

PATHOLOGICAL SUBGRADIENT DYNAMICS*

ARIS DANILIDIS[†] AND DMITRIY DRUSVYATSKIY[‡]

Abstract. We construct examples of Lipschitz continuous functions, with pathological subgradient dynamics in both continuous and discrete time. In both settings, the iterates generate bounded trajectories and yet fail to detect any (generalized) critical points of the function.

Key words. subgradient algorithm, Lipschitz function, Clarke subdifferential, splitting set

AMS subject classifications. 90C30, 49J52, 65K10

DOI. 10.1137/19M1298147

1. Introduction. The subgradient method plays a central role in large-scale optimization and its numerous applications. The primary goal of the method for nonsmooth and nonconvex optimization is to find generalized critical points. For example, for a locally Lipschitz continuous function f , we may be interested in finding a point x satisfying the inclusion $0 \in \partial f(x)$, where the symbol $\partial f(x)$ denotes the Clarke subdifferential.¹ The main difficulty in analyzing subgradient-type methods is that it is unclear how to construct a Lyapunov potential for the iterates when the target function is merely Lipschitz continuous. One popular strategy to circumvent this difficulty is to pass to continuous time, where a Lyapunov function may be more apparent. Indeed, for reasonable function classes, the objective itself decreases along the continuous time subgradient trajectories of the function. For example, this is the path classically followed by Benaïm, Hofbauer, and Sorin [3, 4], Borkar [8], Ljung [27], and more recently by Davis et al. [17] and Duchi and Ruan [18].

Setting the formalism, consider the task of minimizing a Lipschitz continuous function f on \mathbb{R}^d by the subgradient method. It is intuitively clear that the asymptotic performance of the algorithm is dictated by the long term behavior of the absolutely continuous trajectories $\gamma: [0, \infty) \rightarrow \mathbb{R}^d$ of the associated subgradient dynamical system

$$(1) \quad -\dot{\gamma}(t) \in \partial f(\gamma(t)) \quad \text{for a.e. } t \in [0, \infty).$$

Asymptotic convergence guarantees for the subgradient method, such as in the seminal work of Nurminkii [31] and Norkin [32], and the more recent works of Duchi and Ruan [18] and Davis et al. [17], rely either explicitly or implicitly on the following assumption.

- (Lyapunov) For any trajectory $\gamma(\cdot)$ emanating from a noncritical point of f , the composition $f \circ \gamma$ must strictly decrease on some small interval $[0, T)$.

*Received by the editors November 6, 2019; accepted for publication February 14, 2020; published electronically May 4, 2020.

<https://doi.org/10.1137/19M1298147>

Funding: The first author's research was supported by grants CMM-AFB170001, FONDECYT 1171854 (Chile), and PGC2018-097960-B-C22 (MICINN (Spain) and ERDF (EU)). The research of the second author was supported by NSF grants DMS 1651851 and CCF 1740551.

[†]Department of Mathematical Engineering, DIM-CMM, UMI CNRS 2807, Beauchef 851, FCFM, Universidad de Chile, Santiago, 8370459, Chile (arisd@dim.uchile.cl, <http://www.dim.uchile.cl/~arisd>).

[‡]Department of Mathematics, University of Washington, Seattle, WA 98195 (ddrusv@uw.edu, <http://www.math.washington.edu/~ddrusv>).

¹The subdifferential $\partial f(x)$ is the convex hull of all limits of gradients taken at points approaching x , and at which the function is differentiable.

For example, it is known that the Lyapunov property holds for any convex [11, 12], subdifferentially regular [17, 28], semismooth, and Whitney stratifiable functions [17]. Since this property holds for such a wide class of functions, it is natural to ask the following question.

– Is the Lyapunov property simply true *for all* Lipschitz continuous functions?

In this work, we show that the answer is negative (cf. subsection 3.1). Indeed, we will show that there exist pathological Lipschitz functions that generate subgradient curves (1) with surprising behavior. For the first example, we construct (see Proposition 3) a Lipschitz continuous function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ and a subgradient curve $\gamma: [0, T) \rightarrow \mathbb{R}^2$ emanating from a noncritical point $\gamma(0)$, such that $f \circ \gamma$ *increases* α -linearly (for some $\alpha > 0$):

$$f(\gamma(t)) - f(\gamma(0)) \geq \alpha t \quad \text{for all } t \in [0, T].$$

In particular, the Lyapunov property clearly fails. Our second example presents a Lipschitz function $f: \mathbb{R}^4 \rightarrow \mathbb{R}$ and a *periodic* subgradient curve $\gamma: [0, \infty) \rightarrow \mathbb{R}^d$ that contains no critical points of f . In particular, the limiting set of the trajectory $\gamma(\cdot)$ is disjoint from the set of critical points of the function f (see Theorem 5 in subsection 3.2). Our final example returns to the discrete subgradient method with the usual nonsummable but square summable stepsize. We construct a Lipschitz function $f: \mathbb{R}^4 \rightarrow \mathbb{R}$, for which the subgradient iterates form a limit cycle that is disjoint from the critical point set of f (cf. subsection 3.3). Thus the method fails to find any critical points of the constructed function.

The examples we construct are built from so-called interval splitting sets (Definition 1). These are the subsets of the real line, whose restriction to any interval is neither zero- nor full-measure. Splitting sets have famously been used by Ekeland and Lebourg [20] and Rockafellar [35] to construct a pathological Lipschitz function, for which the Clarke subdifferential is the unit interval $\partial f = [-1, 1]$ everywhere. Later, it was shown that functions with such pathologically large subdifferentials are topologically [10] and algebraically generic [16] (see section 2). Notice the function above does not directly furnish a counterexample to the Lyapunov property, since every point in its domain is critical. Nonetheless, in this work, we borrow the general idea of using splitting sets to define Lipschitz functions with large Clarke subdifferentials. The pathological subgradient dynamics then appear by an adequate selection of subgradients that yields a smooth vector field with simple dynamics. It is worthwhile to note that, in contrast to the aforementioned works, the functions we construct trivially satisfy the conclusion of the Morse–Sard theorem: the set of the Clarke critical values has zero-measure.

Although this work is of purely theoretical interest, it does identify a limitation of the subgradient method and the differential inclusion approach in nonsmooth and nonconvex optimization. In particular, this work supports the common practice of focusing on alternative techniques (e.g., smoothing [23, 30], gradient sampling [13]), or explicitly targeting better behaved function classes (e.g., weakly convex [1, 31, 36], amenable [33], prox-regular [34], generalized differentiable [22, 29, 32], and semi-algebraic [7, 25]).

2. Notation. Throughout, we let \mathbb{R}^d denote the standard d -dimensional Euclidean space with inner product $\langle \cdot, \cdot \rangle$ and the induced norm $\|x\| = \sqrt{\langle x, x \rangle}$. The symbol \mathbb{B} will stand for the closed unit ball in \mathbb{R}^d . For any set $A \subset \mathbb{R}^d$, we let χ_A denote the function that evaluates to one on A and to zero elsewhere. Throughout, we let $m(\cdot)$ denote the Lebesgue measure in \mathbb{R} .

A function $f: \mathcal{U} \rightarrow \mathbb{R}$, defined on an open set $\mathcal{U} \subset \mathbb{R}^d$, is called *Lipschitz continuous* if there exists a real $L > 0$ such that the following estimate holds:

$$(2) \quad |f(x) - f(y)| \leq L\|x - y\| \quad \text{for all } x, y \in \mathcal{U}.$$

The infimum of all constants $L > 0$ satisfying (2) is called the Lipschitz modulus of f and will be denoted by $\|f\|_{\text{Lip}}$. For any Lipschitz continuous function f , we let $\mathcal{D}_f \subset \mathcal{U}$ denote the set of points where f is differentiable. The classical Rademacher's theorem guarantees that \mathcal{D}_f has full Lebesgue measure in \mathcal{U} . The *Clarke subdifferential* of f at x is then defined to be the set

$$(3) \quad \partial f(x) := \text{conv} \left\{ \lim_{k \rightarrow \infty} \nabla f(x_k) : x_k \rightarrow x \text{ and } \{x_k\} \subset \mathcal{D}_f \right\},$$

where the symbol conv denotes the convex hull operation. It is important to note that in the definition, the set \mathcal{D}_f can be replaced by any full-measure subset $\mathcal{D} \subset \mathcal{D}_f$; see [14, Chapter 2] for details. It is easily seen that $\partial f(x)$ is a nonempty convex compact set, whose elements are bounded in norm by $\|f\|_{\text{Lip}}$. A point x is called (*Clarke*) *critical* if the inclusion $0 \in \partial f(x)$ holds. We will denote the set of all critical points of f by $\text{Crit}(f)$. A real number $r \in \mathbb{R}$ is called a *critical value* of f whenever there exists some point $x \in \text{Crit}(f)$ satisfying $f(x) = r$.

Though the definition of the Clarke subdifferential is appealingly simple, the behavior of the (set-valued) map $x \mapsto \partial f(x)$ can be quite pathological; see, e.g., [9, 15, 19]. For example, there exists a 1-Lipschitz function $f: \mathbb{R} \rightarrow \mathbb{R}$ having maximal possible subdifferential $\partial f(x) = [-1, 1]$ at every point on the real line. The first example of such *Clarke-saturated* functions appears in [20, Proposition 1.9] and is based on interval splitting sets.

DEFINITION 1 (splitting set). *A measurable set $A \subseteq \mathbb{R}$ is said to split intervals if for every nonempty interval $I \subset \mathbb{R}$ it holds that*

$$(4) \quad 0 < m(A \cap I) < m(I),$$

where m denotes the Lebesgue measure.

The first definition and construction of a splitting set goes back to [26], while the first examples of Clarke saturated functions can be found in [20, 35]. The basic construction proceeds as follows. For any fixed set $A \subseteq \mathbb{R}$ that splits intervals, define the univariate function

$$f(t) = \int_0^t \chi_A(\tau) - \chi_{A^c}(\tau) \, d\tau \quad \text{for all } t \in \mathbb{R}.$$

An easy computation shows the f is Clarke saturated, namely, $\partial f(t) = [-1, 1]$ for all $t \in \mathbb{R}$. We will use this observation throughout. Famously, the papers [9, 10] established that, for the uniform topology, a “generic” Lipschitz function (in the Baire sense) is Clarke saturated. Although Clarke saturation is not generic under the $\|\cdot\|_{\text{Lip}}$ -topology, it has been recently proved in [16] that the set of Clarke saturated functions contains a nonseparable Banach space.² We refer the reader to [2, 5, 6, 21] for recent results on the topic.

²Consequently, in the notation of [24], the set of Clarke saturated functions is “spaceable.” Moreover, if we endow the space of Lipschitz continuous functions with the $\|\cdot\|_{\text{Lip}}$ -seminorm, the space of pathological functions contains an isometric copy of ℓ^∞ .

3. Main results. In this section we construct the three pathological examples announced in the introduction. Our first example will make use of a splitting set that satisfies an auxiliary property. The construction is summarized in Lemma 2. The proof is essentially standard, and therefore we have placed it in the appendix.

LEMMA 2 (controlled splitting). *For every $\lambda \in (\frac{1}{2}, 1)$, there exists a measurable set $A \subset \mathbb{R}$ that splits intervals and satisfies*

$$(5) \quad m(A \cap [0, t]) \geq \lambda t \quad \text{for every } t > 0.$$

3.1. Nondecreasing subgradient trajectories. The following proposition answers in the negative the first question of the introduction, revealing that the Lyapunov property for a subgradient trajectory may fail.

PROPOSITION 3 (linear increase along orbits). *Let $\alpha > 0$ be arbitrary. Then there exists a Lipschitz continuous function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ and a subgradient orbit $\gamma: [0, +\infty) \rightarrow \mathbb{R}^d$ emanating from a noncritical point, meaning*

$$(6) \quad \begin{cases} -\dot{\gamma}(t) \in \partial f(\gamma(t)) \text{ for a.e. } t \in [0, +\infty), \\ \gamma(0) \notin \text{Crit}(f) \end{cases}$$

and satisfying the linear increase guarantee

$$f(\gamma(t)) - f(\gamma(0)) \geq \alpha t \quad \text{for every } t \in [0, +\infty).$$

Proof. According to Lemma 2, there exists a constant $\lambda > \frac{1}{2}$ and a set $A \subset \mathbb{R}$ satisfying (5). Define the constant $\mu := \sqrt{\frac{\alpha+1}{2\lambda-1}}$ and define the function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ by

$$f(x, y) := -x + \mu \int_0^y (\chi_A(\tau) - \chi_{A^c}(\tau)) \, d\tau.$$

It is easily seen that f is Lipschitz continuous and the Clarke subdifferential of f is given by

$$\partial f(x, y) = \{-1\} \times [-\mu, \mu] \quad \text{for every } (x, y) \in \mathbb{R}^2.$$

Notice that $\partial f(x, y)$ contains the direction $u = -(1, \mu)$ at every point (x, y) . Taking into account $\text{Crit}(f) = \emptyset$, we deduce that the curve $\gamma(t) = -tu = (t, \mu t)$ satisfies the system (6). Moreover, we have the estimate

$$f(\gamma(t)) = f(t, \mu t) = -t + \mu (m(A \cap [0, \mu t]) - m(A^c \cap [0, \mu t])) \geq ((2\lambda - 1)\mu^2 - 1)t \geq \alpha t.$$

The proof is complete. \square

3.2. (Periodic) subgradient orbits without critical points. Next, we present the second example given in the introduction, namely, a Lipschitz continuous function along with a periodic subgradient curve that contains no critical points of the function. We begin with the following intermediate construction. Henceforth, the symbol \mathbb{B}_∞ will denote the closed unit ℓ_∞ -ball in \mathbb{R}^d , and i will denote the imaginary unit.

THEOREM 4. *Fix an arbitrary real $M > 0$ and $b \in (0, \frac{M}{2})$, and let $A \subset \mathbb{R}$ be a measurable subset that splits intervals. Define the function $\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}$ by*

$$\Phi(x, y) := xy + M \int_0^y (\chi_A(\tau) - \chi_{A^c}(\tau)) \, d\tau.$$

Then the following are true:

- (i) The function Φ is $2M$ -Lipschitz continuous when restricted to the ball $b\mathbb{B}_\infty$.
- (ii) Equality holds: $b\mathbb{B}_\infty \cap \text{Crit}(\Phi) = [-b, b] \times \{0\}$.
- (iii) For any real $r < b$ and $\theta \in \mathbb{R}$, the curve $\gamma(t) = re^{i(t+\theta)}$ is a subgradient orbit of Φ , that is, $-\dot{\gamma}(t) \in \partial\Phi(\gamma(t))$ for all $t > 0$.

Proof. The standard sum rule yields the expression for the subdifferential

$$(7) \quad \partial\Phi(x, y) = \{(y, x + h) : h \in [-M, M]\} \quad \text{for all } (x, y) \in \mathbb{R}^2.$$

The first claim then follows immediately by noting

$$\max_{(x, y) \in b\mathbb{B}_\infty} \max_{v \in \partial\Phi(x, y)} \|v\| \leq 2M.$$

The second claim also follows immediately from the expression (7). Next, for any $(x, y) \in b\mathbb{B}_\infty$ we can take $h = -2x$ in (7), yielding the selection $(-y, x) \in -\partial\Phi(x, y)$. Thus the inclusion $\dot{\gamma} \in -\partial\Phi(\gamma)$ holds as long as the curve $\gamma(t) = (x(t), y(t))$ satisfies the ODE

$$(8) \quad \dot{x}(t) = -y(t), \quad \dot{y}(t) = x(t) \quad \text{for all } t.$$

Clearly, the curve $\gamma(t) := re^{i(t+\theta)}$ indeed satisfies (8). \square

Thus Theorem 4 provides an example of a periodic subgradient curve $\gamma(\cdot)$ for a Lipschitz continuous function $\Phi: \mathbb{R}^d \rightarrow \mathbb{R}$. The deficiency of the construction is that γ does pass through some critical points of f . We will now see that by doubling the dimension, we can ensure that the periodic curve never passes through the critical point set.

THEOREM 5 (periodic subgradient orbits without critical points). *There exists a Lipschitz continuous function $f: \mathcal{U} \rightarrow \mathbb{R}$, defined on an open set $\mathcal{U} \subset \mathbb{R}^4$, and a periodic analytic curve $\gamma: \mathbb{R} \rightarrow \mathcal{U}$, satisfying*

$$\dot{\gamma}(t) \in -\partial f(\gamma(t)) \quad \text{for all } t \quad \text{and} \quad \text{Im}(\gamma) \cap \text{Crit}(f) = \emptyset.$$

Proof. Let M , b , A , and Φ be as defined in Theorem 4. Set $\mathcal{U} := bB_\infty \times bB_\infty$ and define the function

$$(9) \quad \begin{cases} f: \mathcal{U} \rightarrow \mathbb{R}, \\ f(x_1, x_2, x_3, x_4) = \Phi(x_1, x_2) + \Phi(x_3, x_4). \end{cases}$$

It follows easily from Theorem 4 that the critical point set is given by

$$(10) \quad \text{Crit}(f) = \{(x_1, 0, x_3, 0) : \max\{|x_1|, |x_3|\} \leq b\}.$$

Define the curve $\gamma: \mathbb{R} \rightarrow \mathcal{U}$ by $\gamma(t) = \frac{b}{2}(e^{it}, e^{i(t+\frac{\pi}{2})})$. Theorem 4 immediately guarantees the inclusion $\dot{\gamma}(t) \in -\partial f(\gamma(t))$ for all $t \in \mathbb{R}$, while the expression (10) implies $\text{Im}(\gamma) \cap \text{Crit}(f) = \emptyset$. See Figure 1 for an illustration. \square

It is worthwhile to note that the function f in Theorem 5 trivially satisfies the conclusion of the Morse–Sard theorem, since $f(\text{Crit}(f)) = \{0\}$.

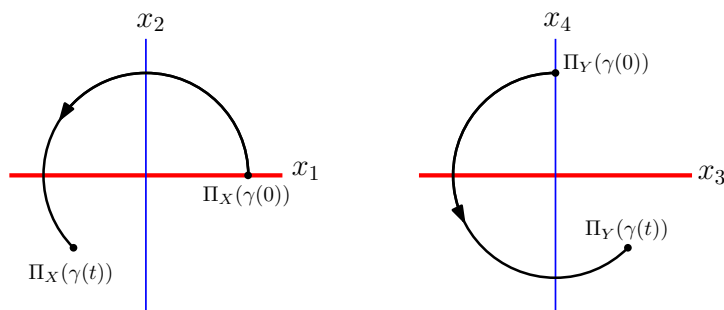


FIG. 1. Subgradient curve $\gamma(\cdot)$ generated by f is depicted in the figure with an arrow; the set of critical points of f is the product of the x_1 and x_3 axes. The picture on the left shows the projection of γ onto the coordinate $X = (x_1, x_2)$, while the picture on the right shows the projection of γ onto the coordinates $Y = (x_3, x_4)$.

3.3. Subgradient sequences without reaching critical points. We next present the final example announced in the introduction. We exhibit a Lipschitz continuous function f such that the subgradient method, which can access f only by querying subgradients, fails to detect critical points in any sense, under any choice of (nonsummable, square summable) steps $\{t_n\}_{n \geq 1}$. As the initial attempt at the construction, one may try applying the subgradient method to the function f constructed in Theorem 5, since it has periodic subgradient orbits in continuous time. The difficulty is that when applied to this function, the subgradient iterates (in discrete time) quickly grow unbounded. Therefore, as part of the construction, we will modify the function f from Theorem 5 by exponentially damping its slope.

THEOREM 6 (subgradient method does not detect critical points). *There exists a Lipschitz continuous function $f: \mathbb{R}^4 \rightarrow \mathbb{R}$, a subgradient selection $G(X) \in \partial f(X)$, and initial condition $\bar{X} \in \mathbb{R}^4$ such that*

$$(11) \quad \text{for every sequence } \{t_n\}_n \subset (0, +\infty) \text{ with } \sum_{n \geq 1} t_n = +\infty \text{ and } \sum_{n \geq 1} t_n^2 < +\infty$$

the subgradient algorithm

$$(12) \quad \begin{cases} X_{n+1} = X_n - t_n G(X_n) \\ X_1 = \bar{X} \end{cases}$$

generates a bounded sequence $\{X_n\}_{n \geq 1}$ whose accumulation points do not meet $\text{Crit}(f)$.

Proof. Let us first define the function $\rho: \mathbb{R}^2 \rightarrow \mathbb{R}$ by

$$\rho(x, y) = \exp(-y^2) \int_0^x \exp(-\tau^2) d\tau,$$

and notice that

$$\nabla \rho(x, y) = \left(e^{-(x^2+y^2)}, -2y \cdot \rho(x, y) \right).$$

Then for $\delta > 0$ and $M \geq \frac{\delta}{2}(\sqrt{\pi} + 1)$, define the function $\phi: \mathbb{R}^2 \rightarrow \mathbb{R}$ by

$$(13) \quad \phi(x, y) := \delta y \rho(x, y) + M \int_0^y (\chi_A(\tau) - \chi_{A^c}(\tau)) d\tau.$$

An easy calculation shows that ϕ is Lipschitz continuous and its subdifferential is given by

$$(14) \quad \partial\phi(x, y) = \left\{ \delta y e^{-(x^2+y^2)} \right\} \times \left\{ \left(\delta (1 - 2y^2) e^{-y^2} \int_0^x e^{-\tau^2} d\tau \right) + [-M, M] \right\}.$$

It follows from (14) that a point $(x, y) \in \mathbb{R}^2$ is Clarke critical for ϕ if and only if $y = 0$, that is,

$$\text{Crit}(\phi) = \mathbb{R} \times \{0\}.$$

We claim that for every $(x, y) \in \mathbb{R}^2$, we have

$$(15) \quad g(x, y) := \delta e^{-(x^2+y^2)}(y, -x) \in \partial\phi(x, y).$$

To see this, denoting by $\pi_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$ the projection to the second coordinate, we observe that

$$\begin{aligned} \left| \pi_2(g(x, y)) - \delta (1 - 2y^2) e^{-y^2} \int_0^x e^{-\tau^2} d\tau \right| &= \delta e^{-y^2} \left| (1 - 2y^2) \int_0^x e^{-\tau^2} d\tau - x e^{-x^2} \right| \\ &\leq \delta e^{-y^2} |1 - 2y^2| \cdot \frac{\sqrt{\pi}}{2} + |x| e^{-x^2} \\ &< \delta \cdot \frac{\sqrt{\pi}}{2} + \frac{1}{2} \leq M. \end{aligned}$$

Notice that the integral curves of the above vector field,

$$(\dot{x}, \dot{y}) = g(x, y) = \delta e^{-(x^2+y^2)}(y, -x),$$

are the homocentric cycles $x^2 + y^2 = r^2$ for any $r \geq 0$.

Now we consider applying a subgradient method to ϕ . Namely, let $\gamma_n := (x_n, y_n)$, and consider the subgradient sequence

$$(16) \quad \gamma_{n+1} = \gamma_n - t_n g_n,$$

where we set $g_n := g(x_n, y_n) \in \partial\phi(x_n, y_n)$. Then the norms $r_n := \|\gamma_n\|$ satisfy

$$(17) \quad \|g_n\| = \delta r_n e^{-r_n^2} \leq \frac{\delta}{\sqrt{2e}}.$$

Since g_n is tangent at γ_n to the homocentric cycle centered at 0 with radius r_n , we deduce easily from (16) that the sequence $\{r_n\}_{n \geq 1}$ is strictly increasing. On the other hand, by the Pythagorean theorem and (17), we deduce

$$r_{n+1}^2 = r_n^2 + t_n^2 \|g_n\|^2 \leq r_n^2 + \left(\frac{\delta^2}{2e} \right) t_n^2,$$

and by induction

$$r_{n+1}^2 \leq r_1^2 + \left(\frac{\delta^2}{2e} \right) \sum_{n \geq 1} t_n^2 < +\infty.$$

Therefore, $\{r_n\}_{n \geq 1}$ is bounded and the sequence $\gamma_n := (x_n, y_n)$ has accumulation points.

The proof is not yet complete, since in principle the accumulation points of the sequence (x_n, y_n) may be critical. To eliminate this possibility, we proceed as in the proof of Theorem 5 by doubling the dimension. Namely, define the function

$$(18) \quad \begin{cases} f: \mathbb{R}^4 \rightarrow \mathbb{R}, \\ f(x, y, z, w) = \phi(x, y) + \phi(z, w) \end{cases}$$

and observe that f is Lipschitz continuous and equality holds:

$$\text{Crit}(f) = \mathbb{R} \times \{0\} \times \mathbb{R} \times \{0\}.$$

We shall now prescribe a subgradient selection:

$$(19) \quad G(x, y, z, w) = (g(x, y), g(z, w)) \in \partial f(x, y, z, w),$$

where $g(\cdot, \cdot)$ is defined in (15). Let us also prescribe the initial condition

$$X_1 \equiv \bar{X} = (1, 0, 0, 1) \in \mathbb{R}^4.$$

Then (12) generates a bounded sequence

$$X_n := (x_n, y_n, z_n, w_n) \quad \text{with} \quad G_n := G(X_n) \in \partial f(X_n)$$

which splits in $\mathbb{R}^2 \times \mathbb{R}^2$ as follows:

$$X_n = (\gamma_n, \tilde{\gamma}_n), \quad G_n = (g_n, \tilde{g}_n) \in \partial f(X_n) \quad \text{with} \quad \gamma_n := (x_n, y_n) \quad \text{and} \quad \tilde{\gamma}_n := (z_n, w_n)$$

such that

$$(20) \quad \left\{ \begin{array}{l} \gamma_{n+1} = \gamma_n - t_n g_n \\ g_n \in \partial \phi(\gamma_n) \end{array} \right\} \quad \text{and} \quad \left\{ \begin{array}{l} \tilde{\gamma}_{n+1} = \tilde{\gamma}_n - t_n \tilde{g}_n \\ \tilde{g}_n \in \partial \phi(\tilde{\gamma}_n) \end{array} \right\},$$

respectively. Thanks to the initial condition, for every $n \geq 1$ the vector $\tilde{\gamma}_n$ is a $\frac{\pi}{2}$ -rotation of the vector γ_n . Taking into account this rotational symmetry, we deduce that all limit points of $\{X_n\}_{n \geq 1}$ lie outside of $\text{Crit}(f)$. \square

It is worthwhile to note again that the function f in Theorem 6 trivially satisfies the conclusion of the Morse–Sard theorem, since $f(\text{Crit}(f)) = \{0\}$.

Appendix: Proof of Lemma 2. We will first need the following lemma, outlining a standard construction of a fat Cantor set. We will impose an extra property, which will play a key role in establishing Lemma 2.

LEMMA 7 (α -fat Cantor set with attributes). *Fix any $\alpha \in (0, 1)$. Then for every interval $I = [a, b] \subset \mathbb{R}$ there exists an α -fat Cantor subset $\mathcal{F}_\alpha(I) \subset I$, that is, a Cantor-type set of total measure*

$$m(\mathcal{F}_\alpha(I)) = \alpha m(I).$$

Moreover, for any $\frac{1}{2} < \lambda < 1$, taking $\alpha > \frac{3\lambda}{2+\lambda}$, we ensure that

$$(21) \quad m(\mathcal{F}_\alpha(I) \cap [a, x]) \geq \lambda \cdot m([a, x]) \quad \text{for all } x \in [a, b].$$

Proof. The construction is standard and is sketched for the reader's convenience. Fix any $\alpha \in (0, 1)$. Let us denote $I^0 = (a, b)$, $\ell_0 = m(I^0) = b - a$ and let us set

$$(22) \quad \delta_0 := \left(\frac{1 - \alpha}{2} \right) \cdot \ell_0 > 0.$$

We shall construct the fat Cantor set by removing, successively, countably many intervals, indexed by a dyadic tree. To this end, we start by removing from our initial set I^0 the interval

$$\Delta^0 := \left[\frac{b + a - \delta_0}{2}, \frac{b + a + \delta_0}{2} \right],$$

that is, an interval of length δ_0 centered at $\frac{a+b}{2}$ (the midpoint of I^0). Then from each of the two remaining intervals $I_0^1 := (a, \frac{b+a-\delta_0}{2})$ and $I_1^1 := (\frac{b+a+\delta_0}{2}, b)$ we subtract intervals Δ_0^1 and Δ_1^1 of length

$$(23) \quad \delta_1 := \frac{1}{2} \cdot \frac{\delta_0}{2}$$

centered at the midpoints of I_0^1 and I_1^1 , respectively. Setting

$$m(I_0^1) = m(I_1^1) := \ell_1 = \frac{\ell_0 - \delta_0}{2},$$

we observe that

$$\frac{m(\Delta^0)}{m(I^0)} = \frac{\delta_0}{\ell_0} = \frac{1 - \alpha}{2} \quad \text{and} \quad \frac{m(\Delta_i^1)}{m(I_i^1)} = \frac{\delta_1}{\ell_1} = \frac{(\delta_0/4)}{(\ell_0 - \delta_0)/2} = \frac{1}{2} \left(\frac{1 - \alpha}{1 + \alpha} \right) \quad \text{for } i \in \{0, 1\},$$

and consequently

$$(24) \quad \frac{m(\Delta_i^1)}{m(I_i^1)} = \left(\frac{1}{1 + \alpha} \right) \frac{m(\Delta^0)}{m(I^0)} < \frac{m(\Delta^0)}{m(I^0)} \quad \text{for } i \in \{0, 1\}.$$

The above relation reveals that in the second step, we subtract a proportionally smaller part of each of the intervals I_i^1 , $i \in \{0, 1\}$, compared with what we subtract from I^0 in the first step.

We continue by induction, subtracting at step n intervals Δ_k^n , $k \in \{0, 1, \dots, 2^n - 1\}$, of length

$$\delta_n = \frac{1}{2} \cdot \frac{\delta_{n-1}}{2} = \frac{1}{2^n} \cdot \frac{\delta_0}{2^2}$$

centered at the midpoints of the intervals I_k^n , where $m(I_k^n) = \ell_n = (\ell_{n-1} - \delta_{n-1})/2$, so that for all $k' \in \{0, 1, \dots, 2^{n-1} - 1\}$ and $k \in \{0, 1, \dots, 2^n - 1\}$ we have

$$(25) \quad \frac{m(\Delta_k^n)}{m(I_k^n)} = \frac{\delta_n}{\ell_n} = \frac{(\delta_{n-1}/4)}{(\ell_{n-1} - \delta_{n-1})/2} < \frac{\delta_{n-1}}{\ell_{n-1}} = \frac{m(\Delta_{k'}^{n-1})}{m(I_{k'}^{n-1})}.$$

The above relation says that at step $n \geq 2$ we remove a proportionally smaller part of each of the intervals I_k^n , $k \in \{0, 1, \dots, 2^n - 1\}$, compared to what we did in the previous step to the intervals $I_{k'}^{n-1}$, $k' \in \{0, 1, \dots, 2^{n-1} - 1\}$.

Now let $\mathcal{F}_\alpha(I)$ be the complement of the union of all extracted intervals. Clearly $\mathcal{F}_\alpha(I)$ cannot contain any interval (i.e., it is a Cantor-type set), and in view of (22) its total length is

$$m(\mathcal{F}_\alpha(I)) = m(I) - \left(\delta + 2 \left(\frac{1}{2} \cdot \frac{\delta}{2} \right) + 2^2 \left(\frac{1}{2^2} \cdot \frac{\delta}{2^2} \right) + \dots \right) = m(I) - \delta \sum_{n \geq 0} \frac{1}{2^n} = \alpha m(I).$$

It remains to prove that (21) holds. To this end, we start by treating the case $x \in [a, b] \setminus \mathcal{F}_\alpha(I)$.

Let us first assume

$$x \in \Delta^0 := \left[\frac{b+a-\delta_0}{2}, \frac{b+a+\delta_0}{2} \right].$$

In this case (which is the less favorable case) we have

$$(26) \quad m([a, x]) \leq \frac{b-a+\delta_0}{2} = \frac{\ell_0+\delta_0}{2} \quad \text{and} \quad m(\mathcal{F}_\alpha(I) \cap [a, x]) = \frac{m(\mathcal{F}_\alpha(I))}{2} = \frac{\alpha\ell_0}{2}.$$

Therefore (21) holds for $\alpha > 3\lambda(2+\lambda)^{-1}$.

Assume now that $x \in \Delta_0^1 \cup \Delta_1^1$. By construction, using (22), (23), and (24), we deduce that for $i \in \{0, 1\}$, the set $\mathcal{F}_\alpha(I^0) \cap I_i^1$ is an α_1 -fat Cantor set in I_i^1 , denoted $\mathcal{F}_{\alpha_1}(I_i^1)$, where α_1 satisfies

$$\frac{1-\alpha_1}{2} := \frac{\delta_1}{\ell_1} = \left(\frac{1}{1+\alpha} \right) \left(\frac{\delta_0}{\ell_0} \right) = \left(\frac{1}{1+\alpha} \right) \left(\frac{1-\alpha}{2} \right), \text{ yielding that } \alpha_1 = \frac{2\alpha}{1+\alpha} > \alpha.$$

The above guarantees that for $x \in \Delta_0^1 \subset I_0^1 = (a, \frac{b+a-\delta_0}{2})$ we have

$$m(\mathcal{F}_\alpha(I^0) \cap [a, x]) = m(\mathcal{F}_{\alpha_1}(I_0^1) \cap [a, x]) \geq \lambda \cdot m([a, x])$$

by the previous step. If now

$$x \in \Delta_1^1 \subset I_1^1 = \left(\frac{b+a+\delta_0}{2}, b \right),$$

then by (26) and the fact that $\mathcal{F}_{\alpha_1}(I_1^1) := \mathcal{F}_\alpha(I^0) \cap I_1^1$ is an α_1 -fat Cantor set in I_1^1 with $\alpha_1 > \alpha$ we deduce that

$$\begin{aligned} m(\mathcal{F}_\alpha(I^0) \cap [a, x]) &= \frac{m(\mathcal{F}_\alpha(I))}{2} + m\left(\mathcal{F}_{\alpha_1}(I_1^1) \cap \left[\frac{b+a+\delta_0}{2}, x\right]\right) \\ &\geq \lambda \cdot \left(m\left(\left[a, \frac{b+a+\delta_0}{2}\right]\right) + m\left(\left[\frac{b+a+\delta_0}{2}, x\right]\right) \right), \end{aligned}$$

that is, (21) holds. Continuing, we deduce that (21) holds for all $x \in [a, b] \setminus \mathcal{F}_\alpha(I)$.

Now let $x \in \mathcal{F}_\alpha(I)$. Then there exists $\{x_n\} \subset [a, b] \setminus \mathcal{F}_\alpha(I)$ with $\{x_n\} \rightarrow x$. Then by (21) we have

$$m(\mathcal{F}_\alpha(I^0) \cap [a, x_n]) \geq \lambda \cdot m([a, x_n]) \quad \text{for all } n \geq 1.$$

The result follows by passing to the limit as $n \rightarrow \infty$. The proof is complete. \square

We are now ready to complete the proof of Lemma 2.

Proof of Lemma 2. Let us fix $\lambda \in (\frac{1}{2}, 1)$ and choose $\alpha, \theta < 1$ (close to 1) such that

$$\alpha \cdot \theta > \lambda \quad \text{and} \quad \alpha > 3\lambda(2+\lambda)^{-1}.$$

For any interval $I = [a, b]$, we set

$$I^+ := [a, a + \theta(b-a)] \quad \text{and} \quad I^- := [a + \theta(b-a), b]$$

and define the operators

- $\mathcal{T}(I) := \mathcal{F}_\alpha(I^+)$ (partial α -fat Cantor subset of I of measure $\alpha \cdot \theta \cdot m(I)$), and
- $\mathcal{N}(I) = \mathcal{F}_{\frac{1}{2}}(I^-)$ (partial $\frac{1}{2}$ -fat Cantor subset of I of measure $\frac{1}{2}(1 - \theta)m(I)$).

Notice that it is sufficient to construct $A_0 \subset [0, 1]$ (splitting the family of intervals of $(0, 1)$) satisfying (5) for $x \in [0, 1]$. Indeed, translating the construction by n , we obtain $A_n \subset [n, n + 1]$ and observe that the set $A = \cup_n A_n$ has the desired property.

To this end, set $I_0 = [0, 1]$ and consider an enumeration $\{I_n\}_{n \geq 1}$ of all strict subintervals $I_n := [p_n, q_n]$ of $[0, 1]$ with rational endpoints $p_n, q_n \in \mathbb{Q}$. Set further $T_0 := \mathcal{T}(I_0)$ and $N_0 := \mathcal{N}(I_0)$ and notice that $(T_0 \cup N_0)$ contains no intervals. Let $I_1 := (p_1, q_1)$ be the first interval of the above enumeration. Then there exists a subinterval $I'_1 = [p'_1, q'_1]$ of I_1 contained in $I_0 \setminus (T_0 \cup N_0)$. We set $T_1 = \mathcal{T}(I'_1)$ and $N_1 = \mathcal{N}(I'_1)$. Similarly, we find a subinterval $I'_2 \subset I_2$ contained in

$$I_0 \setminus [(T_0 \cup T_1) \cup (N_0 \cup N_1)]$$

and set $T_2 = \mathcal{T}(I'_2)$ and $N_2 = \mathcal{N}(I'_2)$ and continue by induction. Then set

$$N = \bigcup_{i=0}^{\infty} N_i \quad \text{and} \quad A_0 = I_0 \setminus N.$$

It is easily seen that A_0 splits the family of intervals of $I_0 = [0, 1]$. Let us show that (5) holds.

Let $t \in [0, 1] = I_0$. If $t \in I_0^+ = [0, \theta)$, then since $T_0 = \mathcal{F}_\alpha(I_0^+)$ we conclude by Lemma 7 that

$$m(A \cap [0, t]) \geq m(T_0 \cap [0, t]) \geq \lambda t.$$

If $t \in I_0^- = [\theta, 1]$, then

$$m(A \cap [0, t]) \geq m(T_0 \cap [0, t]) = m(T_0) = \alpha \cdot \theta > \lambda \geq \lambda t.$$

Therefore, A_0 satisfies (5) for all $t \in [0, 1]$, and, consequently, so does the set $A := \bigcup_{n=-\infty}^{+\infty} (A_0 + n)$ for all $t > 0$. \square

Acknowledgment. A major part of this work was done during a research visit of the first author to the University of Washington (July 2019). This author thanks the host institute for its hospitality.

REFERENCES

- [1] P. ALBANO AND P. CANNARSA, *Singularities of semiconcave functions in Banach spaces*, in Stochastic Analysis, Control, Optimization and Applications, Systems Control Found. Appl., Birkhäuser Boston, Boston, MA, 1999, pp. 171–190.
- [2] R. ARON, V. GURARIY, AND J. SEOANE, *Lineability and spaceability of sets of functions on \mathbb{R}* , Proc. Amer. Math. Soc., 133 (2005), pp. 795–803.
- [3] M. BENAÏM, J. HOFBAUER, AND S. SORIN, *Stochastic approximations and differential inclusions*, SIAM J. Control Optim., 44 (2005), pp. 328–348, <https://doi.org/10.1137/S0363012904439301>.
- [4] M. BENAÏM, J. HOFBAUER, AND S. SORIN, *Stochastic approximations and differential inclusions. II. Applications*, Math. Oper. Res., 31 (2006), pp. 673–695.
- [5] L. BERNAL-GONZALEZ AND M. ORDONEZ-CABRERA, *Lineability criteria with applications*, J. Funct. Anal., 266 (2014), pp. 3997–4025.
- [6] L. BERNAL-GONZALEZ, D. PELLEGRINO, AND J. SEOANE-SEPULVEDA, *Linear subsets of nonlinear sets in topological vector spaces*, Bull. Amer. Math. Soc., 51 (2014), pp. 71–130.
- [7] J. BOLTE, A. DANILIDIS, A. LEWIS, AND M. SHIOTA, *Clarke subgradients of stratifiable functions*, SIAM J. Optim., 18 (2007), pp. 556–572, <https://doi.org/10.1137/060670080>.

- [8] V. S. BORKAR, *Stochastic Approximation. A Dynamical Systems Viewpoint*, Cambridge University Press, Cambridge, Hindustan Book Agency, New Delhi, 2008.
- [9] J. BORWEIN, W. MOORS, AND X. WANG, *Lipschitz functions with prescribed derivatives and subderivatives*, *Nonlinear Anal.*, 29 (1997), pp. 53–63.
- [10] J. BORWEIN AND X. WANG, *Lipschitz functions with maximal subdifferentials are generic*, *Proc. Amer. Math. Soc.*, 128 (2000), pp. 3221–3229.
- [11] H. BRÉZIS, *Opérateurs maximaux monotones et semi-groupes de contraction dans des espaces de Hilbert*, North-Holland Math. Stud. 5, North-Holland, Amsterdam, 1973.
- [12] R. E. BRUCK, *Asymptotic convergence of nonlinear contraction semigroups in Hilbert space*, *J. Funct. Anal.*, 18 (1975), pp. 15–26.
- [13] J. V. BURKE, A. S. LEWIS, AND M. L. OVERTON, *A robust gradient sampling algorithm for nonsmooth, nonconvex optimization*, *SIAM J. Optim.*, 15 (2005), pp. 751–779, <https://doi.org/10.1137/030601296>.
- [14] F. CLARKE, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, 1991.
- [15] M. CSÖRNYEI, D. PREISS, AND J. TIŠER, *Lipschitz functions with unexpectedly large sets of nondifferentiability points*, *Abstr. Appl. Anal.*, 2005 (2005), pp. 361–373.
- [16] A. DANIILIDIS AND G. FLORES, *Linear structure of functions with maximal Clarke subdifferential*, *SIAM J. Optim.*, 29 (2019), pp. 511–521, <https://doi.org/10.1137/18M1211398>.
- [17] D. DAVIS, D. DRUSVYATSKIY, S. KAKADE, AND J. D. LEE, *Stochastic subgradient method converges on tame functions*, *Found. Comput. Math.*, 20 (2020), pp. 119–154.
- [18] J. C. DUCHI AND F. RUAN, *Stochastic methods for composite and weakly convex optimization problems*, *SIAM J. Optim.*, 28 (2018), pp. 3229–3259, <https://doi.org/10.1137/17M1135086>.
- [19] M. DYMOND AND V. KALUŽA, *Highly Irregular Separated Nets*, preprint, <https://arxiv.org/abs/1903.05923>, 2019.
- [20] I. EKELAND AND G. LEBOURG, *Generic differentiability of Lipschitzian functions*, *Trans. Amer. Math. Soc.*, 256 (1979), pp. 125–144.
- [21] P. H. ENFLO, V. GURARIY, AND J. SEOANE-SEPÚLVEDA, *Some results and open questions on spaceability in function spaces*, *Trans. Amer. Math. Soc.*, 366 (2014), pp. 611–625.
- [22] YU. M. ERMOLIEV AND V. I. NORKIN, *Solution of nonconvex nonsmooth stochastic optimization problems*, *Cybernet. Systems Anal.*, 39 (2003), pp. 701–715.
- [23] YU. M. ERMOLIEV, V. I. NORKIN, AND R. J.-B. WETS, *The minimization of semicontinuous functions: Mollifier subgradients*, *SIAM J. Control Optim.*, 33 (1995), pp. 149–167, <https://doi.org/10.1137/S0363012992238369>.
- [24] V. GURARIY, *Subspaces and bases in spaces of continuous functions*, *Dokl. Akad. Nauk SSSR*, 167 (1966), pp. 971–973 (in Russian).
- [25] A. D. IOFFE, *An invitation to tame optimization*, *SIAM J. Optim.*, 19 (2009), pp. 1894–1917, <https://doi.org/10.1137/080722059>.
- [26] R. KIRK, *Sets which split families of measurable sets*, *Amer. Math. Monthly*, 79 (1972), pp. 884–886.
- [27] L. LJUNG, *Analysis of recursive stochastic algorithms*, *IEEE Trans. Automat. Control*, 22 (1977), pp. 551–575.
- [28] S. MAJEWSKI, B. MIAŠOJEDOW, AND E. MOULINES, *Analysis of Nonsmooth Stochastic Approximation: The Differential Inclusion Approach*, preprint, <https://arxiv.org/abs/1805.01916>, 2018.
- [29] R. MIFFLIN, *Semismooth and semiconvex functions in constrained optimization*, *SIAM J. Control Optim.*, 15 (1977), pp. 959–972, <https://doi.org/10.1137/0315061>.
- [30] Y. NESTEROV AND V. G. SPOKOINY, *Random gradient-free minimization of convex functions*, *Found. Comput. Math.*, 17 (2017), pp. 527–566.
- [31] E. A. NURMINSKII, *The quasigradient method for the solving of the nonlinear programming problems*, *Cybernetics*, 9 (1973), pp. 145–150.
- [32] V. I. NORKIN, *Nonlocal minimization algorithms of nondifferentiable functions*, *Cybernet. Systems Anal.*, 14 (1978), pp. 704–707.
- [33] R. A. POLIQUIN AND R. T. ROCKAFELLAR, *Amenable functions in optimization*, in *Nonsmooth Optimization: Methods and Applications* (Erice, 1991), Gordon and Breach, Montreux, 1992, pp. 338–353.
- [34] R. A. POLIQUIN AND R. T. ROCKAFELLAR, *Prox-regular functions in variational analysis*, *Trans. Amer. Math. Soc.*, 348 (1996), pp. 1805–1838.
- [35] R. T. ROCKAFELLAR, *Favorable Classes of Lipschitz Continuous Functions in Subgradient Optimization*, IIASA Working Paper WP-81-001, IIASA, Laxenburg, Austria, 1981.
- [36] R. T. ROCKAFELLAR, *Clarke’s tangent cones and the boundaries of closed sets in \mathbf{R}^n* , *Nonlinear Anal.*, 3 (1979), pp. 145–154.