Nonminimizing Limits of Noisy Clarke Subgradient Descent

Abstract

We disprove the conjecture of Davis, Drusvyatskiy, and Jiang [11] that generic convergence of noisy subgradient descent to local minima remains valid without subdifferential regularity. Our counterexample is an explicit two-dimensional, nonnegative, coercive, globally Lipschitz, semialgebraic continuous piecewise-affine objective. For every linear tilt in a nonempty open set, every initialization in a fixed positive-area rectangle, and every $\gamma \in (1/2, 1)$, the noisy Clarke-subgradient iterates with steps $\alpha_k = (k+8)^{-\gamma}$ converge almost surely to the origin, which is not a local minimizer. The conclusion holds for every admissible Clarke choice made before the current noise is sampled, under conditionally absolutely continuous perturbations bounded coordinatewise by 10^{-2}. In particular, a fixed Borel selection with independent and identically distributed uniform disk noise satisfies the original stochastic-approximation assumptions and directly falsifies the conjecture; the open bad-tilt set rules out an almost-everywhere guarantee. The limit is Clarke-critical but not limiting-critical, identifying subdifferential convexification as the obstruction. A shrinking pathwise invariant gives an $O(\alpha_{k-1})$ convergence envelope. Scaling and separable embedding extend the construction to arbitrary finite uniform-noise radii and every dimension at least two. Thus definability, coercivity, generic linear perturbations, and continuous noise do not suffice to guarantee local-minimum convergence for this recursion.

Note. This paper was generated entirely by AI, including MiMo, using an automated research pipeline developed by Chenghua Liu and Hanyu Li.

1 Introduction

Noise is a standard device for avoiding unstable stationary points. For smooth stochastic approximation, classical nonconvergence theorems show that sufficiently excited noise prevents convergence to linearly unstable equilibria [15]. Related results now cover important nonsmooth models, especially weakly convex and convex-composite objectives with active strict saddles [2, 9, 10, 16]. These works support the principle that sufficiently excited stochastic first-order methods should avoid suitably structured saddles.

Davis et al. [11] (DDJ) established this principle for subdifferentially regular definable objectives. Their theorem combines two facts. Generic linear tilts classify limiting-critical points as local minima or active strict saddles, and uniform noise rules out convergence to the latter. They conjectured that the same conclusion remains valid after subdifferential regularity is removed. The issue matters for nonsmooth optimization theory: definability controls geometric complexity, but it does not by itself prevent the Clarke subdifferential from containing convexified vectors absent from the limiting subdifferential.

Throughout, a property holds *generically* if it holds for Lebesgue-almost every parameter, equivalently on a conull parameter set. A function is *tame* here when it is definable in a fixed o-minimal structure; every semialgebraic function is tame.

This paper gives a negative answer using a nonnegative, coercive, continuous piecewise-affine function in $\mathbb{R}^2$. The failure holds throughout an open set of tilts and for every history-dependent Clarke subgradient chosen before the current perturbation is sampled. Closely related static geometry already appears in Example 2 of Bianchi et al. [2]; the new point is a coercive objective together with a standard noisy recursion that actually converges to the nonminimizing critical point.

Contributions. The main result combines the following features.

- *Global and variational geometry.* The objective and every bad tilt are nonnegative and coercive, while the trapped origin is Clarke-critical but not limiting-critical or locally minimizing.
- *Oracle-robust dynamics.* The trap holds for every $\gamma \in (1/2, 1)$, every initialization in a positive-area basin, every pre-noise Clarke oracle, and conditionally absolutely continuous bounded noise. Independence and centering are not used by the invariant.
- *A direct contradiction.* A fixed Borel selection, an ordinary differentiability-point initialization, iid uniform disk noise, and the displayed power-law steps satisfy the published assumptions. The open bad-tilt set therefore rules out a conull good-tilt set.
- *A quantitative mechanism and extensions.* A three-transition pathwise invariant gives an $O(\alpha_{k-1})$ envelope. Scaling and separable embedding extend the construction to every finite noise radius and every dimension $d \geq 2$; we make no minimal-dimension claim.

Organization. Section 2 formalizes the target and states the counterexample. Sections 3 to 5 establish its geometry, singular-set avoidance, and pathwise invariant; Section 6 assembles the proof. Section 7 states the finite-radius and higher-dimensional consequences, and Section 8 gives interpretation and comparisons. Technical robustness details and extension proofs are collected after the references in Appendix A.

2 Setup and main result

This section fixes the variational and probabilistic conventions, states the DDJ target with all algorithmic quantifiers bound, and then gives both the oracle-robust theorem and its fixed-selection specialization.

2.1 Subdifferentials and noisy Clarke oracles

For a locally Lipschitz function $h : \mathbb{R}^d \to \mathbb{R}$, the Fréchet subdifferential is

$$\widehat{\partial} h(z) := \left\{ w : \liminf_{\substack{u \to 0 \\ u \neq 0}} \frac{h(z+u) - h(z) - \langle w, u \rangle}{\|u\|_2} \geq 0 \right\}.$$

The limiting, or Mordukhovich, subdifferential is

$$\partial h(z) := \left\{ w : \begin{array}{l} \text{there exist } z_i \to z,\ h(z_i) \to h(z), \text{ and } w_i \to w \\ \text{with } w_i \in \widehat{\partial} h(z_i) \end{array} \right\}.$$

Writing D_h for the differentiability set of h, the Clarke subdifferential is

$$\partial_{\mathrm{c}} h(z) := \operatorname{cl\,conv} \left\{ \lim_{i \to \infty} \nabla h(z_i) : z_i \to z,\ z_i \in D_h \right\}.$$

Here cl conv denotes the closed convex hull. These definitions agree with the standard variational-analysis conventions [7, 18]. Adding a C^1 function translates all three subdifferentials. Thus, for $v \in \mathbb{R}^d$ and $h_v(z) := h(z) - \langle v, z \rangle$,

$$\mathcal{D} h_v(z) = \mathcal{D} h(z) - v \quad \text{for } \mathcal{D} \in \{\widehat{\partial}, \partial, \partial_{\mathrm{c}}\}. \tag{2.1}$$

We use

$$\mathrm{Crit}(h) := \{z : 0 \in \partial h(z)\}, \qquad \mathrm{Crit}_{\mathrm{c}}(h) := \{z : 0 \in \partial_{\mathrm{c}} h(z)\},$$

for the limiting- and Clarke-critical sets, and $\mathrm{LocMin}(h)$ for the set of local minimizers. We call h *subdifferentially regular* at z when its epigraph is Clarke regular at $(z, h(z))$ in the standard epigraphical sense of Rockafellar and Wets [18]; for locally Lipschitz functions this implies agreement of the Fréchet, limiting, and Clarke subdifferentials at z. For $\rho \geq 0$, the function is *locally ρ-weakly convex* near z if $h + \frac{\rho}{2}\|\cdot\|_2^2$ is convex on some convex neighborhood of z.

Fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_k)_{k\geq 0}, \mathbb{P})$. The sigma-field $\mathcal{F}_k$ contains the initialization, all past perturbations, and any oracle randomness used before the perturbation at step k is sampled. For fixed h and v, an *admissible pre-noise Clarke oracle* is a sequence (w_k) such that z_k and w_k are $\mathcal{F}_k$-measurable and

$$w_k \in \partial_{\mathrm{c}} h_v(z_k) \quad \text{almost surely.}$$

The perturbation ν_k and the next iterate are $\mathcal{F}_{k+1}$-measurable, and ν_k is sampled after w_k. The update for positive step sizes (α_k) is

$$z_{k+1} = z_k - \alpha_k(w_k + \nu_k). \tag{2.2}$$

A fixed Borel selection $s_v(z) \in \partial_{\mathrm{c}} h_v(z)$, used through $w_k = s_v(z_k)$, is a special case. No jointly measurable policy family is required: every result below is stated for one fixed measurable oracle process. We write $B_r(0) := \{u : \|u\|_2 \leq r\}$ for the closed Euclidean ball centered at zero. The notation $\mathrm{Law}(X \mid \mathcal{F}_k)$ denotes a version of the conditional law of X given $\mathcal{F}_k$, Leb_2 denotes two-dimensional Lebesgue measure, and $\mu \ll \lambda$ means that μ is absolutely continuous with respect to λ.

2.2 The conjectural target

In arXiv version 2, immediately after their informal main theorem, Davis et al. [11, p. 2] conjecture that its conclusion remains valid without subdifferential regularity. The sentence does not repeat the algorithmic quantifiers, so we state the natural uniform reading explicitly. The fixed-selection specialization below already refutes the conjecture and does not rely on this stronger uniform formulation.

For an algorithmic instance, define the bounded-orbit and local-minimum-limit events

$$\mathcal{B} := \left\{ \sup_{k\geq 0} \|z_k\|_2 < \infty \right\}, \qquad \mathcal{L}_v(h) := \{\exists z_\star \in \mathrm{LocMin}(h_v) : z_k \to z_\star\}. \tag{2.3}$$

Thus conditional convergence to a local minimizer means $\mathbb{P}(\mathcal{B} \setminus \mathcal{L}_v(h)) = 0$, equivalently $\mathbb{P}(\mathcal{B}^c \cup \mathcal{L}_v(h)) = 1$.

Target statement 2.1 (Generic convergence without subdifferential regularity)**.** Let $h : \mathbb{R}^d \to \mathbb{R}$ be locally Lipschitz and definable in a fixed o-minimal structure. Without assuming subdifferential regularity, does there exist a Lebesgue-conull set $G_h \subseteq \mathbb{R}^d$, fixed independently of subsequent algorithmic choices, with the following property? For every $v \in G_h$, every Borel Clarke selection, every deterministic initial point, every fixed radius $r > 0$ with iid noise $\nu_k \sim \mathrm{Unif}(B_r(0))$, and every positive deterministic step sequence for which there exist $0 < c_1 \leq c_2 < \infty$ and $\gamma \in (1/2, 1)$ satisfying

$$c_1(k+1)^{-\gamma} \leq \alpha_k \leq c_2(k+1)^{-\gamma} \qquad (k \geq 0),$$

do the resulting iterates satisfy $\mathbb{P}(\mathcal{B} \setminus \mathcal{L}_v(h)) = 0$?

In DDJ's notation, a fixed Borel selection gives their measurable map in Algorithm (4.5) and Assumption E1; the displayed step condition is E2. The iid centered uniform-ball perturbations have bounded fourth moments and hence satisfy E3 and Uniform-noise Assumption F, while our finite-valued objective satisfies E4. This gives the promised direct correspondence with the published stochastic-approximation framework.

2.3 A coercive counterexample

Define the local core and its coercive extension by

$$g(x,y) := \min\{|x| + 2\,|y|\,,\, 7\,|x| - |y|\}, \tag{2.4}$$

$$F(x,y) := 8 + \max\{g(x,y),\, |x| + |y| - 8\} = \max\{g(x,y) + 8,\, |x| + |y|\}. \tag{2.5}$$

Let

$$U := (-10^{-2}, 10^{-2})^2, \qquad F_v(z) := F(z) - \langle v, z\rangle. \tag{2.6}$$

For $\gamma \in (1/2, 1)$, put

$$\alpha_k := (k+8)^{-\gamma}, \qquad D_\star := \left\{(x,y) : |x| < \frac{1}{80},\ |y| < \frac{1}{20}\right\}. \tag{2.7}$$

Since $k + 8 \le 8(k+1)$, these steps satisfy the bounds in target statement 2.1 with $c_1 = 8^{-\gamma}$ and $c_2 = 1$.

The next theorem states a pathwise-robust result that is stronger than needed for falsification: after the first step, every subsequent oracle value is forced by the realized iterate.

Theorem 2.2 (Coercive trapping on an open set of tilts)**.** *Fix any $v \in U$, $\gamma \in (1/2, 1)$, and deterministic $z_0 \in D_\star$. Let (w_k) be any admissible pre-noise Clarke oracle for F_v. Suppose that, for every k, almost surely*

$$\mathrm{Law}(\nu_k \mid \mathcal{F}_k) \ll \mathrm{Leb}_2, \qquad \|\nu_k\|_\infty \le 10^{-2}. \tag{2.8}$$

Writing $z_k = (x_k, y_k)$, the iterates (2.2), with the steps in (2.7), satisfy almost surely

$$|x_k| \le 7.2\alpha_{k-1}, \qquad |y_k| \le 3.9\alpha_{k-1} \qquad (k \ge 1). \tag{2.9}$$

Consequently $z_k \to 0$.

Moreover, F_v is nonnegative and coercive, the origin is Clarke-critical but not limiting-critical for F_v, and the origin is not a local minimizer. In particular, for iid $\nu_k \sim \mathrm{Unif}(B_{10^{-2}}(0))$ and every fixed Borel Clarke selection,

$$\mathbb{P}(z_k \to 0) = 1, \qquad \mathbb{P}(\mathcal{B} \setminus \mathcal{L}_v(F)) = 1. \tag{2.10}$$

The set $D_\star$ is independent of γ and has area 1/400; the phenomenon is therefore not tied to the exact initialization $z_0 = 0$. For each fixed pre-noise oracle and conditional noise kernel, the probability-one event in the theorem may depend on that pair; no common event over an uncountable policy class is asserted.

For each $v \in U$, define the concrete fixed selection

$$s_v^\circ(z) := \operatorname*{argmin}_{u \in \partial_{\mathrm{c}} F_v(z)} \|u\|_2\,. \tag{2.11}$$

The value is unique because $\partial_{\mathrm{c}} F_v(z)$ is nonempty, compact, and convex. For a function with finitely many affine pieces, the graph of its Clarke multifunction is semialgebraic. The graph of (2.11) is then semialgebraic by quantifier elimination, so s_v° is Borel measurable.

The following specialization makes the contradiction with the DDJ conjecture independent of the stronger oracle quantifier in Theorem 2.2.

Corollary 2.3 (A fixed-selection counterexample in the DDJ framework)**.** *Fix $v \in U$ and $\gamma \in (1/2, 1)$. Starting from $z_0 = (1/160, 1/100)$, use the Borel selection s_v°, iid noise $\nu_k \sim \mathrm{Unif}(B_{10^{-2}}(0))$, and $\alpha_k = (k+8)^{-\gamma}$. Then $z_k \to 0$ almost surely although $0 \notin \mathrm{LocMin}(F_v)$. This instance satisfies DDJ's Algorithm (4.5), Assumptions E1–E4, and Uniform-noise Assumption F.*

Proof. The initialization lies in $D_\star$, so Theorem 2.2 gives the stated limit and shows that it is not a local minimizer. The assumptions, including the step bounds after (2.7), were matched immediately after target statement 2.1. □

Remark 2.4 (Differentiable and random initializations). Fix $v \in U$ and $\gamma \in (1/2, 1)$. The initialization

$$z_0 = (1/160,\ 1/100)$$

lies in $D_\star$ and is a differentiability point of F. With iid disk noise and any admissible pre-noise oracle, almost surely every finite-time iterate is a differentiability point and the method evaluates the unique ordinary gradient at every step, yet converges to the nonminimum at zero. If instead z_0 is uniform on $D_\star$, independently of the noise and any exogenous oracle randomization, conditioning on z_0 gives the same result.

Corollary 2.5 (Failure of a conull good-tilt set)**.** *The conull set requested in target statement 2.1 does not exist, even when restricted to nonnegative coercive globally Lipschitz semialgebraic continuous piecewise-affine objectives.*

Proof strategy. Inside the ℓ_1-ball of radius 4, the coercive extension F equals $g + 8$. Its two branch families meet on $|y| = 2\,|x|$. In the broad horizontal sector, sign-gradient steps attract both coordinates toward zero. In the narrow vertical sector, the horizontal gradient has magnitude 7: one step crosses the vertical axis and lands strictly in the horizontal sector. Three explicit transitions between a tight envelope and a recovery envelope keep the orbit inside an $O(\alpha_{k-1})$ bound. Conditional absolute continuity ensures that positive-time iterates never hit an axis or branch boundary, so after the first step every Clarke oracle is forced to return the unique gradient.

3 Geometry of the objective

We first verify the global function class and coercivity, and then compute the local limiting and Clarke subdifferentials. The latter calculation identifies the criticality gap that permits convergence to a nonminimum.

3.1 Global coercivity and the local core

Here a continuous piecewise-affine function means a continuous function that is affine on every cell of a finite polyhedral decomposition. The coercive extension in (2.5) preserves this finite structure while leaving the core unchanged near the origin.

Lemma 3.1 (Function class and coercive tilts)**.** *The function F is nonnegative, coercive, globally Lipschitz, semialgebraic, and continuous piecewise affine. For every $v \in U$,*

$$F_v(z) \geq (1 - \|v\|_\infty)\,\|z\|_1 > 0.99\,\|z\|_1 \quad (z \neq 0), \tag{3.1}$$

and hence F_v is nonnegative and coercive. Furthermore,

$$F(z) = g(z) + 8 \qquad \textit{whenever}\ \ \|z\|_1 < 4. \tag{3.2}$$

Proof. Both branches defining g are globally Lipschitz, semialgebraic, and continuous piecewise linear. Finite minimum and maximum operations preserve global Lipschitz continuity, semialgebraicity, and the continuous piecewise-affine property. Also

$$g(z) \geq -|y| \geq -\|z\|_1 .$$

If $\|z\|_1 < 4$, then

$$g(z) + 8 \geq 8 - \|z\|_1 > \|z\|_1 ,$$

which proves (3.2). Globally, $F(z) \geq \|z\|_1$. Hölder's inequality therefore gives

$$F_v(z) \geq \|z\|_1 - |\langle v, z\rangle| \geq (1 - \|v\|_\infty) \|z\|_1 .$$

Since $\|v\|_\infty < 0.01$, this proves (3.1), nonnegativity, and coercivity. □

Write

$$p(x,y) := |x| + 2|y| , \qquad q(x,y) := 7|x| - |y| .$$

Their comparison is

$$p - q = 3|y| - 6|x| . \tag{3.3}$$

The finite singular arrangement

$$N := \{x = 0\} \cup \{y = 0\} \cup \{|y| = 2|x|\} \tag{3.4}$$

contains every nondifferentiability point of g. Outside N, the local core has horizontal and vertical branch families

$$H := \{(x,y) : |y| < 2|x|\}, \qquad \nabla g(x,y) = (\operatorname{sgn} x,\ 2\operatorname{sgn} y), \qquad (x,y) \in H \setminus N, \tag{3.5}$$

$$V := \{(x,y) : |y| > 2|x|\}, \qquad \nabla g(x,y) = (7\operatorname{sgn} x,\ -\operatorname{sgn} y), \qquad (x,y) \in V \setminus N. \tag{3.6}$$

Here $\operatorname{sgn} t \in \{-1, 1\}$ for $t \neq 0$. In H, both coordinate gradients point away from their axes, so a negative-gradient step attracts the coordinates toward zero. In V, the larger horizontal component drives the one-step axis crossing used below.

3.2 Exact subdifferentials

The main theorem requires more than Clarke criticality: its limit must fall outside the limiting-critical classification used in the generic theory. We therefore compute both subdifferentials exactly.

Proposition 3.2 (Limiting and Clarke geometry at the origin)**.** *At the origin,*

$$\partial F(0) = \partial g(0) = \big(\{-1,1\} \times [-2,2]\big) \cup \big([-7,7] \times \{-1,1\}\big), \tag{3.7}$$

whereas

$$\partial_{\mathrm{c}} F(0) = \partial_{\mathrm{c}} g(0) = \operatorname{conv}(\{(\sigma, 2\tau) : \sigma, \tau \in \{-1,1\}\} \cup \{(7\sigma, \tau) : \sigma, \tau \in \{-1,1\}\}) . \tag{3.8}$$

In particular,

$$[-1,1] \times [-2,2] \subseteq \partial_{\mathrm{c}} F(0). \tag{3.9}$$

Proof. Local equality (3.2) shows that all three subdifferentials of F and g agree at the origin. We compute those of g.

At a nonzero point of the horizontal axis, the local formula is p, so

$$\widehat{\partial} g(x,0) = \{\operatorname{sgn} x\} \times [-2,2] \qquad (x \neq 0). \tag{3.10}$$

At a nonzero point of the vertical axis, the local formula is q, so

$$\widehat{\partial} g(0,y) = [-7,7] \times \{-\operatorname{sgn} y\} \qquad (y \neq 0). \tag{3.11}$$

Now fix a nonzero point on $|y| = 2|x|$. In a sufficiently small neighborhood contained in one open orthant, g is the minimum of two distinct affine functions with gradients $a \neq b$. If w were a Fréchet subgradient, applying the defining inequality along td and $-td$ as $t \downarrow 0$, for a direction d with $\langle a-b, d\rangle \neq 0$, would give

$$\langle w, d\rangle \leq \min\{\langle a, d\rangle, \langle b, d\rangle\}, \qquad \langle w, d\rangle \geq \max\{\langle a, d\rangle, \langle b, d\rangle\},$$

a contradiction. Hence the Fréchet subdifferential is empty on every nonzero branch-boundary ray. Finally, $g(0,t) = -|t|$. The positive and negative vertical directions would require simultaneously $w_y \leq -1$ and $w_y \geq 1$, so $\widehat{\partial} g(0) = \varnothing$.

The finite stratification now makes the limiting calculation exact. The only nonempty Fréchet values on nondifferentiable strata approaching zero are (3.10) and (3.11); all differentiable gradients are endpoints of those segments. Conversely, $(\sigma,\eta) \in \{-1,1\} \times [-2,2]$ is realized along $z_i = (\sigma/i, 0)$, while $(\xi,\tau) \in [-7,7] \times \{-1,1\}$ is realized along $z_i = (0, -\tau/i)$. This proves (3.7).

The Clarke subdifferential is the closed convex hull of the eight gradients realized in the open sectors, giving (3.8). The four gradients from H alone have convex hull $[-1,1] \times [-2,2]$, which proves (3.9). $\square$

The exact formulas place the origin in the Clarke–limiting gap for every tilt in the open square U, while a vertical displacement gives an explicit descent direction.

Corollary 3.3 (Criticality gap at a nonminimum)**.** *For every $v \in U$,*

$$0 \in \partial_c F_v(0), \qquad 0 \notin \partial F_v(0). \tag{3.12}$$

The origin is not a local minimizer of F_v. Moreover, F is not subdifferentially regular at the origin and is not locally ρ-weakly convex there for any $\rho \geq 0$.

Proof. By (2.1), the two criticality statements are equivalent to $v \in \partial_c F(0)$ and $v \notin \partial F(0)$. The first follows from (3.9). The second follows from (3.7), because neither coordinate of $v \in U$ equals 1 or -1.

For $0 < |t| < 4$, local equality gives

$$F_v(0,t) = 8 - |t| - v_y t < 8 = F_v(0,0), \tag{3.13}$$

since $|v_y| < 1$. Thus zero is not a local minimizer. The mismatch between the Fréchet and limiting subdifferentials rules out subdifferential regularity. Finally, for any $\rho \geq 0$, the restriction of $F + \frac{\rho}{2}\|\cdot\|_2^2$ to the vertical axis near zero is $8 - |t| + \frac{\rho}{2}t^2$. Its left and right derivatives at zero are 1 and -1, respectively, violating the necessary convexity condition that the left derivative not exceed the right derivative. Hence no such local weak-convexity parameter exists. $\square$

4 Conditional absolute continuity avoids the singular locus

We show that every positive-time iterate avoids the singular arrangement N from (3.4), where the Clarke oracle could otherwise be set-valued. This is the only part of the proof that uses conditional absolute continuity; the support bound instead closes the deterministic invariant.

Lemma 4.1 (Oracle-independent first-step bound)**.** *Under the hypotheses of Theorem 2.2, uniformly over every admissible pre-noise value $w_0 \in \partial_c F_v(z_0)$, the first iterate satisfies almost surely*

$$|x_1| < 7.2\alpha_0, \qquad |y_1| < 2.5\alpha_0. \tag{4.1}$$

Proof. Since $\gamma < 1$, we have $\alpha_0 = 8^{-\gamma} > 1/8$. Hence $z_0 \in D_\star$ implies

$$|x_0| < 0.1\alpha_0, \qquad |y_0| < 0.4\alpha_0, \qquad \|z_0\|_1 < \frac{1}{16} < 4.$$

Thus $F = g + 8$ near z_0. The closed convex hull of the finite cell gradients shows that every Clarke value of g has first coordinate in $[-7, 7]$ and second coordinate in $[-2, 2]$. Hence

$$|w_{0,x}| < 7.01, \qquad |w_{0,y}| < 2.01.$$

Together with $\|\nu_0\|_\infty \le 0.01$ and the preceding bounds, the first update gives

$$|x_1| < (0.1 + 7.02)\alpha_0 < 7.2\alpha_0, \qquad |y_1| < (0.4 + 2.02)\alpha_0 < 2.5\alpha_0.$$

□

Lemma 4.2 (Almost-sure avoidance of the singular locus)**.** *Under the hypotheses of Theorem 2.2, fix an admissible pre-noise oracle and a noise process satisfying* (2.8)*. Then*

$$\mathbb{P}(z_k \notin N \text{ for every } k \ge 1) = 1. \tag{4.2}$$

On this event, whenever $k \ge 1$ and $\|z_k\|_1 < 4$,

$$w_k = \nabla g(z_k) - v, \qquad z_{k+1} = z_k - \alpha_k\big(\nabla g(z_k) + e_k\big), \qquad e_k := \nu_k - v, \tag{4.3}$$

and $\|e_k\|_\infty < \delta := 0.02$.

Proof. Conditional on $\mathcal{F}_k$, both z_k and w_k are fixed, while ν_k has an absolutely continuous law. Since $\alpha_k > 0$, the conditional law of

$$z_{k+1} = z_k - \alpha_k w_k - \alpha_k \nu_k$$

is absolutely continuous. Therefore $\mathbb{P}(z_{k+1} \in N \mid \mathcal{F}_k) = 0$ almost surely. Taking expectations and then a countable union proves (4.2).

For $k \ge 1$ with $\|z_k\|_1 < 4$, $F = g + 8$ locally. Outside N, the core is locally affine, so $\partial_c F_v(z_k) = \{\nabla g(z_k) - v\}$. The oracle is therefore forced to take the value in (4.3). Finally, $\|\nu_k - v\|_\infty < 0.02$. □

Remark 4.3 (A weaker noise condition). The noise assumption in Theorem 2.2 separates two roles: conditional absolute continuity supplies singular-locus avoidance, while the support bound supplies the pathwise invariant. Absolute continuity can be weakened to the direct conditional-hit condition

$$\mathbb{P}(z_k - \alpha_k w_k - \alpha_k \nu_k \in N \mid \mathcal{F}_k) = 0 \quad \text{almost surely for every } k.$$

This condition handles the history-dependent affine preimage of N. Marginal nullity for the fixed set N is insufficient because a conditional law could concentrate on that preimage after observing the past. Conversely, neither iid sampling nor mean zero is required.

For each realized path on the event in (4.2), all remaining estimates are deterministic and use only the coordinate bound.

5 The shrinking two-sector trap

After removing the singular-locus event, the proof is deterministic. We first record a scalar sign-step estimate and then use it to close a two-envelope invariant through three transitions.

The state at time $k \geq 1$ is normalized by the preceding step $r_k := \alpha_{k-1}$; the update at time k must produce bounds at scale α_k. This convention explains the index shift in (2.9).

Lemma 5.1 (Attracting sign step)**.** *If $u \geq 0$, $a > \delta \geq 0$, $\alpha > 0$, and $|e| \leq \delta$, then*

$$|u - \alpha(a+e)| \leq \max\{u - (a-\delta)\alpha,\, (a+\delta)\alpha\}. \tag{5.1}$$

Proof. If the step does not cross zero, the first term bounds the remaining distance. If it crosses zero, the overshoot is at most $(a+\delta)\alpha$. Equivalently, one may maximize the convex function $e \mapsto |u - \alpha(a+e)|$ over $[-\delta, \delta]$. $\square$

For $k \geq 1$, call z_k *tight* when

$$|x_k| \leq 7.2 r_k, \qquad |y_k| \leq 2.5 r_k, \tag{5.2}$$

and call it a *recovery* state when

$$|x_k| \leq 7.2 r_k, \qquad |y_k| \leq 3.9 r_k, \qquad |y_k| < 2\,|x_k|\,. \tag{5.3}$$

Proposition 5.2 (Two-state pathwise invariant)**.** *Let $\{\alpha_k\}_{k\geq 0}$ be positive and nonincreasing, with $\alpha_k \to 0$, and suppose*

$$\frac{\alpha_{k-1}}{\alpha_k} \leq \kappa := 1.13 \qquad (k \geq 1). \tag{5.4}$$

Consider the core dynamics

$$z_{k+1} = z_k - \alpha_k(\nabla g(z_k) + e_k), \qquad z_k \notin N, \quad \|e_k\|_\infty \leq 0.02 \qquad (k \geq 1).$$

Writing $z_k = (x_k, y_k)$, suppose

$$|x_1| \leq 7.2\alpha_0, \qquad |y_1| \leq 2.5\alpha_0. \tag{5.5}$$

Then $|x_k| \leq 7.2\alpha_{k-1}$ and $|y_k| \leq 3.9\alpha_{k-1}$ for every $k \geq 1$, and $z_k \to 0$. More precisely, if this trajectory is tight and lies in V, its next iterate lies strictly in H; hence two consecutive V-steps are impossible.

Proof. Every recovery state lies in H, and the initial state is tight. We verify three transitions; the two predicates need not be mutually exclusive.

Tight state in H. Multiplying each coordinate update by the sign of that coordinate reduces it to Lemma 5.1. Using $a = 1$ horizontally, $a = 2$ vertically, and $r_k \leq \kappa\alpha_k$, we obtain

$$|x_{k+1}| \leq \max\{(7.2\kappa - 0.98)\alpha_k,\, 1.02\alpha_k\} = 7.156\alpha_k < 7.2\alpha_k, \tag{5.6}$$

$$|y_{k+1}| \leq \max\{(2.5\kappa - 1.98)\alpha_k,\, 2.02\alpha_k\} = 2.02\alpha_k < 2.5\alpha_k. \tag{5.7}$$

The next state is tight, regardless of its new sector.

Tight state in V. The sector inequality and (5.2) give

$$|x_k| < \frac{|y_k|}{2} \leq 1.25 r_k \leq 1.4125\alpha_k. \tag{5.8}$$

Writing $\sigma_x = \operatorname{sgn} x_k$ and $\sigma_y = \operatorname{sgn} y_k$, the update in V is

$$\sigma_x x_{k+1} = |x_k| - \alpha_k(7 + \sigma_x e_{k,x}),$$
$$\sigma_y y_{k+1} = |y_k| + \alpha_k(1 - \sigma_y e_{k,y}).$$

The first expression is negative because $7 - 0.02 > 1.4125$; hence the horizontal coordinate crosses zero. The second is positive, so the vertical coordinate keeps its sign. Moreover,

$$5.5675\alpha_k < |x_{k+1}| < 7.02\alpha_k < 7.2\alpha_k, \tag{5.9}$$
$$|y_{k+1}| \le (2.5\kappa + 1.02)\alpha_k = 3.845\alpha_k < 3.9\alpha_k. \tag{5.10}$$

Since $3.845 < 2(5.5675)$, the next state lies strictly in H and is in the recovery envelope.

Recovery state in H. The horizontal estimate (5.6) is unchanged. For the vertical coordinate, Lemma 5.1 and the looser bound in (5.3) give

$$|y_{k+1}| \le \max\{(3.9\kappa - 1.98)\alpha_k,\, 2.02\alpha_k\} = 2.427\alpha_k < 2.5\alpha_k. \tag{5.11}$$

Thus the next state is tight.

The transitions close the induction. Both state types satisfy the asserted bounds; since $\alpha_k \to 0$, the trajectory converges to zero. The tight-V calculation also proves the final assertion. $\square$

6 Proof of the main theorem

We now combine first-step capture, singular-locus avoidance, and the pathwise invariant. The proof also checks that the invariant remains inside the neighborhood where the coercive extension equals the local core.

Proof of Theorem 2.2. Fix v, γ, z_0, an admissible pre-noise oracle, and a noise process satisfying (2.8). By Lemma 4.1, (4.1) holds and hence so does the initial condition (5.5) of Proposition 5.2. Work on the probability-one event from Lemma 4.2.

For every $k \ge 1$,

$$\frac{\alpha_{k-1}}{\alpha_k} = \left(\frac{k+8}{k+7}\right)^{\gamma} < \frac{k+8}{k+7} \le \frac{9}{8} < 1.13. \tag{6.1}$$

We now apply the transitions of Proposition 5.2 inductively. There is one local-to-global point to check. At every stage of the transition induction, the tight or recovery bounds give

$$\|z_k\|_1 \le 11.1\alpha_{k-1} \le 11.1\alpha_0 < \frac{11.1}{\sqrt{8}} < 4. \tag{6.2}$$

Thus the induction hypothesis itself places z_k in the region where $F = g + 8$. Since $z_k \notin N$, Lemma 4.2 forces the oracle update to be the core dynamics with $\|e_k\|_\infty < 0.02$. The appropriate transition in Proposition 5.2 then produces the next state. This proves (2.9), simultaneously verifies the local formula, and closes the induction. It also proves $z_k \to 0$ and boundedness.

Lemma 3.1 proves that F_v is nonnegative and coercive, while Corollary 3.3 proves that the limit is Clarke-critical, not limiting-critical, and not a local minimizer. This establishes $\mathbb{P}(z_k \to 0) = 1$; since zero is not in $\operatorname{LocMin}(F_v)$, it also gives $\mathbb{P}(\mathcal{B} \setminus \mathcal{L}_v(F)) = 1$, proving (2.10). $\square$

Proof of Corollary 2.5. The open square U has Lebesgue measure $4 \cdot 10^{-4} > 0$. If a conull good set G_F existed, then $G_F \cap U$ would have positive measure. For every $v \in G_F \cap U$, however, the explicit algorithmic instance in Corollary 2.3 is admissible and has $\mathbb{P}(\mathcal{B} \setminus \mathcal{L}_v(F)) = 1$, a contradiction. $\square$

7 Further consequences

The counterexample extends to every finite uniform-noise radius and every dimension at least two. The statements are recorded here; their technical proofs appear in Appendix A.

7.1 Arbitrary finite uniform-noise radius

Corollary 7.1 (Scaling to every finite noise radius). *Fix $R > 0$ and $\gamma \in (1/2, 1)$. There is a nonnegative coercive globally Lipschitz semialgebraic piecewise-affine counterexample for iid noise uniform on $B_R(0)$. More precisely, choose $C \geq 100R$, set $\widehat{F} := CF$, use any tilt $\widetilde{v} := Cv$ with $v \in U$, let $a_k := (k+8)^{-\gamma}$, and use steps*

$$\beta_k := C^{-1} a_k.$$

For every fixed Borel selection $\widehat{s}_{\widetilde{v}}(z) \in \partial_c \widehat{F}_{\widetilde{v}}(z)$, every initialization in $D_\star$ converges almost surely to the nonminimizing origin.

7.2 Embedding in every dimension

The separable embedding preserves both the two-dimensional oracle robustness and the Clarke–limiting gap. It uses only absolute continuity of the first two-dimensional noise marginal, not independence of coordinates.

Corollary 7.2 (Counterexamples in all dimensions $d \geq 2$). *For $d \geq 2$, define*

$$F^{(d)}(z) := F(z_1, z_2) + 2 \sum_{j=3}^{d} |z_j| .$$

Let $U_d := U \times (-10^{-2}, 10^{-2})^{d-2}$, and for $v \in U_d$ set $F_v^{(d)} := F^{(d)} - \langle v, \cdot \rangle$. For every $v \in U_d$, $\gamma \in (1/2, 1)$, iid noise uniform on the Euclidean ball of radius 10^{-2} in $\mathbb{R}^d$, and every fixed Borel Clarke selection of $\partial_c F_v^{(d)}$, the iterates with $z_0 = 0$ and $\alpha_k = (k+8)^{-\gamma}$ converge almost surely to the nonminimizing origin.

8 Interpretation and related work

8.1 The exact gap in the generic theory

In arXiv version 2 of Davis et al. [11], Theorem 3.30 and Corollary 4.2 classify limiting-critical points after generic tilting, while Corollary 6.5 gives Clarke-critical convergence on the bounded-orbit event under the standing step and noise assumptions. Subdifferential regularity identifies these notions so that the classification and saddle-avoidance argument applies. Our origin instead has $0 \in \partial_c F_v(0)$ but $0 \notin \partial F_v(0)$. The example therefore does not contradict the limiting-critical classification; it shows that definability, coercivity, generic tilting, and continuous noise do not replace regularity for a Clarke-subgradient algorithm.

8.2 Why noise does not force escape

Nonzero vertical displacements near the origin decrease F_v, but conditional absolute continuity makes positive-time hits of the vertical axis null. Inside V, the gradient forces an axis crossing whose overshoot obeys (5.9)–(5.10). Continuous-time differential inclusions and asymptotic-pseudotrajectory heuristics [1] do not encode these one-step bounds.

8.3 Relation to prior work

Example 2 of Bianchi et al. [2] already gives the static open-tilt Clarke–limiting gap for $f(y,z) = -|y| + |z|$, but neither coercivity nor a noisy orbit converging to the origin. Those are the global and dynamical additions here. The tame convergence theorem of Davis et al. [12] places limit points in the Clarke critical set, while conservative-field theory gives analogous endpoint guarantees for automatic differentiation [4]. Neither classifies every Clarke-critical point as a local minimum.

Daniilidis and Drusvyatskiy [8] construct a four-dimensional limit cycle using a nondefinable interval-splitting function and a specially chosen Clarke selection. Ríos-Zertuche [17] use infinitely many shrinking modifications to obtain deterministic nonconvergence around a critical circle, and Bolte et al. [5] study oscillatory occupation measures. By contrast, our construction gives convergence to a nonminimum for a two-dimensional semialgebraic piecewise-affine objective under bounded conditionally absolutely continuous noise and every pre-noise Clarke oracle. Appendix C of Bianchi et al. [2] also discusses the invalid asymptotic-pseudotrajectory implication behind the withdrawn preprint of Schechtman [19].

Recent positive results impose assumptions or reach endpoints that do not close the present gap. In the active-geometry regimes of Lai and Song [14], bounded momentum trajectories converge for steps of order k^{-a} with $1/2 < a \leq 1$, and bounded stochastic subgradient trajectories converge almost surely for $2/3 < a \leq 1$, while bounded nonconvergence can occur at $a = 1/2$. Related work [6, 13] likewise classifies critical rather than local-minimum endpoints, and Bolte et al. [3] analyze inexact semialgebraic subgradient methods. For representation-dependent piecewise gradients, Traoré [20] obtains conservative-critical convergence; its Clarke-critical upgrade also requires a null interface, generic step scaling and initialization, and the stated dimension–smoothness condition. Qiu et al. [16] treat smooth and mirror methods and convex-composite normal-map proximal methods under an active strict saddle property. None covers arbitrary Clarke selections at the gap in (3.12).

8.4 Scope

The result concerns the selected Clarke-subgradient recursion (2.2), not momentum, adaptive preconditioning, or proximal normal-map algorithms. It also leaves intact escape theorems assuming weak convexity, active-manifold regularity, or an active strict saddle property.

9 Conclusion

Subdifferential regularity cannot be removed without replacement from the DDJ theorem for selected Clarke-subgradient recursions. For every $\gamma \in (1/2, 1)$, the iterates generated with the schedule $\alpha_k = (k+8)^{-\gamma}$ converge from an explicit positive-area basin to a nonminimizing Clarke-critical point on an open tilt set, almost surely and for every pre-noise Clarke oracle.

References

[1] M. Benaïm, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. *SIAM Journal on Control and Optimization*, 44(1):328–348, 2005. doi:10.1137/S0363012904439301.

[2] P. Bianchi, W. Hachem, and S. Schechtman. Stochastic subgradient descent escapes active strict saddles on weakly convex functions. *Mathematics of Operations Research*, 49(3):1761–1790, 2024. doi:10.1287/moor.2021.0194.

[3] J. Bolte, T. Le, E. Moulines, and E. Pauwels. Inexact subgradient methods for semialgebraic functions. *Mathematical Programming*, online first, 2025. doi:10.1007/s10107-025-02245-w.

[4] J. Bolte and E. Pauwels. Conservative set valued fields, automatic differentiation, stochastic gradient methods and deep learning. *Mathematical Programming*, 188(1):19–51, 2021. doi:10.1007/s10107-020-01501-5.

[5] J. Bolte, E. Pauwels, and R. 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.

[6] E. Chzhen and S. Schechtman. On convergence rates of subgradient descent on semialgebraic functions. *arXiv preprint arXiv:2604.17060v1*, 2026. https://arxiv.org/abs/2604.17060v1.

[7] F. H. Clarke. *Optimization and Nonsmooth Analysis*. SIAM, Philadelphia, second edition, 1990. doi:10.1137/1.9781611971309.

[8] A. Daniilidis and D. Drusvyatskiy. Pathological subgradient dynamics. *SIAM Journal on Optimization*, 30(2):1327–1338, 2020. doi:10.1137/19M1298147.

[9] D. Davis, M. Díaz, and D. Drusvyatskiy. Escaping strict saddle points of the Moreau envelope in nonsmooth optimization. *SIAM Journal on Optimization*, 32(3):1958–1983, 2022. doi:10.1137/21M1430868.

[10] D. Davis and D. Drusvyatskiy. Proximal methods avoid active strict saddles of weakly convex functions. *Foundations of Computational Mathematics*, 22(2):561–606, 2022. doi:10.1007/s10208-021-09516-w.

[11] D. Davis, D. Drusvyatskiy, and L. Jiang. Active manifolds, stratifications, and convergence to local minima in nonsmooth optimization. *Foundations of Computational Mathematics*, 26(2):779–861, 2026. doi:10.1007/s10208-025-09691-0. https://arxiv.org/abs/2108.11832v2.

[12] D. Davis, D. Drusvyatskiy, S. M. Kakade, and J. 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.

[13] L. Lai and M. Song. On the diameter of subgradient sequences in o-minimal structures. *arXiv preprint arXiv:2511.06868v3*, 2025; revised 2026. https://arxiv.org/abs/2511.06868v3.

[14] L. Lai and M. Song. Convergence of difference inclusions: A diameter criterion and step-size conditions. *arXiv preprint arXiv:2605.14345v2*, 2026. https://arxiv.org/abs/2605.14345v2.

[15] R. Pemantle. Nonconvergence to unstable points in urn models and stochastic approximations. *The Annals of Probability*, 18(2):698–712, 1990. doi:10.1214/aop/1176990853.

[16] J. Qiu, B. Ma, A. Milzarek, and J. Zhang. Stochastic saddle avoidance beyond unit excitation and smoothness: A pathwise Lyapunov–Perron framework. *arXiv preprint arXiv:2608.03001v2*, 2026. https://arxiv.org/abs/2608.03001v2.

[17] R. Ríos-Zertuche. Examples of pathological dynamics of the subgradient method for Lipschitz path-differentiable functions. *Mathematics of Operations Research*, 47(4):3184–3206, 2022. doi:10.1287/moor.2021.1241.

[18] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis.* Springer, Berlin, 1998. doi:10.1007/978-3-642-02431-3.

[19] S. Schechtman. Stochastic subgradient descent on a generic definable function converges to a minimizer. *arXiv preprint arXiv:2109.02455v3*, 2021; withdrawn in 2022. https://arxiv.org/abs/2109.02455v3.

[20] C. Traoré. Piecewise smooth functions and conservative fields: Calculus for nonsmooth nonconvex optimization beyond stratification. *arXiv preprint arXiv:2607.13973v3*, 2026. https://arxiv.org/abs/2607.13973v3.

A Robustness and extensions

This appendix records the strict parameter inequalities behind the pathwise invariant, noise rescaling, and a separable high-dimensional embedding. These technical extensions are not needed for the main falsification.

A.1 A parameter-family criterion

The numerical choices are convenient rather than isolated. The next criterion exposes the inequalities behind the mechanism.

Proposition A.1 (Parameterized two-sector trap)**.** *Let*

$$g_{A,B}(x,y) := \min\{|x| + A\,|y|\,,\, B\,|x| - |y|\}, \qquad m := \frac{B-1}{A+1},$$

where $0 \le \delta < 1$*,* $A > \delta$*, and* $B > 1$*. Suppose* $\kappa \ge 1$ *and* $X, Y, Z > 0$ *satisfy*

$$B + \delta < X, \qquad A + \delta < Y, \qquad X(\kappa - 1) < 1 - \delta, \tag{A.1}$$

$$Y(\kappa - 1) < A - \delta, \qquad \frac{Y\kappa}{m} < B - \delta, \qquad Y\kappa + 1 + \delta < Z, \tag{A.2}$$

$$Z < m(B - \delta) - Y\kappa, \qquad Z\kappa - (A - \delta) < Y. \tag{A.3}$$

Define the singular arrangement for this parameterized core by

$$N_{A,B} := \{x = 0\} \cup \{y = 0\} \cup \{|y| = m\,|x|\}.$$

Call $|y| < m\,|x|$ *and* $|y| > m\,|x|$ *the horizontal and vertical sectors, respectively. Let positive steps satisfy* $\alpha_k \to 0$ *and* $\alpha_{k-1}/\alpha_k \le \kappa$ *for every* $k \ge 1$*. Every trajectory*

$$z_{k+1} = z_k - \alpha_k(\nabla g_{A,B}(z_k) + e_k), \qquad z_k \notin N_{A,B}, \quad \|e_k\|_\infty \le \delta \qquad (k \ge 1),$$

with $z_k = (x_k, y_k)$ *and* $|x_1| \le X\alpha_0$ *and* $|y_1| \le Y\alpha_0$ *obeys*

$$|x_k| \le X\alpha_{k-1}, \qquad |y_k| \le Z\alpha_{k-1} \qquad (k \ge 1),$$

and converges to zero. Whenever $|x_k| \le X\alpha_{k-1}$*,* $|y_k| \le Y\alpha_{k-1}$*, and* $|y_k| > m\,|x_k|$ *at some* $k \ge 1$*, its next iterate lies in the horizontal sector.*

Proof. Use tight bounds (X, Y) and recovery bounds (X, Z). In the horizontal sector, Lemma 5.1 closes the x-coordinate by the first and third inequalities in (A.1), and closes the y-coordinate by the second inequality in (A.1) and the first inequality in (A.2). In a tight vertical state,

$$|x_k| < \frac{Y\alpha_{k-1}}{m} \le \frac{Y\kappa}{m}\alpha_k.$$

The second inequality in (A.2) forces the horizontal step to cross zero. Its new magnitude is below $(B+\delta)\alpha_k < X\alpha_k$ and above $(B-\delta-Y\kappa/m)\alpha_k$. The vertical magnitude is below $(Y\kappa+1+\delta)\alpha_k < Z\alpha_k$. The first inequality in (A.3) places this landing point strictly in the horizontal sector. Finally, the last inequality in (A.3), together with $A + \delta < Y$, returns a recovery state to the tight y-bound. These are the same three transitions as in Proposition 5.2. □

Our constants

$$(A, B, \delta, \kappa, X, Y, Z) = (2, 7, 0.02, 1.13, 7.2, 2.5, 3.9)$$

satisfy all inequalities strictly. Consequently an open neighborhood of the displayed parameter tuple preserves the same pathwise core trapping invariant. After singular-locus avoidance, the statement is entirely pathwise: no independence, centering, summability, martingale argument, or continuous-time approximation is used.

A.2 Finite-radius scaling

Proof of Corollary 7.1. Write $\widehat{F}_{\widetilde{v}} = CF_v$ and $W_k := \widehat{s}_{\widetilde{v}}(z_k)$. Then $C^{-1}W_k$ is a Clarke value for F_v at z_k. The update becomes

$$z_{k+1} = z_k - \beta_k(W_k + \nu_k) = z_k - a_k\left(C^{-1}W_k + \frac{\nu_k}{C}\right).$$

The scaled noise has a density and radius $R/C \le 0.01$, so Theorem 2.2 applies. Constant rescaling preserves the admissible power-law order. Finally, $\widehat{F}_{\widetilde{v}} = CF_v$ remains nonnegative and coercive. □

A.3 High-dimensional embedding

Proof of Corollary 7.2. For $z = (z_1, \ldots, z_d)$, separability and the definition by gradient limits give the product formula

$$\partial_{\mathrm{c}} F_v^{(d)}(z) = \partial_{\mathrm{c}} F_{(v_1,v_2)}(z_1, z_2) \times \prod_{j=3}^{d}\big(2\partial_{\mathrm{c}}\,|\cdot|\,(z_j) - v_j\big). \tag{A.4}$$

Consequently the first two coordinates of any d-dimensional pre-noise oracle form an $\mathcal{F}_k$-measurable two-dimensional pre-noise oracle. The first two coordinates of uniform noise on a d-ball have an absolutely continuous two-dimensional marginal and are bounded in absolute value by 10^{-2}. Thus Theorem 2.2 applies to the first two coordinates; no coordinate independence is used.

Writing $z_{j,k}$ and $\nu_{j,k}$ for coordinate j of z_k and ν_k, respectively, the full singular set is a finite union of hyperplanes. Thus the argument of Lemma 4.2 gives $z_{j,k} \neq 0$ for every $3 \le j \le d$ and $k \ge 1$, almost surely. With $e_{j,k} := \nu_{j,k} - v_j$, the positive-time update is

$$z_{j,k+1} = z_{j,k} - \alpha_k(2\,\mathrm{sgn}\, z_{j,k} + e_{j,k}), \qquad |e_{j,k}| < 0.02.$$

At the origin the extra-coordinate Clarke value lies in $[-2, 2] - v_j$, so the first step satisfies $|z_{j,1}| < 2.02\alpha_0$. If $|z_{j,k}| \le 2.1\alpha_{k-1}$, Lemma 5.1 gives

$$|z_{j,k+1}| \le \max\{(2.1 \cdot 1.13 - 1.98)\alpha_k,\ 2.02\alpha_k\} < 2.1\alpha_k.$$

Thus all extra coordinates converge to zero.

For completeness, Fréchet subdifferentials factor for sums on orthogonal coordinates. Passing to limits yields one inclusion in the limiting-product identity below. For the reverse inclusion, synchronize a sequence realizing the chosen element of $\partial F(0)$ with constant zero sequences in the extra coordinates and fixed Fréchet subgradients in $[-2, 2]$. Hence

$$\partial F^{(d)}(0) = \partial F(0) \times [-2, 2]^{d-2},$$

and convexification similarly yields

$$\partial_{\mathrm{c}} F^{(d)}(0) = \partial_{\mathrm{c}} F(0) \times [-2, 2]^{d-2}.$$

Translation by $v \in U_d$ now shows that the origin remains Clarke-critical but not limiting-critical, and the vertical descent from (3.13) persists. Finally,

$$F_v^{(d)}(z) \geq 0.99 \, \|(z_1, z_2)\|_1 + 1.99 \sum_{j=3}^{d} |z_j| \, ,$$

so every tilt in U_d remains nonnegative and coercive. $\square$