{-14}$. These full-Jacobian/adjoint checks do not increase network coverage.

The scalar smoothing control uses $F(u) = .2 + 1.8[1 + \tanh(u/2)]/2$ on $[-20, 20]$, with 25 logarithmic values each of η, τ from 10^{-5} to 10^{-1}: 625 float64 cells. The transfer error is $|J - 10|$ and the truncation floor is $2.04 \cdot 10^{-7}$. This fixed profile supports the scalar sufficient resolution condition, not a universal phase boundary.

K Prior work and the mathematical increment

Primary-text locations. For Stewart & Anitescu (2010), we inspected the authors' June 6, 2005 preprint ANL/MCS-P1258-0605: printed pp. 4–5, Section 2, (2.3)–(2.4), and pp. 7–8, Lemma 3.1. These are preprint pages, not an inferred offset into the journal version, pp. 653–695, whose subscription full text was unavailable for version comparison. The publisher abstract confirms the crossing-sensitivity obstruction. For Nurkanović et al. (2024), we inspected the published Theorem 19 and proof, pp. 1150–1152, and Section 4.6/Figure 9, pp. 1152–1154.

Stewart–Anitescu's crossing dynamics is $\dot{x} \in (1+\alpha) - \mathrm{Sgn}(x)$, $\alpha > 0$. They explicitly calculate the flow sensitivity $\alpha/(2+\alpha)$ and persistent errors of differentiated discretizations. Their theta-method includes implicit and partly implicit branches; their derivatives are not all the Euler regional limit. Lemma 3.1 explicitly obtains a nonpositive normal field jump from one-sided Lipschitzness. The obstruction and attenuation geometry therefore have direct precedents.

Convex-potential deduction and what strong convexification changes. Our algebraic deduction from their field is

$$V_\alpha(x) = |x| - (1+\alpha)x = 2\,\mathrm{ReLU}(x) - (2+\alpha)x, \qquad -\partial V_\alpha = (1+\alpha) - \mathrm{Sgn}(x).$$

This potential is convex, but unbounded below for $\alpha > 0$, with zero regional Hessians. It is a deduction, not a convex-ReLU theorem stated by the prior authors. Expanding our residuals gives exactly

$$\mathcal{L}_{\alpha+2,-2}(x) = V_\alpha(x) + \tfrac{1}{2}x^2 + \tfrac{1}{2}\,\mathrm{ReLU}(x)^2 + C_\alpha, \qquad C_\alpha = \tfrac{1}{2}(\alpha+2)^2 + 2. \tag{158}$$

The added term is C^1, globally 1-strongly convex, and has zero derivative at zero. It preserves the interface, jump magnitude 2, one-sided speeds $(\alpha+2, \alpha)$, and saltation ratio. It changes the regional fields from constants to $\alpha + 2 - x$ and $\alpha - 2x$, and Hessians from $(0, 0)$ to $(1, 2)$. The crossing clock changes from $-x_0/(\alpha+2)$ to $\log((\alpha+2-x_0)/(\alpha+2))$; regional propagation is no longer the identity. The new potential is coercive with unique minimizer $\alpha/2$. Thus the full trajectories and discrete branches are changed. Every reciprocal ratio $R > 1$ already occurs as $R = (\alpha+2)/\alpha$; our fixed-speed family also rescales the instance. Strong convexity and the explicit conditioning calculation are extensions of the example, not a new inconsistency mechanism.

Mathematical versus application distinctions. Fixed-time GD refinement is explicit Euler refinement of a discontinuous field. Calling depth T/η optimization time changes the learning interpretation, not that numerical limit. Fixed-program AD has different quantifiers: it identifies the returned derivative at each finite program without uniformly controlling shrinking cells (Lee et al., 2020, Section 3). Fixed-point differentiation genuinely changes the limiting object: a fixed map is iterated to equilibrium, rather than an η-dependent map evolved for time T (Bolte et al., 2022, Assumption 1 and Theorem 2, pp. 3 and 5). Architectural Euler refinement can share the same numerical structure; substantive differences require specifying fields, parameter sharing, regularity and derivative targets. Smooth high-order ODE-net corrections (Xu et al., 2023, Propositions 2.2–2.3) concern different discretization weights. Gao et al.'s smooth gradient theorem (Gao et al., 2025, Proposition 2, p. 5)

and ReLU experiment do not identify this optimizer sensitivity, but terminology alone supplies no priority. Event-aware ReLU neural modeling also has a direct precedent in Brändle et al. (2026).

K.1 Finite perturbation analysis and the mathematical increment

Classical perturbation analysis already propagates finite changes along a sample path, including changes of subsequent event decisions (Ho et al., 1983; Ho & Cao, 1991; Ho, 1992; Wardi et al., 2017). For example, the Lindley recursion in Wardi et al. (2017, author-hosted preprint, pp. 3084–3085, Eqs. (1)–(3)) gives exactly

$$\delta D_k = \delta Z_k + \max\{I_k, \delta D_{k-1}\} - \max\{I_k, 0\}, \qquad I_k = A_k - D_{k-1}.$$

An upstream finite perturbation therefore changes a downstream branch when it crosses the corresponding idle-time threshold. This mechanism, and the distinction between finite propagation and its branchwise infinitesimal approximation, are not new here. Hybrid IPA likewise propagates state and endogenous event-time derivatives recursively (Cassandras et al., 2010, pp. 645–647, Eqs. (6), (7), (11)). The ordered variational and saltation factors are classical (Burden et al., 2016, Sec. 7.1).

Encoding each Euler step as a clock event places the entire finite training program within this broad perturbation-analysis framework. Consequently, exact finite propagation alone cannot establish novelty. The additional result concerns a changing family of these programs: the step size and perturbation radius vanish together. We derive the integer crossing indices from grid signs, including the smooth Euler bias and the earlier crossings' effects on later boundaries, and prove a uniform local endpoint expansion away from the resulting recursive boundaries. For the contractive affine subclass, verified branch words also support an explicit finite-grid remainder bound. Neither the continuous-flow saltation formula nor the exact finite Lindley rule provides these mesh-uniform estimates without this additional analysis. This is a numerical asymptotic result within the scope of perturbation analysis, not a new general framework for event propagation.

The full 1983 article, the 1985 event-domain article, and the full 1991 FPA chapter were not accessible in this audit. Their available primary abstracts and bibliographic records support historical attribution but do not support theorem-level exclusion claims. In particular, we make no claim that no formulation of classical FPA can encode the recurrence.

K.2 Finite outer learning and the resolution question

Maclaurin et al. (2015) differentiate a complete finite training run (PDF pp. 1–2, Algorithms 1–2). Their Section 4/Figure 10 (PDF p. 7) already shows that exact hypergradients can fail to describe a broader objective trend under unstable learning dynamics. The warning that a derivative is local is therefore not an additional contribution here. Our transverse-event analysis gives a quantitative mesh-dependent response law, including strongly convex examples, without requiring that instability mechanism.

Franceschi et al. (2018) explicitly define the finite-inner objective in Eqs. (5)–(10) (PDF pp. 3–5), while their convergence theorems compare increasing inner iteration counts with an inner optimum. This differs from refining the Euler step at fixed physical horizon. Finn et al. (2017) update an initialization through a finite adaptation step (Eq. (1), Algorithms 1–2, PDF pp. 2–4). Ren et al. (2018) obtain validation-guided example weights through a one-step training surrogate and subsequent rectification/normalization (Eqs. (4)–(9), PDF pp. 3–4). These sources motivate finite outer actions; they do not establish the current event geometry for their stochastic, momentum or online algorithms. Continuous outer parameters can be incorporated under the augmented-state conditions in Appendix F.4. Discrete changes of architecture or program length require a separate model.

The contribution boundary is thus precise. Smooth Euler bias, chronological propagation, and converting an error bound into a sign test are standard ingredients. The main independent argument derives the selected integers and closes their uniform endpoint expansion together, with a phase-dependent mesh threshold. The affine theorem supplies an explicit verified remainder. The three-radius statement and scalar phase moments are deductions used to interpret those results, not extra novelty claims.

Table 10: Closest prior results and the restricted mathematical increment. Locators refer to the inspected versions; detailed access records and the full comparison matrix accompany the evidence package.

Source	Object and locator	Already established	Limit considered	Additional proof here	Access
Ho et al.; Ho–Cao	FPA, 1983; book Ch. 6, pp. 185–227	Finite sample-path reconstruction; event-order changes	Finite perturbation of a model	Uniform refining-grid theorem not verified there	Abstract and contents only
Ho (1992)	GSMP and FPA, pp. 231–238	Clock-event representation; cut-and-paste propagation	Finite or infinitesimal path changes	Bias-corrected crossing-index asymptotics	Full text
Wardi et al. (2017)	Lindley rule, preprint pp. 3084–3085	Exact finite thresholds; downstream feedback	Finite queue perturbation	Uniform Euler-scale expansion	Full text
Cassandras et al. (2010)	Hybrid IPA, pp. 645–647	Recursive endoge-nous/induced event sensitivities	Infinitesimal parameter change	Derive integer events and surviving mesh phases	Relevant pages
Burden et al. (2016)	Sec. 7.1, pp. 1250–1254	Ordered saltation; linearized guards; flow selections	Directional flow derivative	Discontinuous, nonhomoge-neous grid-scale response	Relevant sections
Zhang & Chan (2007)	Thms. 1–3; Sec. 3	Local critical regions and finite error analysis	Queue perturbation within a basis	Recursive grid cells and cross-cell response	Full text
Stewart–Anitescu; FESD	Crossing example; Thm. 19	Discretization sensitivity failure; event-aligned recovery	Fixed-grid or adapted ODE dis-cretization	Coupled finite-response law and verified comparison	Full relevant proofs

L ADDITIONAL DISTRIBUTIONS AND SUPPORTING FIGURES

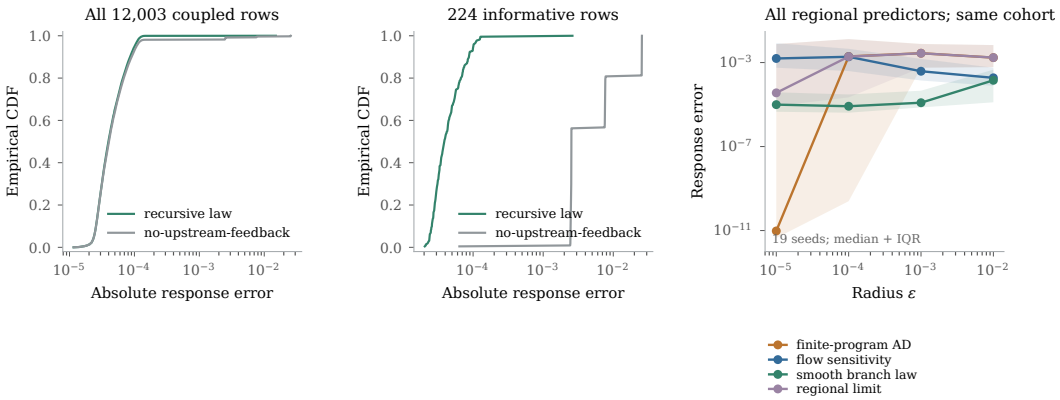

Figure 5: **Full error distributions and the common network cohort.** Left: the empirical error CDF includes all 12,003 coupled dense-grid rows, including four failed words. Middle: the 224-row subset is defined by a difference in the predicted signed second-event crossing count, without consulting endpoint objectives. It retains the one failed word where the ablation is more accurate. Right: the same 19 reference seeds and both meshes at every radius, now also showing the regional limit. Bands are interquartile ranges across 38 seed–mesh observations per radius. The smooth branch predictor is evaluated even when its proposed word fails; it has no finite-grid network certificate. Mean, median, tail errors and all failure counts are reported in the accompanying tables rather than replacing the cohort with successful subsets.

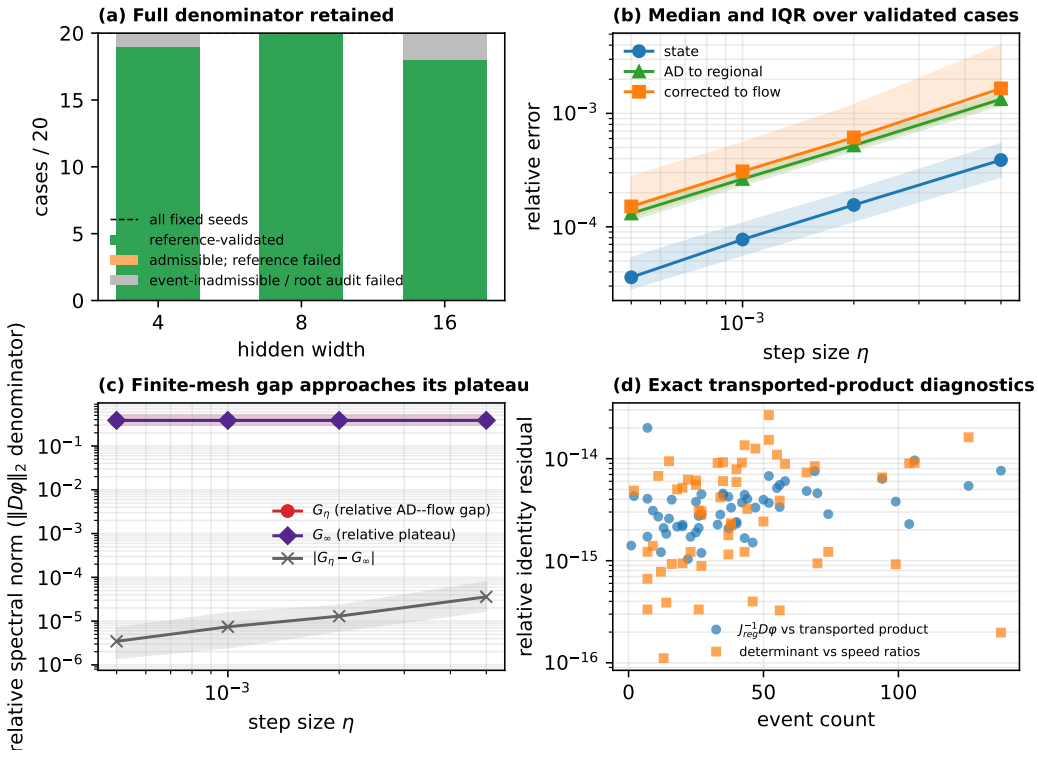

Figure 6: **Root-audited all-seed diagnostics.** (a) Exact counts retain all 20 prespecified seeds per width; no Wilson or other sampling interval is attached. (b) Over the 57 reference-validated cases, state, AD-to-regional, and coarse-path corrected-to-flow errors decrease under mesh refinement; thin curves show individual cases and thick curves show medians with IQR bands. (c) The relative finite-mesh gap G_η approaches the event-predicted plateau G_∞, while Δ_G decreases. (d) The transported-product and determinant/speed-ratio identities hold to numerical precision. Reference coverage is not full-Jacobian independent coverage. The 63/228 coarse itinerary mismatches remain in these pooled summaries; Appendix J.5 separately gives the same 31 matched configurations across all meshes. This controlled grid is not a prevalence study.

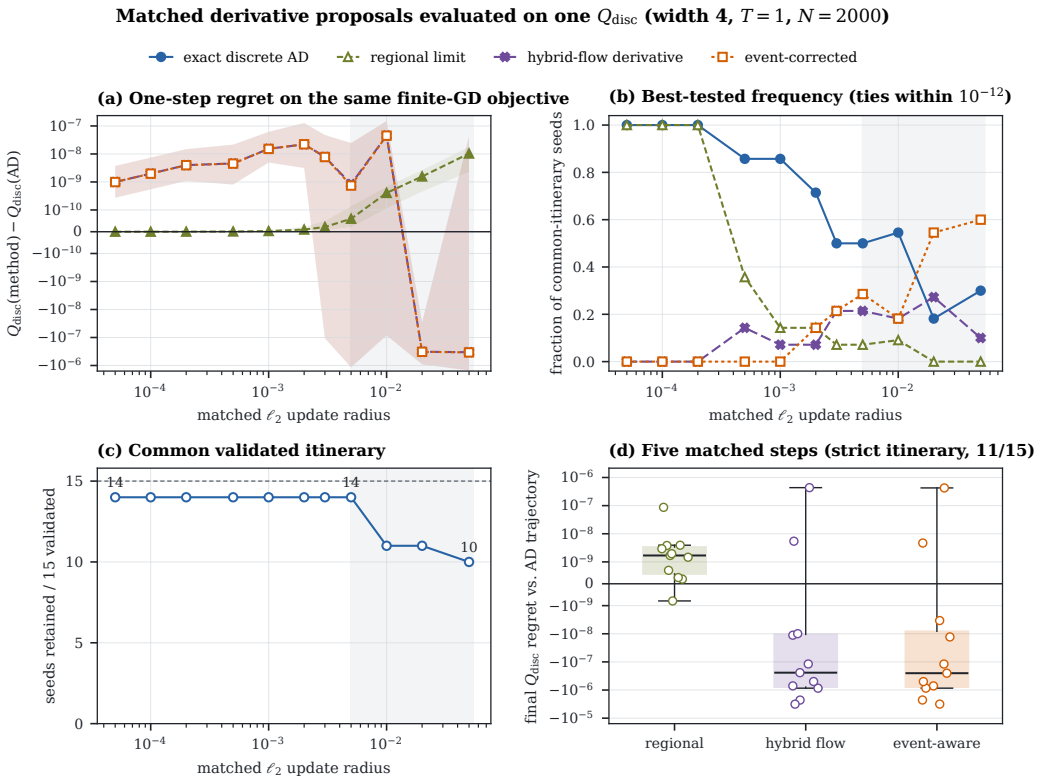

Figure 7: **All derivative proposals evaluated on the same finite-GD objective.** Signed differences are measured against the exact discrete-AD proposal; negative values mean that a method happened to achieve a lower finite-radius loss. The local exact-gradient advantage is measurable but small, and it does not imply a multi-step performance ordering. Shading and denominators mark the strict common-itinerary subset.