CONVERGENCE ANALYSIS OF NORMALIZED GRADIENT DESCENT FOR DEFINABLE FUNCTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Normalized gradient descent (NGD) updates model parameters in the negative gradient direction with a prescribed step length. Its invariance to positive rescaling and its ability to move through regions with small gradients make it an attractive alternative to gradient descent. We study when diminishing stepsizes make NGD iterates bounded and convergent. For locally Lipschitz definable objectives and every positive vanishing nonsummable schedule, a bounded NGD trajectory has convergent objective values and only critical accumulation points. For a coercive definable loss, all trajectories of NGD from a bounded initial set are bounded when all the stepsizes are below a uniform threshold. We also prove that stepsizes $\alpha_k = c/(k+1)$, with $c > 0$, guarantee convergence of every bounded trajectory to a critical point for locally Lipschitz losses definable in a polynomially bounded o-minimal structure. This class includes squared-loss objectives for finite ReLU (Nair & Hinton, 2010) networks on finite datasets. We also show a sharp dependence on the stepsize schedule and dimension: for every $1/2 < p < 1$ and $c > 0$, a real-analytic semialgebraic objective in three dimensions admits a bounded nonconvergent NGD trajectory with $\alpha_k = c(k+1)^{-p}$. In two dimensions, every bounded NGD trajectory converges throughout $1/2 < p \leq 1$ for locally Lipschitz losses definable in a polynomially bounded o-minimal structure. Thus, even when the loss converges to a critical value and the distance to the set of critical points tends to zero, parameter convergence can depend on dimension and stepsize decay.

1 INTRODUCTION

Training a model involves choosing both an update direction and a step length. In gradient descent, the step length equals the learning rate multiplied by the gradient norm. Normalized gradient descent (NGD) follows the same negative gradient direction while prescribing the step length independently of the gradient norm:

$$x_{k+1} = x_k - \alpha_k \frac{\nabla f(x_k)}{\|\nabla f(x_k)\|}, \tag{1}$$

where x_k denotes the model parameters and f the training objective. Throughout, ∂f denotes the Clarke subdifferential of a locally Lipschitz function f (Clarke, 1990, Propositions 2.1.2 and 2.1.5). For a nonsmooth locally Lipschitz objective, we can choose a nonzero subgradient instead:

$$x_{k+1} = x_k - \alpha_k \frac{v_k}{\|v_k\|}, \qquad 0 \neq v_k \in \partial f(x_k). \tag{2}$$

When $\partial f(x_k) = \{0\}$, we set $x_{k+1} = x_k$. For simplicity, we will use NGD for both (1) and (2). Every nonzero update of NGD has length α_k, and the norm of the selected gradient or subgradient does not affect the step length. This can help NGD traverse nearly flat regions where small gradients slow GD. In a variational quantum optimization simulation for the H_2 molecule, (Suzuki et al., 2021, Section IV.B and Figure 3) report that NGD leaves a plateau at about iteration 50, compared with about iteration 450 for GD, with both methods using the constant learning-rate parameter $\eta = 0.05$.

A second motivation is implicit regularization: NGD can favor particular solutions among those that fit the training data. For separable multiclass linear classification with bounded features and cross-entropy loss, (Fan et al., 2025, Section 2 and Theorem 1) prove that, with a decreasing schedule $\eta_k = \Theta(k^{-1/2})$, the margin of the k-th NGD iterate divided by its norm converges to the largest

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.

normalized margin over all linear classifiers. In a local analysis near a manifold of minimizers, (Arora et al., 2022, Theorem 4.4) relate perturbed constant-step NGD to a flow that decreases sharpness, measured by the largest Hessian eigenvalue, as the stepsize tends to zero. (Josz, 2026a, Theorem 3.32 and Example 4.6) develops Lyapunov conditions for implicit regularization with sufficiently slowly diminishing stepsizes and proves attraction to flat minima in semialgebraic examples. These results connect the long-run dynamics of NGD to the geometry of learned solutions and motivate its study in relation to generalization.

These observations lead to two convergence questions: when are the iterates bounded, and when do they converge to one point? Existing convergence results for NGD are concentrated in more structured regimes. For convex and quasiconvex objectives, normalized subgradient schemes have classical guarantees for objective values and, under additional regularity and stepsize conditions, for the iterates (Kiwiel, 2001). In locally quasiconvex and stochastic settings, (Hazan et al., 2015) give finite-time best-iterate guarantees. For smooth strict-saddle objectives, (Levy, 2016) give finite-time guarantees for a noise-injected variant of NGD, while (Murray et al., 2019) analyze saddle avoidance and escape times for the continuous normalized flow. More recently, (Josz, 2026a) introduced a graph-closed normalized subdifferential and established discrete stability and attraction to selected flat minima under objective-dependent Lyapunov and regularity conditions. The present paper studies global boundedness and whole-sequence convergence for general nonconvex NGD trajectories.

A parallel literature studies discrete first-order dynamics for tame objectives without normalization. For bounded stochastic subgradient sequences, (Davis et al., 2020) establish asymptotic stationarity under their stepsize and noise assumptions. (Bolte et al., 2024, Theorem 5) prove objective-value convergence and criticality of every accumulation point for the subgradient method under positive non-increasing, vanishing, nonsummable steps. Their Remark 6 points to the asymptotic-pseudotrajectory framework of (Benaïm et al., 2005). For gradient descent, (Josz, 2023) relate small-step boundedness to boundedness of continuous gradient trajectories. (Josz & Lai, 2024) develop a global stability framework for constant-step methods that are approximated by subgradient trajectories of coercive tame functions. Most closely related to whole-sequence convergence, (Lai & Song, 2026a) prove that bounded subgradient sequences converge for decreasing steps of order $1/k$. Their subsequent work (Lai & Song, 2026b) gives an abstract diameter criterion for difference inclusions and applies it to inexact, momentum, and stochastic subgradient methods. In NGD, normalization introduces the state-dependent factor $1/\|v_k\|$, singular near critical points, and requires new stratified diameter estimates.

Taken together, the literature leaves open when general nonconvex NGD iterates are bounded and, conditional on boundedness, which stepsize schedules force convergence to one point. We address these questions for power-law stepsizes

$$\alpha_k = c(k+1)^{-p}, \qquad c > 0, \tag{3}$$

and relate the answers to the geometry of the objective.

Our setting includes nonconvex and nonsmooth training objectives. For point convergence, we consider locally Lipschitz functions definable in a polynomially bounded o-minimal structure (van den Dries & Miller, 1996). This means that every definable $g : (a, \infty) \to \mathbb{R}$ in this o-minimal structure satisfies $|g(t)| \leq t^N$ for all sufficiently large t and some $N \in \mathbb{N}$. This class includes semialgebraic losses and more general globally subanalytic losses, such as ReLU networks trained on finite datasets with squared loss and polynomial regularization (Davis et al., 2020, Section 5.2). Our theory also covers those activations whose bounded restriction is definable in a polynomially bounded o-minimal structure, which includes most practical activations such as sigmoid (Rumelhart et al., 1986), tanh (LeCun et al., 1998), softplus (Dugas et al., 2000), SiLU (Elfwing et al., 2018), ELU (Clevert et al., 2016), and exact GELU (Hendrycks & Gimpel, 2016).

Our contributions. We first prove a boundedness theorem based on the coercivity of an auxiliary function (Theorem 3.1). It includes every coercive definable locally Lipschitz loss and also covers noncoercive least-squares matrix factorization through the conserved quantity (Corollary 3.3).

For a locally Lipschitz definable objective definable and every positive vanishing nonsummable stepsizes schedule, each bounded NGD trajectory has convergent objective values and only critical accumulation points (Theorem 2.1). If the objective is definable in a polynomially bounded o-minimal structure, then every bounded trajectory with harmonic stepsizes $\alpha_k = c/(k+1)$ converges to a critical point (Theorem 2.3).

The convergence guarantee changes with dimension when the stepsizes decrease more slowly. In two dimensions, every bounded NGD trajectory converges for every locally Lipschitz loss definable in a polynomially bounded o-minimal structure under positive nonsummable, square-summable stepsizes (Theorem 2.4). In three dimensions, for every $1/2 < p < 1$ and $c > 0$, we construct a real-analytic semialgebraic objective with a bounded nonconvergent trajectory for (3) (Theorem 2.5). Along this trajectory the loss converges and the gradient norm tends to zero, while the parameters keep moving around a circle of critical points. Thus, diminishing steps and first-order stationarity can coexist without parameter convergence for NGD with square summable stepsizes and real analytic semialgebraic objective.

Table 1 summarizes the main guarantees and counterexamples. Section 6 establishes harmonic convergence for bounded trajectories of losses whose restriction to every bounded parameter box is definable in one common polynomially bounded o-minimal structure. It applies this result to quadratically regularized finite networks, proving boundedness and convergence for $\alpha_k = c/(k+1)$ with sufficiently small c, including ReLU, sigmoid, and exact GELU activations. Proofs of the main convergence theorems and the counterexample construction are in the appendices.

2 Setup and main results

Notation. $\|\cdot\|$ denotes the Euclidean norm, $\langle\cdot,\cdot\rangle$ denotes the Euclidean inner product, and $\|\cdot\|_F$ is the Frobenius norm. For nonempty sets $A, B \subseteq \mathbb{R}^n$, set

$$d(A,B) := \inf_{a\in A,\, b\in B} \|a-b\|, \qquad d(x,A) = \operatorname{dist}(x,A) := d(\{x\},A),$$

and let $B(A,r) := \{x \in \mathbb{R}^n : d(x,A) < r\}$. For a point a, $\overline{B}(a,r)$ denotes the corresponding closed ball. For a nonempty closed set $C \subseteq \mathbb{R}^n$,

$$P_C(x) := \operatorname*{argmin}_{y\in C} \|x-y\|$$

is its Euclidean projection mapping. If M is a C^1 embedded submanifold of $\mathbb{R}^n$ and $x \in M$, then $T_xM \subseteq \mathbb{R}^n$ is its tangent space. When $f|_M$ is C^1, $\nabla_M f(x)$ is its Riemannian gradient for the induced Euclidean metric on the manifold, and the rank of $f|_M$ means the rank of $D(f|_M)$. For $f \in C^1$, $\operatorname{Crit} f := \{x \in \mathbb{R}^n : \nabla f(x) = 0\}$. The class $C^{1,1}_{\mathrm{loc}}$ consists of C^1 functions with locally Lipschitz gradient. A function $V : \mathbb{R}^n \to \mathbb{R}$ is coercive if $V(x) \to +\infty$ as $\|x\| \to \infty$.

Concerning the definition of o-minimal structures we refer to (van den Dries & Miller, 1996). Here we only talk about the definition of the definability of mappings and functions. Definability should be understood as definability in a certain o-minimal structure, but for simplicity we will omit the o-minimal structure unless necessary. A set is said to be definable in an o-minimal structure if it belongs to that o-minimal structure. A set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is said to be definable, if its graph $\operatorname{gph} F := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m :\ y \in F(x)\}$ is definable. The definability of a function is defined similarly.

A point x is said to be critical if $0 \in \partial f(x)$. Define

$$\mathcal{N}_C f(x) = \begin{cases} \{v/\|v\| : 0 \neq v \in \partial f(x)\}, & \partial f(x) \neq \{0\}, \\ \{0\}, & \partial f(x) = \{0\}. \end{cases} \tag{4}$$

The NGD iteration (2) can equivalently be written as

$$x_{k+1} \in x_k - \alpha_k \mathcal{N}_C f(x_k). \tag{5}$$

For $f \in C^1$, (5) reduces to the smooth iteration (1). Write $\Omega(x) = \{z : x_{k_j} \to z \text{ for some } k_j \to \infty\}$ for the set of accumulation points of (x_k).

Theorem 2.1 (General value and stationarity convergence)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and definable. Let (x_k) be any bounded trajectory of* (5) *with*

$$\alpha_k > 0, \qquad \alpha_k \to 0, \qquad \sum_{k=0}^{\infty} \alpha_k = \infty. \tag{6}$$

Then, for some $\ell \in \mathbb{R}$,

$$f(x_k) \to \ell, \qquad \Omega(x) \subseteq \{x \in \mathbb{R}^n : 0 \in \partial f(x)\} \cap \{x \in \mathbb{R}^n : f(x) = \ell\}. \tag{7}$$

In particular, $\operatorname{dist}(x_k, \{x \in \mathbb{R}^n : 0 \in \partial f(x)\}) \to 0$.

Table 1: Stepsize and dimension results for bounded trajectories of definable objectives. Stepsizes are positive and $c > 0$. The first three rows use (5). The last row uses smooth NGD. The first row holds in every o-minimal structure. The second and third rows hold in every polynomially bounded o-minimal structure. The counterexamples in the last row are semialgebraic.

Regularity	Dimension	Stepsizes	Result
Locally Lipschitz	Any	$\alpha_k \to 0,\ \sum \alpha_k = \infty$	Values converge; accumulation points are critical
Locally Lipschitz	Any	$c/(k+1)$	Every trajectory converges
Locally Lipschitz	2	$\sum \alpha_k = \infty,\ \sum \alpha_k^2 < \infty$	Every trajectory converges
Analytic semialgebraic	$n \geq 3$	$c(k+1)^{-p},\ 1/2 < p < 1$	A bounded nonconvergent trajectory exists

Remark 2.2. Theorem 2.1 follows from two structural properties of f: the Clarke chain rule along absolutely continuous curves, and the weak Sard condition

$$\operatorname{int} f\big(\{x \in \mathbb{R}^n : 0 \in \partial f(x)\}\big) = \varnothing.$$

Consequently, the same conclusion holds for every locally Lipschitz function satisfying these two properties. Locally Lipschitz definable functions satisfy both (Bolte et al., 2007, Proposition 4 and Corollary 9). Thus Theorem 2.1 is valid in every o-minimal structure. The point-convergence theorems below use polynomial boundedness.

The next three theorems upgrade this limit-set statement to point convergence for harmonic steps, extend point convergence in the plane, and give counterexamples to point convergence in dimension three.

Theorem 2.3 (Convergence with harmonic stepsizes)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and definable in a polynomially bounded o-minimal structure. Let $c > 0$, and (x_k) be any bounded trajectory of* (5) *with $\alpha_k = c/(k+1)$. Then x_k converges to one point x^*, and $0 \in \partial f(x^*)$.*

Theorem 2.3 applies in every dimension. In $\mathbb{R}^2$, square summability is sufficient.

Theorem 2.4 (Planar convergence)**.** *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be locally Lipschitz and definable in a polynomially bounded o-minimal structure. Every bounded trajectory of* (5) *converges to a point x^* satisfying $0 \in \partial f(x^*)$ if*

$$\alpha_k > 0, \qquad \sum_k \alpha_k = \infty, \qquad \sum_k \alpha_k^2 < \infty. \tag{8}$$

In particular, convergence holds for (3) *whenever $1/2 < p \leq 1$.*

Theorem 2.5 (Analytic counterexample in $\mathbb{R}^3$)**.** *For every $p \in (1/2, 1)$ and $c > 0$, there exists a real-analytic semialgebraic function $F_{p,c} : \mathbb{R}^3 \to \mathbb{R}$ which admits a bounded nonstationary NGD orbit for $\alpha_k = c(k+1)^{-p}$ whose set of accumulation points is a circle.*

The conclusion extends to every dimension $n > 3$. For $z \in \mathbb{R}^3$ and $w \in \mathbb{R}^{n-3}$, set $\widetilde{F}_{p,c}(z, w) = F_{p,c}(z)$. This function is real-analytic and semialgebraic, and $\nabla \widetilde{F}_{p,c} = (\nabla F_{p,c}, 0)$, so adjoining any fixed w to the orbit in Theorem 2.5 preserves the NGD recursion and its set of accumulation points.

The proofs of these theorems are in Appendices B to D, respectively.

3 Boundedness of the iterates

For gradient descent, (Josz, 2023, Theorem 1) characterizes boundedness for all sufficiently small steps through boundedness of continuous gradient trajectories. For NGD, we use the coercive auxiliary function $G := f + \|I\|^2$, where I is constant along the continuous subgradient dynamics. The mapping I is usually called a conserved quantity (Josz, 2026a, Section 3.4). The next theorem covers the coercive f case (by setting $I \equiv 0$) and the case where a conserved quantity supplies the missing coercivity, as in (Josz, 2026a, Example 1.1).

Theorem 3.1 (Boundedness from a coercive auxiliary function). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and definable. Let $I : \mathbb{R}^n \to \mathbb{R}^s$ be continuously differentiable, and suppose that*

$$DI(x)v = 0 \qquad \big(x \in \mathbb{R}^n,\ v \in \partial f(x)\big). \tag{9}$$

Assume that

$$G(x) := f(x) + \|I(x)\|^2 \tag{10}$$

is coercive. For every nonempty bounded set $X_0 \subseteq \mathbb{R}^n$ and every $E \geq 0$, there are $\bar{\alpha} > 0$ and a compact set $K \subseteq \mathbb{R}^n$ such that every trajectory of (5) *satisfying*

$$x_0 \in X_0, \qquad 0 < \alpha_k \leq \bar{\alpha} \quad (k \geq 0), \qquad \sup_{k \geq 0} \|I(x_k) - I(x_0)\| \leq E$$

is contained in K.

In particular, taking $I \equiv 0$ and $E = 0$ covers every coercive definable locally Lipschitz objective. Moreover, if DI is globally Lipschitz with constant L_I, then the conservation identity (9) and $\sum_{k=0}^{\infty} \alpha_k^2 \leq S < \infty$ imply the drift bound with $E = L_I S/2$. Indeed, write $h_k = -\alpha_k v_k/\|v_k\|$ on a nonconstant step. Taylor's integral formula and (9) give

$$\|I(x_{k+1}) - I(x_k)\| = \left\| \int_0^1 \big(DI(x_k + t h_k) - DI(x_k)\big) h_k \, dt \right\| \leq \frac{L_I}{2} \alpha_k^2.$$

Thus square-summable steps give bounded discrete drift.

Remark 3.2. For GD on a differentiable definable objective with locally Lipschitz gradient, (Josz, 2023, Theorem 1) proves that boundedness of all continuous gradient trajectories is equivalent to the following discrete property: for every initial point, there is a positive step cap under which every variable-step GD trajectory is bounded. The implication from continuous to discrete boundedness fails for NGD. The smooth semialgebraic function $f(x, y) = y^2/(1+x^2)$ has bounded gradient trajectories but, from $(1, 1)$, admits an unbounded NGD trajectory with $\alpha_k \to 0$ below every prescribed positive step cap (Proposition A.2).

Corollary 3.3. *Fix $M \in \mathbb{R}^{m \times n}$ and an integer $r \geq 1$, and set*

$$f(X, Y) = \|XY^T - M\|_F^2, \qquad X \in \mathbb{R}^{m \times r}, \quad Y \in \mathbb{R}^{n \times r}. \tag{11}$$

For every bounded set Z_0 of initial pairs and every $E > 0$, there are $\bar{\alpha} > 0$ and a compact set K such that every joint Euclidean NGD trajectory satisfying

$$(X_0, Y_0) \in Z_0, \qquad 0 < \alpha_k \leq \bar{\alpha}, \qquad \sum_{k=0}^{\infty} \alpha_k^2 \leq E$$

is contained in K. Consequently, there is $\bar{c} > 0$ such that NGD with harmonic stepsizes as in Theorem 2.3 converges to a critical point for every $(X_0, Y_0) \in Z_0$ and $0 < c \leq \bar{c}$.

Proof. If $Z_0 = \varnothing$, take $\bar{\alpha} = \bar{c} = 1$ and $K = \{(0, 0)\}$. Suppose $Z_0 \neq \varnothing$ and set $I(X, Y) = X^T X - Y^T Y$ and $G = f + \|I\|_F^2$. For the same objective, Josz (2023, Example 1) establishes the balance identity $DI(X, Y)[\nabla_X f, \nabla_Y f] = 0$ and also proves the coercivity of G.

The quadratic map I has globally Lipschitz derivative with modulus L_I. The above argument then shows that $\sup_{k \geq 0} \|I(x_k) - I(x_0)\| \leq L_I E/2$. Consequently, Theorem 3.1 gives the first assertion. The conclusion for the harmonic stepsizes follows from Theorem 2.3. □

4 Point convergence of bounded trajectories

We sketch the main idea of the proofs of Theorems 2.3 and 2.4 in this section. For Theorem 2.3, we project consecutive NGD steps onto strata and control the resulting lengths. We construct an L'-regular stratification using (Lai & Song, 2026a, Proposition 3.2) such that f has constant rank on each stratum and every incident pair satisfies the strict Thom (w_f) condition (Lemma B.3). For a stratum M, the projection formula (Bolte et al., 2007, Proposition 4) gives

$$P_{T_y M} v = \nabla_M f(y), \qquad y \in M, \quad v \in \partial f(y). \tag{12}$$

Fix a stratum $M = M_i$ and a consecutive block $s, \dots, t$ such that $x_k \in N_i(\alpha_k)$ for every $s \le k \le t$. For sufficiently small $a > 0$, $N_i(a)$ consists of points x in a fixed compact ball containing the trajectory with $d(x, M_i) < c_i a^{\beta_i}$ and $d(x, M_j) \ge c_j a^{\gamma_j}$ for every frontier stratum $M_j \subseteq \overline{M}_i \setminus M_i$, where the constants are positive ((47)). The projection mapping P_M is also single valued there by Lemma B.5. Put $y_k = P_M x_k$ and $r_k = \|\nabla_M f(y_k)\|$. On a nonconstant step, choose $0 \ne v_k \in \partial f(x_k)$ with $x_{k+1} = x_k - \alpha_k v_k/\|v_k\|$, and set $q_k = \|v_k\|$.

If $f|_M$ has rank one, then $\nabla_M f(y) \ne 0$ for every $y \in M$, so $r_k > 0$. On each compact subset of M separated from its frontier $\overline{M} \setminus M$, continuity gives a positive minimum for $\|\nabla_M f\|$. The neighborhoods $N_i(\alpha_k)$ keep y_k at least a positive constant times a power of α_k away from this frontier and make the projected-gradient and projection-derivative errors a fixed fraction of r_k and the Taylor remainder a fixed fraction of $\alpha_k r_k^2/q_k$ (Lemmas B.4 and B.5). In particular, $\|P_{T_{y_k}M} v_k - \nabla_M f(y_k)\| \le r_k/10$. Projected descent and the projection estimate then give

$$\begin{aligned} f(y_{k+1}) - f(y_k) &\le -\frac{1}{2}\alpha_k r_k^2/q_k, \\ \|y_{k+1} - y_k\| &\le \frac{11}{5}\alpha_k r_k/q_k. \end{aligned} \tag{13}$$

Their common factor q_k^{-1} disappears when the projected length is compared with the objective decrease:

$$\|y_{k+1} - y_k\| \le \frac{22}{5r_k}\big(f(y_k) - f(y_{k+1})\big).$$

By Theorem 2.1, $f(x_k)$ converges to a critical value ℓ. Write $s_k = f(y_k) - \ell$. The signed Kurdyka–Łojasiewicz inequality on the finitely many rank-one strata gives $A > 0$ and $\theta \in (0, 1)$ such that $\psi(z) = A\,\mathrm{sgn}(z)|z|^{1-\theta}$ satisfies $\psi'(s_k) r_k \ge 12$ whenever $s_k \ne 0$ is sufficiently close to zero ((39)). The definition of $N_i(\alpha_k)$ gives $\|x_k - y_k\| < c_i \alpha_k^{\beta_i} \to 0$. Local Lipschitzness of f and $f(x_k) \to \ell$ therefore give $s_k \to 0$, so the inequality applies on sufficiently late blocks ((67)). Concavity of ψ on the positive half-line, convexity on the negative half-line, and $r_{k+1} \le 2r_k$ give

$$\psi(s_k) - \psi(s_{k+1}) \ge \begin{cases} 6\alpha_k r_k/q_k, & s_k > s_{k+1} \ge 0, \\ 3\alpha_k r_k/q_k, & 0 \ge s_k > s_{k+1}. \end{cases}$$

Both bounds exceed the projected length in (13). A decreasing block crosses ℓ at most once, and that step has projected length at most $2\alpha_k$. Thus Lemma B.6 gives

$$\sum_{k=s}^{t-1} \|y_{k+1} - y_k\| \le \psi(f(y_s) - \ell) - \psi(f(y_t) - \ell) + 2 \max_{s \le k < t} \alpha_k.$$

If $f|_M$ has rank zero, then $\nabla_M f = 0$ on M. Strict Thom regularity and the scale-dependent neighborhoods bound the tangential component of the normalized direction, and the projection estimate gives

$$\left\| P_{T_{y_k}M} \frac{v_k}{q_k} \right\| \le C\alpha_k^{\delta_M}, \qquad \|y_{k+1} - y_k\| \le 2C\alpha_k^{1+\delta_M}, \qquad \delta_M > 0. \tag{14}$$

Since M is connected and $\nabla_M f = 0$, the restriction $f|_M$ is constant. Summing the projected-length bound gives Lemma B.7. Harmonic steps satisfy $\sum_k \alpha_k^{1+\delta_M} < \infty$ for every $\delta_M > 0$. Finiteness of the stratification gives a common $\delta > 0$. The two block bounds, together with the separation of the scale-dependent neighborhoods, verify the stratified-descent conditions of (Lai & Song, 2026b, Definition 2.3) for $\mathcal{N}_C f$. The diameter estimate (Lai & Song, 2026b, Theorem 2.4) bounds each late trajectory segment by a ψ-drop, a tail of $\sum_k \alpha_k^{1+\delta}$, and a positive power of its first stepsize ((72)). The resulting tail diameter tends to zero uniformly over terminal indices because $\psi(f(x_k) - \ell) \to 0$, $\sum_k \alpha_k^{1+\delta} < \infty$, and $\alpha_k \to 0$. Hence (x_k) is Cauchy and converges to the critical point given by Theorem 2.1; Appendix B gives the complete argument.

For Theorem 2.4, the accumulation set is connected (Bolte et al., 2014, Lemma 3.5(ii)–(iii) and Remark 3.3) and consists of critical points by (7). If it contains more than one point, it must meet a one-dimensional critical stratum S. The zero-dimensional strata contain only finitely many points,

while an accumulation point in a two-dimensional critical stratum would force an iterate to enter that open stratum and stop.

Choose an accumulation point $x^* \in S$ and a short arc around it. This arc separates a thin neighborhood into two connected sides, each lying in one two-dimensional stratum. Let P be the projection mapping onto S. On a side where the gradient is nonzero, strict Thom regularity bounds the component of the normalized gradient along $T_{P(x)}S$ by $C\,\mathrm{dist}(x, S)$, as in (81). Thus the NGD direction is almost perpendicular to the curve. It varies continuously on each such side, and its normal component stays away from zero. It therefore points either toward S throughout that side or away from S throughout it.

Suppose that arbitrarily late iterates near x^* enter a side where the NGD direction points away from S. Their distance to S then increases by at least a positive multiple of α_k per step. Since $\sum_k \alpha_k = \infty$, they must reach a positive distance from S. The small movement along the curve keeps their projections within a compact arc until then. This would produce iterates a positive distance from the critical set at arbitrarily late times, contradicting (75).

Consequently, sufficiently late iterates near x^* lie on S or on a side where NGD moves toward S. Put $d_k = \mathrm{dist}(x_k, S)$. For these local steps, including steps starting on S, Appendix C gives constants $\nu, C > 0$ such that

$$d_{k+1}^2 \le d_k^2 - \nu\alpha_k d_k + C\alpha_k^2, \qquad \|P(x_{k+1}) - P(x_k)\| \le C(\alpha_k d_k + \alpha_k^2).$$

On any segment $[m, n]$ that stays in this neighborhood, summation bounds both $\max_{m\le j\le n} d_j^2$ and the total movement of the projected iterates by $C(d_m^2 + \sum_{k=m}^{\infty} \alpha_k^2)$. This bound becomes small when m is large and x_m is close to x^*. It controls both the movement along the curve and the distance from it, so the iterates cannot travel far enough to leave the neighborhood. The estimates then apply to the entire tail. Along a subsequence $x_{m_j} \to x^*$, the bound tends to zero, making all iterates after m_j arbitrarily close to one another. Hence the trajectory converges.

5 A BOUNDED NONCONVERGENT TRAJECTORY FOR AN ANALYTIC OBJECTIVE

In this section, we give some intuition on the counterexample constructed in Appendix D. Our idea is to first design the motion of a bounded nonconvergent trajectory and then construct a function whose NGD iterates follow this motion. For $c = 1$, we work near the circle $\mathcal{C} = \{(x_1, x_2, 0) : x_1^2 + x_2^2 = 1\} \subseteq \mathbb{R}^3$, whose normal plane is two-dimensional. The construction behind Theorem 2.5 uses this plane for an eight-step loop: at each fixed angle on $\mathcal{C}$, the leading normal directions in the last four steps are the opposites of those in the first four. Their leading displacements therefore cancel over a block. At the same time, the gradient also produces a smaller but nonsummable displacement along $\mathcal{C}$ after each loop, which keeps the iterates moving intead of converging to a single point. Figure 1 shows the normal loop and the resulting motion around $\mathcal{C}$.

For $\alpha_k = (k+1)^{-p}$ with $1/2 < p < 1$, choose an integer $m \ge \max\{3, p/(1-p)\}$ and set $\tau_k = (k+1)^{-p/m}$. The orbit stays at distance comparable to τ_k from $\mathcal{C}$, while one normal loop has diameter of order $\alpha_k = \tau_k^m$. The angular gain of each eight-step block is comparable to $\alpha_k\tau_k$. The scale relation is

$$\tau_k \to 0, \qquad \sum_{k\equiv 0 \pmod 8} \alpha_k\tau_k = \sum_{k\equiv 0 \pmod 8} (k+1)^{-p(1+1/m)} = \infty,$$

because $p(1 + 1/m) \le 1$. Thus the orbit approaches the critical circle while its accumulated rotation is unbounded. The block rotations tend to zero, so every point of the circle is a limit point.

The appendix realizes these directions as gradients of a real-analytic semialgebraic function and proves that an exact NGD orbit stays in the narrowing region shown in the figure. For $c > 0$ it suffcies to rescale the function.

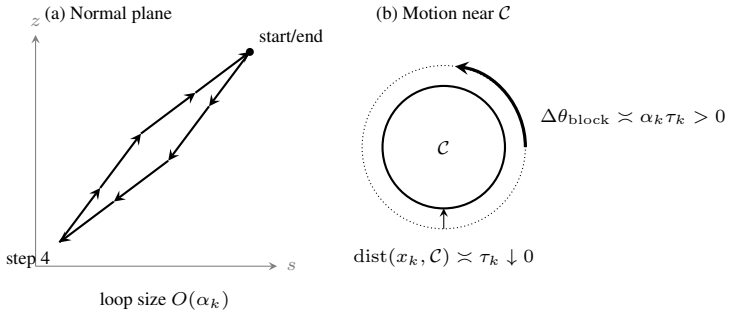


Figure 1: The eight-step mechanism. The transverse directions close a loop at leading order; a smaller positive angular gain remains after each block. The transverse loop is enlarged in panel (a).

6 Learning implications and discussion

In this section, we apply the convergence results to practical learning problems. Theorem 2.3 assumes that the objective, as a function of all trainable parameters, is definable in a polynomially bounded o-minimal structure. Common smooth activations can violate this global condition. For example, the unrestricted sigmoid $\sigma(t) = (1 + e^{-t})^{-1}$ defines $e^t = \sigma(t)/(1 - \sigma(t))$ through its graph and therefore does not belong to any polynomially bounded o-minimal structure.

For a bounded NGD trajectory with harmonic stepsizes, the subdifferential at each iterate depends on the loss in a neighborhood of that iterate. It therefore suffices for its point convergence that the complete loss be locally Lipschitz and that its restriction to every bounded parameter box be definable in one common polynomially bounded o-minimal structure $\mathcal{R}$. The next proposition uses a cutoff to produce a globally $\mathcal{R}$-definable loss agreeing with the original loss near a bounded trajectory.

Proposition 6.1 (Harmonic convergence from local definability)**.** *Theorem 2.3 remains valid when its definability hypothesis is replaced by the condition that $f|_{[-M,M]^n}$ is definable in one common polynomially bounded o-minimal structure for every $M > 0$.*

Proof. Let $\mathcal{R}$ denote the structure in the hypothesis. Fix such a trajectory and choose $M > 0$ with $w_k \in (-M/2, M/2)^n$ for every k. Define the semialgebraic Lipschitz cutoff function

$$\rho_M(t) = \max\{0, \min\{1, 2 - |t|/M\}\}, \qquad \chi(w) = \prod_{i=1}^{n} \rho_M(w_i),$$

and define $\widetilde{f}(w) = \chi(w) f(w)$. Then, in $(-M, M)^n$, f and $\tilde{f}$ coincide, and so do their subdifferentials. It is now sufficient to apply Theorem 2.3 to the o-minimal structure $\mathcal{R}$. □

The next result extends the above argument to finite feed-forward networks.

Corollary 6.2 (Regularized finite networks)**.** *Let h_w be a finite feed-forward network with an affine output layer, where $w \in \mathbb{R}^d$ collects every trainable weight and bias. Suppose its activations are locally Lipschitz and definable in one common o-minimal structure $\mathcal{S}$. Assume there is a polynomially bounded o-minimal structure $\mathcal{R}$ in which the restriction of every activation to every bounded interval is $\mathcal{R}$-definable. For a finite dataset $\{(z_i, y_i)\}_{i=1}^N$, $N \geq 1$, and $\lambda > 0$, set*

$$L_\lambda(w) = \frac{1}{N} \sum_{i=1}^{N} \|h_w(z_i) - y_i\|^2 + \lambda \|w\|^2. \tag{15}$$

For every nonempty bounded set $W_0 \subseteq \mathbb{R}^d$, there is $c_0 > 0$ such that, for every $w_0 \in W_0$ and $0 < c \leq c_0$, every trajectory (w_k) of (5) with $f = L_\lambda$ and $\alpha_k = c/(k + 1)$ converges to w^* with $0 \in \partial L_\lambda(w^*)$.

Proof. The activation hypotheses make L_λ locally Lipschitz and $\mathcal{S}$-definable. Since $L_\lambda(w) \geq \lambda \|w\|^2$, it is coercive.

Fix $M > 0$ and $B = [-M, M]^d$. For each training input z_i, factor $w \mapsto h_w(z_i)$ as $g_L \circ \cdots \circ g_1$, where the layer maps carry w and the current outputs. The images of B under the continuous partial compositions are compact. Choose a bounded box B_ℓ containing each image, with $B_0 = B$. On $B_{\ell-1}$, the weights, biases, and previous outputs are bounded, so each preactivation $\sum_j w_j a_j + b$ has bounded range. The assumed activation restrictions make $g_\ell|_{B_{\ell-1}}$ $\mathcal{R}$-definable. Finite composition, summation, squaring, and square regularization then make $L_\lambda|_B$ $\mathcal{R}$-definable. With $I \equiv 0$, $E = 0$, and $X_0 = W_0$, Theorem 3.1 gives a cap $c_0 > 0$ and a common compact bound for the trajectories. For $c \le c_0$, the harmonic steps satisfy $\alpha_k \le c_0$, so Proposition 6.1 gives convergence to a critical point. □

ReLU, leaky ReLU (Maas et al., 2013), hard-tanh (Collobert et al., 2011), and continuous finite piecewise-polynomial activations are semialgebraic and locally Lipschitz. Sigmoid, tanh, softplus, SiLU, and ELU are globally definable in the o-minimal structure $\mathbb{R}_{\mathrm{an,exp}}$ (van den Dries & Miller, 1994). Their graphs over bounded intervals are bounded and subanalytic, hence globally subanalytic (Bolte et al., 2006, Section 2.1). Globally subanalytic sets form the polynomially bounded o-minimal structure $\mathbb{R}_{\mathrm{an}}$ (Rainer, 2025, Section 6.1). Exact GELU $t \mapsto t\Phi(t)$, where Φ is the Gaussian distribution function, is definable in the o-minimal Pfaffian structure $\mathbb{R}_{\mathrm{Pfaff}}$ (Kranz et al., 2025, Theorem 2.3 and Example 2.4(v)). The Gaussian $g(t) = e^{-t^2/2}$ is an entire function, and hence $\Phi(t) = 1/2 + (2\pi)^{-1/2} \int_0^t g(u)\, du$ is also an entire function. Therefore, its restriction on every bounded set is globally subanalytic, and hence belongs to the polynomially bounded o-minimal structure $\mathbb{R}_{\mathrm{an}}$. These activations satisfy Corollary 6.2 with the common polynomially bounded structure $\mathbb{R}_{\mathrm{an}}$ (van den Dries & Miller, 1996) on every bounded interval.

7 Conclusion and limitations

We established conditions for boundedness and convergence of NGD with diminishing stepsizes. For locally Lipschitz objectives definable in a polynomially bounded o-minimal structure, harmonic steps guarantee convergence of every bounded trajectory to a stationary point. The planar convergence result and the three-dimensional analytic counterexamples show that the stepsize conditions for point convergence depend on dimension. In particular, the harmonic schedule is sharp as a uniform worst-case guarantee among nonsummable power-law schedules in dimensions $n \ge 3$.

Our point-convergence guarantees characterize the limit through Clarke stationarity. Experiments by Arora et al. (2022, Section 6) show NGD moving toward regions of lower sharpness, suggesting an implicit preference for flatter solutions. A natural next step is to determine which types of stationary points are selected by NGD and under what conditions its iterates converge to flat local minima. This would connect our convergence results with the implicit regularization observed empirically.

AI USE STATEMENT

The authors formulated the research questions and directed the study. Generative AI tools assisted in finding relevant literature, drafting and polishing portions of the manuscript, and developing mathematical proofs. In particular, an AI assistant constructed the three-dimensional analytic counterexample in Theorem 2.5 and supplied its proof. The authors reviewed and revised the AI-assisted arguments, checked the final proofs and cited references, and take responsibility for the paper.

ETHICS STATEMENT

This work develops theoretical convergence results for normalized gradient descent. It involves no human subjects, personal data, or experiments, and we do not identify an immediate ethical risk arising from the stated results.

REPRODUCIBILITY STATEMENT

The main theorem statements specify their assumptions, update rules, and conclusions. Appendices A–D contain the proofs of the main theorems, the auxiliary boundedness results, and the explicit constructions and verifications of the counterexamples.

REFERENCES

Charalambos D. Aliprantis and Kim C. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, 3rd edition, 2006. doi: 10.1007/3-540-29587-9.

Sanjeev Arora, Zhiyuan Li, and Abhishek Panigrahi. Understanding gradient descent on the edge of stability in deep learning. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 948–1024. PMLR, 2022.

Michel Benaïm, Josef Hofbauer, and Sylvain Sorin. Stochastic approximations and differential inclusions. *SIAM Journal on Control and Optimization*, 44(1):328–348, 2005. doi: 10.1137/S0363012904439301.

Jacek Bochnak, Michel Coste, and Marie-Françoise Roy. *Real Algebraic Geometry*, volume 36 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer, 1998. doi: 10.1007/978-3-662-03718-8.

Jérôme Bolte, Aris Daniilidis, and Adrian S. Lewis. A nonsmooth Morse–Sard theorem for subanalytic functions. *Journal of Mathematical Analysis and Applications*, 321(2):729–740, 2006. doi: 10.1016/j.jmaa.2005.07.068.

Jérôme Bolte, Aris Daniilidis, Adrian S. Lewis, and Masahiro Shiota. Clarke subgradients of stratifiable functions. *SIAM Journal on Optimization*, 18(2):556–572, 2007. doi: 10.1137/060670080.

Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1–2):459–494, 2014. doi: 10.1007/s10107-013-0701-9.

Jérôme Bolte, Edouard Pauwels, and Rodolfo Ríos-Zertuche. Long term dynamics of the subgradient method for Lipschitz path differentiable functions. *Journal of the European Mathematical Society*, 26(7):2533–2563, 2024. doi: 10.4171/JEMS/1285.

Frank H. Clarke. *Optimization and Nonsmooth Analysis*, volume 5 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1990. doi: 10.1137/1.9781611971309.

Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (ELUs). In *International Conference on Learning Representations*, 2016.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12 (76):2493–2537, 2011.

Michel Coste. An introduction to o-minimal geometry, 1999. RAAG Notes, Institut de Recherche Mathématique de Rennes.

Damek Davis, Dmitriy Drusvyatskiy, Sham Kakade, and Jason D. Lee. Stochastic subgradient method converges on tame functions. *Foundations of Computational Mathematics*, 20(1):119–154, 2020. doi: 10.1007/s10208-018-09409-5.

Maciej P. Denkowski and Paulina Pełszyńska. On definable multifunctions and Łojasiewicz inequalities. *Journal of Mathematical Analysis and Applications*, 456(2):1101–1122, 2017. doi: 10.1016/j.jmaa.2017.07.019.

Ewa Dudek and Konstanty Holly. Nonlinear orthogonal projection. *Annales Polonici Mathematici*, 59(1):1–31, 1994. doi: 10.4064/ap-59-1-1-31.

Charles Dugas, Yoshua Bengio, François Bélisle, Claude Nadeau, and René Garcia. Incorporating second-order functional knowledge for better option pricing. In *Advances in Neural Information Processing Systems*, volume 13, pp. 472–478, 2000.

Stefan Elfwing, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks*, 107:3–11, 2018. doi: 10.1016/j.neunet.2017.12.012.

Chen Fan, Mark Schmidt, and Christos Thrampoulidis. Implicit bias of spectral descent and Muon on multiclass separable data. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-1322.

Albert Fathi. Weak KAM theorem in Lagrangian dynamics. Preliminary Version 10, 2008. URL `https://www.math.u-bordeaux.fr/~pthieull/Recherche/KamFaible/Publications/Fathi2008_01.pdf`.

Elad Hazan, Kfir Y. Levy, and Shai Shalev-Shwartz. Beyond convexity: Stochastic quasi-convex optimization. In *Advances in Neural Information Processing Systems*, volume 28, pp. 1594–1602, 2015.

Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (GELUs), 2016. arXiv:1606.08415.

Ziwei Ji and Matus J. Telgarsky. Directional convergence and alignment in deep learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 17176–17186, 2020.

Cédric Josz. Global convergence of the gradient method for functions definable in o-minimal structures. *Mathematical Programming*, 202:355–383, 2023. doi: 10.1007/s10107-023-01937-5.

Cédric Josz. Implicit regularization of normalized gradient descent, 2026a. arXiv:2602.08177v1.

Cédric Josz. Lyapunov stability of the Euler method, 2026b. arXiv:2509.11415v3.

Cédric Josz and Lexiao Lai. Global stability of first-order methods for coercive tame functions. *Mathematical Programming*, 207(1–2):551–576, 2024. doi: 10.1007/s10107-023-02020-9.

Krzysztof C. Kiwiel. Convergence and efficiency of subgradient methods for quasiconvex minimization. *Mathematical Programming*, 90(1):1–25, 2001. doi: 10.1007/PL00011414.

Julian Kranz, Davide Gallon, Steffen Dereich, and Arnulf Jentzen. SAD neural networks: Divergent gradient flows and asymptotic optimality via o-minimal structures. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-0138.

Lexiao Lai and Mingzhi Song. On the diameter of subgradient sequences in o-minimal structures, 2026a. arXiv:2511.06868v3.

Lexiao Lai and Mingzhi Song. Convergence of difference inclusions: A diameter criterion and step-size conditions, 2026b. arXiv:2605.14345v2.

Ta Lê Loi. Verdier and strict Thom stratifications in o-minimal structures. *Illinois Journal of Mathematics*, 42(2):347–356, 1998. doi: 10.1215/ijm/1256045049.

Ta Lê Loi. Stratification of families of functions definable in o-minimal structures. *Acta Mathematica Vietnamica*, 27(2):239–244, 2002.

Yann LeCun, Léon Bottou, Genevieve B. Orr, and Klaus-Robert Müller. Efficient BackProp. In Genevieve B. Orr and Klaus-Robert Müller (eds.), *Neural Networks: Tricks of the Trade*, volume 1524 of *Lecture Notes in Computer Science*, pp. 9–50. Springer, 1998. doi: 10.1007/3-540-49430-8_2.

John M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, 2nd edition, 2013. doi: 10.1007/978-1-4419-9982-5.

Kfir Y. Levy. The power of normalization: Faster evasion of saddle points, 2016. arXiv:1611.04831.

Andrew L. Maas, Awni Y. Hannun, and Andrew Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *ICML Workshop on Deep Learning for Audio, Speech, and Language Processing*, 2013.

John N. Mather. Notes on topological stability. *Bulletin of the American Mathematical Society*, 49 (4):475–506, 2012. doi: 10.1090/S0273-0979-2012-01383-6.

Ryan Murray, Brian Swenson, and Soummya Kar. Revisiting normalized gradient descent: Fast evasion of saddle points. *IEEE Transactions on Automatic Control*, 64(11):4818–4824, 2019. doi: 10.1109/TAC.2019.2914998.

Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning*, pp. 807–814, 2010.

Ulrich Pinkall and Oliver Gross. *Differential Geometry: From Elastic Curves to Willmore Surfaces*. Compact Textbooks in Mathematics. Birkhäuser, Cham, 2024. doi: 10.1007/978-3-031-39838-4.

Armin Rainer. Arc-smooth functions and cuspidality of sets. *Journal d'Analyse Mathématique*, 155 (1):1–41, 2025. doi: 10.1007/s11854-024-0337-0.

R. Tyrrell Rockafellar and Roger J. B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer, 1998. doi: 10.1007/978-3-642-02431-3.

David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. doi: 10.1038/323533a0.

Yudai Suzuki, Hiroshi Yano, Rudy Raymond, and Naoki Yamamoto. Normalized gradient descent for variational quantum algorithms. In *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pp. 1–9. IEEE, 2021. doi: 10.1109/QCE52317.2021.00015.

Lou van den Dries and Chris Miller. On the real exponential field with restricted analytic functions. *Israel Journal of Mathematics*, 85(1–3):19–56, 1994. doi: 10.1007/BF02758635.

Lou van den Dries and Chris Miller. Geometric categories and o-minimal structures. *Duke Mathematical Journal*, 84(2):497–540, 1996. doi: 10.1215/S0012-7094-96-08416-1.

A Boundedness and convergence from arbitrary initialization

This appendix proves Theorem 2.1 and Theorem 3.1.

A.1 A GENERAL LIMIT-SET THEOREM

We first construct the differential inclusion used both for Theorem 2.1 and for the boundedness result. At a critical point, we insert the zero velocity before taking the graph closure:

$$\mathcal{A}_C(x) := -\mathcal{N}_C f(x) \cup \begin{cases} \{0\}, & 0 \in \partial f(x), \\ \varnothing, & 0 \notin \partial f(x). \end{cases} \tag{16}$$

Define its graph-closed convexification by

$$\mathcal{F}(x) := \operatorname{co}\big\{u \in \mathbb{R}^n : (x,u) \in \overline{\operatorname{gph} \mathcal{A}_C}\big\}. \tag{17}$$

For every critical point x, the inserted zero belongs to $\mathcal{F}(x)$, so $z(t) \equiv x$ solves $\dot{z} \in \mathcal{F}(z)$.

Comparison with the normalized subdifferential of (Josz, 2026a). Let $\mathcal{D}_f$ be the differentiability set of f. (Josz, 2026a, Definition 3.1) first defines

$$\widetilde{\nabla} f(x) = \begin{cases} \{\nabla f(x)/\|\nabla f(x)\|\}, & x \in \mathcal{D}_f,\ \nabla f(x) \neq 0, \\ \{0\}, & f \text{ is constant near } x, \\ \varnothing, & \text{otherwise}, \end{cases} \qquad \operatorname{gph} \widehat{\nabla} f = \overline{\operatorname{gph} \widetilde{\nabla} f}.$$

Thus $\widehat{\nabla} f$ normalizes differentiable gradients before taking the graph closure. When zero is absent from the Bouligand subdifferential, (Josz, 2026a, Lemma 3.3) shows that $\widehat{\nabla} f$ coincides with the normalized Bouligand subgradients. In contrast, $\mathcal{N}_C f$ normalizes every nonzero vector after the Clarke convexification, and $\mathcal{A}_C$ additionally inserts zero at every critical point. Since $\operatorname{gph}(-\widetilde{\nabla} f) \subseteq \operatorname{gph} \mathcal{A}_C$, we have

$$-\operatorname{co} \widehat{\nabla} f(x) \subseteq \mathcal{F}(x).$$

Every admissible choice $0 \neq g_k \in \partial f(x_k)$ generates a velocity $-g_k/\|g_k\| \in \mathcal{A}_C(x_k)$. The smaller field $-\widehat{\nabla} f(x_k)$ can exclude this velocity. The inclusion can be strict even at a noncritical point. For $f(s,t) = s + |t|$ at the origin,

$$\widehat{\nabla} f(0,0) = \{2^{-1/2}(1,1), 2^{-1/2}(1,-1)\}, \qquad (1,0) \in \mathcal{N}_C f(0,0),$$

so $(-1,0) \in \mathcal{F}(0,0)$ but $(-1,0) \notin -\operatorname{co} \widehat{\nabla} f(0,0)$. Accordingly, our analysis uses $\mathcal{N}_C f$, which contains all directions admitted by (2).

We next show that $\mathcal{F}$ is definable. Put

$$G := \operatorname{gph} \partial f = \{(x,v) \in \mathbb{R}^n \times \mathbb{R}^n : v \in \partial f(x)\}.$$

The set G is definable by (Ji & Telgarsky, 2020, Supplementary Material, Lemma B.8). In terms of this graph,

$$\begin{aligned} \operatorname{gph} \mathcal{N}_C f = \{(x,u) \in \mathbb{R}^n \times \mathbb{R}^n : \|u\| = 1,\ \exists t > 0 \text{ s.t. } (x,tu) \in G\} \\ \cup \{(x,0) \in \mathbb{R}^n \times \mathbb{R}^n : \nexists v \neq 0 \text{ s.t. } (x,v) \in G\}. \end{aligned}$$

By (Coste, 1999, Theorem 1.13), the set satisfying any first-order formula in the language of the o-minimal structure is definable in that o-minimal structure. Since the Euclidean norm is semialgebraic, the above formula therefore shows that $\operatorname{gph} \mathcal{N}_C f$ is definable. Consequently

$$\operatorname{gph} \mathcal{A}_C = \{(x,-u) \in \mathbb{R}^n \times \mathbb{R}^n : (x,u) \in \operatorname{gph} \mathcal{N}_C f\} \cup \{(x,0) \in \mathbb{R}^n \times \mathbb{R}^n : (x,0) \in G\}$$

is definable. (Coste, 1999, Proposition 1.12) then shows that

$$H := \overline{\operatorname{gph} \mathcal{A}_C}$$

is definable. For $x \in \mathbb{R}^n$, write $H_x := \{h \in \mathbb{R}^n : (x,h) \in H\}$. The Carathéodory argument in (Denkowski & Pełszyńska, 2017, Section 7) applies to each section H_x: every point of $\operatorname{co} H_x$ is a convex combination of at most $n+1$ points of H_x. More precisely,

$$\begin{aligned} (x,w) \in \operatorname{gph} \mathcal{F} \quad \Longleftrightarrow \quad & \exists h_0, \dots, h_n \in \mathbb{R}^n, \quad \exists \lambda_0, \dots, \lambda_n \in \mathbb{R} \text{ such that} \\ & (x,h_i) \in H \quad (0 \le i \le n), \\ & \lambda_i \ge 0 \quad (0 \le i \le n), \qquad \sum_{i=0}^{n} \lambda_i = 1, \qquad w = \sum_{i=0}^{n} \lambda_i h_i. \end{aligned}$$

This formula defines $\mathcal{F}$ in the same structure as f.

We next record the topological properties of $\mathcal{F}$. By (Clarke, 1990, Proposition 2.1.2(a)) and the assumption that f is locally Lipschitz, we know that ∂f is locally bounded with nonempty compact convex values. Hence $\mathcal{A}_C(x) \neq \varnothing$ for every x, all of its values are contained in $\overline{B}(0,1)$. Therefore, the same holds for $\mathcal{F}$, that is

$$\sup_{u \in \mathcal{F}(x)} \|u\| \leq 1, \qquad -\mathcal{N}_C f(x) \subseteq \mathcal{F}(x), \tag{18}$$

Therefore, the closure of $\mathcal{A}_C$ is locally bounded, and hence upper semicontinuous by (Rockafellar & Wets, 1998, Theorem 5.19). Thus, $\mathcal{F}$ is also nonempty compact convex valued and upper semicontinuous by (Aliprantis & Border, 2006, Theorem 17.35). Moreover,

$$0 \in \mathcal{F}(x) \quad \Longleftrightarrow \quad 0 \in \partial f(x). \tag{19}$$

The implication $0 \in \partial f(x) \implies 0 \in \mathcal{F}(x)$ follows from the zero insertion in (16). We next prove the implication $0 \in \mathcal{F}(x) \implies 0 \in \partial f(x)$. Assume that $0 \notin \partial f(x)$ and set $p = P_{\partial f(x)}(0)$. The characterization of the Euclidean projection onto the closed convex set $\partial f(x)$ gives

$$\langle 0 - p, v - p \rangle \leq 0, \qquad v \in \partial f(x),$$

or equivalently,

$$\langle p, v \rangle \geq \|p\|^2, \qquad v \in \partial f(x).$$

The open set

$$V := \left\{ v \in \mathbb{R}^n : \langle p, v \rangle > \tfrac{1}{2}\|p\|^2 \right\}$$

contains $\partial f(x)$ by the preceding projection inequality. Upper semicontinuity of the subdifferential (Clarke, 1990, Proposition 2.1.5(d)) therefore gives a neighborhood U of x such that

$$\partial f(y) \subseteq V \qquad \text{for every } y \in U.$$

Local boundedness of ∂f shows that we may assume that

$$\|v\| \leq L \qquad \text{for every } y \in U \text{ and every } v \in \partial f(y).$$

Set $\gamma := \|p\|^2/(2L) > 0$. If $y \in U$ and $v \in \partial f(y)$, then the definition of V yields $\langle p, v \rangle > \|p\|^2/2 > 0$. In particular, $v \neq 0$ and $0 \notin \partial f(y)$. Hence $\mathcal{A}_C(y) = -\mathcal{N}_C f(y)$, and every $u \in \mathcal{A}_C(y)$ has the form $u = -v/\|v\|$ for some such v. Together with the local bound $\|v\| \leq L$, this gives

$$\langle p, u \rangle = -\frac{\langle p, v \rangle}{\|v\|} \leq -\frac{\|p\|^2}{2L} =: -\gamma.$$

Now let $h \in H_x$, where $H = \overline{\operatorname{gph} \mathcal{A}_C}$. By the definition of graph closure, there are sequences $y_j \to x$ and $u_j \to h$ with $u_j \in \mathcal{A}_C(y_j)$. For all sufficiently large j, one has $y_j \in U$, and hence the preceding bound gives $\langle p, u_j \rangle \leq -\gamma$. Passing to the limit yields

$$\langle p, h \rangle \leq -\gamma, \qquad h \in H_x.$$

Finally, each $w \in \mathcal{F}(x) = \operatorname{co} H_x$ is a finite convex combination $w = \sum_{i=1}^q \lambda_i h_i$, where $h_i \in H_x$, $\lambda_i \geq 0$, and $\sum_{i=1}^q \lambda_i = 1$, and hence $\langle p, w \rangle \leq -\gamma$. Thus $0 \notin \mathcal{F}(x)$. This yields contradiction and completes the proof of (19).

We next turn to the proof of Theorem 2.1. The main idea is to compare the affine interpolation of the discrete dynamics with the continuous flow $\dot{z} \in \mathcal{F}(z)$. The following lemma supplies the descent property of the continuous flow.

Lemma A.1. *Every absolutely continuous solution of $\dot{z}(t) \in \mathcal{F}(z(t))$ satisfies $f(z(t_2)) \leq f(z(t_1))$ for $t_2 \geq t_1$. More precisely, if $Q \subseteq \mathbb{R}^n$ is nonempty and compact and*

$$L_Q := \max\{1, \sup\{\|v\| : x \in \mathbb{R}^n,\ d(x,Q) \leq 1,\ v \in \partial f(x)\}\}, \qquad m(x) := d(0, \partial f(x)),$$

then

$$(f \circ z)'(t) \leq -\frac{m(z(t))^2}{L_Q} \quad \textit{for a.e. } t \textit{ such that } z(t) \in Q. \tag{20}$$

Proof. Notice that L_Q is finite by the local boundedness of ∂f. Fix $x \in Q$ with $m(x) > 0$, and let

$$C_x := \partial f(x), \qquad p := P_{C_x}(0).$$

The set C_x is nonempty, compact, and convex, so p is uniquely defined and

$$\|p\| = d(0, C_x) = m(x).$$

Fix $h \in H_x$. By the definition of closure, there are sequences $x_j \to x$ and $u_j \to h$ such that $u_j \in \mathcal{A}_C(x_j)$ for every j. Since $0 \notin \partial f(x)$, we know that $0 \notin \partial f(x_j)$ for all sufficiently large j due to the upper semicontinuity of ∂f. After discarding finitely many indices, we therefore have

$$\mathcal{A}_C(x_j) = -\mathcal{N}_C f(x_j), \qquad u_j = -\frac{v_j}{\|v_j\|}$$

for some $0 \neq v_j \in \partial f(x_j)$.

Since $x_j \to x \in Q$, we have $d(x_j, Q) \leq 1$ for all sufficiently large j, and the definition of L_Q gives $\|v_j\| \leq L_Q$. Passing to a subsequence we may assume $v_j \to v$. The closed graph of ∂f gives $v \in \partial f(x) = C_x$. Consequently,

$$\|v\| \geq d(0, C_x) = m(x) > 0, \qquad \|v\| \leq L_Q,$$

and continuity of $v \mapsto v/\|v\|$ at this nonzero vector yields $h = -v/\|v\|$. The projection characterization for $p = P_{C_x}(0)$ states that

$$\langle p, v \rangle \geq \|p\|^2, \qquad v \in C_x.$$

Combining this inequality with $\|p\| = m(x)$, $\|v\| \leq L_Q$, and $h = -v/\|v\|$ gives

$$\langle p, h \rangle = -\frac{\langle p, v \rangle}{\|v\|} \leq -\frac{\|p\|^2}{L_Q} = -\frac{m(x)^2}{L_Q}.$$

This estimate holds for every $h \in H_x$ and hence for all $h \in \operatorname{co} H_x = \mathcal{F}(x)$. We now apply this pointwise estimate along a solution z. Definable locally Lipschitz functions are path differentiable (Davis et al., 2020, Theorem 5.8). Consequently, there is a null set N such that, for every $t \notin N$, the derivatives $z'(t)$ and $(f \circ z)'(t)$ exist, the inclusion $z'(t) \in \mathcal{F}(z(t))$ holds, and

$$(f \circ z)'(t) = \langle q, z'(t) \rangle \qquad \text{for every } q \in \partial f(z(t)).$$

Fix such a time t with $z(t) \in Q$, and put $x = z(t)$. If $m(x) > 0$, choose $p = P_{\partial f(x)}(0)$ in the chain rule and choose $w = z'(t)$ in the convex-hull estimate above. This gives

$$(f \circ z)'(t) = \langle p, z'(t) \rangle \leq -\frac{m(z(t))^2}{L_Q}.$$

If $m(x) = 0$, then $0 \in \partial f(x)$, and the chain rule gives

$$(f \circ z)'(t) = \langle 0, z'(t) \rangle = 0 = -\frac{m(z(t))^2}{L_Q}.$$

Thus (20) holds for almost every t such that $z(t) \in Q$. Finally, for the nonincreasing property, it suffices to set $Q = \overline{B}(0, R)$, where $R > \max_{t \in [t_1, t_2]} \|z(t)\|$, and then apply (20) on Q. □

Proof of Theorem 2.1. Suppose first that $\partial f(x_{k_0}) = \{0\}$ for some index k_0. Then $\mathcal{N}_C f(x_{k_0}) = \{0\}$, and it is clear that (x_k) will stay at x_{k_0} and the required assertions hold trivially. We may therefore assume that $\partial f(x_k) \neq \{0\}$ for all $k \in \mathbb{N}$. Set

$$t_0 = 0, \qquad t_{k+1} = t_k + \alpha_k,$$

and let $X : [0, \infty) \to \mathbb{R}^n$ be the affine interpolation with $X(t_k) = x_k$. Since $\sum_k \alpha_k = \infty$, the time grid is unbounded. For $t \in [t_k, t_{k+1})$,

$$X'(t) = \frac{x_{k+1} - x_k}{\alpha_k} \in \mathcal{F}(x_k), \qquad \|X(t) - x_k\| \leq \alpha_k. \tag{21}$$

By (18), X is 1-Lipschitz, and boundedness of (x_k) makes X bounded. Define $\delta : \mathbb{R}_+ \to \mathbb{R}_+$ by $\delta(t) = 2\alpha_k$ on $[t_k, t_{k+1})$. Since $t_k \to \infty$ and $\alpha_k \to 0$, $\delta(t) \to 0$. For $t \in (t_k, t_{k+1})$, $X'(t) \in \mathcal{F}(x_k)$ and $\|X(t) - x_k\| \le \alpha_k < \delta(t)$. Thus X is a perturbed solution of $\dot{z}(t) \in \mathcal{F}(z(t))$ in the sense of (Benaïm et al., 2005, Definition II). (Benaïm et al., 2005, Theorem 3.6) implies that $\Omega(X)$, the set of limits of $X(t_j)$ along $t_j \to \infty$, is internally chain transitive (Benaïm et al., 2005, Definition VI). Since $\operatorname{diam}(X([t_k, t_{k+1}])) \le \alpha_k \to 0$, $\Omega(X) = \Omega(x)$.

Set $S := \{x \in \mathbb{R}^n : 0 \in \partial f(x)\}$ be the critical set of f, which is the equilibrium set of $\dot{z} \in \mathcal{F}(z)$ by (19). The graph of ∂f is closed, so S is closed. By Lemma A.1, f is nonincreasing along every solution. Fix $x \notin S$ and put $\mu = m(x) := d(0, \partial f(x))) > 0$. There is a neighborhood U of x on which $m \ge \mu/2$ by the upper semicontinuity of ∂f. For any solution z with $z(0) = x$, continuity gives $\tau > 0$ with $z([0, \tau]) \subseteq U$. The set $Q = z([0, \tau])$ is compact. The absolute continuity and rate estimate in Lemma A.1 give, for $0 < t \le \tau$,

$$f(z(t)) - f(x) = \int_0^t (f \circ z)'(s)\, ds \le -\frac{1}{L_Q} \int_0^t m(z(s))^2\, ds \le -\frac{\mu^2 t}{4L_Q} < 0.$$

For $t > \tau$, nonincrease gives $f(z(t)) \le f(z(\tau)) < f(x)$. Thus f is a Lyapunov function for S in the sense of (Benaïm et al., 2005, p. 341).

The set $f(S)$ of critical values is finite by (Bolte et al., 2007, Corollary 9). Since $\Omega(x)$ is internally chain transitive, (Benaïm et al., 2005, Proposition 3.27) gives $\Omega(x) \subseteq S$ and $f|_{\Omega(x)} \equiv \ell$ for some $\ell \in \mathbb{R}$. If $f(x_k)$ failed to converge to ℓ, there would be $\varepsilon > 0$ and indices $k_j \to \infty$ with $|f(x_{k_j}) - \ell| \ge \varepsilon$. Boundedness gives a further subsequence $x_{k_{j_l}} \to x^* \in \Omega(x)$. Continuity of f then gives $f(x_{k_{j_l}}) \to f(x^*) = \ell$, a contradiction. If $d(x_k, S)$ failed to tend to zero, choose indices $m_j \to \infty$ with $d(x_{m_j}, S) \ge \varepsilon$ for some $\varepsilon > 0$. A further bounded subsequence converges to $x^* \in \Omega(x) \subseteq S$, whereas $d(x_{m_j}, S) \le \|x_{m_j} - x^*\| \to 0$ along that subsequence. This proves the distance assertion. □

A.2 Coercivity and uniformly small stepsizes

This subsection proves Theorem 3.1 and constructs an objective with bounded gradient flows and an unbounded NGD trajectory. For the boundedness theorem, coercivity makes the outer set C compact. Solutions of $\dot{z} \in \mathcal{F}(z)$ starting in C enter an inner set D within a fixed time. The finite-time Euler approximation (Josz, 2026b, Lemma 3.8) then controls each discrete block before the next restart.

Proof of Theorem 3.1. Put

$$J := E + \max_{x \in \overline{X}_0} \|I(x)\|, \qquad S_J := \{x \in \mathbb{R}^n : \|I(x)\| \le J\}.$$

The maximum is finite because $\overline{X}_0$ is compact and I is continuous. The drift bound gives

$$\|I(x_k)\| \le \|I(x_0)\| + E \le J \quad \forall k \in \mathbb{N}. \tag{22}$$

The set of critical values is finite by (Bolte et al., 2007, Corollary 9). Hence the following maximum is well defined:

$$M := \max\left(f(\overline{X}_0) \cup \{f(x) : x \in \mathbb{R}^n,\ 0 \in \partial f(x)\}\right).$$

Fix $e > 0$, choose $\ell > M + e$, and define

$$C := \{x \in \mathbb{R}^n : f(x) \le \ell,\ \|I(x)\| \le J\}, \qquad D := \{x \in C : f(x) \le \ell - e\}.$$

The set $C = f^{-1}((-\infty, \ell]) \cap I^{-1}(\overline{B}(0, J))$ is closed because f and I are continuous. It is bounded since $C \subseteq G^{-1}((-\infty, \ell + J^2])$ and G is coercive, so C is compact. The set $D = C \cap f^{-1}((-\infty, \ell - e])$ is compact as a closed subset of C. Moreover, $x_0 \in C$ because $f(x_0) \le M < \ell$ and $\|I(x_0)\| \le J$.

For $u \in \mathcal{A}_C(x)$, (9) and (16) give $DI(x)u = 0$. If $(x_j, u_j) \to (x, u)$ with $u_j \in \mathcal{A}_C(x_j)$, continuity of DI gives $DI(x)u = 0$. Taking the convex hull in (17) gives $DI(x)u = 0$ for every $u \in \mathcal{F}(x)$. Let $z : [0, R] \to \mathbb{R}^n$ be a solution of $\dot{z} \in \mathcal{F}(z)$ with $z(0) \in C$. Since $I \in C^1(\mathbb{R}^n)$, and in particular, it is locally Lipschitz, $I \circ z$ is absolutely continuous as a composition of a locally Lipschitz function and an absolutely continuous mapping (Fathi, 2008, Lemma 3.1.3(1)). For each $t \in [0, R]$, the fundamental theorem of calculus gives $I(z(t)) - I(z(0)) = \int_0^t (I \circ z)'(s)\, ds = 0$. By Lemma A.1, $f(z(t)) \le f(z(0))$ on this interval. Hence $z(t) \in C$ for every $t \in [0, R]$.

Adopt the convention $\inf \varnothing = +\infty$ and put

$$\delta := \min\left\{1, \inf_{x\in C,\ f(x)\geq \ell - e} m(x)\right\}.$$

The set $C \cap \{f \geq \ell - e\}$ is compact and contains no critical point because $\ell - e > M$. If it is empty, $\delta = 1$. Otherwise, suppose its infimum of m were zero. Choose $x_j \in C \cap \{f \geq \ell - e\}$ with $m(x_j) \to 0$ and a minimum-norm $v_j \in \partial f(x_j)$, so $\|v_j\| = m(x_j)$. Compactness gives a subsequence $x_j \to x$ in the same set, while $v_j \to 0$. The closed graph of ∂f gives $0 \in \partial f(x)$, contradicting $f(x) \geq \ell - e > M$. Thus $\delta > 0$.

Let L_C be the constant given by Lemma A.1 for the compact set C, and choose

$$T > \frac{eL_C}{\delta^2}.$$

Every solution starting in C satisfies

$$z(t) \in D \quad \forall t \geq T \tag{23}$$

whenever it is defined. Indeed, $I \circ z$ is invariant and $f \circ z$ is nonincreasing by Lemma A.1, so $z(t) \in C$ for all $t \geq 0$. Suppose that $f(z(t)) > \ell - e$ for some $t \geq T$. Since $f \circ z$ is nonincreasing, $f(z(s)) > \ell - e$ for every $s \in [0, T]$. Thus $m(z(s)) \geq \delta$ on $[0, T]$. The absolute continuity of $f \circ z$ and (20) give

$$f(z(T)) \leq f(z(0)) - \frac{\delta^2 T}{L_C} \leq \ell - \frac{\delta^2 T}{L_C} < \ell - e.$$

This contradicts $f(z(T)) \geq f(z(t)) > \ell - e$.

The set D is nonempty because $x_0 \in D$, and it is compact. If no $\eta > 0$ satisfied the inclusion below, there would be points y_j with $d(y_j, D) \leq 1/j$, $\|I(y_j)\| \leq J$, and $f(y_j) > \ell$. Choose $z_j \in D$ with $\|y_j - z_j\| = d(y_j, D)$. A subsequence of z_j converges to $z \in D$, and the corresponding y_j also converges to z. Continuity of f would give $f(z) \geq \ell$, whereas $f(z) \leq \ell - e$ by the definition of D. Consequently, some $\eta > 0$ satisfies

$$\{y \in \mathbb{R}^n : d(y, D) \leq \eta,\ \|I(y)\| \leq J\} \subseteq C. \tag{24}$$

For every Euler sequence $y_{i+1} = y_i + h_i u_i$ with $y_0 \in C$, $u_i \in \mathcal{F}(y_i)$, and $0 < h_i \leq T$, (18) gives $d(y_i, C) \leq \sum_{j<i} h_j =: s_i \leq 2T$ whenever $s_i \leq 2T$. Thus $y_i \in B(C, 2T + 1)$ whenever $s_i \leq 2T$. Apply (Josz, 2026b, Lemma 3.8) with initial set C, time horizon $2T$, radius $2T + 1$, preliminary step cap T, and tracking tolerance η. Its upper semicontinuity and compact-value hypotheses hold for $\mathcal{F}$ by (18) and the graph-closure construction above. The lemma gives $\bar{\alpha} \in (0, T]$ such that, for all sequences $\{h_i\}_{i\in\mathbb{N}} \subseteq (0, \bar{\alpha}]$ and $\{y_i\}_{i\in\mathbb{N}} \subseteq \mathbb{R}^n$ with $y_0 \in C$ and $y_{i+1} - y_i \in h_i \mathcal{F}(y_i)$ for every $i \in \mathbb{N}$, there exists an absolutely continuous $z : [0, 2T] \to \mathbb{R}^n$ satisfying

$$\begin{aligned} &\dot{z}(t) \in \mathcal{F}(z(t)) \quad \text{for a.e. } t \in (0, 2T), \qquad z(0) \in C, \\ &s_i \leq 2T \quad \Longrightarrow \quad \|y_i - z(s_i)\| \leq \eta \quad \text{for every } i \in \mathbb{N}. \end{aligned} \tag{25}$$

Here the lemma's initial point belongs to C because C is closed, and its differential inclusion uses $\mathcal{F}$ because $\operatorname{co} \mathcal{F}(x) = \mathcal{F}(x)$ for every x. The NGD update lies in $x_k + \alpha_k \mathcal{F}(x_k)$ by (18). Fix a starting index p with $x_p \in C$ and write $s_{p,q} := \sum_{k=p}^{q-1} \alpha_k$ for $q \geq p$. Applying the tracking conclusion (25) to the sequence starting at x_p, we know the tracking solution z remains in C, so

$$s_{p,q} \leq 2T \quad \Longrightarrow \quad d(x_q, C) \leq \eta. \tag{26}$$

If $T \leq s_{p,q} \leq 2T$, then (23) gives $z(s_{p,q}) \in D$. Hence $d(x_q, D) \leq \eta$, and (22) and (24) gives $x_q \in C$.

If the set $\{q \in \mathbb{N} : q \geq p,\ s_{p,q} \leq 2T\}$ has a largest element $\bar{q}$, then $s_{p,\bar{q}+1} > 2T$ and $\alpha_{\bar{q}} \leq \bar{\alpha} \leq T$, so $T < s_{p,\bar{q}} \leq 2T$. Since $s_{p,p+1} = \alpha_p \leq T$, one has $\bar{q} \geq p + 1$. Hence $x_{\bar{q}} \in C$, and we can repeat this argument at $\bar{q}$. If this set has no largest element, it contains every $q \geq p$, so $q \geq p$ satisfies (26). Each restart index is a strictly larger integer than the preceding one. If restarts continue indefinitely, their indices tend to infinity, and (26) covers every finite index. If a block has no largest index, that estimate covers its entire tail. Starting from $x_0 \in C$, every iterate therefore belongs to the compact set

$$K := \{x \in \mathbb{R}^n : d(x, C) \leq \eta\}.$$

The set K is compact because C is compact. This proves the assertion. □

Proposition A.2. *The smooth semialgebraic function*

$$f(x,y)=\frac{y^2}{1+x^2}$$

has uniformly bounded gradient-flow trajectories. In particular, for every $r>0$, every gradient flow of f initialized in $\overline{B}(0,r)$ remains in $\overline{B}(0,r)$. From $z_0=(1,1)$, however, every $\bar{\alpha}>0$ admits positive steps satisfying

$$\alpha_k\le\bar{\alpha},\qquad \alpha_k\to 0,$$

for which the corresponding NGD trajectory is unbounded.

Proof. The gradient flow is

$$\dot{x}=\frac{2xy^2}{(1+x^2)^2},\qquad \dot{y}=-\frac{2y}{1+x^2}.$$

Along every solution,

$$\frac{d}{dt}(x^2+y^2)=-\frac{4y^2}{(1+x^2)^2}\le 0.$$

Thus $\|(x(t),y(t))\|\le\|(x(0),y(0))\|$ whenever it is defined. Therefore, every gradient flow of f initialized in $\overline{B}(0,r)$ remains in $\overline{B}(0,r)$. For $y\neq 0$, set

$$D(x,y)=\sqrt{x^2y^2+(1+x^2)^2}.$$

For $t\in\mathbb{R}$, let $\operatorname{sgn}(t)$ be -1, 0, or 1 according as $t<0$, $t=0$, or $t>0$. One NGD step is

$$x^+=x+\alpha\frac{x|y|}{D(x,y)},\qquad y^+=y-\alpha\frac{\operatorname{sgn}(y)(1+x^2)}{D(x,y)}. \tag{27}$$

Fix $\bar{\alpha}>0$ and put $\delta=\min\{1/2,\bar{\alpha}/4\}$. A finite sequence of positive-y steps takes $(1,1)$ to some (X,δ), with $X\ge 1$, while keeping every step below $\bar{\alpha}$. Indeed, to decrease y by d, take

$$\alpha=d\,\frac{D(x,y)}{1+x^2},\qquad \frac{D(x,y)}{1+x^2}\le\sqrt{1+\frac{y^2}{4}}\le\frac{\sqrt{5}}{2},$$

and partition $1-\delta$ into positive decrements no larger than $2\bar{\alpha}/\sqrt{5}$. Starting at (X,δ), relabel the tail so that $(x_0,y_0)=(X,\delta)$. Set

$$r_j=\frac{\delta}{\sqrt{j+1}},\qquad \alpha_j=(r_j+r_{j+1})\frac{D(x_j,y_j)}{1+x_j^2}. \tag{28}$$

It holds that $y_0=\delta=r_0$, and we will use induction to show $y_j=(-1)^jr_j$ for all $j\in\mathbb{N}$. Indeed, if this identity holds at index j, then $\operatorname{sgn}(y_j)=(-1)^j$, and by the recursion in (27)

$$y_{j+1}=(-1)^jr_j-(r_j+r_{j+1})(-1)^j=(-1)^{j+1}r_{j+1}.$$

Since $x/(1+x^2)\le 1/2$,

$$\alpha_j\le 2r_j\sqrt{1+\frac{\delta^2}{4}}\le 2\delta\sqrt{1+\frac{\delta^2}{4}}\le\bar{\alpha},\qquad \alpha_j\to 0,$$

where we have used the fact that $r_j\to 0$. Using $|y_j|=r_j$ in (27) also gives

$$x_{j+1}=x_j+\frac{x_j}{1+x_j^2}r_j(r_j+r_{j+1}).$$

Since $x_0=X\ge 1$, this gives $x_j\ge 1$ for every j. Squaring both sides yields

$$x_{j+1}^2=x_j^2+2\frac{x_j^2}{1+x_j^2}r_j(r_j+r_{j+1})+\left(\frac{x_j}{1+x_j^2}r_j(r_j+r_{j+1})\right)^2.$$

Therefore,

$$x_{j+1}^2-x_j^2\ge 2\frac{x_j^2}{1+x_j^2}r_j(r_j+r_{j+1})\ge r_j^2.$$

The series $\sum_j r_j^2$ diverges, so $x_j\to+\infty$. □

B HARMONIC CONVERGENCE FOR NGD

Fix f, c, and (x_k) as in Theorem 2.3, and write $a_k = c/(k+1)$. For each k, fix $u_k \in \mathcal{N}_C f(x_k)$ such that $x_{k+1} = x_k - a_k u_k$. Objective-value convergence and stationarity follow from Theorem 2.1. For point convergence, we apply the stratified diameter bound of (Lai & Song, 2026b, Theorem 2.4). The geometric construction uses (Lai & Song, 2026a, Proposition 3.2, Lemma 3.3, Proposition 3.5, and Corollary 4.2). The argument below verifies that $\mathcal{N}_C f$ satisfies the required projected-length estimates. In particular, the normalization factor must cancel on rank-one strata and produce a higher-order tangential term on rank-zero strata.

B.1 A SIMULTANEOUS L'-REGULAR AND STRICT-THOM STRATIFICATION

Let us first introduce the notion of stratification. A finite C^p stratification of a set $X \subseteq \mathbb{R}^n$ is a finite partition $\mathcal{S}$ of X into nonempty connected locally closed embedded C^p submanifolds of $\mathbb{R}^n$, called strata, satisfying the frontier condition (Mather, 2012, Section 5): for distinct $S, R \in \mathcal{S}$,

$$S \cap \overline{R} \neq \varnothing \quad \Longrightarrow \quad S \subseteq \overline{R} \setminus R \quad \text{and} \quad \dim S < \dim R.$$

The stratification is definable if every stratum is definable. The ordered pair (S, R) is incident if $S \subseteq \overline{R} \setminus R$. A stratification $\mathcal{S}$ refines a partition $\mathcal{P}$ of X if every $S \in \mathcal{S}$ is contained in some $P \in \mathcal{P}$. The stratification is compatible with $A \subseteq X$ if A is a union of its strata; it is compatible with a family of subsets if this holds for every member. Its d-skeleton is the union of its strata of dimension at most d, and the frontier condition makes this union closed in X.

Let $R_{\mathrm{orb}} = \sup_k \|x_k\| < \infty$, and define the closed ball

$$K^+ = \overline{B}(0, R_{\mathrm{orb}} + 1). \tag{29}$$

The orbit lies in $\overline{B}(0, R_{\mathrm{orb}}) \subseteq \operatorname{int} K^+$. Every segment of length at most one starting in $\overline{B}(0, R_{\mathrm{orb}})$ lies in K^+. Local boundedness of ∂f (Clarke, 1990, Proposition 2.1.2(a)) and compactness of K^+ give $1 \leq Q < \infty$ such that

$$\|v\| \leq Q \quad (x \in K^+,\ v \in \partial f(x)). \tag{30}$$

Since K^+ is convex, the Clarke mean value inequality gives $|f(x) - f(y)| \leq Q\|x - y\|$ for all $x, y \in K^+$.

Our results build on proper stratification of K^+. We next introduce the stratification used in Lai & Song (2026a).

Definition B.1 (L'-regular stratification)**.** We use the terminology of (Lai & Song, 2026a, Definition 2.3). Standard L-regular cells are defined recursively. In $\mathbb{R}$, they are the singletons and open intervals. If $B \subseteq \mathbb{R}^{d-1}$ is a standard L-regular cell and $\xi, \xi_1, \xi_2 : B \to \mathbb{R}$ are definable Lipschitz functions with $\xi_1 < \xi_2$, then the following are standard L-regular cells in $\mathbb{R}^d$:

$$\begin{aligned}
\Gamma(\xi) &= \{(z, s) \in B \times \mathbb{R} : s = \xi(z)\}, \\
(\xi_1, \xi_2)_B &= \{(z, s) \in B \times \mathbb{R} : \xi_1(z) < s < \xi_2(z)\}, \\
(\xi, +\infty)_B &= \{(z, s) \in B \times \mathbb{R} : s > \xi(z)\}, \\
(-\infty, \xi)_B &= \{(z, s) \in B \times \mathbb{R} : s < \xi(z)\}.
\end{aligned}$$

An L-regular cell is carried to a standard L-regular cell by a linear orthogonal change of coordinates. A set $M \subseteq \mathbb{R}^n$ is L'-regular if it is bi-Lipschitz homeomorphic to an L-regular cell, possibly in another Euclidean space. A finite stratification is L'-regular if each of its strata is L'-regular (Lai & Song, 2026a, p. 9).

Let the critical set of f be

$$C_f = \{x \in \mathbb{R}^n : 0 \in \partial f(x)\}. \tag{31}$$

It is definable by (Ji & Telgarsky, 2020, Supplementary Material, Lemma B.8) and closed by the upper semicontinuity of ∂f (Clarke, 1990, Proposition 2.1.5(d)). Apply (Bolte et al., 2007, Lemma 8) to the lower semicontinuous definable function f on $\mathbb{R}^n$, with smoothness order three and the prescribed definable set C_f. It gives a finite definable C^3 Whitney stratification $\{G_1, \ldots, G_J\}$ of $\operatorname{gph} f$ whose projected strata D_j stratify $\mathbb{R}^n$, are compatible with C_f, and satisfy the nonvertical condition of that lemma. By (Bolte et al., 2007, Proposition 4),

$$P_{T_x D_j} v = \nabla_{D_j} f(x) \quad (x \in D_j,\ v \in \partial f(x)).$$

The graph $\mathrm{gph}(f|_{K^+})$ is compact and definable, and $f|_{K^+}$ is Q-Lipschitz by (30) and the mean value inequality. Apply (Lai & Song, 2026a, Theorem 2.5) to this graph with smoothness order three, prescribing its intersections with every G_j. The resulting Lipschitz L-regular C^3 graph stratification projects, as in (Lai & Song, 2026a, Proposition 3.2, proof), to a finite definable L'-regular C^3 stratification $\mathcal{P}$ of K^+ on which f is C^3. For each $P \in \mathcal{P}$, compatibility with the prescribed intersections gives an index j such that $P \subseteq D_j$. Thus $\mathcal{P}$ is compatible with C_f. For $x \in P$, the inclusion $P \subseteq D_j$ gives $T_xP \subseteq T_xD_j$ and $\nabla_P f(x) = P_{T_xP}\nabla_{D_j} f(x)$. Projecting the preceding identity onto T_xP gives, for every $v \in \partial f(x)$,

$$P_{T_xP} v = \nabla_P f(x). \tag{32}$$

L'-regularity supplies controlled inner curves, projection tubes, and Lipschitz estimates away from stratum frontiers. Near the critical set, $\|v_k\|$ in (2) can approach zero, making its reciprocal unbounded. We therefore add a condition on the tangent spaces to the level sets of f.

Definition B.2 (Constant-rank strict Thom (w_f) stratification)**.** Following (Lê Loi, 1998, Definition 2.1), let $\mathcal{S}$ be a finite definable C^p stratification of a definable set X, where $f|_S$ is C^p and has constant rank[1] for every $S \in \mathcal{S}$. Since f is real-valued, this rank is zero or one. For $x \in S$, set

$$T^S_{x,f} := \ker D(f|_S)(x),$$

the tangent space at x to the level set of $f|_S$. For linear subspaces $E, F \subseteq \mathbb{R}^n$, define

$$\delta(E,F) = \sup\{\mathrm{dist}(e,F) : e \in E,\ \|e\| = 1\}, \qquad \delta(\{0\},F) = 0.$$

The stratification $\mathcal{S}$ is a strict Thom (w_f) stratification if, for every incident pair $S \subseteq \overline{R} \setminus R$ and every $y_0 \in S$, there are constants $C, r > 0$ such that

$$\delta\big(T^S_{y,f}, T^R_{x,f}\big) \le C\|x - y\| \quad \begin{array}{l} y \in S \cap B(y_0, r), \\ x \in R \cap B(y_0, r). \end{array} \tag{33}$$

The two rank cases use these properties differently. If $f|_M$ has rank one and $C \subseteq M$ is compact, then

$$\mu_C := \min_{y \in C} \|\nabla_M f(y)\| > 0.$$

For every x sufficiently near C, every $y \in P_C(x)$, and every $v \in \partial f(x)$, one has

$$\|v\| \ge \|P_{T_yM} v\| \ge \frac{1}{2}\mu_C.$$

To obtain one neighborhood valid for all these choices, suppose the second inequality failed along $d(x_j, C) \to 0$. Choose $y_j \in P_C(x_j)$ and $v_j \in \partial f(x_j)$ witnessing the failure. Compactness of C gives a subsequence $y_j \to y \in C$ and then $x_j \to y$. Local boundedness of ∂f gives a further subsequence $v_j \to v$; its closed graph gives $v \in \partial f(y)$. Continuity of the tangent spaces and (38) yield $P_{T_{y_j}M} v_j \to \nabla_M f(y)$, whose norm is at least μ_C. This contradicts the assumed failure. The estimate gives a positive local lower bound on the norm of every selected subgradient. If $f|_M$ has rank zero, then $\nabla_M f = 0$ and $T^M_{y,f} = T_yM$ for all $y \in M$. If x lies on an incident stratum R and $u = v/\|v\|$ with $0 \ne v \in \partial f(x)$, set $E = T_yM$ and $F = T^R_{x,f}$. For every $w \in F$, one has $w \in T_xR$ and $D(f|_R)(x)[w] = 0$. The projection identity on R therefore gives

$$\langle u, w\rangle = \frac{1}{\|v\|}\langle P_{T_xR} v, w\rangle = \frac{1}{\|v\|}\langle \nabla_R f(x), w\rangle = \frac{1}{\|v\|} D(f|_R)(x)[w] = 0.$$

Thus $u \perp F$. If $E = \{0\}$, then $\|P_E u\| = 0 = \delta(E,F)$. If $E \ne \{0\}$, then $\|u\| = 1$ and

$$\|P_E u\| = \sup_{\substack{e \in E \\ \|e\| = 1}} |\langle u, e\rangle| = \sup_{\substack{e \in E \\ \|e\| = 1}} |\langle u, e - P_F e\rangle| \le \sup_{\substack{e \in E \\ \|e\| = 1}} \|e - P_F e\| = \delta(E,F).$$

The strict Thom (w_f) estimate now gives

$$\|P_{T_yM} u\| \le \delta(T_yM, T^R_{x,f}) \le C\|x - y\|. \tag{34}$$

This controls motion along the rank-zero stratum. At the same time, L'-regularity must be retained in order to use the estimates of (Lai & Song, 2026a, Proposition 2.4 and Lemmas 3.3–3.4). We therefore need the following simultaneous refinement.

[1]Here, the word constant rank means its differential has constant rank, i.e., in the sense of differential geometry; see (Lee, 2013, Chapter 4).

Lemma B.3. *For every closed ball K^+ and its parent stratification $\mathcal{P}$ constructed above, there is a finite definable C^3 stratification*

$$\mathcal{M} = \{M_1, \dots, M_T\} \tag{35}$$

of K^+ that refines $\mathcal{P}$ and has the following properties:

1. *Every M_i is an L-regular cell and hence L'-regular.*
2. *Every $f|_{M_i}$ has constant rank zero or one.*
3. *Every incident pair of strata satisfies strict Thom (w_f).*
4. *The stratification is compatible with C_f.*

Proof. We first record the strict-Thom genericity fact used below. For a nonempty incident constant-rank pair Γ_1, Γ_2, define

$$P_{w_f}(\Gamma_1, \Gamma_2) = \{y \in \Gamma_1 : (\Gamma_1, \Gamma_2) \text{ satisfies (33) at } y\}.$$

The condition is local and definable. Since f is continuous and definable and both restrictions have constant rank, (Lê Loi, 1998, Proposition 2.7) applies and gives

$$P_{w_f}(\Gamma_1, \Gamma_2) \text{ definable}, \qquad \dim\big(\Gamma_1 \setminus P_{w_f}(\Gamma_1, \Gamma_2)\big) < \dim \Gamma_1. \tag{36}$$

Set $\mathcal{P}_n = \mathcal{P}$. We construct $\mathcal{P}_{n-1}, \dots, \mathcal{P}_{-1}$ by downward induction. At the beginning of stage $d \in \{n, n-1, \dots, 0\}$, the strata of $\mathcal{P}_d$ of dimension greater than d have been fixed: they are L-regular, f has constant rank on each of them, and every incident pair among them satisfies (w_f). Let Z^d be the d-skeleton of $\mathcal{P}_d$. It is closed in K^+ by the frontier condition.

The first refinement at this stage creates constant rank. Apply (Lê Loi, 2002, Theorem 2.1) to the continuous definable map $f : Z^d \to \mathbb{R}$ with $p = 3$ to obtain a finite definable C^3 map stratification $\mathcal{X}^d$ of Z^d that refines $\mathcal{P}_d$ on Z^d. By the definition of a map stratification (Lê Loi, 2002, p. 240), for each $S \in \mathcal{X}^d$ there is a stratum $Y_S \subseteq \mathbb{R}$ of $\mathbb{R}$ such that

$$f|_S : S \longrightarrow Y_S$$

is a C^3 submersion. Consequently,

$$\operatorname{rank} D(f|_S)(x) = \dim Y_S \in \{0, 1\} \qquad (x \in S). \tag{37}$$

For every d-dimensional $S \in \mathcal{X}^d$ and every fixed higher-dimensional stratum R with $S \subseteq \overline{R}$, form

$$B_{S,R} = S \setminus P_{w_f}(S, R).$$

By (36), $B_{S,R}$ is definable and has dimension smaller than d.

The second refinement at this stage creates L-regularity. Apply (Lai & Song, 2026a, Theorem 2.5) to the compact definable set Z^d to obtain an L-regular C^3 stratification that is compatible with every $S \in \mathcal{X}^d$ and every set $B_{S,R}$. Let Λ^d be the resulting L-regular C^3 stratification of Z^d. Since Λ^d is compatible with the partition $\mathcal{X}^d$, each $M \in \Lambda^d$ is contained in a unique $S \in \mathcal{X}^d$. Hence Λ^d refines $\mathcal{P}_d$ on Z^d.

Replace the strata of $\mathcal{P}_d$ contained in Z^d by Λ^d and leave its strata of dimension greater than d unchanged. Denote the resulting partition by $\mathcal{P}_{d-1}$. This partition satisfies the frontier condition. Indeed, let $M, Q \in \mathcal{P}_{d-1}$ be distinct and assume $M \cap \overline{Q} \neq \varnothing$. If both strata have dimension greater than d, the frontier condition of $\mathcal{P}_d$ applies; if both are in Λ^d, the frontier condition of Λ^d applies. Suppose $M \in \Lambda^d$ and $Q \in \mathcal{P}_d$ has dimension greater than d. Let $P \in \mathcal{P}_d$ contain M. Then $P \cap \overline{Q} \neq \varnothing$, so the frontier condition of $\mathcal{P}_d$ gives $P \subseteq \overline{Q} \setminus Q$ and $\dim P < \dim Q$. Hence $M \subseteq \overline{Q} \setminus Q$ and $\dim M \leq \dim P < \dim Q$. If instead M has dimension greater than d and $Q \in \Lambda^d$, then $\overline{Q} \subseteq Z^d$ because Z^d is closed, whereas $M \cap Z^d = \varnothing$. Thus $M \cap \overline{Q} = \varnothing$, contrary to the assumption.

It remains to identify the strata that can be fixed at this stage. Let $M \in \Lambda^d$ have dimension d, and let $S \in \mathcal{X}^d$ be the unique stratum containing M. Then $\dim S = d$, and hence M is an open embedded submanifold of S, which means that

$$T_x M = T_x S, \qquad T^M_{x,f} = T^S_{x,f} \quad (x \in M),$$

while (37) gives constant rank for $f|_M$. Fix any higher-dimensional stratum R with $M \subseteq \overline{R}$. Let $P \in \mathcal{P}_d$ contain S. Since $M \subseteq P \cap \overline{R}$, the frontier condition for $\mathcal{P}_d$ gives $P \subseteq \overline{R}$, hence $S \subseteq \overline{R}$. Thus $B_{S,R}$ was included in the second refinement. Compatibility of Λ^d with $B_{S,R}$ implies that M is either contained in $B_{S,R}$ or disjoint from it. The first alternative is impossible because $\dim B_{S,R} < d = \dim M$. Thus $M \subseteq P_{w_f}(S,R)$. At each $y_0 \in M$, the strict Thom estimate for (S,R) supplied by the definition of $P_{w_f}(S,R)$ restricts to (M,R) because $T^M_{y,f} = T^S_{y,f}$ for $y \in M$.

The d-dimensional strata of Λ^d are fixed in all later stages. When $d \geq 1$, the next stage refines the closed $(d-1)$-skeleton of $\mathcal{P}_{d-1}$. At $d = 0$ every set $B_{S,R}$ is empty by (36). Finally, every stratum of $\mathcal{P}_{-1}$ was fixed as an L-regular stratum of some Λ^d, has constant rank, and satisfies (w_f) with every incident stratum. Moreover, $\mathcal{P}_{d-1}$ refines $\mathcal{P}_d$ at every stage $d = n, n-1, \ldots, 0$. Set $\mathcal{M} = \mathcal{P}_{-1}$. Then $\mathcal{M}$ refines $\mathcal{P}$ and has the first three properties. Since $\mathcal{P}$ is compatible with C_f, every stratum of its refinement $\mathcal{M}$ is either contained in or disjoint from C_f. This gives the fourth property. □

If $M_i \subseteq P$ and $x \in M_i$, then $T_x M_i \subseteq T_x P$. Projecting (32) onto $T_x M_i$ gives

$$P_{T_x M_i} v = P_{T_x M_i} P_{T_x P} v = P_{T_x M_i} \nabla_P f(x) = \nabla_{M_i} f(x) \quad \big(v \in \partial f(x)\big). \tag{38}$$

By Theorem 2.1, $f(x_k) \to \ell$, and every accumulation point x_* satisfies $0 \in \partial f(x_*)$ and $f(x_*) = \ell$. The bounded trajectory has an accumulation point, so ℓ is a critical value of f. There are $A > 0$, $\theta \in (0,1)$, and $\epsilon > 0$ such that, for every rank-one stratum M_i and every $y \in M_i \cap K^+$ with $0 < |f(y) - \ell| \leq \epsilon$, the signed gradient inequality is

$$\frac{12}{\psi'(f(y)-\ell)} \leq \|\nabla_{M_i} f(y)\|, \qquad \psi(s) = A\ \mathrm{sgn}(s)|s|^{1-\theta}. \tag{39}$$

For each rank-one stratum M_i, the signed gradient inequality (Lai & Song, 2026a, Section 4.1, (4.1)) applies to the bounded definable C^3 manifold M_i and the C^1 definable function $g_i = f|_{M_i} - \ell$. It gives $\epsilon_i > 0$, $\theta_i \in [0,1)$, and $\mu_i > 0$ such that

$$\|\nabla_{M_i} f(y)\| \geq \mu_i |f(y) - \ell|^{\theta_i} \quad \text{whenever } y \in M_i, \quad 0 < |f(y) - \ell| \leq \epsilon_i.$$

If there are rank-one strata, choose $\theta \in (\max_i \theta_i, 1)$, $\epsilon = \min\{1, \min_i \epsilon_i\}$, $\mu = \min_i \mu_i$, and $A = 12/[\mu(1-\theta)]$. For $0 < |s| \leq \epsilon$, one has $|s|^{\theta_i} \geq |s|^{\theta}$; moreover $12/\psi'(s) = \mu|s|^{\theta}$. These inequalities give (39) for every rank-one stratum. If there are no rank-one strata, choose $\theta = 1/2$, $\epsilon = 1$, and $A = 1$; the quantified inequality in (39) then has no instances.

Write

$$\mathrm{fr}\, M_i := \overline{M_i} \setminus M_i, \qquad \rho_i(y) := d(y, \mathrm{fr}\, M_i), \qquad K_i(t) := \{y \in M_i : \rho_i(y) \geq t\}, \tag{40}$$

with $\rho_i \equiv 1$ when the stratum frontier is empty. For $t > 0$ and $\mathrm{fr}\, M_i \neq \varnothing$,

$$K_i(t) = \{y \in \overline{M}_i : d(y, \mathrm{fr}\, M_i) \geq t\}.$$

When $\mathrm{fr}\, M_i \neq \varnothing$, this is a closed subset of the compact set $\overline{M}_i \subseteq K^+$, since distance to a set is continuous. If $\mathrm{fr}\, M_i = \varnothing$, then $M_i = \overline{M}_i$ is compact, and $K_i(t)$ is either M_i or empty. Thus $K_i(t)$ is compact for every $t > 0$.

Lemma B.4. *Let $\mathcal{M}$ be given as in Lemma B.3. For each $i \in \{1, \ldots, T\}$, the following statements hold:*

1. *There are constants $c_i^P, C_i > 0$ and exponents $b_i, m_i \geq 0$ such that, for every $0 < t \leq 1$, the projection mapping $P_i = P_{M_i}$ is single valued, C^2, and 2-Lipschitz on $B(K_i(t), c_i^P t^{b_i})$. On the smaller set $B(K_i(t), c_i^P t^{b_i}/2)$, the Lipschitz moduli of DP_i, $\nabla_{M_i} f \circ P_i$, and $\nabla((f|_{M_i}) \circ P_i)$ are at most $C_i t^{-m_i}$. The same constants can be chosen so that*

$$P_i(y + w) = y \quad \big(y \in K_i(t),\ w \in (T_y M_i)^{\perp},\ \|w\| < c_i^P t^{b_i}\big). \tag{41}$$

2. *If $f|_{M_i}$ has rank one, define*

$$\mu_i(t) = \min_{y \in K_i(t)} \|\nabla_{M_i} f(y)\| \tag{42}$$

and set $\mu_i(t) = 1$ when $K_i(t)$ is empty. There are constants $c_i^{\mu}, c_i^{E} > 0$ and exponents $p_i, e_i \geq 0$ such that, for every $0 < t \leq 1$, $\mu_i(t) \geq c_i^{\mu} t^{p_i}$ and

$$\|P_{T_y M_i} v - \nabla_{M_i} f(y)\| \leq \frac{1}{10}\mu_i(t) \tag{43}$$

whenever $y \in K_i(t)$, $x \in K^+$, $\|x - y\| \leq c_i^{E} t^{e_i}$, and $v \in \partial f(x)$.

3. *If $f|_{M_i}$ has rank zero, there are constants $\widehat{c}_i, \widehat{C}_i > 0$ and exponents $b_i^{w}, s_i \geq 0$ such that, for every $0 < t \leq 1$,*

$$\delta(T_y M_i, T_{x,f}) \leq \widehat{C}_i t^{-s_i}\|x - y\|, \qquad T_{x,f} = \ker D(f|_R)(x), \tag{44}$$

for every stratum $R \neq M_i$ with $M_i \subseteq \overline{R}$, $y \in K_i(t)$, $x \in R$, and $0 < \|x - y\| < \widehat{c}_i t^{b_i^{w}}$.

Proof. Every M_i is bounded, L-regular (hence L'-regular), and C^3 by Lemma B.3. In (Lai & Song, 2026a, Lemmas 3.3–3.4), their ∂M_i equals $\operatorname{fr} M_i$. We denote their neighborhood $A + \overline{B}(0, r)$ by $\overline{B}(A, r)$ in the citations below. For $0 < t \leq 1$,

$$K_i(t) \subseteq M_i \setminus \overline{B}(\operatorname{fr} M_i, t/2).$$

Indeed, if the frontier is nonempty, every $y \in K_i(t)$ satisfies $d(y, \operatorname{fr} M_i) \geq t > t/2$; if it is empty, the set on the right is M_i.

Apply their Lemma 3.4 with $L = 2$, denoting its constants by $\varrho_i \in (0, 1]$, $\vartheta_i \geq 1$, $\overline{\eta}_i \geq \vartheta_i$, and $\overline{r}_i > 0$. Set

$$b_i = \vartheta_i \overline{\eta}_i, \qquad r_i = \overline{r}_i \varrho_i^{\overline{\eta}_i}, \qquad c_i^{P} = 2^{-b_i} r_i.$$

For $0 < s \leq 1$, define

$$\Omega_{i,s} = B\big(M_i \setminus \overline{B}(\operatorname{fr} M_i, \varrho_i s^{\vartheta_i}), r_i s^{b_i}\big).$$

Their Lemma 3.4(2), with truncation parameter $\varrho_i s^{\vartheta_i}$, exponent $\overline{\eta}_i$, and radius constant $\overline{r}_i$, makes P_i single valued, C^2, and 2-Lipschitz on $\Omega_{i,s}$, since $r_i s^{b_i} = \overline{r}_i(\varrho_i s^{\vartheta_i})^{\overline{\eta}_i}$. For $s = t/2$, one has $\varrho_i s^{\vartheta_i} \leq s$ and $c_i^{P} t^{b_i} = r_i s^{b_i}$, so the inclusion above gives

$$B(K_i(t), c_i^{P} t^{b_i}) \subseteq \Omega_{i,t/2}.$$

This proves the asserted projection properties. The C^2 regularity of P_i and the C^3 regularity of $f|_{M_i}$ make the three maps

$$DP_i, \qquad \nabla_{M_i} f \circ P_i, \qquad \nabla((f|_{M_i}) \circ P_i)$$

C^1 and definable on $\bigcup_{0<s\leq 1} \Omega_{i,s}$. Apply their Lemma 3.4(3) to these maps with parameter $s = t/2$, exponent $b_i \geq \overline{\eta}_i$, and radius constants $r_1 = r_i/2 < r_2 = r_i \leq \overline{r}_i$. The inclusion of $K_i(t)$ above restricts its estimates to $B(K_i(t), c_i^{P} t^{b_i}/2)$. For the three maps, the resulting bounds are $C_{ij}(t/2)^{-m_{ij}}$, $j = 1, 2, 3$, with $C_{ij} > 0$ and $m_{ij} \geq 0$. Taking $m_i = \max_j m_{ij}$ and $C_i = \max_j 2^{m_{ij}} C_{ij}$ gives the bound $C_i t^{-m_i}$ for $0 < t \leq 1$.

For a positive definable function $h : (0, 1] \to (0, \infty)$, (van den Dries & Miller, 1996, Growth Dichotomy 4.12), applied to $h(1/s)$ and its reciprocal, gives constants $c, C > 0$ and exponents $a, b \geq 0$ such that $ct^{a} \leq h(t) \leq Ct^{-b}$ for all sufficiently small $t > 0$.

We establish (41). If $\dim M_i = 0$, then connectedness makes M_i a singleton and P_i is constant on $\mathbb{R}^n$. For $\dim M_i > 0$, the C^3 regularity of M_i gives the locally Lipschitz tangent-space condition (Dudek & Holly, 1994, (3.6)). For each nonempty compact $K_i(t)$, (Dudek & Holly, 1994, Theorem 3.2) supplies a positive radius on which projection returns every point on the normal fiber at its base. Let $\nu_i(t)$ be the supremum of such radii in $(0, 1]$, and set $\nu_i(t) = 1$ if $K_i(t)$ is empty. The condition defining an admissible radius quantifies over the definable sets $K_i(t)$ and $\{(y, w) : y \in M_i,\ w \perp T_y M_i\}$ and the definable projection mapping P_i. Thus $\nu_i(t)$ is positive and definable. Since $K_i(t') \subseteq K_i(t)$ for $t' \geq t$, ν_i is nondecreasing. The growth bound above gives $c, d > 0$ with $\nu_i(t) \geq ct^{d}$ for sufficiently small t. Monotonicity and positivity extend this estimate to $0 < t \leq 1$ after reducing c. Increasing b_i to at least d and decreasing c_i^{P} to at most $c/2$ preserve the projection and Lipschitz bounds already obtained and give (41).

On a rank-one stratum, the minimum in (42) is positive on every nonempty $K_i(t)$ because $D(f|_{M_i})$ has rank one. It is definable as a function of t since $K_i(t)$ and $\nabla_{M_i} f$ are definable. The above growth

bound gives $\mu_i(t) \geq c_i^{\mu} t^{p_i}$ for sufficiently small $t > 0$. Choose $t_0 \in (0, 1]$ so this inequality holds for $0 < t \leq t_0$. For $t \in [t_0, 1]$, one has $K_i(t) \subseteq K_i(t_0)$; hence $\mu_i(t) \geq \mu_i(t_0)$ when $K_i(t)$ is nonempty, and $\mu_i(t) = 1$ otherwise. Since $t^{p_i} \leq 1$, decreasing c_i^{μ} to at most $\min\{\mu_i(t_0), 1\}$ gives the bound for every $0 < t \leq 1$.

It remains to prove (43). For a rank-one stratum define

$$\Omega_i(t, s) = \sup \left\{ \|P_{T_y M_i} v - \nabla_{M_i} f(y)\| : \begin{array}{l} y \in K_i(t),\ x \in K^+,\ \|x - y\| \leq s, \\ v \in \partial f(x) \end{array} \right\}. \tag{45}$$

When $K_i(t)$ is empty, set $\Omega_i(t, s) = 0$. For fixed $t > 0$, one has $\Omega_i(t, s) \to 0$ as $s \downarrow 0$. Indeed, otherwise compactness, (30), and passing to a subsequence would give $x_j, y_j \to y \in K_i(t)$ and $v_j \to v \in \partial f(y)$ while $\liminf_{j\to\infty} \|P_{T_{y_j} M_i} v_j - \nabla_{M_i} f(y_j)\| > 0$. Continuity of the tangent spaces, the Riemannian gradient and (38) make the limit zero, which yields contradiction. Moreover, the function in (45) is definable. Consequently the set of $(t, s) \in (0, 1]^2$ satisfying $\Omega_i(t, s) \leq \mu_i(t)/10$ is definable and has a nonempty section at every t. Definable choice (Coste, 1999, Theorem 3.1) gives a positive definable selection $s(t)$ from these sections. The growth bound above gives $s(t) \geq c_i^E t^{e_i}$ for sufficiently small $t > 0$. Since $\Omega_i(t, s)$ is nondecreasing in s, $\Omega_i(t, c_i^E t^{e_i}) \leq \mu_i(t)/10$ for sufficiently small $t > 0$. Decrease t_0 so that this holds for $0 < t \leq t_0$. For $t \in [t_0, 1]$ with $K_i(t) \neq \varnothing$,

$$\Omega_i(t, s(t_0)) \leq \Omega_i(t_0, s(t_0)) \leq \mu_i(t_0)/10 \leq \mu_i(t)/10.$$

If $K_i(t)$ is empty, then $\Omega_i(t, s) = 0$. Decreasing c_i^E to at most $s(t_0)$ gives $c_i^E t^{e_i} \leq s(t_0)$ on $[t_0, 1]$, and monotonicity in s gives $\Omega_i(t, c_i^E t^{e_i}) \leq \mu_i(t)/10$ for all $t \in (0, 1]$.

For item 3, rank zero gives $T_{y,f}^{M_i} = T_y M_i$. Fix $t > 0$. If $K_i(t)$ is empty or M_i has no incident higher-dimensional stratum, any positive pair (r, C) is admissible. Otherwise fix an incident stratum R. At each $y_0 \in K_i(t)$, the strict Thom condition (33) gives a radius $\epsilon_{y_0} > 0$ and a constant $C_{y_0} > 0$ for points of M_i and R in $B(y_0, \epsilon_{y_0})$. Choose finitely many centers $y_1, \ldots, y_m$ such that the balls $B(y_j, \epsilon_{y_j}/3)$ cover $K_i(t)$, and set $r_R = \min_j \epsilon_{y_j}/3$ and $C_R = \max_j C_{y_j}$. If $y \in K_i(t)$ and $x \in R$ satisfy $\|x - y\| < r_R$, choose j with $y \in B(y_j, \epsilon_{y_j}/3)$. Then both y and x lie in $B(y_j, \epsilon_{y_j})$, so the strict Thom estimate holds with r_R, C_R. Taking the minimum of these radii and maximum of these constants over the finitely many incident strata R makes the pair (r_R, C_R) valid for all of them. The set of admissible triples (t, r, C) is definable because the strata, $K_i(t)$, the tangent map $T_x M$ and $T_{x,f}^M$ are definable. Definable choice (Coste, 1999, Theorem 3.1) gives positive definable selections $r(t), C(t)$. The growth bound above gives $r(t) \geq \widehat{c}_i t^{b_i^w}$ and $C(t) \leq \widehat{C}_i t^{-s_i}$ for sufficiently small $t > 0$. For $0 < \|x - y\| < \widehat{c}_i t^{b_i^w}$, the selected pair is admissible, so $\delta(T_y M_i, T_{x,f}) \leq C(t)\|x - y\| \leq \widehat{C}_i t^{-s_i} \|x - y\|$. Choose $t_1 \in (0, 1]$ so these power bounds hold for $0 < t \leq t_1$. For $t \in [t_1, 1]$, the inclusion $K_i(t) \subseteq K_i(t_1)$ makes $r(t_1), C(t_1)$ admissible for every incident stratum. Decrease $\widehat{c}_i$ to at most $r(t_1)$ and increase $\widehat{C}_i$ to at least $C(t_1)$. As $t^{b_i^w} \leq 1 \leq t^{-s_i}$, (44) then holds for all $t \in (0, 1]$. □

B.2 Scale-dependent neighborhoods

The next lemma supplies the scale-dependent neighborhoods and Lipschitz bounds used to verify the stratified-descent condition of (Lai & Song, 2026b, Definition 2.3). (Lai & Song, 2026a, Proposition 3.5) establishes corresponding neighborhood and derivative bounds for the stratification furnished by its Proposition 3.2. We apply these geometric estimates to the constant-rank strict-Thom refinement $\mathcal{M}$ from Lemma B.3. The next lemma also gives a relative error bound on rank-one strata and controls projected normalized subgradient directions on rank-zero strata.

For $E \subseteq \mathbb{R}^n$ and $F : E \to \mathbb{R}^m$, write

$$\mathrm{Lip}_E(F) = \sup_{\substack{x,y \in E \\ x \neq y}} \frac{\|F(x) - F(y)\|}{\|x - y\|},$$

with value zero when E has at most one point.

For each i, let

$$J_i = \{j \in \{1, \ldots, T\} : M_j \subseteq \mathrm{fr}\, M_i\}.$$

Lemma B.5. *Fix any compact definable set $X \subseteq \overline{B}(0, R_{\mathrm{orb}})$. There are constants $C > 0$ and $c_0 \in (0, 1]$, constants $c_i, \beta_i, \gamma_i > 0$ and $\kappa_i \in (0, 1]$ for each i, a threshold $\bar{a} \in (0, 1]$, and numbers $\delta_i > 0$ for the rank-zero strata such that, with*

$$\Gamma_i := \max\big(\{0\} \cup \{\gamma_j : j \in J_i\}\big),$$

one has

$$0 < \beta_i < \gamma_i(1-\theta) < 1-\theta, \qquad M_j \subseteq \operatorname{fr} M_i \Longrightarrow \beta_i > \gamma_j. \tag{46}$$

Define

$$N_i(a) = X \cap \left(B(M_i, c_i a^{\beta_i}) \setminus \bigcup_{j \in J_i} B(M_j, c_j a^{\gamma_j}) \right), \qquad 0 < a \le \bar{a}, \tag{47}$$

The threshold $\bar{a}$ also satisfies the cross-scale separation

$$N_i(a) \cap N_j(a') = \varnothing \quad (0 < a, a' \le \bar{a}) \quad \text{if } M_i \not\subseteq \overline{M}_j \text{ and } M_j \not\subseteq \overline{M}_i. \tag{48}$$

For every i, every $0 < a \le \bar{a}$, and every $x \in N_i(a)$, the following also hold:

1. *The projection $y = P_i(x)$ is well defined. If R is the stratum containing x, then*

$$y \in K_i(\kappa_i a^{\Gamma_i}), \qquad M_i \subseteq \overline{R}, \qquad \|x - y\| \le C a^{\beta_i}. \tag{49}$$

2. *Put $t = \kappa_i a^{\Gamma_i}$ and $\mathcal{T}_i(a) = B(K_i(t), c_i^P t^{b_i}/2)$. Then*

$$[x, y] \cup \overline{B}(x, a) \subseteq B(y, c_i^P t^{b_i}/2) \subseteq \mathcal{T}_i(a).$$

3. *If $f|_{M_i}$ has rank one, then, with $r = \|\nabla_{M_i} f(y)\|$, every $v \in \partial f(x)$ satisfies*

$$\|P_{T_y M_i} v - \nabla_{M_i} f(y)\| \le r/10, \tag{50}$$

$$Q\|DP_i(x) - P_{T_y M_i}\| \le r/10, \tag{51}$$

$$a \operatorname{Lip}_{\mathcal{T}_i(a)}(\nabla((f|_{M_i}) \circ P_i))\, Q \le r^2/10, \tag{52}$$

$$a \operatorname{Lip}_{\mathcal{T}_i(a)}(\nabla_{M_i} f \circ P_i) \le c_0 r \tag{53}$$

4. *If $f|_{M_i}$ has rank zero, then every nonzero $v \in \partial f(x)$ satisfies*

$$\left\| P_{T_y M_i} \frac{v}{\|v\|} \right\| \le C a^{\delta_i}. \tag{54}$$

Proof. Choose $H \ge 1$ larger than every applicable exponent among $b_i, e_i, 4p_i, 4m_i, b_i^w, s_i$ appeared in Lemma B.4 and large enough that, for every pair of distinct strata M_i, M_j whose closures meet but neither lies in the closure of the other, there is $A_{ij} > 0$ with

$$d(y, M_j) \ge A_{ij} \rho_i(y)^H \qquad (y \in M_i).$$

The definable distance inequality in (Lai & Song, 2026a, Proposition 3.5, proof of Step 3) gives a power bound for each pair. There are finitely many pairs and each ρ_i is bounded on M_i, so increasing the exponents to one common H only changes the constants A_{ij}. Choose $\lambda \in (0, 1)$ with $H\lambda < 1-\theta$ and define

$$\gamma_i = \lambda^{n - \dim M_i + 1}.$$

If $j \in J_i$, the frontier condition gives $\dim M_j < \dim M_i$, thus $\gamma_j \le \lambda \gamma_i$ and $\Gamma_i \le \lambda \gamma_i$. Consequently, for every i one can choose

$$H\Gamma_i < \beta_i < (1-\theta)\gamma_i.$$

Since $H \ge 1$, these choices give $\beta_i > \gamma_j$ for $j \in J_i$ and $0 < \beta_i < \gamma_i(1-\theta) < 1-\theta$, as in (46). They also give $a^{\beta_i} = o(a^{H\Gamma_i})$ as $a \downarrow 0$. Choose the constants $c_i > 0$ as in (Lai & Song, 2026a, Proposition 3.5, Steps 2–3), with ambient set K^+, using the distance inequality above, the frontier condition, and the projection domains in Lemma B.4(1). Define

$$N_0(i, a) = X \cap B\left(M_i \setminus \bigcup_{j \in J_i} B(M_j, c_j a^{\gamma_j}/2), 2c_i a^{\beta_i} \right).$$

Initially set $\bar{a} = 1$. Intersect the neighborhoods in the cited construction with X and write

$$\widetilde{N}_0(i,s) = X \cap \overline{B}\left(M_i \setminus \bigcup_{j \in J_i} \overline{B}(M_j, c_j s^{\gamma_j}/2), 2c_i s^{\beta_i} \right).$$

For $x \in N_0(i,a)$, choose a point z in its defining central set with $\|x - z\| < 2c_i a^{\beta_i}$. Since $\beta_i > 0$, choose

$$\left(\frac{\|x-z\|}{2c_i} \right)^{1/\beta_i} < s < a.$$

Then $\|x - z\| < 2c_i s^{\beta_i}$. For every $j \in J_i$, positivity of γ_j gives

$$d(z, M_j) \geq c_j a^{\gamma_j}/2 > c_j s^{\gamma_j}/2.$$

Thus z belongs to the central set defining $\widetilde{N}_0(i,s)$, and $x \in \widetilde{N}_0(i,s)$. Consequently,

$$N_0(i,a) \subseteq \bigcup_{0<s<a} \widetilde{N}_0(i,s).$$

(Lai & Song, 2026a, Proposition 3.5, Step 4) shows that P_i is single valued on $\bigcup_{0<s\leq 1} \widetilde{N}_0(i,s)$. (Lai & Song, 2026a, Proposition 3.5, Step 5) gives

$$\left(\bigcup_{0<s\leq 1} \widetilde{N}_0(i,s) \right) \cap \left(\bigcup_{0<s'\leq 1} \widetilde{N}_0(j,s') \right) = \varnothing$$

when neither M_i nor M_j lies in the closure of the other. (Lai & Song, 2026a, Proposition 3.5, Step 6) gives $M_i \subseteq \overline{R}$ whenever $x \in \widetilde{N}_0(i,s) \cap R$ for some $0 < s \leq 1$. The inclusion above gives single-valued projection and this incidence property on $N_0(i,a)$; applying it to both i, a and j, a' gives $N_0(i,a) \cap N_0(j,a') = \varnothing$ for the indicated pairs and all $0 < a, a' \leq \bar{a}$. Fix i, $0 < a \leq \bar{a}$, and $x \in N_i(a)$. Choose $z \in M_i$ with $\|x - z\| < 2c_i a^{\beta_i}$. Since x lies outside every deleted ball and $\beta_i > \gamma_j$ for $j \in J_i$, decreasing $\bar{a}$ gives

$$d(z, M_j) \geq c_j a^{\gamma_j} - 2c_i a^{\beta_i} \geq \frac{c_j}{2} a^{\gamma_j}.$$

Thus $N_i(a) \subseteq N_0(i,a)$, so the three conclusions above hold on $N_i(a)$. Applying this inclusion for both i, a and j, a' to the cross-scale separation of N_0 proves (48). Moreover, $\|x - P_i(x)\| \leq c_i a^{\beta_i}$. For every $j \in J_i$, reducing $\bar{a}$ gives

$$d(P_i(x), M_j) \geq c_j a^{\gamma_j} - c_i a^{\beta_i} \geq \frac{c_j}{2} a^{\gamma_j}.$$

The frontier condition gives $\operatorname{fr} M_i = \bigcup_{j \in J_i} M_j$. Choose $\kappa_i \in (0,1]$ with $\kappa_i \leq c_j/2$ for all $j \in J_i$. Since $\gamma_j \leq \Gamma_i$ and $a \leq 1$, these inequalities give $P_i(x) \in K_i(\kappa_i a^{\Gamma_i})$, completing (49). The parameter t from item 2 satisfies $0 < t \leq \kappa_i \leq 1$, since $\Gamma_i \geq 0$, $0 < a \leq \bar{a} \leq 1$, and $0 < \kappa_i \leq 1$. Since $b_i \Gamma_i < \beta_i < 1$, one has $\|x - y\| + a \leq c_i a^{\beta_i} + a = o(t^{b_i})$. Decrease $\bar{a}$ so that this upper bound is less than $c_i^P t^{b_i}/2$. Every $z \in [x,y] \cup \overline{B}(x,a)$ then satisfies $d(z, K_i(t)) \leq \|z - y\| \leq \|x - y\| + a < c_i^P t^{b_i}/2$, which proves both inclusions in item 2.

The choice $H\Gamma_i < \beta_i$ gives, whenever the corresponding exponents are defined,

$$e_i \Gamma_i < \beta_i, \qquad p_i \Gamma_i < \frac{\beta_i}{4}, \qquad m_i \Gamma_i < \frac{\beta_i}{4}, \qquad b_i^w \Gamma_i < \beta_i, \qquad s_i \Gamma_i < \beta_i.$$

On a rank-one stratum, put $r = \|\nabla_{M_i} f(y)\|$. By item 2 of Lemma B.4, $r \geq \mu_i(t) \geq c_i^{\mu} \kappa_i^{p_i} a^{p_i \Gamma_i}$. Since $\|x - y\| \leq c_i a^{\beta_i}$ and $e_i \Gamma_i < \beta_i$, decreasing $\bar{a}$ ensures $\|x - y\| \leq c_i^E t^{e_i}$. Thus (43) gives (50). On $\mathcal{T}_i(a)$, set $L_i = C_i \kappa_i^{-m_i}$. Item 1 of Lemma B.4 bounds the relevant Lipschitz moduli by $L_i a^{-m_i \Gamma_i}$. Since $DP_i(y) = P_{T_y M_i}$, it follows that

$$Q \|DP_i(x) - P_{T_y M_i}\| \leq Q L_i c_i a^{\beta_i - m_i \Gamma_i},$$

$$a \operatorname{Lip}_{\mathcal{T}_i(a)} (\nabla((f|_{M_i}) \circ P_i))\, Q \leq Q L_i a^{1 - m_i \Gamma_i},$$

$$a \operatorname{Lip}_{\mathcal{T}_i(a)} (\nabla_{M_i} f \circ P_i) \leq L_i a^{1 - m_i \Gamma_i}.$$

The first bound is at most $r/10$ because $\beta_i > (m_i + p_i)\Gamma_i$; the second is at most $r^2/10$ because $1 > (m_i + 2p_i)\Gamma_i$; and the third is at most $c_0 r$ because $1 > (m_i + p_i)\Gamma_i$, after decreasing $\bar{a}$. These are (51) to (53).

For rank-zero M_i, if the stratum R containing x equals M_i, then (38) gives $P_{T_y M_i} v = 0$. Otherwise, for $u = v/\|v\|$, the same identity gives $u \in T_{x,f}^{\perp}$. Since $\|x - y\| \le c_i a^{\beta_i}$ and $\beta_i > b_i^w \Gamma_i$, decrease $\bar{a}$ until $c_i a^{\beta_i} < \widehat{c}_i t^{b_i^w}$. Then (44) gives

$$\|P_{T_y M_i} u\| \le \delta(T_y M_i, T_{x,f}) \le \widehat{C}_i t^{-s_i} \|x - y\| \le \widehat{C}_i c_i \kappa_i^{-s_i} a^{\beta_i - s_i \Gamma_i}.$$

Set $\delta_i = \beta_i - s_i \Gamma_i > 0$. Since there are finitely many strata, choose $C > 0$ with $C \ge c_i$ for every i and $C \ge \widehat{C}_i c_i \kappa_i^{-s_i}$ for every rank-zero i. These choices give (49) and (54). □

B.3 Projected block estimates

Lemma B.6. *Let ℓ and ϵ be as in* (39)*. Let (a_k) be any sequence with $0 < a_k \le \bar{a}$, and let (x_k) be any trajectory of* (5) *with stepsizes (a_k). Let $0 \le p < q$, and suppose $x_k \in N_i(a_k)$ for $p \le k \le q$, where $f|_{M_i}$ has rank one. Put $y_k = P_i(x_k)$ and $s_k = f(y_k) - \ell$, and assume $|s_k| \le \epsilon$ for $p \le k \le q$. Then*

$$\sum_{k=p}^{q-1} \|y_{k+1} - y_k\| \le \psi(s_p) - \psi(s_q) + 2 \max_{p \le k < q} a_k. \tag{55}$$

Proof. For $p \le k < q$, put $u_k = (x_k - x_{k+1})/a_k \in \mathcal{N}_C f(x_k)$. If $x_{k+1} = x_k$, then $y_{k+1} = y_k$. Otherwise, the definition of $\mathcal{N}_C f(x_k)$ supplies $0 \ne v_k \in \partial f(x_k)$ with $u_k = v_k/\|v_k\|$. Write

$$g_k = \nabla_{M_i} f(y_k), \quad r_k = \|g_k\|, \quad q_k = \|v_k\|. \tag{56}$$

Write $T_k = T_{y_k} M_i$ and $D_k = DP_i(x_k)$. By (50), define $e_k = P_{T_k} v_k - g_k$. Since $x_k \in N_i(a_k) \subseteq X \subseteq K^+$, (30) applies to v_k. The reverse triangle inequality and the norm bound for an orthogonal projection give

$$\|e_k\| \le \frac{r_k}{10}, \qquad \frac{9}{10} r_k \le r_k - \|e_k\| \le \|g_k + e_k\| = \|P_{T_k} v_k\| \le q_k \le Q. \tag{57}$$

Since $f|_{M_i}$ has rank one, $r_k > 0$ and hence $q_k > 0$. Set $h_i = (f|_{M_i}) \circ P_i$. The chain rule gives $\nabla h_i(x_k) = D_k^{\top} g_k$.

$$\begin{aligned} \langle \nabla h_i(x_k), u_k \rangle &= \langle g_k, D_k u_k \rangle = \langle g_k, P_{T_k} u_k \rangle + \langle g_k, (D_k - P_{T_k}) u_k \rangle \\ &\overset{\text{(a)}}{=} \frac{\langle g_k, g_k + e_k \rangle}{q_k} + \langle g_k, (D_k - P_{T_k}) u_k \rangle \\ &\overset{\text{(b)}}{\ge} \frac{r_k^2 - r_k \|e_k\|}{q_k} - r_k \|D_k - P_{T_k}\| \\ &\overset{\text{(c)}}{\ge} \frac{9 r_k^2}{10 q_k} - \frac{r_k^2}{10 Q} \overset{\text{(d)}}{\ge} \frac{9 r_k^2}{10 q_k} - \frac{r_k^2}{10 q_k} = \frac{4}{5} \frac{r_k^2}{q_k}. \end{aligned} \tag{58}$$

Here (a) uses $u_k = v_k/q_k$ and $P_{T_k} v_k = g_k + e_k$. Inequality (b) follows from $r_k = \|g_k\|$, the Cauchy–Schwarz inequality, and $\|u_k\| = 1$. In (c), $\|e_k\| \le r_k/10$ by (57), while $Q\|D_k - P_{T_k}\| \le r_k/10$ by (51). Inequality (d) uses $q_k \le Q$ by (30), which gives $r_k^2/(10Q) \le r_k^2/(10 q_k)$.

The segment $[x_k, x_{k+1}]$ lies in $\mathcal{T}_i(a_k)$ by item 2 of Lemma B.5, and $x_{k+1} - x_k = -a_k u_k$ has length a_k. Put $L_k = \mathrm{Lip}_{\mathcal{T}_i(a_k)}(\nabla h_i)$. By (52) and $q_k \le Q$,

$$\frac{a_k^2 L_k}{2} \le \frac{a_k r_k^2}{20 Q} \le \frac{a_k r_k^2}{20 q_k}.$$

Descent lemma and (58) now yield

$$\begin{aligned} s_{k+1} - s_k = h_i(x_{k+1}) - h_i(x_k) &\le -a_k \langle \nabla h_i(x_k), u_k \rangle + \frac{a_k^2 L_k}{2} \\ &\le -\frac{4}{5} a_k \frac{r_k^2}{q_k} + \frac{1}{20} a_k \frac{r_k^2}{q_k} = -\frac{3}{4} a_k \frac{r_k^2}{q_k} \le -\frac{1}{2} a_k \frac{r_k^2}{q_k}. \end{aligned} \tag{59}$$

Notice that the normal space $N_{y_k}M_i = (T_{y_k}M_i)^\perp$ and we set

$$z_k = x_k - a_k P_{N_{y_k}M_i} u_k. \tag{60}$$

Since $\|z_k - x_k\| \le a_k$, item 2 of Lemma B.5 puts z_k in $B(y_k, c_i^P t^{b_i}/2)$, where $t = \kappa_i a_k^{\Gamma_i}$. First-order optimality of squared distance at $y_k = P_i(x_k)$ gives $x_k - y_k \perp T_{y_k}M_i$. The definition of z_k then gives $z_k - y_k \perp T_{y_k}M_i$. Equation (41) yields $P_i(z_k) = y_k$. Both z_k and x_{k+1} lie in $\overline{B}(x_k, a_k) \subseteq \mathcal{T}_i(a_k)$. Moreover, (57) gives $\|P_{T_{y_k}M_i} u_k\| \le 11 r_k/(10 q_k)$. The 2-Lipschitz bound for P_i therefore gives

$$\|y_{k+1} - y_k\| \le 2a_k \|P_{T_{y_k}M_i} u_k\| \le \frac{11}{5} a_k \frac{r_k}{q_k}. \tag{61}$$

Since $x_k, x_{k+1} \in \mathcal{T}_i(a_k)$, the gradient variation bound in (53) gives

$$|r_{k+1} - r_k| \le \|\nabla_{M_i} f(P_i(x_{k+1})) - \nabla_{M_i} f(P_i(x_k))\| \le a_k \operatorname{Lip}_{\mathcal{T}_i(a_k)}(\nabla_{M_i} f \circ P_i) \le c_0 r_k.$$

Since $c_0 \le 1$, this yields

$$r_{k+1} \le 2 r_k. \tag{62}$$

On a nonconstant edge, (59) gives $s_{k+1} < s_k$. If $s_k \ge s_{k+1} \ge 0$, then $s_k > 0$. Concavity of ψ on the positive half-line and (39) and (59) give

$$\psi(s_k) - \psi(s_{k+1}) \ge \psi'(s_k)(s_k - s_{k+1}) \ge \frac{12}{r_k} \frac{a_k r_k^2}{2 q_k} = 6 a_k \frac{r_k}{q_k}. \tag{63}$$

If $0 \ge s_k \ge s_{k+1}$, then $s_{k+1} < 0$. Convexity on the negative half-line, the gradient inequality at y_{k+1}, and (62) give

$$\psi(s_k) - \psi(s_{k+1}) \ge \psi'(s_{k+1})(s_k - s_{k+1}) \ge \frac{12}{r_{k+1}} \frac{a_k r_k^2}{2 q_k} \ge 3 a_k \frac{r_k}{q_k}.$$

By (61), the projected length on either kind of edge is at most $11 a_k r_k/(5 q_k)$, hence at most the drop in ψ. A decreasing scalar block has at most one edge with $s_k > 0 > s_{k+1}$. On that edge the first inequality of (61) and $\|u_k\| = 1$ give $\|y_{k+1} - y_k\| \le 2a_k$. Stopped edges have zero projected length and zero change in ψ. Since ψ is increasing, discarding the drop on the crossing edge and summing the remaining drops gives (55). □

Lemma B.7. *Let C be the constant given in Lemma B.5. Let (a_k) be any sequence with $0 < a_k \le \bar{a}$, and let (x_k) be any trajectory of* (5) *with stepsizes (a_k). Let $0 \le p < q$ be nonnegative integers and suppose that $x_k \in N_i(a_k)$ for $p \le k \le q$, where $f|_{M_i}$ has rank zero. Then*

$$\sum_{k=p}^{q-1} \|P_i(x_{k+1}) - P_i(x_k)\| \le 2C \sum_{k=p}^{q-1} a_k^{1+\delta_i}. \tag{64}$$

Proof. Put $y_k = P_i(x_k)$. If $x_{k+1} = x_k$, then $y_{k+1} = y_k$. On a nonconstant step, put $u_k = (x_k - x_{k+1})/a_k \in \mathcal{N}_C f(x_k)$ and set $z_k = x_k - a_k P_{(T_{y_k}M_i)^\perp} u_k$. Item 2 of Lemma B.5 puts z_k and x_{k+1} in $B(y_k, c_i^P t^{b_i}/2)$, where $t = \kappa_i a_k^{\Gamma_i}$. The condition $y_k = P_i(x_k)$ gives $x_k - y_k \perp T_{y_k}M_i$, so $z_k - y_k \perp T_{y_k}M_i$. (41) yields $P_i(z_k) = y_k$. The projection is 2-Lipschitz there, so (54) yields

$$\|P_i(x_{k+1}) - P_i(x_k)\| \le 2a_k \|P_{T_{y_k}M_i} u_k\| \le 2C a_k^{1+\delta_i}. \tag{65}$$

On a constant step, $y_{k+1} = y_k$, so $\|P_i(x_{k+1}) - P_i(x_k)\| = 0 \le 2C a_k^{1+\delta_i}$. Summing this last bound from (65) over $k = p, \dots, q-1$ yields (64). □

B.4 The diameter estimate and convergence

Set $\eta = \epsilon/4$, where ℓ, ϵ are fixed in (39), and choose

$$X = \overline{B}(0, R_{\mathrm{orb}}) \cap [|f - \ell| \le 3\eta]. \tag{66}$$

This set is compact and definable. Let C and $\bar{a}$ be the constants given by Lemma B.5 for this X. Since there are finitely many strata and each $\beta_i > 0$, decrease $\bar{a}$ so that

$$C a^{\beta_i} \le \min\{1, \eta/Q\} \quad \text{for every } i \text{ and } 0 < a \le \bar{a}.$$

For $x \in N_i(a)$, put $y = P_i(x)$. Then $\|x - y\| \le Ca^{\beta_i} \le 1$, so $x \in X \subseteq \overline{B}(0, R_{\mathrm{orb}})$ implies $y \in K^+$. For every rank-one M_i,

$$|f(y) - \ell| \le |f(y) - f(x)| + |f(x) - \ell| \le QCa^{\beta_i} + 3\eta \le 4\eta = \epsilon.$$

Thus the chosen X and $\bar{a}$ satisfy

$$|f(P_i(x)) - \ell| \le \epsilon \quad \text{for every rank-one } M_i, \quad 0 < a \le \bar{a}, \quad x \in N_i(a). \tag{67}$$

Since $X \subseteq \overline{B}(0, R_{\mathrm{orb}})$, $\overline{B}(x, a)$ lies in K^+ whenever $x \in X$ and $0 < a \le \bar{a} \le 1$.

Let

$$\mathsf{F}(x, a) = -a\mathcal{N}_C f(x). \tag{68}$$

Fix any step sequence (a_k) with $0 < a_k \le \bar{a}$ and any realization (x_k) of the difference inclusion $x_{k+1} \in x_k + \mathsf{F}(x_k, a_k)$. The two block lemmas above apply to every such realization. The Lipschitz potential supplied to (Lai & Song, 2026b, Definition 2.3) is $\widetilde{f} = (f - \ell)|_{K^+}$, and the ambient compact set in that definition is X. The function $\widetilde{f}$ is Lipschitz on the compact ball K^+, and its projected values are the scalars s_k used above. Set

$$\delta = \min\big(\{1\} \cup \{\delta_i : i \in \{1, \dots, T\},\ f|_{M_i} \text{ has rank zero}\}\big) > 0 \tag{69}$$

Set $Q_0 = 2C$. For $0 < a \le \bar{a} \le 1$ and every rank-zero stratum, $\delta \le \delta_i$ gives $2Ca^{1+\delta_i} \le Q_0 a^{1+\delta}$. For the error functions in (Lai & Song, 2026b, Definition 2.3), define

$$g_1((b_m), (e_m)) = 0, \qquad g_2((b_m), (e_m)) = Q_0 b_0^{1+\delta}. \tag{70}$$

The perturbation sequence is zero.

The increment bound is

$$\|x_{k+1} - x_k\| \le a_k, \tag{71}$$

so item (i) of (Lai & Song, 2026b, Definition 2.3) holds with $\tau = 1$ and $\iota = 2$; the sequence $g_{1,k}$ is zero. The cross-scale separation is (48); the exponent and projection requirements are (46) and (49). Lemma B.6 gives the projected block bound with the remainder $2 \max a_k^\tau$. On a rank-zero stratum M_i, $\nabla_{M_i} f = 0$. Given $y, y' \in M_i$, connectedness supplies a piecewise C^1 path $\gamma : [0, 1] \to M_i$ from y to y'. On each smooth piece, $(f \circ \gamma)' = \langle \nabla_{M_i} f(\gamma), \dot{\gamma} \rangle = 0$. The fundamental theorem of calculus on each piece and continuity at the finitely many joins give $f(y) = f(y')$. Hence the ψ drop along a rank-zero block is zero. The bound in Lemma B.7 is at most the accumulated $g_{2,k}$ term by (70). Hence (68) exhibits stratified descent in X with the displayed data.

Return to the harmonic trajectory fixed at the start of this section. Its steps $a_k = c/(k + 1)$ satisfy $\sup_{j \ge i} a_j / a_i = 1$ for every i. Let $\underline{\beta} = \min_i \beta_i$. (Lai & Song, 2026b, Theorem 2.4), with step-ratio bound $c_d = 1$, gives a step cap $\alpha_{\mathrm{diam}} > 0$ and a constant $D_{\mathrm{diam}} > 0$. Every segment $x_{[p,q]} = \{x_p, \dots, x_q\} \subseteq X$ with $a_k \le \min\{\bar{a}, \alpha_{\mathrm{diam}}\}$ for $k \ge p$ satisfies

$$\begin{aligned} \operatorname{diam} x_{[p,q]} &\le 2\{\psi(f(x_p) - \ell) - \psi(f(x_q) - \ell)\} \\ &\quad + 2Q_0 \sum_{k=p}^{q-1} a_k^{1+\delta} + D_{\mathrm{diam}} a_p^{\underline{\beta}(1-\theta)}. \end{aligned} \tag{72}$$

Moreover,

$$\sum_{k=0}^{\infty} a_k^{1+\delta} < \infty, \qquad a_k^{\underline{\beta}(1-\theta)} \to 0. \tag{73}$$

Proof of Theorem 2.3. The harmonic steps satisfy (6). The set X and threshold $\bar{a}$ were chosen in (66) and (67). By (7) and $a_k \to 0$, there is an index p_0 such that, for every $k \ge p_0$, $|f(x_k) - \ell| \le 2\eta$ and $a_k \le \min\{\bar{a}, \alpha_{\mathrm{diam}}\}$. By the definition of R_{orb}, every x_k lies in $\overline{B}(0, R_{\mathrm{orb}})$, so $x_k \in X$ for $k \ge p_0$. Thus (72) applies to every finite segment $x_{[p,q]}$ with $q > p \ge p_0$. Uniformly in $q > p$, its three terms tend to zero as $p \to \infty$:

$$\psi(f(x_p) - \ell) - \psi(f(x_q) - \ell) \to 0, \qquad \sum_{k=p}^{q-1} a_k^{1+\delta} \le \sum_{k=p}^{\infty} a_k^{1+\delta} \to 0,$$

and $a_p^{\beta(1-\theta)} \to 0$. Thus

$$\sup_{q>p} \operatorname{diam} x_{[p,q]} \to 0.$$

The sequence is Cauchy and therefore converges to an accumulation point x_*. The stationarity of x_* follows from (7). □

C PLANAR SQUARE-SUMMABLE CONVERGENCE

In the plane, bounded NGD trajectories converge under the stepsize conditions (8). The proof uses the local geometry near a one-dimensional stratum of the critical set.

Proof of Theorem 2.4. If $\partial f(x_{k_0}) = \{0\}$ at some index k_0, then (5) gives $x_k = x_{k_0}$ for every $k \geq k_0$. We therefore consider the case $\partial f(x_k) \neq \{0\}$ for every k. Square summability gives $\alpha_k \to 0$.

Step 1: the critical set and its strata. Let C_f be the critical set defined in (31). Theorem 2.1 gives $\ell \in \mathbb{R}$ such that

$$f(x_k) \to \ell, \qquad \Omega(x) \subseteq C_f \cap \{x \in \mathbb{R}^2 : f(x) = \ell\}, \tag{74}$$

and

$$d(x_k, C_f) \to 0. \tag{75}$$

The trajectory is bounded, and $\|x_{k+1} - x_k\| \leq \alpha_k \to 0$. (Bolte et al., 2014, Lemma 3.5(ii)–(iii) and Remark 3.3) therefore gives a nonempty compact connected set $\Omega(x)$ of accumulation points with $d(x_k, \Omega(x)) \to 0$.

Choose a closed ball K^+ containing the orbit closure in its interior. On this ball take the C_f-compatible stratification $\mathcal{P}$ specified before Lemma B.3, and apply that lemma. The resulting finite C^3 stratification $\mathcal{M}$ is compatible with C_f, has constant rank on each stratum, and satisfies strict Thom (w_f) on incident pairs. Apply Lemma B.4 to $\mathcal{M}$ to fix c_j^P, b_j for every stratum and $\widehat{c}_j, \widehat{C}_j, b_j^w, s_j$ for its rank-zero strata.

If $\Omega(x)$ meets a two-dimensional stratum $M \subseteq C_f$, that stratum is open in $\mathbb{R}^2$ near that accumulation point z. There f is C^3 and $\partial f = \{\nabla f\} = \{0\}$ near z. Then any x_k close to z would satisfy $\partial f(x_k) = \{0\}$, which has been excluded in the beginining. Thus $\Omega(x)$ meets no two-dimensional critical stratum. If it also meets no one-dimensional stratum contained in C_f, (74) shows that $\Omega(x)$ is a finite union of zero-dimensional critical strata. Connectedness makes $\Omega(x)$ a singleton, and $d(x_k, \Omega(x)) \to 0$ gives convergence. We now fix

$$x^* \in \Omega(x) \cap S, \tag{76}$$

where $S = M_i \subseteq C_f$ is a one-dimensional stratum of $\mathcal{M} = \{M_1, \ldots, M_T\}$ for some i.

Step 2: coordinates near a one-dimensional critical stratum. Choose a C^3 embedding $\gamma : (-3, 3) \to S$ with $\gamma(0) = x^*$ and $\gamma([-2, 2]) \subseteq \operatorname{int} K^+$, and set $S_0 = \gamma([-1, 1])$. Since $\gamma([-2, 2])$ is a compact subset of S, its distance from $\operatorname{fr} S$ is positive when this frontier is nonempty. The definition of $K_i(t)$ therefore gives $\gamma([-2, 2]) \subseteq K_i(t_0)$ for some $t_0 \in (0, 1]$, and when the frontier is empty, take $t_0 = 1$. The frontier condition and $\dim S = 1$ give $\overline{M}_j \cap S = \varnothing$ for every $j \neq i$ with $\dim M_j \leq 1$. Since $\mathcal{M}$ is finite and $\gamma([-2, 2])$ is compact, with $d(A, \varnothing) = +\infty$ we have

$$\delta_{\text{low}} := d\left(\gamma([-2, 2]), \bigcup_{\substack{1 \leq j \leq T,\ j \neq i \\ \dim M_j \leq 1}} \overline{M}_j\right) > 0. \tag{77}$$

With $J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, take the C^2 unit normal $n(t) = J\gamma'(t)/\|\gamma'(t)\|$ along γ (Pinkall & Gross, 2024, Section 3.1), and set

$$\Phi(t, v) = \gamma(t) + vn(t).$$

The formula for n gives $\det D\Phi(t, 0) = \|\gamma'(t)\| > 0$. The minimum of this determinant on $[-2, 2]$ is positive. Uniform continuity of $\det D\Phi$ on $[-2, 2] \times [-1, 1]$ gives $\eta_1 > 0$ such that $D\Phi(t, v)$ is invertible whenever $t \in [-2, 2]$ and $|v| \leq \eta_1$.

There is also $\eta_2 > 0$ such that Φ is injective on $[-2,2] \times (-\eta_2, \eta_2)$. Otherwise, for each integer $j \geq 1$ there are distinct pairs $(t_j, v_j), (s_j, w_j)$ in $[-2,2] \times (-1/j, 1/j)$ with equal images under Φ. After passing to subsequences, compactness gives $t_j \to t$ and $s_j \to s$ in $[-2,2]$. Continuity yields $\gamma(t) = \gamma(s)$, so $t = s$ because γ is injective. Both pairs then converge to $(t, 0)$, where the inverse function theorem gives a neighborhood on which Φ is injective. For large j both pairs lie in this neighborhood, contradicting their equal images.

Choose $\rho_0 > 0$ so that $2\rho_0$ is smaller than $\eta_1, \eta_2, c_i^P t_0^{b_i}, \delta_{\text{low}}$ and $d(\gamma([-2,2]), \mathbb{R}^2 \setminus \operatorname{int} K^+)$. Decrease ρ_0 further so that

$$\rho_0 < \widehat{c}_i t_0^{b_i^w}, \qquad \widehat{C}_i t_0^{-s_i} \rho_0 \leq \frac{1}{2}.$$

Define $V_{\text{wide}} = \Phi((-2,2) \times (-2\rho_0, 2\rho_0))$ and $V = \Phi((-1,1) \times (-\rho_0, \rho_0))$. The inverse Φ^{-1} is C^2 on V_{wide}. With $\pi_2(t,v) = v$, set $P = P_i|_{V_{\text{wide}}}$ and

$$r := \pi_2 \circ \Phi^{-1} : V_{\text{wide}} \to (-2\rho_0, 2\rho_0).$$

Then

$$P(\Phi(t,v)) = \gamma(t), \qquad \nabla r(\Phi(t,v)) = n(t), \qquad x = P(x) + r(x)\nabla r(x). \tag{78}$$

The first identity follows from (41). For the second, the chain rule and $r \circ \Phi(t,v) = v$ give

$$D\Phi(t,v)^\top \nabla r(\Phi(t,v)) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Since $\langle \gamma'(t), n(t) \rangle = 0$ and $2\langle n'(t), n(t) \rangle = (\|n(t)\|^2)' = 0$, the columns $\gamma'(t) + vn'(t)$ and $n(t)$ of $D\Phi(t,v)$ give

$$D\Phi(t,v)^\top n(t) = \begin{bmatrix} \langle \gamma'(t) + vn'(t), n(t) \rangle \\ \|n(t)\|^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Invertibility of $D\Phi(t,v)$ yields $\nabla r(\Phi(t,v)) = n(t)$. The third identity is $x = \Phi(t,v) = \gamma(t) + vn(t)$ with $\gamma(t) = P(x)$ and $v = r(x)$. For $\sigma \in \{-1, 1\}$, the sets $V_\sigma = \Phi((-1,1) \times \{v : 0 < \sigma v < \rho_0\})$ are the two connected components of $V \setminus S$. No zero- or one-dimensional stratum other than S meets V by $2\rho_0 \leq \delta_{\text{low}}$ and the definition of δ_{low} in (77). Every point of V_σ therefore lies on a two-dimensional stratum in $\mathcal{M}$. Each two-dimensional stratum is open in $\mathbb{R}^2$, so the sets $V_\sigma \cap M_j$ form a disjoint open partition of V_σ. Connectedness makes exactly one of these sets nonempty, and hence $V_\sigma \subseteq M_{i_\sigma}$ for some i_σ. Compatibility of $\mathcal{M}$ with C_f makes V_σ either critical or noncritical.

Since $0 \in \partial f(y)$ for $y \in S$, (38) gives

$$\nabla_S f(y) = 0, \qquad P_{T_y S} v = 0 \quad \big(v \in \partial f(y)\big). \tag{79}$$

If V_σ is noncritical, set $R = M_{i_\sigma}$. Since $\gamma((-1,1)) \subseteq \overline{V_\sigma} \subseteq \overline{R}$, the frontier condition gives $S \subseteq \overline{R}$. At each $x \in V_\sigma$, f is C^3 near x since $V_\sigma \subseteq M_{i_\sigma}$ and $f \in C^3(M_{i_\sigma})$. Since V_σ is noncritical, $0 \notin \partial f(x) = \{\nabla f(x)\}$, so $\nabla f(x) \neq 0$. Set

$$u(x) = \frac{\nabla f(x)}{\|\nabla f(x)\|}. \tag{80}$$

For $x \in V_\sigma$, (38) gives $u(x) \perp T^R_{x,f}$. The first inequality of (34), with $M = S$ and $u = u(x)$, applies. With $y = P(x)$ and $\|x - y\| = |r(x)|$, (44) gives the second inequality:

$$\big\|P_{T_y S} u(x)\big\| \leq \delta(T_y S, T^R_{x,f}) \leq \widehat{C}_i t_0^{-s_i} |r(x)|. \tag{81}$$

The choice of ρ_0 gives $\widehat{C}_i t_0^{-s_i} |r(x)| \leq 1/2$ on V. Since $\|u(x)\| = 1$, the normal component $A(x) := \langle u(x), \nabla r(x) \rangle$ satisfies $|A(x)| \geq \sqrt{3}/2$ and has constant sign on each noncritical V_σ. The set $W = \Phi([-3/2, 3/2] \times [-3\rho_0/2, 3\rho_0/2])$ is compact, satisfies $\overline{V} \subseteq \operatorname{int} W$ and $W \subseteq V_{\text{wide}}$. Set $\eta = d(\overline{V}, \mathbb{R}^2 \setminus W) > 0$ and choose k_0 so that $\alpha_k < \eta$ for $k \geq k_0$. If $k \geq k_0$ and $x_k \in V_\sigma$ with V_σ noncritical, put $h_k = -\alpha_k u(x_k)$. For $s \in [0,1]$,

$$d(x_k + sh_k, \overline{V}) \leq s\alpha_k < \eta,$$

so the segment $[x_k, x_{k+1}]$ lies in W. For $g \in \{r, P\}$, Taylor's integral formula on this segment gives

$$g(x_k + h_k) = g(x_k) + Dg(x_k)h_k + \int_0^1 (1-s) D^2 g(x_k + sh_k)[h_k, h_k]\, ds.$$

The first-order term for r is $-\alpha_k A(x_k)$. Writing $x_k = \Phi(t_k, v_k)$, differentiation of $P(\Phi(t,v)) = \gamma(t)$ in v gives $DP(x_k)n(t_k) = 0$. Since $n(t_k)$ spans the normal line, $DP(x_k)u(x_k) = DP(x_k)P_{T_{P(x_k)}S}u(x_k)$. By (81),

$$\alpha_k\|DP(x_k)u(x_k)\| \le \alpha_k\|DP(x_k)\|\|P_{T_{P(x_k)}S}u(x_k)\| \le \alpha_k\|DP(x_k)\|\widehat{C}_i t_0^{-s_i}|r(x_k)|.$$

Since W is compact and r, P are C^2 on V_{wide}, we can choose $L \ge 1$ such that

$$L \ge \sup_{z\in W}\max\{\|Dr(z)\|, \|DP(z)\|, \|D^2r(z)\|, \|D^2P(z)\|\},$$
$$L \ge \widehat{C}_i t_0^{-s_i}\sup_{z\in W}\|DP(z)\|.$$

The integral remainders have norm at most $L\alpha_k^2/2$. Thus

$$r(x_{k+1}) = r(x_k) - \alpha_k A(x_k) + e_k, \qquad |e_k| \le L\alpha_k^2, \tag{82}$$
$$\|P(x_{k+1}) - P(x_k)\| \le L\alpha_k|r(x_k)| + L\alpha_k^2. \tag{83}$$

If $x_k \in S$, choose the nonzero $v_k \in \partial f(x_k)$ that generates the update and put $u_k = v_k/\|v_k\|$. (78) and (79) give $u_k = \pm\nabla r(x_k)$. Once $\alpha_k < \rho_0$, the identity (78) gives

$$P(x_{k+1}) = x_k, \qquad |r(x_{k+1})| = \alpha_k. \tag{84}$$

Step 3: excluding one sign of the normal component. For $A \subseteq B \subseteq \mathbb{R}^2$, write $\mathrm{int}_B A$ for the interior of A in the subspace topology of B. Choose compact subarcs $J_0, J_1 \subseteq S_0$ with $x^* \in \mathrm{int}_S J_0$, $J_0 \subseteq \mathrm{int}_S J_1$, and $J_1 \subseteq \mathrm{int}_S S_0$. Put $\delta = d(J_0, \gamma([-3/2, 3/2]) \setminus \mathrm{int}_S J_1) > 0$. If V_σ is noncritical, then A has constant sign and $|A| > 1/2$ on V_σ, so either $\sigma A \le -1/2$ or $\sigma A \ge 1/2$ throughout V_σ. Write $Q_m = \sum_{k=m}^{\infty}\alpha_k^2$. Choose $\rho > 0$ so that $2\rho < \rho_0$, $\{y + v\nabla r(y) : y \in J_1,\ |v| \le 2\rho\} \subseteq \mathrm{int}\, W$ and $8L\rho^2 < \delta/4$.

Suppose V_σ is noncritical, $\sigma A \le -1/2$ on V_σ, and infinitely many iterates $x_m \in V_\sigma$ satisfy $P(x_m) \in J_0$ and $0 < |r(x_m)| < \rho$. Along these iterates, $|r(x_m)| \to 0$ since $d(x_k, C_f) \to 0$. Choose a sufficiently large m with $x_m \in V_\sigma$ and

$$|r(x_m)| < \rho/4, \qquad LQ_m < \delta/4, \tag{85}$$

and, for every $k \ge m$, $L\alpha_k \le 1/4$, $\alpha_k + L\alpha_k^2 < \rho$, and the line segment $[x_k, x_{k+1}]$ lies in V whenever $P(x_k) \in J_1$ and $|r(x_k)| \le \rho$. The last condition is possible by choosing sufficiently large m. Indeed, every x_k satisfying these two conditions lies in the compact set $\{y + v\nabla r(y) : y \in J_1,\ |v| \le \rho\} \subseteq V$. Its distance to $\mathbb{R}^2 \setminus V$ is positive, and $\|x_k + s(x_{k+1} - x_k) - x_k\| \le \alpha_k \to 0$ for every $s \in [0,1]$. As long as $x_k \in V_\sigma$, $|r(x_k)| < \rho$, and $P(x_k) \in J_1$, we have $\sigma r(x_k) = |r(x_k)|$. Using (82), $\sigma A(x_k) \le -1/2$, and $L\alpha_k \le 1/4$, we obtain

$$\sigma r(x_{k+1}) = |r(x_k)| - \alpha_k\sigma A(x_k) + \sigma e_k \ge |r(x_k)| + \tfrac{1}{2}\alpha_k - L\alpha_k^2 \ge |r(x_k)| + \tfrac{1}{4}\alpha_k > 0.$$

The inclusion $[x_k, x_{k+1}] \subseteq V$ gives $x_{k+1} \in V$. Since $\sigma r(x_{k+1}) > 0$, the definition of V_σ gives $x_{k+1} \in V_\sigma$ and $|r(x_{k+1})| = \sigma r(x_{k+1})$. Therefore

$$|r(x_{k+1})| - |r(x_k)| \ge \tfrac{1}{4}\alpha_k > 0. \tag{86}$$

Since $\sum_k \alpha_k = \infty$, there is a first $n > m$ with $|r(x_n)| \ge \rho$ or $P(x_n) \notin J_1$. The step-size bound gives $|r(x_n)| \le 2\rho$. Using $2|r(x_k)|(|r(x_{k+1})| - |r(x_k)|) \le r(x_{k+1})^2 - r(x_k)^2$, we obtain

$$\sum_{k=m}^{n-1}\alpha_k|r(x_k)| \le 4\sum_{k=m}^{n-1}|r(x_k)|(|r(x_{k+1})| - |r(x_k)|) \le 2(r(x_n)^2 - r(x_m)^2) \le 8\rho^2.$$

Consequently, (83) gives

$$\|P(x_n) - P(x_m)\| \le 8L\rho^2 + LQ_m < \delta/2.$$

The definition of δ then gives $P(x_n) \in \mathrm{int}_{S_0} J_1$, and hence $|r(x_n)| \ge \rho$. Since $|r(x_{n-1})| < \rho$, (82) and $|A| \le 1$ show

$$0 \le |r(x_n)| - \rho \le \alpha_{n-1} + L\alpha_{n-1}^2 \longrightarrow 0$$

along any sufficiently large choices of m. The compact set $E_\sigma = \{y + \sigma\rho\nabla r(y) : y \in J_1\}$ is contained in V_σ. Since V_σ is noncritical, $\eta_\sigma := d(E_\sigma, C_f) > 0$. (78) and $P(x_n) \in J_1$ give $d(x_n, E_\sigma) \le |r(x_n)| - \rho \to 0$, but $d(x_n, C_f) \ge \eta_\sigma - d(x_n, E_\sigma)$ contradicts (75).

Choose an open neighborhood U_2 of x^* with $\overline{U}_2 \subseteq V$, $P(U_2) \subseteq \operatorname{int}_{S_0} J_0$, and $|r| < \rho/4$ on U_2. For each $\sigma \in \{-1, 1\}$ such that V_σ is noncritical and $\sigma A \le -1/2$ on V_σ, the preceding contradiction gives an index after which no iterate in U_2 belongs to V_σ. If V_σ is critical, then $\partial f = \{0\}$ on V_σ, so $x_k \notin V_\sigma$ for every k.

Step 4: squared-distance estimates and convergence. Since $\overline{U}_2 \subseteq \operatorname{int} W$ and $\alpha_k \to 0$, $[x_k, x_{k+1}] \subseteq W$ whenever k is sufficiently large and $x_k \in U_2$. Write $d_k = |r(x_k)|$ whenever $x_k \in W$, and set $H = (L+1)^2 + 2L\rho_0 \ge 1$. For all sufficiently large k with $x_k \in U_2 \cap V_\sigma$ and V_σ noncritical, Step 3 excludes $\sigma A(x_k) \le -1/2$. Since $|A(x_k)| > 1/2$, one has $\sigma A(x_k) \ge 1/2$, so $r(x_k)A(x_k) = d_k \sigma A(x_k) \ge \frac{1}{2} d_k$. For such k, $|r(x_k)| < \rho_0$, $|A(x_k)| \le 1$, and $\alpha_k \le 1$. Expanding (82) and using $|e_k| \le L\alpha_k^2$ gives

$$\begin{aligned} d_{k+1}^2 &= d_k^2 - 2\alpha_k r(x_k)A(x_k) + \alpha_k^2 A(x_k)^2 + 2\big(r(x_k) - \alpha_k A(x_k)\big)e_k + e_k^2 \\ &\le d_k^2 - 2\alpha_k r(x_k)A(x_k) + H\alpha_k^2 \le d_k^2 - \alpha_k d_k + H\alpha_k^2. \end{aligned}$$

If $x_k \in S$, then $d_k = 0$ and (84) gives $d_{k+1}^2 = \alpha_k^2$, which is bounded by $H\alpha_k^2$. At such a step, $P(x_{k+1}) = x_k = P(x_k)$, so the bound in (83) also holds when $x_k \in S$. For every sufficiently large index m and every $n \ge m$ such that $x_m, \dots, x_n \in U_2$ and $[x_k, x_{k+1}] \subseteq W$ for $m \le k < n$, sum the squared-distance inequality. Since $d_n^2 \ge 0$,

$$\sum_{k=m}^{n-1} \alpha_k d_k \le d_m^2 + HQ_m, \qquad \max_{m \le j \le n} d_j^2 \le d_m^2 + HQ_m.$$

Summing the bound in (83) over $k = m, \dots, n-1$ and setting $T = L(1+H)$ gives

$$\sum_{k=m}^{n-1} \|P(x_{k+1}) - P(x_k)\| \le L d_m^2 + L(1+H)Q_m \le T(d_m^2 + Q_m). \tag{87}$$

For $m \le j, l \le n$, (78) and the triangle inequality give $\|x_j - x_l\| \le d_j + d_l + \|P(x_j) - P(x_l)\|$. Hence

$$\operatorname{diam}\{x_m, \dots, x_n\} \le 2\sqrt{d_m^2 + HQ_m} + T(d_m^2 + Q_m) =: \varepsilon_m. \tag{88}$$

Choose bounded open neighborhoods U_0, U_1 of x^* with $\overline{U}_0 \subseteq U_1$ and $\overline{U}_1 \subseteq U_2$, and put $\delta_0 = d(\overline{U}_0, \mathbb{R}^2 \setminus U_1) > 0$. Choose indices $m_j \to \infty$ with $x_{m_j} \to x^*$. For all sufficiently large j, $x_{m_j} \in U_0$, and $d_{m_j} = |r(x_{m_j})| \to |r(x^*)| = 0$. Also $Q_{m_j} \to 0$, so $\varepsilon_{m_j} \to 0$. Choose one of these indices $m = m_j$ so large that $x_m \in U_0$, $\varepsilon_m < \delta_0$, and, for every $k \ge m$,

$$\alpha_k < d(\overline{U}_1, \mathbb{R}^2 \setminus U_2), \qquad x_k \in U_2 \Longrightarrow [x_k, x_{k+1}] \subseteq W.$$

If $n > m$ were the first index with $x_n \notin U_1$, then $x_m, \dots, x_{n-1} \in U_1 \subseteq U_2$ and $\|x_n - x_{n-1}\| \le \alpha_{n-1} < d(\overline{U}_1, \mathbb{R}^2 \setminus U_2)$, so $x_n \in U_2$. The line segments $[x_k, x_{k+1}]$ lie in W for $m \le k < n$, and (88) applies, yet $\operatorname{diam}\{x_m, \dots, x_n\} \ge \delta_0 > \varepsilon_m$. Thus $x_k \in U_1$ for all $k \ge m$. Since U_1 can be chosen arbitrarily small, $x_k \to x^*$, and (74) gives $0 \in \partial f(x^*)$. □

D ANALYTIC COUNTEREXAMPLES IN DIMENSION THREE

A real-analytic semialgebraic function is called a Nash function (Bochnak et al., 1998). We begin with $c = 1$. Fix $p \in (1/2, 1)$ and choose an integer

$$m \ge \max\left\{3, \frac{p}{1-p}\right\}, \qquad \tau_k = (k+1)^{-p/m}, \qquad h_k = \tau_k^{m-1}, \qquad \alpha_k = \tau_k^m, \tag{89}$$

The choice of m gives $p(m+1)/m \le 1$, hence $\sum_k \tau_k^{m+1} = \infty$. Also $h_k \le \tau_k^2$ because $m \ge 3$.

D.1 AN EIGHT-STEP PERIODIC CONSTRUCTION

Let δ, ε be positive rational parameters in $(0, 1/20]$, to be fixed below. For $\theta \in \mathbb{R}/(2\pi\mathbb{Z})$, put

$$\begin{aligned} \lambda_1 &= 1 + \delta\cos\theta, \quad \lambda_2 = 1 - \delta\cos\theta, \\ \lambda_3 &= 1 - \delta\sin\theta, \quad \lambda_4 = 1 + \delta\sin\theta \end{aligned}$$

and

$$\begin{aligned} B_1(\theta;\delta,\varepsilon) &= \tfrac{3}{5}(1+\delta\cos\theta)(1+\varepsilon\sin\theta), \quad B_2(\theta;\delta,\varepsilon) = \tfrac{3}{5}(1-\delta\cos\theta)(1-\varepsilon\sin\theta), \\ B_3(\theta;\delta,\varepsilon) &= \tfrac{4}{5}(1-\delta\sin\theta)(1+\varepsilon\cos\theta), \quad B_4(\theta;\delta,\varepsilon) = \tfrac{4}{5}(1+\delta\sin\theta)(1-\varepsilon\cos\theta). \end{aligned} \tag{90}$$

Set $P = [-1/20, 1/20]^2$. Every factor $1 \pm \delta\cos\theta$ and $1 \pm \delta\sin\theta$ lies in $[19/20, 21/20]$ on P, and the same bounds hold for the factors with ε. Hence

$$\frac{1083}{2000} \le B_r(\theta;\delta,\varepsilon) \le \frac{441}{500} < 1 \qquad (1 \le r \le 4)$$

for every $\theta \in \mathbb{R}/(2\pi\mathbb{Z})$ and $(\delta, \varepsilon) \in P$. We suppress the arguments of B_r below. Define

$$Y_r = (1 - B_r^2)^{1/2}, \qquad D_r = (B_r, Y_r) \qquad (1 \le r \le 4).$$

Extend the coordinate functions to the remaining four indices by

$$B_{r+4} = -B_r, \qquad Y_{r+4} = -Y_r \qquad (1 \le r \le 4),$$

For $1 \le j \le 8$, set $E_j = (B_j, Y_j)$. The eight unit vectors, in order, are

$$(E_1, \dots, E_8) = (D_1, D_2, D_3, D_4, -D_1, -D_2, -D_3, -D_4). \tag{91}$$

Set $\lambda_5 = \cdots = \lambda_8 = 1$. The polygon vertices are

$$(\beta_1, q_1) = (0, 0), \qquad (\beta_{j+1}, q_{j+1}) = (\beta_j, q_j) - E_j \quad (1 \le j \le 8). \tag{92}$$

In coordinates, $\beta_{j+1} = \beta_j - B_j$ and $q_{j+1} = q_j - Y_j$. Since $E_{r+4} = -E_r$ for $1 \le r \le 4$, $\sum_{j=1}^{8} E_j = 0$. The recursion therefore gives $(\beta_9, q_9) = (\beta_1, q_1)$; vertex indices are read cyclically below. At $\delta = \varepsilon = 0$, the first four vectors are

$$D_1 = D_2 = \left(\frac{3}{5}, \frac{4}{5}\right), \qquad D_3 = D_4 = \left(\frac{4}{5}, \frac{3}{5}\right).$$

The corresponding scalar coordinates are

$$0, -\frac{4}{5}, -\frac{8}{5}, -\frac{11}{5}, -\frac{14}{5}, -2, -\frac{6}{5}, -\frac{3}{5}. \tag{93}$$

The minimum gap between these eight numbers is $1/5$, and $q_j \le -3/5$ for $2 \le j \le 8$. Uniform continuity of the nodes on the circle times P lets us restrict (δ, ε) to a closed rectangle around $(0, 0)$ on which every $q_j(\theta;\delta,\varepsilon)$ differs from its value in (93) by less than $1/20$. Thus we may take $d_q = 1/10$ and $q_* = 1/2$ in

$$|q_i(\theta) - q_j(\theta)| \ge d_q \quad (i \ne j), \qquad q_j(\theta) \le -q_* \quad (2 \le j \le 8)$$

for every θ and every parameter pair in the rectangle.

Choose an integer $R > m + 3$, put $D = 8(R + 1) - 1$, and choose an integer L so large that

$$L + 1 - mD \ge 4. \tag{94}$$

Define

$$H_j = \frac{\lambda_j B_j}{L+1}, \qquad A_j = \frac{H_j'}{\lambda_j} \qquad (1 \le j \le 8). \tag{95}$$

For $1 \le j \le 4$, the definitions and $\lambda_{j+4} = 1$ give

$$(L+1)(A_j + A_{j+4}) = \frac{(\lambda_j B_j)'}{\lambda_j} - B_j' = B_j \frac{\lambda_j'}{\lambda_j}.$$

The first two such terms sum to $-6\delta\varepsilon\sin^2\theta/5$, and the last two sum to $-8\delta\varepsilon\cos^2\theta/5$. Thus

$$\sum_{j=1}^{8} A_j = -\frac{2\delta\varepsilon}{L+1}\left(\frac{3}{5}\sin^2\theta + \frac{4}{5}\cos^2\theta\right) \le -\frac{6\delta\varepsilon}{5(L+1)} < 0. \tag{96}$$

Set

$$\nu_j = B_j^{-2} \quad (1 \le j \le 4), \qquad \nu_j = 0 \quad (5 \le j \le 8). \tag{97}$$

Define

$$M_j = 1 - \nu_j B_j^2 = \begin{cases} 0, & 1 \le j \le 4, \\ 1, & 5 \le j \le 8. \end{cases} \tag{98}$$

Consider the eight-step periodic system

$$S_{j+1} = M_j S_j + mY_j \quad (1 \le j \le 8), \qquad S_9 = S_1, \tag{99}$$

where S_9 is the state after the $j = 8$ update. For $1 \le j \le 4$, $M_j = 0$, so the first four equations determine

$$S_2 = mY_1, \qquad S_3 = mY_2, \qquad S_4 = mY_3, \qquad S_5 = mY_4,$$

independently of S_1. For $5 \le j \le 8$, $M_j = 1$ and $Y_{r+4} = -Y_r$ for $1 \le r \le 4$. The remaining equations give

$$\begin{aligned} S_6 &= S_5 + mY_5 = m(Y_4 - Y_1), \\ S_7 &= S_6 + mY_6 = m(Y_4 - Y_1 - Y_2), \\ S_8 &= S_7 + mY_7 = m(Y_4 - Y_1 - Y_2 - Y_3), \\ S_9 &= S_8 + mY_8 = -m(Y_1 + Y_2 + Y_3). \end{aligned}$$

The periodic condition $S_9 = S_1$ therefore fixes $S_1 = -m(Y_1 + Y_2 + Y_3)$. With this value, the displayed equations satisfy every update and the periodic condition, proving existence and uniqueness. Set

$$G(\theta) = \sum_{j=1}^{8} \nu_j B_j Y_j S_j.$$

At $\delta = \varepsilon = 0$, one has $\nu_j B_j Y_j = Y_j/B_j$ for $1 \le j \le 4$ and $\nu_j = 0$ for $5 \le j \le 8$. Substituting the baseline values of B_j, Y_j and the computed S_j gives

$$\begin{aligned} \frac{G}{m} &= \sum_{j=1}^{4} \frac{Y_j}{B_j}\frac{S_j}{m} \\ &= \frac{4}{3}\left(-\frac{11}{5} + \frac{4}{5}\right) + \frac{3}{4}\left(\frac{4}{5} + \frac{3}{5}\right) = -\frac{49}{60}. \end{aligned} \tag{100}$$

At $\delta = \varepsilon = 0$, the same baseline nodes occur when $(\cos\theta, \sin\theta)$ is replaced by any (u, v) in the closed unit disk $\mathbb{D}$. Uniform continuity on $\mathbb{D}$ lets us shrink the parameter rectangle so that the bounds with $d_q = 1/10$ and $q_* = 1/2$ also hold throughout $\mathbb{D}$. The functions S_j/m and G/m are continuous on the circle times this rectangle and independent of m. Since $G/m = -49/60$ at $\delta = \varepsilon = 0$, shrink the rectangle again so that $G/m \le -49/120$ for every θ. Choose positive rational δ, ε in this rectangle. With $g_* = 49/120$,

$$G(\theta) \le -mg_* < 0 \tag{101}$$

for every θ. The angular bound (96) remains strict because $\delta, \varepsilon > 0$.

D.2 Construction of a global Nash function

At the nodes $q_j(\theta)$, prescribe univariate Hermite data

$$\begin{aligned} P_0(\theta, q_j) &= H_j, & \partial_q^r P_0(\theta, q_j) &= 0 & (1 \le r \le R), \\ P_1(\theta, q_j) &= \frac{m\lambda_j q_j Y_j}{L+m}, & \partial_q P_1(\theta, q_j) &= \lambda_j Y_j, & \\ \partial_q^2 P_1(\theta, q_j) &= \nu_j \lambda_j, & \partial_q^r P_1(\theta, q_j) &= 0 & (3 \le r \le R). \end{aligned} \tag{102}$$

There are $8(R+1) = D+1$ conditions for each polynomial. The associated linear map on polynomials of degree at most D is injective: a polynomial in its kernel is divisible by $\prod_{j=1}^{8}(q-q_j)^{R+1}$, which has degree $D+1$. The map is therefore invertible. Denote the resulting polynomials on the circle by $P_i^\circ(\theta, q)$, $i = 0, 1$. Replace $(\cos\theta, \sin\theta)$ by (u, v) in the formulas for $\lambda_j, B_j, Y_j, q_j, H_j$, and ν_j on the closed unit disk $\mathbb{D}$. The bounds on B_j hold since $|u|, |v| \le 1$. The node gaps on $\mathbb{D}$ make the Hermite interpolation matrix invertible at every $(u, v) \in \mathbb{D}$. Its entries are Nash on a neighborhood of $\mathbb{D}$. The absolute value of its determinant has a positive minimum on the compact disk and stays at least half that minimum on some neighborhood of the disk. Cramer's rule gives Nash coefficient functions $c_{ir}(u, v)$ there, and $P_i^\circ(\theta, q) = \sum_{r=0}^{D} c_{ir}(e_\theta)q^r$, where $e_\theta = (\cos\theta, \sin\theta)$.

Set

$$s(X) = \frac{|X|^2 - 1}{2}, \qquad \Phi(X) = \frac{2X}{1 + |X|^2}.$$

The global Nash map Φ takes $\mathbb{R}^2$ into $\mathbb{D}$ and fixes the unit circle. For r near one,

$$1 - |\Phi(re_\theta)| = \frac{(r-1)^2}{1+r^2}.$$

Choose an integer N with $2^N \ge m + 3$ and define

$$p_{ir}(X) = c_{ir}(\Phi^{\circ N}(X)), \qquad P_i(X, q) = \sum_{r=0}^{D} p_{ir}(X)q^r.$$

Every p_{ir} is global Nash. Put $r_0 = \sqrt{1+2s}$ and $r_{l+1} = 2r_l/(1+r_l^2)$. Since $r_0 - 1 = O(s)$ and $1 - r_{l+1} = (r_l - 1)^2/(1+r_l^2)$, induction gives $r_N - 1 = O(s^{2^N})$ and $\Phi^{\circ N}(r_0 e_\theta) = r_N e_\theta$. Analyticity of c_{ir} and $r_N - 1 = O(s^{2^N})$ show that each difference $c_{ir}(r_N e_\theta) - c_{ir}(e_\theta)$ is analytic in (s, θ) and $O(s^{2^N})$. Since $2^N \ge m + 3$, it vanishes to order at least $m + 3$ in s. Factoring that power from its Taylor series gives

$$P_i(s, \theta, q) = P_i^\circ(\theta, q) + s^{m+3}A_i(s, \theta, q), \qquad i = 0, 1, \tag{103}$$

where A_i is polynomial in q with analytic coefficients. On a closed annulus and a fixed bounded q interval containing the nodes, these coefficients and the first derivatives used below are bounded uniformly in θ.

For $s \neq 0$, write $q = z/s^m$ and use the local notation

$$F_0(X, z) = s^{L+1}P_0(X, q) + s^{L+m}P_1(X, q). \tag{104}$$

If $P_i(X, q) = \sum_{r=0}^{D} p_{ir}(X)q^r$, the actual global definition is

$$F_0(X, z) = \sum_{r=0}^{D} p_{0r}(X)z^r s^{L+1-mr} + \sum_{r=0}^{D} p_{1r}(X)z^r s^{L+m-mr}. \tag{105}$$

Every power of s in this formula is at least four by (94). Each summand and each of its first derivatives therefore vanish when $s = z = 0$. Thus F_0 is global Nash, has value zero on

$$\mathcal{C} = \{(X, 0) \in \mathbb{R}^2 \times \mathbb{R} : |X| = 1\} \tag{106}$$

and every point of $\mathcal{C}$ is critical.

Since $ds/dr = r$, in polar coordinates $X = r(\cos\theta, \sin\theta)$ the ambient gradient has components

$$r(F_0)_s, \qquad r^{-1}(F_0)_\theta, \qquad (F_0)_z.$$

Before normalization, direct differentiation of (104) gives

$$\begin{aligned} s^{-L}(F_0)_s &= (L+1)P_0 - mqP_{0q} + sP_{0s} \\ &\qquad + s^{m-1}\big((L+m)P_1 - mqP_{1q} + sP_{1s}\big), \\ s^{-L}(F_0)_z &= s^{1-m}P_{0q} + P_{1q}, \\ s^{-L}(F_0)_\theta &= sP_{0\theta} + s^m P_{1\theta}. \end{aligned} \tag{107}$$

Put $q = q_j(\theta) + \xi$. Taylor expansion in q of the circle polynomials, followed by (103), gives

$$\begin{aligned} P_0 &= H_j + O(|\xi|^{R+1} + s^{m+3}), & P_{0q} &= O(|\xi|^R + s^{m+3}), \\ P_{1q} &= \lambda_j(Y_j + \nu_j\xi) + O(|\xi|^R + s^{m+3}), & P_{0s} &= O(s^{m+2}), \\ P_{0\theta} &= H_j' + O(|\xi|^R + s^{m+3}). \end{aligned} \tag{108}$$

For the last line, differentiating $P_0^\circ(\theta, q) = H_j(\theta) + (q - q_j(\theta))^{R+1} a_j(\theta, q)$ at fixed q gives the stated order $|\xi|^R$. All constants are uniform for θ on the circle, $1 \le j \le 8$, $0 < s \le s_0$, and $|\xi| \le \xi_0$, where s_0, ξ_0 are fixed sufficiently small numbers. This uniformity follows from boundedness of the finitely many analytic coefficient functions and their derivatives on a closed annulus times a compact q interval. In particular, the vertical contribution from the coefficient extension is $s^{1-m}O(s^{m+3}) = O(s^4)$.

On the circle the prescribed value of P_1° gives $(L+m)P_1^\circ(\theta, q_j) = mq_j P_{1q}^\circ(\theta, q_j)$. Define the scaled ambient components

$$V_r = s^{-L} r (F_0)_s, \qquad V_z = s^{-L} (F_0)_z, \qquad V_\theta = s^{-L} r^{-1} (F_0)_\theta.$$

Write $V = (V_r, V_z, V_\theta)$. Using $r = 1 + O(s)$ in (107) gives

$$\begin{aligned} (V_r, V_z) &= \lambda_j(B_j, Y_j + \nu_j\xi) + O\big(s + s^{m-1} + s^{1-m}|\xi|^R\big), \\ V_\theta &= sH_j' + O(s^2 + s|\xi| + s^m). \end{aligned} \tag{109}$$

Here the bounded P_1 radial terms contribute $O(s^{m-1})$; the P_0 terms are bounded by (108). Let $\lambda_* := \min_{j,\theta} \lambda_j(\theta) > 0$. For the ranges $0 < s \le s_0$ and $|\xi| \le \xi_0$ fixed above, (109) gives a fixed $C > 0$ with

$$\|V - \lambda_j(B_j, Y_j, 0)\| \le C\big(|\xi| + s + s^{m-1} + s^{1-m}|\xi|^R\big).$$

Decrease s_0, ξ_0 to lie in $(0, 1]$ and choose $\epsilon_0 > 0$ so that $C(\xi_0 + s_0 + s_0^{m-1} + \epsilon_0) \le \lambda_*/2$. Throughout

$$0 < s \le s_0, \qquad |\xi| \le \xi_0, \qquad s^{1-m}|\xi|^R \le \epsilon_0, \tag{110}$$

the distance of V from $\lambda_j(B_j, Y_j, 0)$ is at most $\lambda_*/2$. In particular, for a constant C_g,

$$\frac{\lambda_*}{2} s^L \le \|\nabla F_0\| \le C_g s^L. \tag{111}$$

The normalization map $v \mapsto v/\|v\|$ has bounded first two derivatives on $\{v : \lambda_*/2 \le \|v\| \le C_g\}$. Its derivative at $(B_j, Y_j, 0)$ applied to $(0, \nu_j\xi, 0)$ is $\nu_j\xi(-B_jY_j, B_j^2, 0)$. Taylor's formula, (109), and $s^{1-m}|\xi|^R \le \epsilon_0$ therefore give, for $U = \nabla F_0/\|\nabla F_0\|$,

$$\begin{aligned} U_r &= B_j - \nu_j B_j Y_j \xi + O\left(s + s^{m-1} + \xi^2 + s^{1-m}|\xi|^R\right), \\ U_z &= Y_j + \nu_j B_j^2 \xi + O\left(s + s^{m-1} + \xi^2 + s^{1-m}|\xi|^R\right), \\ U_\theta &= sA_j + O\left(s^2 + s|\xi| + s^m + s^{2-m}|\xi|^R\right). \end{aligned} \tag{112}$$

For the angular component, divide V_θ by $\|V\|$ and use

$$\|V\| = \lambda_j + O\big(s + |\xi| + s^{m-1} + s^{1-m}|\xi|^R\big).$$

This gives the third line, including the term $s^{2-m}|\xi|^R$. Every remainder constant in (112) is fixed on (110) and is independent of the later choice of $C_\sigma, C_\zeta, C_{\mathrm{tube}}$.

D.3 Estimates for the prescribed stepsizes

The estimates below apply uniformly to every unit vector field $\widetilde{U}$ satisfying

$$\|\widetilde{U} - U\| \le s^2 \tag{113}$$

at the points where the estimates are used. Adding a vector of norm at most s^2 to (112) preserves all three remainder bounds after one fixed increase of their constants. The case $\widetilde{U} = U$ gives the tail for F_0; the uniform statement will also apply to the final potential.

Let $j(k) = 1 + (k \bmod 8)$. For a point with $s > 0$ near $\mathcal{C}$, define σ_k, ξ_k by

$$s_k = \tau_k\big(1 + h_k\beta_{j(k)}(\theta_k) + \sigma_k\big), \qquad \frac{z_k}{s_k^m} = q_{j(k)}(\theta_k) + \xi_k. \tag{114}$$

Put

$$d_k = 1 - \frac{\tau_{k+1}}{\tau_k}.$$

The mean value theorem gives $0 < d_k \le p/(m(k+1))$. Since $(k+1)^{-1} = \tau_k^{m/p}$ and $m(1-p)/p \ge 1$,

$$d_k = O(h_k\tau_k^2), \qquad (k+1)^{-1} = o(h_k), \qquad \frac{\tau_{k+i}}{\tau_k} = 1 + O((k+1)^{-1}) \quad (0 \le i \le 8). \tag{115}$$

The last estimate follows by applying the mean value theorem to $(1 + i/(k+1))^{-p/m}$; its constant is uniform over the nine values of i. The corresponding ratio for h_{k+i}/h_k has the same bound.

Write $r_k = \sqrt{1 + 2s_k}$ and let $n_k = (\cos\theta_k, \sin\theta_k)$ and $t_k = (-\sin\theta_k, \cos\theta_k)$. The planar position is $X_k = r_k n_k$, while the planar part of the step direction is $\widetilde{U}_r n_k + \widetilde{U}_\theta t_k$. Hence

$$X_{k+1} = (r_k - \alpha_k\widetilde{U}_r)n_k - \alpha_k\widetilde{U}_\theta t_k.$$

The vectors n_k, t_k are orthonormal. Since $s(X) = (|X|^2 - 1)/2$, orthogonality gives

$$2(s_{k+1} - s_k) = |X_{k+1}|^2 - |X_k|^2 = -2\alpha_k r_k\widetilde{U}_r + \alpha_k^2(\widetilde{U}_r^2 + \widetilde{U}_\theta^2).$$

The vertical update is $z_{k+1} = z_k - \alpha_k\widetilde{U}_z$. When $r_k - \alpha_k\widetilde{U}_r > 0$, the angle of X_{k+1} relative to n_k satisfies $\tan(\theta_{k+1} - \theta_k) = -\alpha_k\widetilde{U}_\theta/(r_k - \alpha_k\widetilde{U}_r)$. The radial coefficient $r_k - \lambda\alpha_k\widetilde{U}_r$ is affine in λ and positive at $\lambda = 0, 1$, so it is positive for $0 \le \lambda \le 1$. The continuously lifted angle therefore stays in $(-\pi/2, \pi/2)$ relative to n_k. These identities yield

$$\begin{aligned} s_{k+1} &= s_k - \alpha_k r_k\widetilde{U}_r + \frac{\alpha_k^2}{2}(\widetilde{U}_r^2 + \widetilde{U}_\theta^2), \\ z_{k+1} &= z_k - \alpha_k\widetilde{U}_z, \qquad \theta_{k+1} - \theta_k = -\arctan\frac{\alpha_k\widetilde{U}_\theta}{r_k - \alpha_k\widetilde{U}_r}. \end{aligned} \tag{116}$$

Put $e_k = |\sigma_k| + |\xi_k|$. For the one-step estimates, take k large enough that $h_k \max_{j,\theta}|\beta_j(\theta)| \le 1/4$ and assume $e_k \le 1/4$. Then $|h_k\beta_j + \sigma_k| \le 1/2$ and $s_k \in [\tau_k/2, 3\tau_k/2] \subseteq [\tau_k/2, 2\tau_k]$. Thus $r_k \ge 1$. Since $\|U\| = 1$ and (113) gives $|\widetilde{U}_r| \le 1 + s_k^2$, while $\alpha_k = \tau_k^m \to 0$, $r_k - \alpha_k\widetilde{U}_r > 0$ for all sufficiently late steps. The first identity in (116) gives $|s_{k+1} - s_k| \le C\alpha_k$. Dividing by $s_k \ge \tau_k/2$ and using $\alpha_k/\tau_k = h_k$ yields $s_{k+1}/s_k = 1 + O(h_k)$, so $s_{k+1} > 0$ for late k. Taylor expansion on $|h_k\beta_j + \sigma_k| \le 1/2$ gives

$$\frac{\alpha_k}{s_k^m} = 1 - m\sigma_k - mh_k\beta_j + O((|\sigma_k| + h_k)^2)$$

and $(s_k/s_{k+1})^m = 1 + O(h_k)$. Consequently,

$$\begin{aligned} \frac{z_{k+1}}{s_{k+1}^m} &= \left(q_j + \xi_k - \frac{\alpha_k}{s_k^m}\widetilde{U}_z\right)\left(\frac{s_k}{s_{k+1}}\right)^m \\ &= q_j - Y_j + M_j\xi_k + mY_j\sigma_k + O\big(\tau_k + e_k^2 + \tau_k^{1-m}|\xi_k|^R\big). \end{aligned}$$

Here $h_k \le \tau_k^2$, products of σ_k and ξ_k are bounded by e_k^2, and the normalization remainder is supplied by (112). The angular identity gives $|\theta_{k+1} - \theta_k| \le C\tau_k^{m+1}$ on (110). Since q'_{j+1} is bounded and $q_{j+1} = q_j - Y_j$, replacing $q_{j+1}(\theta_k)$ by $q_{j+1}(\theta_{k+1})$ adds at most $C\tau_k^{m+1}$.

For the other error coordinate, divide the radial update by $\tau_{k+1} = \tau_k(1 - d_k)$ and subtract $1 + h_{k+1}\beta_{j+1}(\theta_{k+1})$. The identity $\beta_j - B_j - \beta_{j+1} = 0$ cancels the terms of order h_k. More explicitly, $\alpha_k/\tau_k = h_k$, $r_k = 1 + O(\tau_k)$, and the radial component of (112) give

$$\frac{s_{k+1}}{\tau_k} = 1 + h_k\beta_{j+1}(\theta_k) + \sigma_k + h_k\nu_j B_j Y_j\xi_k + O\big(h_k(\tau_k + e_k^2 + \tau_k^{1-m}|\xi_k|^R)\big).$$

The remaining changes from τ_k to τ_{k+1} and from θ_k to θ_{k+1} are bounded using

$$\begin{aligned} |r_k - 1| &\le C\tau_k, \qquad & |h_{k+1} - h_k| &\le Ch_k d_k, \\ |\beta_{j+1}(\theta_{k+1}) - \beta_{j+1}(\theta_k)| &\le C\tau_k^{m+1}, \qquad & \alpha_k^2/\tau_k &= \tau_k^{2m-1} \le h_k\tau_k, \\ |(1-d_k)^{-1} - 1| &\le 2d_k. \end{aligned}$$

The radial and vertical expansions above, together with these clock and angle bounds, give

$$\begin{aligned} \sigma_{k+1} &= \sigma_k + h_k c_j \xi_k + r_{1,k}, \qquad & c_j &:= \nu_j B_j Y_j, \\ \xi_{k+1} &= M_j \xi_k + m Y_j \sigma_k + r_{2,k}, \\ |r_{1,k}| &\le Ch_k\mathcal{R}_k + Cd_k, \qquad & |r_{2,k}| &\le C\mathcal{R}_k, \\ \mathcal{R}_k &:= \tau_k + e_k^2 + \tau_k^{1-m}|\xi_k|^R. \end{aligned} \tag{117}$$

All coefficient functions are evaluated at θ_k and $j = j(k)$. The constant C is uniform under $e_k \le 1/4$ and (110) for all sufficiently large k. Refining the arctangent expansion in (116) also gives

$$\theta_{k+1} - \theta_k = -\tau_k^{m+1} A_j(\theta_k) + O\left(\tau_k^{m+1}(\tau_k + e_k + \tau_k^{1-m}|\xi_k|^R)\right). \tag{118}$$

Indeed, $s_k/\tau_k = 1 + \sigma_k + h_k\beta_j$ and $r_k = 1 + O(\tau_k)$ give the leading coefficient, while the remainder of $\arctan v$ is $O(v^3)$ near zero.

For every fixed $C_{\mathrm{tube}} > 0$, the region $e_k \le C_{\mathrm{tube}}\tau_k$ lies in (110) and satisfies $e_k \le 1/4$ and $h_k \max_{j,\theta} |\beta_j(\theta)| \le 1/4$ for all sufficiently large k. On this region,

$$\frac{e_k^2 + \tau_k^{1-m}|\xi_k|^R}{\tau_k} \le C_{\mathrm{tube}}^2\tau_k + C_{\mathrm{tube}}^R \tau_k^{R-m} \longrightarrow 0. \tag{119}$$

For each fixed $C_{\mathrm{tube}} > 0$, (119) gives an index $K(C_{\mathrm{tube}})$ after which $e_k^2 + \tau_k^{1-m}|\xi_k|^R \le \tau_k$ on the stated region. Taking $C_b = 2C$ in (117), for $k \ge K(C_{\mathrm{tube}})$ we obtain

$$|r_{1,k}| \le C_b(h_k\tau_k + d_k), \qquad |r_{2,k}| \le C_b\tau_k, \tag{120}$$

where C_b is independent of C_{tube}. The angular remainder on this region is $O_{C_{\mathrm{tube}}}(\tau_k^{m+2})$.

Set

$$\zeta_k = \xi_k - S_{j(k)}(\theta_k)\sigma_k. \tag{121}$$

For $C > 0$ and $k \ge 0$, let $\mathcal{B}_k(C)$ be the set of points with $s > 0$ and $|\sigma_k| + |\xi_k| \le C\tau_k$, where the phase is $j(k)$ and σ_k, ξ_k are evaluated by (114).

Proposition D.1. *There are constants $C_\sigma, C_\zeta, C_{\mathrm{tube}}, a_1 > 0$ and a multiple of eight K_0 with the following property. Fix any multiple of eight $K \ge K_0$, any initial angle θ_K, and any unit vector field $\widetilde{U}$ defined on every $\mathcal{B}_k(C_{\mathrm{tube}})$, $k \ge K$, and satisfying* (113) *on each of these sets. Starting at*

$$x_K = \left(\sqrt{1 + 2\tau_K}(\cos\theta_K, \sin\theta_K), 0\right),$$

the recursion $x_{k+1} = x_k - \alpha_k\widetilde{U}(x_k)$ defines a trajectory for every $k \ge K$ with these properties:

1. *Each $\mathcal{B}_k(C_{\mathrm{tube}})$, $k \ge K$, lies in* (110)*, where $\|\nabla F_0\| \ge (\lambda_*/2)s^L$.*
2. *Every x_k lies in $\mathcal{B}_k(C_{\mathrm{tube}})$. At each block boundary $k \equiv 0 \pmod 8$, $|\sigma_k| \le C_\sigma\tau_k$ and $|\zeta_k| \le C_\zeta\tau_k$.*
3. *One has $s_k, z_k \to 0$ and $\|x_{k+1} - x_k\| = \alpha_k > 0$. For a continuous lift of the angles,*

$$\frac{3\delta\varepsilon}{5(L+1)}\tau_k^{m+1} \le \theta_{k+8} - \theta_k \le a_1\tau_k^{m+1} \quad (k \ge K,\ k \equiv 0 \pmod 8).$$

The set of accumulation points of the trajectory is $\mathcal{C}$.

Proof. At a fixed angle define

$$\mathsf{A}_j(h,\theta) = \begin{bmatrix} 1 & hc_j(\theta) \\ mY_j(\theta) & M_j \end{bmatrix}, \quad \mathsf{P} = \mathsf{A}_8 \cdots \mathsf{A}_1, \quad T = \begin{bmatrix} 1 & 0 \\ S_1(\theta) & 1 \end{bmatrix}, \quad \mathsf{Q} = T^{-1}\mathsf{P}T.$$

At $h = 0$, the first coordinate is fixed and the first zero multiplier removes dependence on the initial second coordinate. The recurrence for S_j therefore gives

$$\mathsf{P}(0,\theta) = \begin{bmatrix} 1 & 0 \\ S_1(\theta) & 0 \end{bmatrix}, \qquad \mathsf{Q}(0,\theta) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

To differentiate Q_{11}, start with $(1, S_1)^\top$ in the original coordinates. Its state before step j at $h = 0$ is $(1, S_j)^\top$. Differentiating the jth matrix inserts $(c_j S_j, 0)^\top$, and every later zero-h matrix preserves its first coordinate. Summing these eight contributions gives $\partial_h \mathsf{Q}_{11}(0,\theta) = \sum_j c_j S_j = G(\theta)$. The entries of Q are polynomials in h with bounded coefficient functions on the compact circle. Hence, uniformly in θ,

$$\mathsf{Q}_{11} = 1 + hG + O(h^2), \qquad \mathsf{Q}_{12} = O(h), \quad \mathsf{Q}_{21} = O(h), \quad \mathsf{Q}_{22} = O(h). \tag{122}$$

Consider a block starting at $k \equiv 0 \pmod 8$, and temporarily suppose that its eight input states satisfy $e_{k+i} \le C_{\mathrm{tube}}\tau_{k+i}$. Write $h = h_k$, $\tau = \tau_k$, and $\theta = \theta_k$. (115) and (118) give $h_{k+i}/h = 1 + O((k+1)^{-1})$ and $|\theta_{k+i} - \theta| \le C\tau^{m+1}$ for $0 \le i \le 8$. The functions c_j, Y_j and their derivatives are bounded; the M_j are constant. Replace h_{k+i} and θ_{k+i} by h_k and θ_k in the linear part of (117). This gives

$$v_{k+i+1} = \mathsf{A}_{i+1}(h,\theta)v_{k+i} + \eta_i, \qquad v_l = (\sigma_l, \xi_l)^\top, \qquad |\eta_{i,1}| \le C(h\tau + d_k), \quad |\eta_{i,2}| \le C\tau.$$

For example, the extra first-coordinate error is bounded by $Ch((k+1)^{-1} + \tau^{m+1})e_{k+i}$, and the second by $C\tau^{m+1}e_{k+i}$; these fit the displayed bounds once $e_{k+i} \le 1$. The eightfold substitution is the exact identity

$$v_{k+8} = \mathsf{P}(h,\theta)v_k + \sum_{i=0}^{7} \mathsf{A}_8 \cdots \mathsf{A}_{i+2}\eta_i,$$

with the identity matrix used for the empty product. Every partial product has bounded entries and upper-right entry $O(h)$, since at $h = 0$ its factors are lower triangular. The sum of propagated errors therefore has first coordinate $O(h\tau + d_k)$ and second coordinate $O(\tau)$. For k beyond a threshold depending on C_{tube}, the temporary assumption gives $e_{k+i} \le 1$ for $0 \le i < 8$, and the one-step bound at $k + 7$ gives $|\sigma_{k+8}| \le 2$. Since S_1' is bounded and $|\theta_{k+8} - \theta_k| \le C\tau^{m+1}$, replacing $S_1(\theta)$ by $S_1(\theta_{k+8})$ changes ζ_{k+8} by at most $C\tau^{m+1} = O(\tau)$, with C independent of C_{tube}. Combining these facts with (122) gives

$$\begin{aligned} |\sigma_{k+8}| &\le (1 + hG + Ch^2)|\sigma_k| + Ch|\zeta_k| + Ch\tau + Cd_k, \\ |\zeta_{k+8}| &\le Ch|\sigma_k| + Ch|\zeta_k| + C\tau. \end{aligned}$$

For large k, $1 + hG \ge 1/2$, $Ch \le mg_*/2$, and $Ch \le 1/4$. Thus there are constants C_0, C_1, chosen independently of C_{tube}, such that

$$\begin{aligned} |\zeta_{k+8}| &\le \frac{1}{4}|\zeta_k| + C_0\tau_k + C_0 h_k|\sigma_k|, \\ |\sigma_{k+8}| &\le (1 - \tfrac{1}{2}mg_*h_k)|\sigma_k| + C_1 h_k|\zeta_k| + C_1 h_k\tau_k + C_1 d_k. \end{aligned} \tag{123}$$

The lower bound on k may depend on C_{tube} through (119). We now fix C_{tube} and verify the containment assumption used in the block estimate.

Put $\gamma = mg_*$. Choose $C_\zeta > 4(C_0 + 1)$ and then C_σ with

$$\frac{\gamma}{4}C_\sigma > C_1(C_\zeta + 2). \tag{124}$$

At block boundaries $k \equiv 0 \pmod 8$, impose

$$|\sigma_k| \le C_\sigma\tau_k, \qquad |\zeta_k| \le C_\zeta\tau_k \tag{125}$$

for $k \equiv 0 \pmod 8$. Let $S_* := \max_{j,\theta} |S_j(\theta)|$ and $C_{\mathrm{start}} := (1 + S_*)C_\sigma + C_\zeta$. Since $\xi_k = \zeta_k + S_{j(k)}(\theta_k)\sigma_k$, $e_k \le (1 + S_*)|\sigma_k| + |\zeta_k| \le C_{\mathrm{start}}\tau_k$. Choose $N_* \ge 1$ bounding the induced ℓ^1 norms of $\mathsf{A}_j(h,\theta)$ for $0 \le h \le 1$. On $e_k \le C_{\mathrm{tube}}\tau_k$, (115) and (120) give, for all sufficiently large k, $|r_{1,k}| + |r_{2,k}| \le C_b(h_k\tau_k + d_k + \tau_k) \le B_*\tau_k$, where B_* depends on the fixed clock constants and C_b but is independent of C_{tube}. Now fix

$$C_{\mathrm{tube}} > 2\left(N_*^8 C_{\mathrm{start}} + B_* \sum_{l=0}^{7} N_*^l + 1\right). \tag{126}$$

Only after these constants are fixed do we choose K_0, a multiple of eight, large enough that every $\mathcal{B}_k(C_{\mathrm{tube}})$ with $k \geq K_0$ lies in (110), (117), (118) and (120) hold there with constants independent of $\widetilde{U}$, and $\tau_{k+i} \geq \tau_k/2$ for $0 \leq i \leq 8$. The gradient estimate (111) then holds on each $\mathcal{B}_k(C_{\mathrm{tube}})$, $k \geq K_0$.

Starting from (125), a finite induction gives

$$e_{k+i} \leq \left(N_*^i C_{\mathrm{start}} + B_* \sum_{l=0}^{i-1} N_*^l \right) \tau_k < \frac{C_{\mathrm{tube}}}{2} \tau_k \leq C_{\mathrm{tube}} \tau_{k+i}, \qquad 0 \leq i \leq 8. \tag{127}$$

The empty sum is zero. At $i = 0$ this follows from the definition of C_{start}. Given the bound at step i, the matrix norm bound and (120), together with $\tau_{k+i} \leq \tau_k$, give

$$e_{k+i+1} \leq N_* e_{k+i} + B_* \tau_{k+i} \leq \left(N_*^{i+1} C_{\mathrm{start}} + B_* \sum_{l=0}^{i} N_*^l \right) \tau_k.$$

The choice (126) makes it smaller than $C_{\mathrm{tube}} \tau_{k+i+1}$, so the next evaluation is in its established domain. This proves the containment assumption used in the block estimate.

Replace K_0 by a larger multiple of eight so that, for every $k \geq K_0$,

$$C_0 h_k C_\sigma \leq C_\zeta/8, \qquad d_k \leq h_k \tau_k, \qquad \frac{\tau_{k+8}}{\tau_k} > \frac{7}{8}, \qquad \frac{\tau_{k+8}}{\tau_k} > 1 - \frac{\gamma}{8} h_k.$$

The first inequality follows from $h_k \to 0$, the second from $d_k = O(h_k \tau_k^2)$, and the last two from (115). Substituting $|\sigma_k| \leq C_\sigma \tau_k$ and $|\zeta_k| \leq C_\zeta \tau_k$ into (123) now yields

$$\begin{aligned} |\zeta_{k+8}| &\leq (C_\zeta/4 + C_0 + C_\zeta/8)\tau_k < \tfrac{5}{8} C_\zeta \tau_k < C_\zeta \tau_{k+8}, \\ |\sigma_{k+8}| &\leq C_\sigma \tau_k - h_k \tau_k (\tfrac{\gamma}{2} C_\sigma - C_1 C_\zeta - 2C_1) \\ &< C_\sigma \tau_k (1 - \tfrac{\gamma}{4} h_k) < C_\sigma \tau_{k+8}. \end{aligned}$$

Thus the inequalities in (125) at k hold strictly at $k + 8$. Equation (127) bounds the seven intervening states by $C_{\mathrm{tube}} \tau_{k+i}$. Fix any multiple of eight $K \geq K_0$ and any initial angle, and take the initial point in the proposition, with $s_K = \tau_K$ and $z_K = 0$. Since $\beta_1 = q_1 = 0$, it has $\sigma_K = \xi_K = \zeta_K = 0$. Induction over blocks defines its forward trajectory for all $k \geq K$. The estimates used in this induction hold for every unit field satisfying (113) on $\bigcup_{k \geq K} \mathcal{B}_k(C_{\mathrm{tube}})$, with the same constants. Since $\|\widetilde{U}(x_k)\| = 1$ and $\alpha_k > 0$, $\|x_{k+1} - x_k\| = \alpha_k > 0$ at every step. At every index,

$$s_k = \tau_k (1 + O(\tau_k)) \to 0, \qquad z_k = s_k^m \big(q_{j(k)}(\theta_k) + O(\tau_k) \big) \to 0.$$

Within eight steps, the relative change of τ_k^{m+1} is $O((k+1)^{-1})$, and the angular change is $O(\tau_k^{m+1})$. The derivatives of A_j are bounded. Summing (118) where $e_k \leq C_{\mathrm{tube}} \tau_k$, and using $(k+1)^{-1} = O(\tau_k)$, gives

$$\theta_{k+8} - \theta_k = -\tau_k^{m+1} \sum_{j=1}^{8} A_j(\theta_k) + O_{C_{\mathrm{tube}}}(\tau_k^{m+2}).$$

Substituting (96) gives

$$\theta_{k+8} - \theta_k = \frac{2\delta\varepsilon}{L+1} \left(\frac{3}{5} \sin^2 \theta_k + \frac{4}{5} \cos^2 \theta_k \right) \tau_k^{m+1} + O_{C_{\mathrm{tube}}}(\tau_k^{m+2}). \tag{128}$$

The leading coefficient lies between $6\delta\varepsilon/[5(L+1)]$ and $8\delta\varepsilon/[5(L+1)]$. Replace K_0 by a larger multiple of eight so that, for every $k \geq K_0$, the remainder has absolute value at most $3\delta\varepsilon\tau_k^{m+1}/[5(L+1)]$. Then

$$\frac{3\delta\varepsilon}{5(L+1)} \tau_k^{m+1} \leq \theta_{k+8} - \theta_k \leq a_1 \tau_k^{m+1} \qquad (k \geq K,\ k \equiv 0 \pmod 8)$$

for a finite constant a_1. Every earlier bound remains valid for $k \geq K_0$, so this estimate and the induction using (125) and (127) apply to every multiple of eight $K \geq K_0$. For $l \geq 1$, $\tau_{K+8l}^{m+1} \geq (K+9)^{-p(m+1)/m} l^{-p(m+1)/m}$. Since $p(m+1)/m \leq 1$, the p-series test gives $\sum_{l \geq 0} \tau_{K+8l}^{m+1} = \infty$.

The lifted block angles are therefore increasing and unbounded, with increments tending to zero. For any $\varphi \in [0, 2\pi)$ and each sufficiently large integer l, set $n_l := \min\{n \geq 1 : \theta_{K+8n} \geq \varphi + 2\pi l\}$. Minimality gives

$$0 \leq \theta_{K+8n_l} - (\varphi + 2\pi l) \leq \theta_{K+8n_l} - \theta_{K+8(n_l-1)} \longrightarrow 0.$$

Here $n_l \to \infty$ because every finite set of lifted angles is bounded. Together with $s_k, z_k \to 0$, this shows that $(\cos\varphi, \sin\varphi, 0)$ is an accumulation point. Conversely, $s_k, z_k \to 0$ puts every accumulation point on $\mathcal{C}$ and makes the tail bounded. Its set of accumulation points is therefore exactly $\mathcal{C}$. □

D.4 Adding the initial iterates

Let K_0 and C_{tube} be as in Proposition D.1. Choose a multiple of eight $K \geq K_0$ such that $s_* := 2\tau_K \leq 1$, $C_{\mathrm{tube}}\tau_K \leq 1$, and $2C_{\mathrm{tube}}s_*^{m+1} < \alpha_{K-1}/2$. The last inequality holds eventually because $\alpha_{K-1} = K^{-p} > \tau_K^m$ and $s_*^{m+1}/\tau_K^m = 2^{m+1}\tau_K \to 0$. Let $x_K = (\sqrt{1+2\tau_K}, 0, 0)$ and $x_i = x_K + (\sum_{r=i}^{K-1} \alpha_r)e_3$ for $0 \leq i < K$. Their vertical coordinates z_i are distinct and satisfy $z_i \geq \alpha_{K-1}$.

For $k \geq K$, every point of $\mathcal{B}_k(C_{\mathrm{tube}})$ has $s \in [\tau_k/2, 2\tau_k]$ and $q = q_{j(k)}(\theta) + \xi$ with $|\xi| \leq C_{\mathrm{tube}}\tau_k$. Put $Q_* := 1 + \max_{j,\theta} |q_j(\theta)|$. Since $C_{\mathrm{tube}}\tau_K \leq 1$ and every $q_j \leq 0$,

$$0 < s \leq s_*, \quad |z| \leq Q_* s_*^m, \quad z \leq C_{\mathrm{tube}}\tau_k s^m \leq 2C_{\mathrm{tube}}s_*^{m+1} < \alpha_{K-1}/2.$$

Define

$$a = -Q_* s_*^m - 1, \quad b_i(z) = \frac{z-a}{z_i - a}, \quad \chi = \frac{\alpha_{K-1}/2 - a}{\alpha_{K-1} - a} \in (0,1).$$

On every $\mathcal{B}_k(C_{\mathrm{tube}})$, $k \geq K$, the bounds above give $z - a \geq 1$, $z < \alpha_{K-1}/2$, and $z_i \geq \alpha_{K-1}$. Thus $0 < b_i(z) \leq (\alpha_{K-1}/2 - a)/(\alpha_{K-1} - a) = \chi$, while $b_i(z_i) = 1$.

Set $Q = L + 3$ and choose an integer $M \geq 1$ below. Define

$$L_i(z) = \prod_{\substack{0 \leq r < K \\ r \neq i}} \left(\frac{z - z_r}{z_i - z_r}\right)^2, \qquad \phi_i(X, z) = \left(\frac{s(X)}{\tau_K}\right)^Q L_i(z) b_i(z)^M.$$

These are polynomials. At x_i, $\phi_i = 1$; at x_r for $r \neq i$, $\phi_i = \nabla\phi_i = 0$ because L_i has a double zero. Put $v_i = e_3 - \nabla F_0(x_i)$ and

$$\Psi(x) = \sum_{i=0}^{K-1} \big(v_i \cdot (x - x_i)\big)\phi_i(x).$$

The product rule gives $\nabla(F_0 + \Psi)(x_i) = e_3$ for $0 \leq i < K$.

Write $\Psi = s^Q \sum_i a_i(x) b_i(z)^M$, where $a_i(x) = \tau_K^{-Q} L_i(z)\, v_i \cdot (x - x_i)$. The union $\bigcup_{k \geq K} \mathcal{B}_k(C_{\mathrm{tube}})$ lies in the compact set $|X| \leq \sqrt{3}$, $|z| \leq Q_* s_*^m$. The finitely many a_i, their gradients, and $\nabla s = (X, 0)$ are bounded there. Also $|b_i^M| \leq \chi^M$ and $\|\nabla b_i^M\| \leq M\chi^{M-1}/(\alpha_{K-1} - a)$. Thus the product rule gives C_K independent of M such that

$$|\Psi| \leq C_K E_M s^Q, \qquad \|\nabla\Psi\| \leq C_K E_M s^{Q-1}, \qquad E_M := \chi^M + M\chi^{M-1} \to 0 \tag{129}$$

on every $\mathcal{B}_k(C_{\mathrm{tube}})$, $k \geq K$.

By item 1 of Proposition D.1, every $\mathcal{B}_k(C_{\mathrm{tube}})$ with $k \geq K$ lies in (110), where $\|\nabla F_0\| \geq b_0 s^L$ with $b_0 = \lambda_*/2$. Choose M so that $4C_K E_M \leq b_0$. As $Q = L + 3$, $\|\nabla\Psi\|/\|\nabla F_0\| \leq s^2/4$. The function $F_p = F_0 + \Psi$ is global Nash. Since $s \leq s_* \leq 1$ on the tube sets, $\|\nabla F_p\| \geq (1 - s^2/4)\|\nabla F_0\| \geq b_0 s^L/2 > 0$ on every $\mathcal{B}_k(C_{\mathrm{tube}})$, $k \geq K$. For nonzero g, $g + e$,

$$\left\| \frac{g+e}{\|g+e\|} - \frac{g}{\|g\|} \right\| \leq \frac{2\|e\|}{\|g\|}.$$

Define $\widetilde{U}_p = \nabla F_p/\|\nabla F_p\|$ on $\bigcup_{k \geq K} \mathcal{B}_k(C_{\mathrm{tube}})$. With $g = \nabla F_0$ and $e = \nabla\Psi$, the above inequality gives $\|\widetilde{U}_p - U\| \leq s^2/2 \leq s^2$, which is (113). The point x_K above is the initial point in Proposition D.1 with $\theta_K = 0$. Apply that proposition to $\widetilde{U}_p$: the recursion $x_{k+1} = x_k - \alpha_k \widetilde{U}_p(x_k)$ is defined

for all $k \geq K$. Item 2 puts each x_k in $\mathcal{B}_k(C_{\text{tube}})$ and bounds $|\sigma_k|$ and $|\zeta_k|$ at block boundaries. Item 3 gives $\|x_{k+1} - x_k\| = \alpha_k > 0$ and identifies $\mathcal{C}$ as its set of accumulation points. The factor s^Q gives $\Psi = 0$ and $\nabla\Psi = 0$ on $\mathcal{C}$; hence $F_p = 0$ and $\nabla F_p = 0$ there.

For $0 \leq i < K$, $\nabla F_p(x_i) = e_3$ and $x_{i+1} = x_i - \alpha_i e_3$. These first K updates and the trajectory from x_K form one exact bounded NGD orbit indexed from zero, with $\mathcal{C}$ as its set of accumulation points. For $k \geq K$, let $y_k = (X_k/|X_k|, 0) \in \mathcal{C}$. The limits $s_k, z_k \to 0$ give $\|x_k - y_k\| \to 0$. Uniform continuity of F_p and ∇F_p on a compact neighborhood containing the tail and $\mathcal{C}$, together with their zero values on $\mathcal{C}$, yields $F_p(x_k) \to 0$ and $\|\nabla F_p(x_k)\| \to 0$. This proves Theorem 2.5 for $c = 1$.

Finally, for arbitrary $c > 0$, let

$$F_{p,c}(x) = cF_p(x/c), \qquad x_k^{(c)} = cx_k. \tag{130}$$

Then $\nabla F_{p,c}(x) = \nabla F_p(x/c)$. At $x_k^{(c)} = cx_k$,

$$x_{k+1}^{(c)} = x_k^{(c)} - c\alpha_k \frac{\nabla F_{p,c}(x_k^{(c)})}{\|\nabla F_{p,c}(x_k^{(c)})\|}.$$

Global Nash regularity is preserved, and the set of accumulation points is $c\mathcal{C}$. Moreover, along the constructed orbit,

$$F_{p,c}(x_k^{(c)}) \to 0, \qquad \|\nabla F_{p,c}(x_k^{(c)})\| \to 0.$$

This completes the proof.