TRAJECTORIES OF DESCENT

D. DRUSVYATSKIY*, A.D. IOFFE†, AND A.S. LEWIS‡

Abstract. Steepest descent drives both theory and practice of nonsmooth optimization. We study slight relaxations of two influential notions of steepest descent curves — curves of maximal slope and solutions to evolution equations. In particular, we provide a simple proof showing that lower-semicontinuous functions that are locally Lipschitz continuous on their domains — functions playing a central role in nonsmooth optimization — admit Lipschitz continuous steepest descent curves in both senses.

Key words. Descent, slope, subdifferential, evolution equation, semi-algebraic

AMS subject classifications. Primary, 26D10; Secondary, 32B20, 49J52, 37B35, 37B35, 14P15

1. Introduction.

The intuitive notion of *steepest descent* plays a central role in theory and practice. So what are steepest descent curves in an entirely nonsmooth setting? In addressing this question, it is worthwhile to consider the classical setting: steepest descent curves for smooth functions $f\colon \mathbf{R}^n \to \mathbf{R}$ are simply smooth curves $x\colon [0,T) \to \mathbf{R}^n$ satisfying the differential equation

$$\dot{x}(t) = -\nabla f(x(t)), \qquad \text{for each } t \in [0,T). \tag{1.1}$$

Evidently this gradient dynamical system is equivalent to the pair of scalar equations

$$\|\dot{x}(t)\| = \|\nabla f(x(t))\|, \quad \frac{d(f\circ x)}{dt}(t) = -\big(\|\nabla f(x(t))\|\big)^2, \quad \text{for each } t \in [0,T). \tag{1.2}$$

A number of authors have taken up the task of rigorously modelling steepest descent in a nonsmooth setting, yielding two influential ideas based on generalizations of (1.1) and (1.2). Namely, generalizing the former we may instead consider absolutely continuous curves satisfying the *evolution equation*

$$\dot{x}(t) \in -\partial f(x(t)), \qquad \text{for a.e. } t \in [0,T),$$

where ∂f is some generalized derivative (or subdifferential). See for example [8, 22]. Alternatively, interpreting the quantity $\|\nabla f(x(t))\|$ in (1.2) as the "slope" of f at $x(t)$ leads to the idea of *curves of maximal slope*. For more details see [22, 12, 1, 13, 14, 22]. These two approaches have yielded major impact on optimization, PDEs, probability theory, and optimal transport. For a recent expository monograph, see [2]. In the current work, we study evolution equations and *curves of near-maximal slope* — a slight relaxation of curves of maximal slope that is well-suited for nonsmooth optimization.

---

*School of Operations Research and Information Engineering, Cornell University, Ithaca, New York, USA; http://people.orie.cornell.edu/dd379/. Work of Dmitriy Drusvyatskiy on this paper has been partially supported by the NDSEG grant from the Department of Defense and the US-Israel Binational Science Foundation Travel Grant for Young Scientists.

† Department of Mathematics, Technion-Israel Institute of Technology, Haifa, Israel 32000; http://www.math.technion.ac.il/Site/people/process.php?id=672. Work of A. D. Ioffe was supported in part by the US-Israel Binational Science Foundation Grant 2008261.

‡School of Operations Research and Information Engineering, Cornell University, Ithaca, New York, USA; http://people.orie.cornell.edu/aslewis/. Research supported in part by National Science Foundation Grant DMS-0806057 and by the US-Israel Binational Scientific Foundation Grant 2008261.

The question concerning existence of such curves is at the core of the subject. Roughly speaking, there are two strategies for constructing steepest descent curves for a function f on $\mathbf{R}^n$. The first one revolves around minimizing f on an increasing sequence of balls around a point x_k until the radius hits a certain threshold, at which point one moves the center to the next iterate and repeats the procedure. Passing to the limit as the thresholds tend to zero, under suitable conditions, yields a curve of maximal slope. The second approach is based on De Georgi's generalized movements [12]. Namely, one builds a piecewise constant curve by declaring the next iterate to be a minimizer of the function f plus a scaling of the squared distance from the previous iterate [2, Chapter 2]. The analysis, in both cases, is highly nontrivial and moreover does not give an intuitive meaning to the parametrization of the curve.

In the current work, we propose an alternate transparent strategy for constructing steepest descent curves. Our central object of study is the so-called *proximal descent curve* of a function f. It is obtained by discretizing the range of f and then building a piecewise linear curve by projecting iterates onto successive lower-level sets. Under reasonable condition, taking the limit as the mesh of the partition tends to zero, the resulting curve is a Lipschitz continuous steepest descent curve in both senses! Furthermore, the parametrization of the curve is entirely intuitive: the values of the function parametrize the curve. In particular, our results are strong enough to deduce existence of steepest descent curves for lower-semicontinuous functions that are Lipschitz continuous on their domains — functions of utmost importance in nonsmooth optimization. From a technical viewpoint, using the function values to parametrize the curve allows for the deep theory of metric regularity to enter the picture [17, 25], thereby yielding a simple and elegant existence proof.

The question concerning when solutions of evolution equations and curves of maximal slope are one and the same has been studied as well. However a major standing assumption that has so far been needed to establish positive answers in this direction is that the *slope* of the function f is itself a lower-semicontinuous function [2, 22] — an assumption that many common functions of nonsmooth optimization (e.g. $f(x) = \min\{x, 0\}$) do not satisfy. In the current work, we study this question in absence of such a continuity condition. As a result, *semi-algebraic functions* — those functions whose epigraph can be written as a finite union of sets, each defined by finitely many polynomial inequalities [11, 29] — come to the fore. For semi-algebraic functions that are locally Lipschitz continuous on their domains, solutions to evolution equations are one and the same as curves of near-maximal slope. Going a step further, using an argument based on the Kurdyka-Łojasiewicz inequality, in the spirit of [19, 21, 6, 5], we show that bounded curves of near-maximal slope for semi-algebraic functions necessarily have finite length. Consequently, such curves defined on maximal domains must converge to a generalized critical point of f.

In the paper we have restricted ourselves to function on $\mathbf{R}^n$, although in a number of key publications [22, 15, 2, 14] on the subject the emphasis is on infinite dimensional situations and applications to calculus of variations. Our choice has been mainly motivated by the desire to make the basic ideas and the techniques as clear as possible and not to obscure them by additional technicalities. We hope to consider infinite dimensional extensions of our approach and results elsewhere. Here we just want to emphasize that for many of them, this is an easy task at least for the case of a separable Hilbert (or even separable Asplund) space, provided the function f is *coercive*, in the sense that its sublevel sets are closed and bounded (hence weak compact).

The outline of the manuscript is as follows. Section 2 is a self-contained treatment of the basics of variational analysis that we will use. In this section, we emphasize the interplay between slopes and subdifferentials, and that the slope provides a very precise way of quantifying error bounds (Lemma 2.12). In Section 3 we define our main objects of interest — proximal descent curves — and analyze their properties and their relationship to curves of maximal slope and evolution equations. In Section 4, we strengthen the existence theory from the previous section by means of a simple viability Lemma (Lemma 4.1). In Section 5, we comment on some of the key assumptions needed to make our existence theory work and on when the two aforementioned notions of steepest descent coincide. This naturally leads to Section 6, where we consider additional properties of steepest descent curves for semi-algebraic functions.

2. Preliminaries. In this section, we summarize some of the fundamental tools used in variational analysis and nonsmooth optimization. We refer the reader to the monographs of Borwein-Zhu [7], Clarke-Ledyaev-Stern-Wolenski [9] Mordukhovich [23], and Rockafellar-Wets [26], and to the survey of Ioffe [17], for more details. Unless otherwise stated, we follow the terminology and notation of [17] and [26].

2.1. Variational analysis. Consider the extended real line $\overline{\mathbf{R}} := \mathbf{R} \cup \{-\infty\} \cup \{+\infty\}$. We say that an extended-real-valued function is proper if it is never $\{-\infty\}$ and is not always $\{+\infty\}$. For a function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$, we define the *domain* of f to be

$$\operatorname{dom} f := \{x \in \mathbf{R}^n : f(x) < +\infty\},$$

and we define the *epigraph* of f to be

$$\operatorname{epi} f := \{(x, r) \in \mathbf{R}^n \times \mathbf{R} : r \geq f(x)\}.$$

Throughout this work, we will only use Euclidean norms. Hence for a point $x \in \mathbf{R}^n$, the symbol $\|x\|$ will denote the standard Euclidean norm of x. A function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is *lower-semicontinuous* (or *lsc* for short) at $\bar{x}$ if the inequality $\liminf_{x \to \bar{x}} f(x) \geq f(\bar{x})$ holds. We will say that f is *locally Lipschitz continuous* at a point $\bar{x} \in \mathbf{R}^n$ *relative* to a set Q, if for some $\kappa \geq 0$ the inequality

$$|f(x) - f(y)| \leq \kappa \|x - y\| \quad \text{holds for all } x, y \in Q \text{ near } \bar{x}.$$

The infimum of such constants κ is the *Lipschitz modulus* of f at $\bar{x}$ *relative* to Q, and we will refer to it by $\operatorname{lip} f\big(\bar{x}; Q\big)$. If the set Q is not explicitly mentioned, then the reader should assume that Q is simply $\mathbf{R}^n$. Henceforth, the symbol $o(\|x - \bar{x}\|)$ will denote a term with the property

$$\frac{o(\|x - \bar{x}\|)}{\|x - \bar{x}\|} \to 0, \quad \text{when } x \to \bar{x} \text{ with } x \neq \bar{x}.$$

The symbols $\operatorname{cl} Q$, $\operatorname{conv} Q$, $\operatorname{cone} Q$, $\operatorname{aff} Q$, and $\operatorname{par} Q$ will denote the topological closure, the convex hull, the (non-convex) conical hull, the affine span, and the parallel subspace of Q respectively. An open ball of radius ϵ around a point $\bar{x}$ will be denoted by $B_\epsilon(\bar{x})$, while the open unit ball will be denoted by $\mathbf{B}$. A primary variational-analytic method for studying nonsmooth functions on $\mathbf{R}^n$ is by means of subdifferentials.

Definition 2.1 (Subdifferentials). Consider a function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a point $\bar{x}$ with $f(\bar{x})$ finite.

1. The *Fréchet subdifferential* of f at $\bar{x}$, denoted $\hat{\partial} f(\bar{x})$, consists of all vectors $v \in \mathbf{R}^n$ satisfying
$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(\|x - \bar{x}\|).$$
2. The *limiting subdifferential* of f at $\bar{x}$, denoted $\partial f(\bar{x})$, consists of all vectors $v \in \mathbf{R}^n$ for which there exist sequences $x_i \in \mathbf{R}^n$ and $v_i \in \hat{\partial} f(x_i)$ with $(x_i, f(x_i), v_i) \to (\bar{x}, f(\bar{x}), v)$.
3. The *horizon subdifferential* of f at $\bar{x}$, denoted $\partial^\infty f(\bar{x})$, consists of all vectors $v \in \mathbf{R}^n$ for which there exists a sequence of real numbers $\tau_i \downarrow 0$ and a sequence of points $x_i \in \mathbf{R}^n$, along with subgradients $v_i \in \partial f(x_i)$, satisfying $(x_i, f(x_i), \tau_i v_i) \to (\bar{x}, f(\bar{x}), v)$.
4. The *Clarke subdifferential* of f at $\bar{x}$, denoted $\partial_c f(\bar{x})$, is obtained by the convexification
$$\partial_c f(\bar{x}) := \operatorname{cl} \operatorname{co} [\partial f(\bar{x}) + \partial^\infty f(\bar{x})].$$

We say that f is *subdifferentiable* at $\bar{x}$ whenever $\partial f(\bar{x})$ is nonempty (equivalently when $\partial_c f(\bar{x})$ is nonempty).

In particular, every locally Lipschitz continuous function is subdifferentiable. For x such that $f(x)$ is not finite, we follow the convention that $\hat{\partial} f(x) = \partial f(x) = \partial^\infty f(x) = \partial_c f(\bar{x}) = \emptyset$.

The subdifferentials $\hat{\partial} f(\bar{x})$, $\partial f(\bar{x})$, and $\partial_c f(\bar{x})$ generalize the classical notion of gradient. In particular, for $\mathbf{C}^1$-smooth functions f on $\mathbf{R}^n$, these three subdifferentials consist only of the gradient $\nabla f(x)$ for each $x \in \mathbf{R}^n$. For convex f, these subdifferentials coincide with the convex subdifferential. The horizon subdifferential $\partial^\infty f(\bar{x})$ plays an entirely different role; namely, it detects horizontal "normals" to the epigraph. In particular, a lsc function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is locally Lipschitz continuous around $\bar{x}$ if and only if we have $\partial^\infty f(\bar{x}) = \{0\}$. Moreover, if f is locally Lipschitz continuous at $\bar{x}$, then we have equality
$$\operatorname{lip} f(\bar{x}) = \max_{v \in \partial f(\bar{x})} \|v\|.$$

For a set $Q \subset \mathbf{R}^n$, we define the *indicator function* of Q, denoted δ_Q, to be zero on Q and plus infinity elsewhere. The geometric counterparts of subdifferentials are normal cones.

Definition 2.2 (Normal cones). Consider a set $Q \subset \mathbf{R}^n$. Then the *Fréchet*, *limiting*, and *Clarke normal cones* to Q at any point $\bar{x} \in \mathbf{R}^n$ are defined by $\hat{N}_Q(\bar{x}) := \hat{\partial}\delta(\bar{x})$, $N_Q(\bar{x}) := \partial\delta(\bar{x})$, and $N_Q^c(\bar{x}) := \partial_c \delta(\bar{x})$ respectively.

A particularly nice situation occurs when all the normal cones coincide.

Definition 2.3 (Clarke regularity of sets). A set $Q \subset \mathbf{R}^n$ is said to be *Clarke regular* at a point $\bar{x} \in Q$ if it is locally closed at $\bar{x}$ and every limiting normal vector to Q at $\bar{x}$ is a Fréchet normal vector, that is the equation $N_Q(\bar{x}) = \hat{N}_Q(\bar{x})$ holds.

The functional version of Clarke regularity is as follows.

Definition 2.4 (Subdifferential regularity). A function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is called *subdifferentially regular* at $\bar{x}$ if $f(\bar{x})$ is finite and $\operatorname{epi} f$ is Clarke regular at $(\bar{x}, f(\bar{x}))$ as a subset of $\mathbf{R}^n \times \mathbf{R}$.

In particular, if $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is subdifferentially regular at a point $\bar{x} \in \operatorname{dom} f$, then equality $\hat{\partial} f(\bar{x}) = \partial f(\bar{x})$ holds ([26, Corollary 8.11]). Shortly, we will need the following result describing normals to lower-level sets [26, Proposition 10.3]. We provide an independent proof for completeness and ease of reference in future work. The reader may safely skip it upon first reading.

PROPOSITION 2.5 (Normals to lower-level sets). *Consider a lsc function* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *and a point* $\bar{x} \in \mathbf{R}^n$ *with* $0 \notin \partial f(\bar{x})$*. Then the inclusion*

$$N_{[f \leq f(\bar{x})]}(\bar{x}) \subset (\text{cone}\, \partial f(\bar{x})) \cup \partial^\infty f(\bar{x}) \quad \textit{holds.}$$

Proof. Define the real number $\bar{\alpha} := f(\bar{x})$ and the sets $\mathcal{L}_{\bar{\alpha}} := \{(x, \alpha) :\ \alpha \leq \bar{\alpha}\}$ and

$$Q_{\bar{\alpha}} := (\text{epi}\, f) \bigcap \mathcal{L}_{\bar{\alpha}} = \{(x, \alpha) :\ f(x) \leq \alpha \leq \bar{\alpha}\}.$$

We first show the implication

$$(x^*, 0) \in N_{Q_{\bar{\alpha}}}(\bar{x}, \bar{\alpha}) \implies x^* \in (\text{cone}\, \partial f(\bar{x})) \bigcup \partial^\infty f(\bar{x}). \tag{2.1}$$

Indeed, consider a vector $(x^*, 0)$. Then Fuzzy calculus [17, Chapter 2.1] implies that there are sequences $(x_{1k}, \alpha_{1k}) \in \text{epi}\, f$, $(x^*_{1k}, \beta_{1k}) \in \hat{N}_{\text{epi}\, f}(x_{1k}, \alpha_{1k})$, $(x_{2k}, \alpha_{2k}) \in \mathcal{L}_{\bar{\alpha}}$ and $(x^*_{2k}, \beta_{2k}) \in \hat{N}_{\mathcal{L}_{\bar{\alpha}}}(x_{2k}, \alpha_{2k})$ satisfying

$$(x_{1k}, \alpha_{1k}) \to (\bar{x}, \bar{\alpha}), \quad (x_{2k}, \alpha_{2k}) \to (\bar{x}, \bar{\alpha}), \quad x^*_{1k} + x^*_{2k} \to x^*, \quad \beta_{1k} + \beta_{2k} \to 0.$$

Observe $x^*_{2k} = 0$ and hence $x^*_{1k} \to x^*$. Furthermore, by nature of epigraphs we have $\beta_{1k} \leq 0$. If up to a subsequence we had $\beta_{1k} = 0$, then (2.1) would follow immediately. Consequently, we may suppose that the inequality $\beta_{1k} < 0$ is valid. Then we have $|\beta_{1k}|^{-1} x^*_{1k} \in \hat{\partial} f(x_{1k})$. Since the norms of x^*_{1k} are uniformly bounded and we have $0 \notin \partial f(\bar{x})$, the sequence β_{1k} must be bounded. Consequently we may assume that β_{1k} converges to some β and (2.1) follows.

Now consider a vector $u^* \in \hat{N}_{[f \leq \bar{\alpha}]}(u)$ for some $u \in [f \leq \bar{\alpha}]$. Consequently the inequality $\langle u^*, h\rangle \leq o(\|h\|)$ holds, whenever h satisfies $f(u + h) \leq \bar{\alpha}$. The latter in turn implies $(u^*, 0) \in \hat{N}_{Q_{\bar{\alpha}}}(u, \bar{\alpha})$. Together with (2.1), taking limits of Fréchet subgradients and applying equation (2.1) completes the proof. ◻

REMARK 2.6. Theorem 2.5 and its proof easily extend to the case when f is defined on certain infinite dimensional spaces (e.g. Hilbert Spaces).

We now record the very useful generalization of the classical Mean Value Theorem to an entirely nonsmooth setting. See for example [9, Theorem 2.4].

THEOREM 2.7 (Lebourg's Mean Value Theorem). *Consider a function* $f\colon \mathbf{R}^n \to \mathbf{R}$ *that is Lipschitz continuous on an open set containing the line segment* (x, y)*. Then there exists a point* u *in* (x, y) *satisfying*

$$f(y) - f(x) \in \langle \partial_c f(u), y - x\rangle.$$

For a set $Q \subset \mathbf{R}^n$ and a point $x \in \mathbf{R}^n$, the *distance* of x from Q is

$$d(x, Q) := \inf_{y \in Q} \|x - y\|,$$

and the *metric projection* of x onto Q is

$$P_Q(x) := \{y \in Q : \|x - y\| = d(x, Q)\}.$$

A fundamental notion in variational analysis is that of *slope*. For more details about slope and its relevance to the theory of metric regularity, see [17].

Definition 2.8 (Slope). Consider a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, and a point $\bar{x} \in \mathbf{R}^n$ with $f(\bar{x})$ finite. The *slope* of f at $\bar{x}$ is

$$|\nabla f|(\bar{x}) := \limsup_{\substack{x \to \bar{x} \\ x \neq \bar{x}}} \frac{(f(\bar{x}) - f(x))^+}{\|\bar{x} - x\|}.$$

The *limiting slope* is

$$\overline{|\nabla f|}(\bar{x}) := \liminf_{x \xrightarrow[f]{} \bar{x}} |\nabla f|(x),$$

where the convergence $x \xrightarrow[f]{} \bar{x}$ means $(x, f(x)) \to (\bar{x}, f(\bar{x}))$.

For $\mathbf{C}^1$-smooth functions f on $\mathbf{R}^n$, the equation $\overline{|\nabla f|}(\bar{x}) = |\nabla f|(\bar{x}) = \|\nabla f(\bar{x})\|$ holds. The following result, which follows from the proofs of [17, Propositions 1 and 2, Chapter 3], establishes an elegant relationship between the slope and subdifferentials. We outline a proof for completeness.

Proposition 2.9 (Slope and subdifferentials). *Consider a lsc function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, and a point $\bar{x} \in \mathbf{R}^n$ with $f(\bar{x})$ finite. Then we have $|\nabla f|(\bar{x}) \leq d(0, \hat{\partial} f(\bar{x}))$, and furthermore the equality*

$$\overline{|\nabla f|}(\bar{x}) = d(0, \partial f(\bar{x})), \qquad \textit{holds.}$$

In particular, the two conditions $\overline{|\nabla f|}(\bar{x}) = 0$ and $0 \in \partial f(\bar{x})$ are equivalent.

Proof. The inequality $|\nabla f|(\bar{x}) \leq d(0, \hat{\partial} f(\bar{x}))$ is immediate from the definition of the Fréchet subdifferential. Now define $m = \overline{|\nabla f|}(\bar{x})$. One may easily check that if m is infinite, then the subdifferential $\partial f(\bar{x})$ is empty, and therefore the result holds trivially. Consequently we may suppose that m is finite.

Fix an arbitrary $\epsilon > 0$, and let x be a point satisfying

$$\|x - \bar{x}\| < \epsilon, \quad |f(x) - f(\bar{x})| < \epsilon, \quad \text{and} \quad |\nabla f|(x) < m + \epsilon.$$

Define the function $g(u) := f(u) + (m + \epsilon)\|u - x\|$. Observe that for all u sufficiently close to x, we have $g(u) \geq f(x)$. We deduce (see e.g. [26, Exercise 10.10])

$$0 \in \partial g(x) \subset \partial f(x) + (m + \epsilon)\mathbf{B}.$$

Hence we obtain the inequality $m + \epsilon \geq d(0, \partial f(x))$. Letting ϵ tend to zero, we deduce $m \geq d(0, \partial f(\bar{x}))$.

To see the reverse inequality, consider a vector $\bar{v} \in \partial f(\bar{x})$ achieving $d(0, \partial f(\bar{x}))$. Then there exist sequences of points x_i and vectors $v_i \in \hat{\partial} f(x_i)$ with $(x_i, f(x_i), v_i) \to (\bar{x}, f(\bar{x}), \bar{v})$. Observe that for each index i, we have $\|v_i\| \geq |\nabla f|(x_i)$. Letting i tend to infinity, the result follows. □

Thus by Proposition 2.9, for any lsc function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and any point $\bar{x} \in \mathbf{R}^n$, we have

$$d(0, \partial f(\bar{x})) = \overline{|\nabla f|}(\bar{x}) \leq |\nabla f|(\bar{x}) \leq d(0, \hat{\partial} f(\bar{x})).$$

In particular, if f is subdifferentially regular at $\bar{x}$, then the slope and the limiting slope are one and the same, that is the equation $\overline{|\nabla f|}(\bar{x}) = |\nabla f|(\bar{x})$ holds.

In our work, it will be useful to extend functions that are Lipschitz continuous on their domains to ones that that are Lipschitz continuous on the whole space. This can be done in a standard way. However, we will need more precise properties of such an extension, which we record below.

LEMMA 2.10 (Inclusion of subdifferentials). *Consider a lsc function* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *that is Lipschitz continuous on its domain with modulus* l. *Choose a real number* $K > l$ *and define the function*

$$g(x) := \inf_{y \in \mathbf{R}^n} \{f(y) + K\|x - y\|\}. \tag{2.2}$$

Then g *is a* K*-Lipschitz continuous function, coinciding with* f *on* $\operatorname{dom} f$, *and having the property that for any point* $\bar{x} \in \operatorname{dom} f$ *the inclusion* $\partial g(\bar{x}) \subset \partial f(\bar{x})$ *holds. Furthermore, we have* $|\nabla g|(\bar{x}) = |\nabla f|(\bar{x})$ *and* $\overline{|\nabla g|}(\bar{x}) = \overline{|\nabla f|}(\bar{x})$ *for any point* $\bar{x} \in \operatorname{dom} f$.

Proof. It is well-known that g is a K-Lipschitz continuous function on $\mathbf{R}^n$ and that it coincides with f on $\operatorname{dom} f$. See for example [16] or [9, Problem 11.6]. Define the function $\phi\colon \mathbf{R}^n \times \mathbf{R}^n \to \overline{\mathbf{R}}$ by $\phi(x,y) = f(y) + K\|x - y\|$. Then g is simply the marginal function

$$g(x) = \min_{y \in \mathbf{R}^n} \phi(x,y).$$

It is not difficult to check that the solution mapping $x \mapsto \operatorname{argmin}_y \phi(x,\cdot)$ is locally bounded around $\bar{x}$ and is outer-semicontinuous (in fact, inner-semicontinuous as well). This verifies the assumptions of [20, Proposition 3.3]. Applying this proposition, we deduce that for any subgradient $v \in \partial g(\bar{x})$, we have

$$(v,0) \in \bigcup_{y \in \operatorname{argmin} \phi(\bar{x},\cdot)} \partial\phi(\bar{x},y) \subset \{0\} \times \partial f(\bar{x}) + K\{(w,-w) : w \in \mathbf{B}\},$$

where the latter inclusion follows from the observation $\{\bar{x}\} = \operatorname{argmin}_{y \in \mathbf{R}^n} \phi(\bar{x},y)$. We deduce existence of a vector $w \in \mathbf{B}$ with $v = Kw \in \partial f(\bar{x})$. Hence we have established the inclusion $\partial g(\bar{x}) \subset \partial f(\bar{x})$. This clearly implies $\overline{|\nabla g|}(\bar{x}) \geq \overline{|\nabla f|}(\bar{x})$. On the other hand, we claim that the equation $|\nabla f|(\bar{x}) = |\nabla g|(\bar{x})$ holds for any $\bar{x}$ in $\operatorname{dom} f$. Indeed, the inequality $|\nabla g|(\bar{x}) \geq |\nabla f|(\bar{x})$ is clear. On the other hand, consider a point $x \notin \operatorname{dom} f$ and a point $y \in \operatorname{dom} f$ with $g(x) = f(y) + K\|x - y\|$. Then we have

$$\frac{(g(\bar{x}) - g(y))^+}{\|\bar{x} - y\|} \geq \frac{(g(\bar{x}) - g(x) + K\|y - x\|)^+}{\|\bar{x} - x\| + \|y - x\|} \geq \frac{(g(\bar{x}) - g(x))^+}{\|\bar{x} - x\|},$$

where the last inequality follows since g is K-Lipschitz continuous. The equality $|\nabla g|(\bar{x}) = |\nabla f|(\bar{x})$ follows. Finally, we deduce

$$\overline{|\nabla g|}(\bar{x}) \leq \liminf_{x \to \bar{x}} \{|\nabla g|(x) : x \in \operatorname{dom} f\} = \overline{|\nabla f|}(\bar{x}) \leq \overline{|\nabla g|}(\bar{x}),$$

and hence we have equality throughout, thereby completing the proof. □

We record below the celebrated Ekeland's variational principle.

THEOREM 2.11 (Ekeland's variational principle). *Consider a lsc function* $g\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *that is bounded from below. Suppose that for some* $\epsilon > 0$ *and* $x \in \mathbf{R}^n$, *we have* $g(x) \leq \inf f + \epsilon$. *Then for any* $\rho > 0$, *there exists a point* $\bar{u}$ *satisfying* $g(\bar{u}) \leq g(x)$, $\|\bar{u} - x\| \leq \rho^{-1}\epsilon$, *and*

$$g(u) + \rho\|u - \bar{u}\| > g(\bar{u}), \qquad \text{for all } u \in \mathbf{R}^n \setminus \{\bar{u}\}.$$

The following consequence of Ekeland's variational principle will play a crucial role in our work [17, Basic Lemma, Chapter 1]. We provide a proof for completeness.

Lemma 2.12 (Error bound). *Consider a lsc function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$. Assume that for some point $x \in \operatorname{dom} f$, there are constants $\alpha < f(x)$ and $r, K > 0$ so that the implication*

$$\alpha < f(u) \le f(x) \text{ and } \|u - x\| \le K \quad \Longrightarrow \quad |\nabla f|(u) \ge r, \quad \text{holds.}$$

If in addition the inequality $f(x) - \alpha < Kr$ is valid, then the lower-level set $[f \le \alpha]$ is nonempty and we have the estimate $d(x, [f \le \alpha]) \le r^{-1}(f(x) - \alpha)$.

Proof. Define a lsc function $g\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ by setting $g(u) := (f(u)-\alpha)^+$, and choose a real number $\rho < r$ satisfying $f(x)-\alpha < K\rho$. By Ekeland's principle (Theorem 2.11), there exists a point $\bar{u}$ satisfying

$$g(\bar{u}) \le g(x), \quad \|\bar{u} - x\| \le \rho^{-1} g(x) \le K$$

and

$$g(u) + \rho\|u - \bar{u}\| \ge g(\bar{u}), \quad \text{for all } u.$$

Consequently we obtain the inequality $|\nabla g|(\bar{u}) \le \rho$. On the other hand, a simple computation shows that this can happen only provided $g(\bar{u}) = 0$, for otherwise we would have $|\nabla g|(\bar{u}) = |\nabla f|(\bar{u}) \ge r$. Hence $\bar{u}$ lies in the level set $[f \le \alpha]$, and we obtain the estimate $d(x, [f \le \alpha]) \le \rho^{-1}(f(x) - \alpha)$. The result now follows by taking ρ arbitrarily close to (and still smaller than) r. □

We conclude this subsection with the following standard result of Linear Algebra.

Lemma 2.13 (Result in Linear Algebra). *Suppose that $\mathbf{R}^n$ can be written as a direct sum $\mathbf{R}^n = V \oplus W$, for some vector subspaces V and W. Then for any vector $b \in \mathbf{R}^n$, the equations*

$$P_V(b) = (b + W) \cap V = \operatorname*{argmin}_{z \in b+W} \|z\|, \quad \textit{hold.}$$

Proof. Observe $b = P_V(b) + P_W(b)$, and consequently the inclusion

$$P_V(b) \in (b + W) \cap V \quad \text{holds.}$$

The reverse inclusion follows from the trivial computation

$$z \in (b + W) \cap V \Longrightarrow z - P_V(b) \in V \cap W \Longrightarrow z = P_V(b).$$

Now observe that first order optimality conditions imply that the unique minimizer $\bar{z}$ of the problem

$$\min_{z \in b+W} \|z\|^2,$$

is characterized by the inclusion $\bar{z} \in (b + W) \cap V$, and hence the result follows. □

2.2. Slope and gradient descent. What is steepest descent in absence of differentiability? As was discussed in the introduction, there are two influential approaches. One approach is immediate: replace the gradient in the differential equation $\dot{x}(t) = -\nabla f(x(t))$ by a subdifferential, thus obtaining a differential inclusion. In other words, we may consider absolutely continuous curves $x\colon [0,T) \to \mathbf{R}^n$ satisfying the *evolution equation*

$$\dot{x}(t) \in -\partial f(x(t)), \quad \text{for almost every } t \in [0,T).$$

Of course, one may replace the limiting subdifferential in the inclusion above by a subdifferential of another kind, though we will not dwell on this issue. For more on the theory of differential inclusions, and this approach in particular, see the classical references [3, 8].

The second approach can be motivated by an elementary calculation showing that $\mathbf{C}^1$-smooth solutions of the dynamical system $\dot{x}(t) = -\nabla f(x(t))$ can be characterized by the pair of equations

$$\|\dot{x}(t)\| = \|\nabla f(x(t))\| \quad \text{and} \quad \frac{d(f\circ x)}{dt}(t) = -\big(\|\nabla f(x(t))\|\big)^2, \quad \text{for each } t \in [0,T).$$

In light of this observation, the ensuing notion arises naturally. The definition we propose follows closely that of curves of maximal slope (discussed below), introduced and extensively studied in [15].

Definition 2.14 (Curve of near-maximal slope). A curve $x(t)$ defined on a segment $[0,T)$ (finite or infinite) is a *curve of near-maximal slope* if we have

(a) $x(t)$ is absolutely continuous;
(b) $\|\dot{x}(t)\| = \overline{|\nabla f|}(x(t))$ a.e. on $[0,T)$;
(c) $(f\circ x)(t) = f(x(t))$ is absolutely continuous, non-increasing, and such that

$$\frac{d(f\circ x)}{dt}(t) = -\big(\overline{|\nabla f|}(x(t))\big)^2 \text{ a.e. on } [0,T).$$

If the limiting slope $\overline{|\nabla f|}(x(t))$ in the definition above is replaced by the slope $|\nabla f|(x(t))$, then the resulting curve is said to be a *curve of maximal slope.* Clearly whenever f is subdifferentially regular, the two notions coincide.

3. Main results. In this section, we analyze when curves of near-maximal slope and solutions of evolution equations exist. In fact, we will construct a curve that (under reasonable conditions) is a steepest descent curve in both senses. To this end, we propose to study the following intuitive notion.

Definition 3.1 (Proximal descent curves). Consider a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, a point $\bar{x}$, and a real number $\eta > 0$. Let $0 = \tau_0 < \tau_1 < \ldots < \tau_k = \eta$ be a partition of $[0,\eta]$ into k equal parts. With this partition, we associate a piecewise linear curve $u_k(\tau)$, for $\tau \in [0,\eta]$, as follows. Set $u_k(0) = \bar{x}$, and inductively define $u_k(\tau_{i+1})$ to be any point belonging to the projection of $u_k(\tau_i)$ onto the lower level set $[f \le f(\bar{x}) - \tau_{i+1}]$, provided that this set is nonempty. Then we will call any limit point $x(\tau)$ of $u_k(\tau)$ in the uniform metric, as k tends to infinity, a *proximal descent curve* of f at $\bar{x}$.

See Figure 3.1 for an illustration of a proximal-descent curve. So when do proximal descent curves exist and what are their properties? We will see that in answering this question, the following condition appears naturally. At the risk of sounding extravagant, we give this property a name.

Definition 3.2 (Level prox-stability). A lsc function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is *level prox-stable* at $\bar{x}$ if there exists a neighborhood U of $\bar{x}$ so that for all sufficiently small $\eta > 0$ and for any $x \in U$ the implication

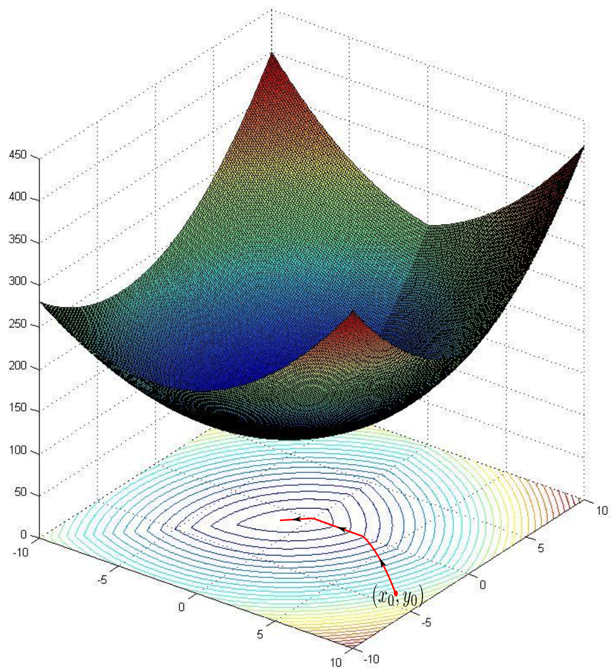

FIG. 3.1. $f(x,y) = \max\{x+y,\, |x-y|\} + x(x+1) + y(y+1) + 100$

$$x \in [f(\bar{x}) - \eta < f < \infty] \quad \text{and} \quad y \in P_{U\cap[f\le f(\bar{x})-\eta]}(x) \Longrightarrow f(y) = f(\bar{x}) - \eta,$$

holds.

Roughly speaking, a function f is level prox-stable, provided that in a local sense, projecting points in the domain of f, lying outside of the lower level sets $[f \le f(\bar{x}) - \eta]$, onto this set is the same as projecting onto the level sets $[f = f(\bar{x}) - \eta]$. In particular, one can easily check that *functions continuous on lines* — those functions f so that for any pair $x, y \in \operatorname{dom} f$, the restriction of f to the line segment joining x and y is finite and continuous — are necessarily level prox-stable. Hence all functions of the form $f = g + \phi$, where $\phi\colon \mathbf{R}^n \to \mathbf{R}$ is a continuous function and $g\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is a lsc, convex function, are level prox-stable [24, Theorem 10.2]. Even though, level prox-stability will be needed to state the main results of this section, we will eliminate this condition all-together in Section 4 using a simple viability argument.

In passing, one may wonder if continuity on the domain is sufficient to guarantee that level prox-stability holds. The following example illustrates the contrary.

EXAMPLE 3.3 (Failure of level prox-stability). Define the set $Q = \{(x,y) \in \mathbf{R}^2 : y \ge 2|x|\}$ and consider the function $f(x,y) = x + \delta_Q(x,y)$. It is easy to check that f is not level prox-stable at the origin.

We now arrive at the main result of this section.

THEOREM 3.4 (Existence and properties of proximal descent curves). *Consider a lsc function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a point $\bar{x} \in \mathbf{R}^n$, with f finite at $\bar{x}$ and $0 \notin \partial f(\bar{x})$. Then for all sufficiently small $\eta > 0$, the following are true.*

Existence: *There exists a proximal descent curve $x\colon [0,\eta] \to \mathbf{R}^n$ of f at $\bar{x}$.*

Regularity: *Any proximal descent curve* $x\colon [0,\eta] \to \mathbf{R}^n$ *of* f *at* $\bar{x}$ *is necessarily Lipschitz continuous and lies fully in the domain of* f*.*

Speed of descent: *Any proximal descent-curve* $x\colon [0,\eta] \to \mathbf{R}^n$ *of* f *at* $\bar{x}$ *satisfies the following properties.*

(a) $f(x(\tau)) \le f(\bar{x}) - \tau$*, for each* $\tau \in [0,\eta]$*.*

(b) $\frac{1}{\|\dot{x}(\tau)\|} \ge \overline{|\nabla f|}(x(\tau))$ *for a.e.* $\tau \in [0,\eta]$*.*

If f *is level prox-stable at* $\bar{x}$ *and is continuous on* $\operatorname{dom} f$*, then the above estimates can be strengthened:*

(a) $f(x(\tau)) = f(\bar{x}) - \tau$*, for each* $\tau \in [0,\eta]$*. In particular* $f \circ x$ *is Lipschitz.*

(b) $\operatorname{lip} f\big(x(\tau); \operatorname{dom} f\big) \ge \frac{1}{\|\dot{x}(\tau)\|} \ge \overline{|\nabla f|}(x(\tau))$ *for a.e.* $\tau \in [0,\eta]$*.*

Differential inclusion: *Assume that* f *is level prox-stable at* $\bar{x}$ *and is continuous on* $\operatorname{dom} f$*. Then for any proximal descent curve* $x\colon [0,\eta] \to \mathbf{R}^n$ *of* f *at* $\bar{x}$*, the inclusion*

$$\dot{x}(\tau) \in -\operatorname{cl}\operatorname{conv}\Big[\big(\operatorname{cone} \partial f(x(\tau))\big) \cup \partial^{\infty} f(x(\tau))\Big], \qquad \textit{holds for a.e. } \tau \in [0,\eta].$$

If f *is subdifferentiable along the curve* $x(\tau)$*, then the inclusion*

$$\dot{x}(\tau) \in -\operatorname{cl}\operatorname{cone} \partial_c f(x(\tau)), \qquad \textit{holds for a.e. } \tau \in [0,\eta].$$

If f *is Lipschitz continuous on its domain, then the closure is not needed in the inclusion above.*

Proof. Since zero is not a subgradient of f at $\bar{x}$, we can find constants $\eta > 0$, $r > 0$ and $C > 0$ such that all conditions of Lemma 2.12 are satisfied for $\bar{x}$, $\alpha = f(\bar{x}) - \eta$ and $K = C$. In particular, shrinking η we may enforce the inequality $\eta < rC$. Let $0 = \tau_0 < \tau_1 < \ldots < \tau_k = \eta$ be a partition of $[0,\eta]$ into k equal parts, and let $\lambda := \frac{\tau_{i+1} - \tau_i}{\eta}$ be the partition size. Finally define the piecewise linear functions $u_k(\tau)$ as in the definition of proximal descent curves. For notational convenience, in the proof we will suppress the index k in $u_k(\tau)$.

We now show that all the points $u(\tau_i)$ (for $i = 0, \ldots, k$) are well-defined, and in terms of the quantities

$$r_i := \inf\{|\nabla f|(u) : \ f(\bar{x}) - \tau_{i+1} < f(u) \le f(u(\tau_i)),\ \|u - u(\tau_i)\| < \lambda C\},$$

we have

$$\|u(\tau_{i+1}) - u(\tau_i)\| \le r_i^{-1}(\tau_{i+1} - \tau_i), \tag{3.1}$$

and

$$r_{i+1} \ge r, \quad \|u(\tau_{i+1}) - \bar{x}\| \le r^{-1}\tau_{i+1}. \tag{3.2}$$

To this end, suppose that equations (3.1) and (3.2) are valid for indices $i = 0, \ldots, j-1$. Observe if we have $f(u(\tau_j)) \le f(\bar{x}) - \tau_{j+1}$, then equality $u(\tau_j) = u(\tau_{j+1})$ holds and the inductive step is true trivially. Hence suppose otherwise. We claim that the conditions of Lemma 2.12 are satisfied with $x = u(\tau_j)$, $\alpha = f(\bar{x}) - \tau_{j+1}$, $K = \lambda C$, and with r_j in place of r.

To this end, we deduce

- $f(u(\tau_j)) - (f(\bar{x}) - \tau_{j+1}) \le \lambda r_j C$;
- $f(\bar{x}) - \tau_{j+1} < f(u) \le f(u(\tau_j))$ and $d(u, u(\tau_j)) \le \lambda C \quad \Longrightarrow \quad |\nabla f|(u) \ge r_j$.

Observe

$$f(u(\tau_j)) - (f(\bar{x}) - \tau_{j+1}) = \tau_{j+1} - \tau_j \le (\tau_{j+1} - \tau_j)\frac{rC}{\eta} = \lambda rC \le \lambda r_j C,$$

which is the first of the desired relations. The second relation follows immediately from the definition of r_j.

Applying Lemma 2.12, we conclude that the lower-level set $[f \leq f(\bar{x}) - \tau_{j+1}]$ is nonempty, and that the inequality $\|u(\tau_{j+1}) - u(\tau_j)\| \leq r_j^{-1}(\tau_{j+1} - \tau_j)$ holds. Consequently, we obtain

$$\|u(\tau_{j+1}) - \bar{x}\| \leq \|u(\tau_{j+1}) - u(\tau_j)\| + \|u(\tau_j) - \bar{x}\| \leq r_j^{-1}(\tau_{j+1} - \tau_j) + r^{-1}\tau_j \leq r^{-1}\tau_{j+1}.$$

Finally we claim that the inequality $r_{j+1} \geq r$ holds. To see this, consider a point u satisfying $f(\bar{x}) - \tau_{j+2} < f(u) \leq f(u(\tau_{j+1}))$ and $\|u - u(\tau_{j+1})\| < \lambda C$. Taking (3.2) into account, along with the inequality $r^{-1} \leq C/\eta$, we obtain

$$\|u - \bar{x}\| \leq \|u - u(\tau_{j+1})\| + \|u(\tau_{j+1}) - \bar{x}\| \leq \frac{\tau_{j+2} - \tau_{j+1}}{\eta}C + \frac{\tau_{j+1}}{r} = \frac{\tau_{j+2}}{\eta}C < C$$

Combining this with the obvious inequality $f(\bar{x}) > f(u) > f(\bar{x}) - \eta$, we deduce $|\nabla f|(u) \geq r$, and consequently $r_{j+1} \geq r$. This completes the induction.

It follows immediately from (3.1) that for varying k, the mappings $u_k(\tau)$ are Lipschitz continuous with a uniform constant r^{-1}. As all of these mappings coincide at $\tau = 0$, the well-known theorem of Arzelà and Ascoli guarantees that a certain subsequence of $u_k(\tau)$ converges uniformly on $[0, \eta]$ to some mapping $x(\tau)$. Furthermore any uniform limit of $u_k(\tau)$ is clearly also Lipschitz continuous with the same constant r^{-1}. This establishes the **existence** part of the theorem.

We now claim that the inequality

$$f(x(\tau)) \leq f(\bar{x}) - \tau, \quad \text{holds for each } \tau \in [0, \eta]. \tag{3.3}$$

Indeed, fix $\tau \in [0, \eta]$ and observe that there exists a sequence $\tau_k \to \tau$ with $f(u_k(\tau_k)) \leq f(\bar{x}) - \tau_k$. Since f is lsc, we deduce

$$f(x(\tau)) \leq \liminf_{k \to \infty} f(u_k(\tau_k)) \leq f(\bar{x}) - \tau,$$

thereby establishing (3.3). In particular, we conclude that the curve $x(\tau)$ lies fully in the domain of f, thereby completing the proof of the **regularity** claim.

Next we claim that almost everywhere on $[0, \eta]$, we have

$$\|\dot{x}(\tau)\| \leq \frac{1}{\overline{|\nabla f|}(x(\tau))}. \tag{3.4}$$

This is almost immediate from (3.1). Indeed fix $\tau \in (0, \eta)$ at which x is differentiable. We can refer to (3.1) to show that for large k, we have

$$\|\dot{u}_k(\tau)\| \leq \frac{1}{r_i^{(k)}},$$

for some $i \in \{0, \ldots, k\}$. Observe that since $\dot{u}(\tau)$ are uniformly bounded, up to a subsequence, the curves $u_k(\tau)$ converge weakly to $\dot{x}(\tau)$ in $L^2[0, \eta]$. Since weak convergence does not increase pointwise the norm almost everywhere, letting k tend to infinity, we deduce the result.

Now suppose that f is level prox-stable at $\bar{x}$. We'll show that we have equality in (3.3). To see this, fix $\tau \in [0, \eta]$ and observe that for any curve u_k obtained from the partition $0 = \tau_0 < \tau_1 < \ldots < \tau_k = \eta$ of the interval $[0, \eta]$, the equality $u_k(\tau_i) = f(\bar{x}) - \tau_i$ holds.

Moreover, for any $\delta > 0$ there exists a sufficiently large integer k and a point $\tau' \in [0, \eta]$, satisfying $f(u_k(\tau')) = f(\bar{x}) - \tau'$ and

$$\max\left\{|\tau - \tau'|, |f(x(\tau)) - f(x(\tau'))|, |f(x(\tau')) - f(u_k(\tau'))|\right\} < \delta.$$

We deduce

$$\begin{aligned}|f(x(\tau)) - (f(\bar{x}) - \tau)| \ &\leq \ |f(x(\tau)) - f(x(\tau'))| + |f(x(\tau') - f(u_k(\tau'))| + \\ &+ |f(u_k(\tau')) - (f(\bar{x}) - \tau)| \leq 3\delta.\end{aligned}$$

Since δ can be arbitrarily small, we conclude that the equation

$$f(x(\tau)) = f(\bar{x}) - \tau, \quad \text{holds for each } \tau \in [0, \eta]. \tag{3.5}$$

Now consider a real $\tau \in (0, \eta)$ at which x is differentiable and a real $\delta > 0$. From the equation above, we deduce that for sufficiently small ϵ, we have

$$\begin{aligned}\epsilon = |f(x(\tau + \epsilon)) - f(x(\tau))| &= |f\big(x(\tau) + \epsilon\dot{x}(\tau) + o(\epsilon)\big) - f(x(\tau))| \\ &\leq \Big(\mathrm{lip}\, f\big(x(\tau); \mathrm{dom}\, f\big) + \delta\Big)\|\epsilon\dot{x}(\tau) + o(\epsilon)\|.\end{aligned}$$

Letting ϵ and δ tend to 0, we conclude that the inequality

$$\frac{1}{\mathrm{lip}\, f\big(x(\tau); \mathrm{dom}\, f\big)} \leq \|\dot{x}(\tau)\|, \quad \text{holds for a.e. } \tau \in [0, \eta]. \tag{3.6}$$

Suppose that f is level prox-stable and continuous on $\mathrm{dom}\, f$. Observe that in light of Proposition 2.5, for any index k and any $\tau \in [\tau_i, \tau_{i+1}]$ (for $i = 1, \ldots, k$) we have $\dot{u}_k(\tau) \in -\big(\mathrm{cone}\, \partial f(u_k(\tau_{i+1}))\big) \cup \partial^\infty f(u_k(\tau_{i+1}))$. Furthermore, recall that restricting to a subsequence we may suppose that $\dot{u}_k$ converges weakly to $\dot{x}(\tau)$ in $L^2[0, \eta]$. Mazur's Lemma then implies that a subsequence of convex combinations of the form $\sum_{n=k}^{N(k)} \alpha_n^k \dot{u}_n$ converges strongly to $\dot{x}$ as k tends to ∞. Since convergence in $L^2[0, \eta]$ implies almost everywhere pointwise convergence, we deduce that for almost every $\tau \in [0, \eta]$, we have

$$\Big\| \sum_{n=k}^{N(k)} \alpha_n^k \dot{u}_n(\tau) - \dot{x}(\tau) \Big\| \to 0.$$

Therefore if the inclusion

$$\dot{x}(\tau) \in -\mathrm{cl}\, \mathrm{conv}\Big[\big(\mathrm{cone}\ \partial f(x(\tau))\big) \cup \partial^\infty f(x(\tau))\Big]$$

did not hold, then we would deduce that there exists a subsequence of vectors $\dot{u}_{n_l}^{k_l}(\tau)$ with $\lim_{l\to\infty} \dot{u}_{n_l}^{k_l}(\tau)$ not lying in the set on the right-hand-side of the inclusion above. This immediately yields a contradiction.

If f is subdifferentiable along the curve $x(\tau)$, then the claimed inclusion follows from the definition of the Clarke subdifferential. If f is Lipschitz continuous on its domain, then Lemma 2.5 implies that f is subdifferentiable on $\mathrm{dom}\, f$. A standard argument using Carathéodory's theorem then implies that the closure operation is not needed in the inclusion above. This completes the proof of the **differential inclusion** claim and the proof of the theorem. □

In Theorem 3.4, in particular we saw that if $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is locally Lipschitz continuous on $\operatorname{dom} f$ at $\bar{x}$, then for any proximal descent curve $x\colon [0,\eta] \to \mathbf{R}^n$ of f at $\bar{x}$, with $\eta > 0$ sufficiently small, there exist multipliers $\lambda(\tau)$ satisfying

$$\dot{x}(\tau) \in -\lambda(\tau)\partial_c f(x(\tau)), \qquad \text{for a.e. } \tau \in [0,\eta].$$

To further our goals, we need a more precise estimate on $\lambda(\tau)$. The following condition comes to the fore.

Assumption 3.5. *For a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and for any absolutely continuous curve $x\colon [0,\eta) \to \operatorname{dom} f$, the function*

$$y \mapsto \langle \dot{x}(\tau), y\rangle \quad \text{ is constant on } \partial_c f(x(\tau)).$$

for almost all $\tau \in [0,\eta)$.

Assumption 3.5 may look a bit strange at first sight. Nevertheless, in Section 5 we will see that the collection of functions satisfying this condition is very large, in particular including all Lipschitz continuous subdifferentially regular functions and all semi-algebraic functions.

Theorem 3.6 (Proximal descent curves and differential inclusions). *Consider a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ that is locally Lipschitz continuous at a point $\bar{x}$ relative to its domain, and that satisfies $0 \notin \partial f(\bar{x})$. Suppose in addition that f is level prox-stable at $\bar{x}$ and that Assumption 3.5 holds. Then for any proximal descent curve $x\colon [0,\eta] \to \mathbf{R}^n$ of f at $\bar{x}$, with $\eta > 0$ sufficiently small, we have*

$$\frac{\dot{x}(\tau)}{\|\dot{x}(\tau)\|^2} \in -\partial f(x(\tau)), \quad x(0) = \bar{x}, \quad \textit{and}$$

$$\|\dot{x}(\tau)\|^{-1} = d(0, \partial f(x(\tau))) = d(0, \partial_c f(x(\tau))),$$

for almost every $\tau \in [0,\eta]$.

Proof. Without loss of generality, we may assume that f is Lipschitz continuous on the whole domain. Let l be its Lipschitz modulus. Let $\tau \in [0,\eta]$ be such that x is differentiable at τ and the function

$$y \mapsto \langle \dot{x}(\tau), y\rangle \quad \text{ is constant on } \partial_c f(x(\tau)).$$

Consider the affine subspaces

$$V_\tau = \operatorname{par} \partial_c f\big(x(\tau)\big),$$
$$W_\tau = \Big(\operatorname{par} \partial_c f\big(x(\tau)\big)\Big)^{\perp},$$

and let b be an arbitrary element of $\partial_c f(x(\tau))$. Then we have

$$\operatorname{aff} \partial_c f(x(\tau)) = b + V_\tau.$$

Combining this with Theorem 3.4, we deduce that there exists a real $\lambda > 0$ satisfying

$$-\dot{x}(\tau) \in \lambda \partial_c f\big(x(\tau)\big) \subset \lambda b + V_\tau.$$

We claim now that the inclusion $\dot{x}(\tau) \in W_\tau$ holds. To see this, observe that for any real λ_i and for vectors $v_i \in \partial_c f\big(x(\tau)\big)$, for $i = 1, \ldots, k$, we have

$$\langle \dot{x}(\tau), \sum_{i=1}^k \lambda_i (v_i - b)\rangle = \sum_{i=1}^k \lambda_i \big[\langle \dot{x}(\tau), v_i\rangle - \langle \dot{x}(\tau), b\rangle\big] = 0.$$

Now using Lemma 2.13, we deduce that $-\frac{1}{\lambda}\dot{x}(\tau)$ achieves the distance of the affine space, $\operatorname{aff}\partial_c f(x(\tau))$, to the origin. On the other hand, the inclusion $-\frac{1}{\lambda}\dot{x}(\tau) \in \partial_c f\big(x(\tau)\big)$ holds, and hence $-\frac{1}{\lambda}\dot{x}(\tau)$ actually achieves the distance of $\partial_c f(x(\tau))$ to the origin. In particular, we deduce

$$\frac{1}{\lambda}\|\dot{x}(\tau)\| = d(0, \partial_c f(x(\tau))).$$

Now for a fixed $K > l$, define the function

$$g(x) := \inf_{y\in\mathbf{R}^n}\{f(y) + K\|x-y\|\}. \tag{3.7}$$

This function is K-Lipschitz continuous on all of $\mathbf{R}^n$ and it coincides with f on $\operatorname{dom} f$. By Lebourg's Mean Value Theorem (Theorem 2.7), we have $\epsilon = \langle y_\epsilon, x(\tau+\epsilon) - x(\tau)\rangle$, for some subgradient $y \in \partial_c g(z_\epsilon)$, where z_ϵ is a point lying on the line segment joining $x(\tau+\epsilon)$ and $x(\tau)$. Letting ϵ tend to 0, we deduce the existence of a subgradient $y \in \partial_c g(x(\tau)) \subset \partial_c f(x(\tau))$ (Lemma 2.10) satisfying

$$\frac{1}{\|\dot{x}(\tau)\|} = \Big\langle y, \frac{\dot{x}(\tau)}{\|\dot{x}(\tau)\|}\Big\rangle = d(0, \partial_c f(x(\tau))), \tag{3.8}$$

where the second equality follows from the computation

$$\Big\langle y, \frac{\dot{x}(\tau)}{\|\dot{x}(\tau)\|}\Big\rangle = \Big\langle y + \operatorname{par}\partial_c f\big(x(t)\big), \frac{\dot{x}(\tau)}{\|\dot{x}(\tau)\|}\Big\rangle = \Big\langle \frac{1}{\lambda}\dot{x}(\tau), \frac{\dot{x}(\tau)}{\|\dot{x}(\tau)\|}\Big\rangle = \frac{1}{\lambda}\|\dot{x}(\tau)\|.$$

Finally combining this with Theorem 3.4, we obtain the chain of inequalities

$$\frac{1}{d(0, \partial f(x(\tau)))} \geq \|\dot{x}(\tau)\| = \frac{1}{d(0, \partial_c f(x(\tau)))} \geq \frac{1}{d(0, \partial f(x(\tau)))}. \tag{3.9}$$

Hence we have equality throughout. Observe now that the equation $\frac{1}{\lambda}\|\dot{x}(\tau)\| = \frac{1}{\|\dot{x}(\tau)\|}$ implies $\lambda = \|\dot{x}(\tau)\|^2$. Combining this with (3.9), the stronger inclusion $\frac{\dot{x}(\tau)}{\|\dot{x}(\tau)\|^2} \in -\partial f(x(\tau))$ easily follows. □

Theorem 3.7 (Natural reparametrization of proximal descent curves). *Consider a function $f\colon \mathbf{R}^n \to \mathbf{R}$ that is locally Lipschitz continuous relative to its domain at a point $\bar{x}$, with $0 \notin \partial f(\bar{x})$. Suppose in addition that f is level prox-stable at $\bar{x}$. Then for any proximal descent curve $x\colon [0,\eta] \to \mathbf{R}^n$ of f at $\bar{x}$, with $\eta > 0$ sufficiently small, there exists a Lipschitz continuous reparametrization $t(\tau)$ sending the interval $[0,\eta]$ to $[0,T]$ and satisfying the following properties.*

(a) $x(t)$ is Lipschitz continuous.

(b) $\operatorname{lip} f(x(t); \operatorname{dom} f) \geq \|\dot{x}(t)\| \geq \overline{|\nabla f|}(x(t))$ a.e. on $[0,T]$.

(c) $(f\circ x)(t) = f(x(t))$ is Lipschitz continuous, strictly decreasing, and satisfies

$$\frac{d(f\circ x)}{dt}(t) \leq -\big(\overline{|\nabla f|}(x(t))\big)^2, \quad \text{a.e. on } [0,T].$$

(d) The inclusion

$$\dot{x}(t) \in -\operatorname{cone}\partial_c f(x(t)), \quad \textit{holds a.e. on } [0,T].$$

If furthermore Assumption 3.5 is satisfied, then we have

$$\dot{x}(t) \in -\partial f(x(t)), \quad x(0) = \bar{x}, \quad \textit{and}$$

$$|\dot{x}(t)| = d(0, \partial f(x(t))) = d(0, \partial_c f(x(t))),$$

for almost every $t \in [0, T]$*, and moreover* $x(t)$ *is a curve of near-maximal slope.*

Proof. Consider the function

$$t(\tau) := \int_0^\tau \|\dot{x}(s)\|^2 ds.$$

and define $T := t(\eta)$. Theorem 3.4 implies that $t(\tau)$ is Lipschitz continuous and $\|\dot{x}(s)\|$ is bounded away from zero. Consequently $t(\tau)$ has a Lipschitz continuous inverse. Let $\tau(t)$ be the inverse of $t(\tau)$, and set $x(t) := x(\tau(t))$ for $t \in [0, T]$. A trivial application of the chain rule, along with Theorems 3.6 and 3.4, yield all of the claimed results. ◻

4. Viability and strengthened existence theory. Observe that the combination of Theorems 3.4 and 3.7 shows that for any lsc, level prox-stable, function $f\colon \mathbf{R}^n \to \mathbf{R}$ that is locally Lipschitz continuous on its domain, and that satisfies the key Assumption 3.5, there exists a proximal descent curve that is both a curve of near-maximal slope and a solution to the corresponding evolution equation. By far, the most stringent of the conditions needed to establish this is level prox-stability. In this section, we completely eliminate this condition. The following viability result will be key. Roughly speaking, it states that given a function f that is Lipschitz continuous on its domain, we can always find a Lipschitz continuous function g on all of $\mathbf{R}^n$, agreeing with f on the domain of f, and having the crucial property that it admits proximal descent curves lying entirely in the domain of f.

Lemma 4.1 (Viability). *Consider a lsc function* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *that is Lipschitz continuous on its domain with modulus* l*. Assume that there is a point* $\bar{x} \in \operatorname{dom} f$ *and a real number* $r > 0$ *with* $\overline{|\nabla f|}(\bar{x}) \geq r$*. Take* $K > \max\{l, r\}$ *and define the function*

$$g(x) := \inf_{y \in \mathbf{R}^n} \{f(y) + K\|x - y\|\}. \tag{4.1}$$

Then g *is a* K*-Lipschitz continuous function, coinciding with* f *on* $\operatorname{dom} f$*, and having the property that there exists a proximal descent curve of* g *at* $\bar{x}$ *lying entirely in* $\operatorname{dom} f$*.*

Proof. It is well-known that g is a K-Lipschitz continuous function on $\mathbf{R}^n$ and that it coincides with f on $\operatorname{dom} f$. Moreover, since f is lsc and Lipschitz continuous on $\operatorname{dom} f$, the infimum in (4.1) is attained for any x. Indeed, it is easy to see that since we have $l < K$, minimizing sequences are bounded.

In the notation of Definition 3.1, define $\epsilon := \tau_{i+1} - \tau_i$. For sufficiently small $\eta > 0$ consider constructing curves $u_k(\tau)$. When building the sequence $u_k(\tau_i)$ for $i \in \{0, k\}$, there are choices to be made, since projections onto lower level sets are not necessarily single-valued. For notational convenience, we drop the index k in u_k and we define points $u_i := u_k(\tau_i)$. We claim that for any fixed k, we may choose the iterates $u(\tau_i)$ so that there is a sequence of points $y_i \in \operatorname{dom} f$ satisfying

$$g(u_i) = f(y_i) + K\|u_i - y_i\| \tag{4.2}$$

$$\|u_i - y_i\| \leq \frac{\epsilon l}{r(K - l)}. \tag{4.3}$$

This is clearly true for $i = 0$. As the inductive hypothesis, suppose this is true for $i = 0, \ldots, j$. We now consider two cases.

Case (a): Suppose first $\|u_j - y_j\| \geq \frac{\epsilon}{K}$. Then for any $\hat{u}$ with $g(\hat{u}) = g(u_i) - \epsilon$, we have

$$\epsilon = g(u_i) - g(\hat{u}) \leq K\|\hat{u} - u_i\|.$$

Consequently we deduce $d(u_i, [g \leq g(u_i) - \epsilon]) \geq \frac{\epsilon}{K}$. On the other hand, if we take $\tilde{u} = u_j + \frac{\epsilon}{K}\frac{y_j - u_j}{\|y_j - u_j\|}$, we obtain

$$g(\tilde{u}) \leq f(y_j) + K\|\tilde{u} - y_j\| = f(y_j) + K\left\|u_j - y_j + \frac{\epsilon}{K}\frac{y_j - u_j}{\|y_j - u_j\|}\right\| = g(u_j) - \epsilon.$$

In conjunction with the equality $\frac{\epsilon}{K} = \|\tilde{u} - u_j\|$, we deduce the inclusion $\tilde{u} \in P_{\mathrm{dom}\, f}(u_j)$. Therefore we may set $u(\tau_{j+1}) = \tilde{u}$ and $y_{j+1} = y_j$. Equations (4.2) and (4.3) then follow from the inductive hypothesis.

Case (b): Suppose now that the inequality $\|u_j - y_j\| < \frac{\epsilon}{K}$ holds. In light of Lemma 2.10, decreasing r slightly we may assume $|\nabla g|(x) > r$ for each x near $\bar{x}$. Therefore we can take any u_{j+1} such that $g(u_{j+1}) = g(u_j) - \epsilon$ and $\|u_{j+1} - u_j\| \leq \frac{\epsilon}{r}$. Let y_{j+1} be a point achieving the minimum in the expression for $g(u_{j+1})$. We then obtain

$$f(y_{j+1}) + K\|u_{j+1} - y_{j+1}\| \leq f(y_j) + K\|u_j - y_j\| - \epsilon \leq f(y_j),$$

In turn, this implies

$$K\|u_{j+1} - y_{j+1}\| \leq l\|y_{j+1} - y_j\| \leq l(\|y_{j+1} - u_{j+1}\| + \|u_{j+1} - y_j\|),$$

that is

$$(K - l)\|u_{j+1} - y_{j+1}\| \leq \frac{\epsilon l}{r}.$$

Equations (4.2) and (4.3) follow for $j + 1$. □

We finally come to the main result of this section.

THEOREM 4.2 (Existence of steepest descent curves). *Consider a function $f : \mathbf{R}^n \to \mathbf{R}$ that is locally Lipschitz continuous relative to its domain at a point $\bar{x}$, with $0 \notin \partial f(\bar{x})$. Then there exists an absolutely continuous curve $x : [0, T] \to \mathbf{R}^n$ with $x(0) = \bar{x}$, satisfying the following properties.*

(a) $x(t)$ is Lipschitz continuous.

(b) $\|\dot{x}(t)\| \geq \overline{|\nabla f|}(x(t))$ a.e. on $[0, T]$.

(c) $(f \circ x)(t) = f(x(t))$ is Lipschitz continuous, strictly decreasing, and satisfies

$$\frac{d(f \circ x)}{dt}(t) \leq -\left(\overline{|\nabla f|}(x(t))\right)^2, \quad \text{a.e. on } [0, T].$$

(d) The inclusion

$$\dot{x}(t) \in -\mathrm{cone}\, \partial_c f(x(t)), \quad \textit{holds a.e. on } [0, T].$$

If furthermore Assumption 3.5 is satisfied, then we have

$$\dot{x}(t) \in -\partial f(x(t)), \quad x(0) = \bar{x}, \quad \textit{and}$$

$$|\dot{x}(t)| = d(0, \partial f(x(t))) = d(0, \partial_c f(x(t))),$$

for almost every $t \in [0, T]$*, and moreover* $x(t)$ *is a curve of near-maximal slope.*

Proof. Without loss of generality, we may assume that f is Lipschitz continuous on the entire $\operatorname{dom} f$. Lemma 4.1 yields a Lipschitz continuous function $g\colon \mathbf{R}^n \to \mathbf{R}$ and a proximal descent curve $x\colon [0, \eta] \to \mathbf{R}^n$ of g at $\bar{x}$ lying entirely in $\operatorname{dom} f$. Then Theorems 3.7 and 2.10 readily imply all of the claims. ◻

5. Comments on Assumption 3.5 and curves of near-maximal slope. In this section, we further analyze Assumption 3.5 and the notion of a curve of near-maximal slope. Our first goal is to prove that locally Lipschitz continuous subdifferentially regular functions satisfy Assumption 3.5. We begin with the following key lemma. A similar result may be found in [8, Lemma 3.3, p.73].

Lemma 5.1 (Calculus). *Consider a function* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *and an absolutely continuous path* $x\colon [0, T) \to \mathbf{R}^n$*. Then for* $t \in (0, T)$ *we have*

$$\frac{d(f \circ x)}{dt}(t) = \langle \hat{\partial} f(x(t)), \dot{x}(t) \rangle,$$

provided that both x *and* $f \circ x$ *are differentiable at* t*, and* $\hat{\partial} f(x(t)) \neq \emptyset$*. Consequently, in this case the vector* $\dot{x}(t)$ *is orthogonal to* $\operatorname{par} \hat{\partial} f(x(t))$*.*

Proof. Observe

$$\begin{aligned} \frac{d(f \circ x)}{dt}(t) &= \lim_{\epsilon \downarrow 0} \frac{f(x(t+\epsilon)) - f(x(t))}{\epsilon} \\ &\geq \langle v, \dot{x}(t) \rangle, \quad \text{ for any } v \in \hat{\partial} f(x(t). \end{aligned}$$

Similarly we have

$$\begin{aligned} \frac{d(f \circ x)}{dt}(t) &= \lim_{\epsilon \downarrow 0} \frac{f(x(t-\epsilon)) - f(x(t))}{-\epsilon} \\ &\leq \langle v, \dot{x}(t) \rangle, \quad \text{ for any } v \in \hat{\partial} f(x(t)). \end{aligned}$$

Hence if we have $\hat{\partial} f(x(t)) \neq \emptyset$, then the equation

$$\frac{d(f \circ x)}{dt}(t) = \langle \hat{\partial} f(x(t)), \dot{x}(t) \rangle, \quad \text{holds,}$$

as claimed. Observe that for any convex set Q, we have $\operatorname{par} Q = \mathbf{R}(Q - Q)$. The fact that $\dot{x}(t)$ is orthogonal to $\operatorname{par} \hat{\partial} f(x(t))$ is now immediate. ◻

Proposition 5.2 (Lipschitz continuous, subdifferentially regular functions). *Consider a function* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *that is subdifferentiable, locally Lipschitz continuous relative to* $\operatorname{dom} f$*, and subdifferentially regular. Then* f *satisfies Assumption 3.5.*

Proof. Consider an absolutely continuous path $x\colon [0, T) \to \operatorname{dom} f$. Then clearly both $f \circ x$ and x are differentiable almost everywhere on $[0, T)$. The result now follows trivially from Lemma 5.1. ◻

The following proposition shows that for a subdifferentially regular function that is locally Lipschitz continuous on its domain, curves of maximal slope are exactly the solutions to the corresponding evolution equations.

Proposition 5.3 (Equivalence for subdifferentially regular Lipschitz functions). *Consider a subdifferentially regular function* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *that is Lipschitz continuous* *relative to its domain. Then* $x\colon [0,T) \to \operatorname{dom} f$ *is a curve of maximal slope if and only if it satisfies*

$$\dot{x}(t) \in -\partial f(x(t)), \qquad \textit{for almost every } t \in [0,T).$$

Proof. Consider an absolutely continuous curve $x\colon [0,T) \to \operatorname{dom} f$. Then clearly both x and $f \circ x$ are differentiable almost everywhere on $[0,T)$. Suppose now that x satisfies the subdifferential inclusion for almost every $t \in [0,T)$. Then according to Lemma 5.1, for such t we have $-\dot{x}(t) \in \partial f(x(t))$ and $-\dot{x}(t) \in (\operatorname{par} \partial f(x(t)))^{\perp}$. Lemma 2.13 then implies $\|\dot{x}(t)\| = d(0, \partial f(x(t))) = |\nabla f|(x(t))$. In turn, Lemma 5.1 shows

$$\frac{d(f \circ x)}{dt}(t) = -\|\dot{x}(t)\|^2 = -(|\nabla f|(x(t)))^2.$$

Thus $x(t)$ is a curve of maximal slope.

Conversely suppose that $x\colon [0,T) \to \mathbf{R}^n$ is a curve of maximal slope and let $v(t) \in \partial f(x(t))$ be a vector achieving $d(0, \partial f(x(t)))$. Then again by Lemma 5.1, for almost every $t \in [0,T)$ we have

$$-(|\nabla f|(x(t)))^2 = \frac{d(f \circ x)}{dt}(t) = \langle v(t), \dot{x}(t)\rangle. \tag{5.1}$$

Consequently the inequality $(|\nabla f|(x(t)))^2 \leq \|\bar{v}\|\|\dot{x}(t)\|$ holds, with equality if and only if $v(t)$ and $\dot{x}(t)$ are collinear. Observe that $v(t)$ and $\dot{x}(t)$ are indeed collinear, since if it were otherwise we would have

$$(|\nabla f|(x(t)))^2 < \|\bar{v}\|\|\dot{x}(t)\| = (|\nabla f|(x(t)))^2,$$

a contradiction. Using the equations

$$\|v(t)\| = \|\dot{x}(t)\| = |\nabla f|(x(t)),$$

and (5.1), we deduce $\dot{x}(t) = -v(t) \in \partial f(x(t))$, as we needed to show. □

6. The semi-algebraic setting. In this section we consider to what extent the results obtained thus far can be strengthened when the function f in question is semi-algebraic. For an extensive discussion on semi-algebraic geometry, see the monographs of Basu-Pollack-Roy [4], Lou van den Dries [28], and Shiota [27]. For a quick survey, see the article of van den Dries-Miller [29] and the surveys of Coste [11, 10]. Unless otherwise stated, we follow the notation of [29] and [11].

A *semi-algebraic* set $S \subset \mathbf{R}^n$ is a finite union of sets of the form

$$\{x \in \mathbf{R}^n : P_1(x) = 0, \ldots, P_k(x) = 0, Q_1(x) < 0, \ldots, Q_l(x) < 0\},$$

where $P_1, \ldots, P_k$ and $Q_1, \ldots, Q_l$ are polynomials in n variables. In other words, S is a union of finitely many sets, each defined by finitely many polynomial equalities and inequalities. A function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is *semi-algebraic* if $\operatorname{epi} f \subset \mathbf{R}^{n+1}$ is a semi-algebraic set.

Our immediate goal is to show that Assumption 3.5 holds for all semi-algebraic functions. We first record the following simple lemma, which in fact has nothing to do with semi-algebraicity.

LEMMA 6.1. *Consider a set* $M \subset \mathbf{R}^n$ *and a path* $x\colon [0,\eta) \to \mathbf{R}^n$ *that is differentiable almost everywhere on* $[0,\eta)$. *Then for almost every* $t \in [0,\eta)$, *the implication*

$$x(t) \in M \Longrightarrow \dot{x}(t) \in T_M(x(t)), \qquad \textit{holds.}$$

The following is a key property of semi-algebraic functions that we will use.

THEOREM 6.2 (Projection formula). *Consider a semi-algebraic function* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$. *Then there exists a partition of* $\operatorname{dom} f$ *into finitely many* $\mathbf{C}^2$*-manifolds* $\{M_i\}$ *so that for any manifold* M_i *and any point* $x \in M_i$, *the inclusion*

$$\operatorname{par} \partial_c f(x) \subset N_{M_i}(x),$$

holds.

PROPOSITION 6.3 (Semi-algebraic functions satisfy Assumption 3.5). *Consider a semi-algebraic function* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$. *Then* f *satisfies Assumption 3.5.*

Proof. Consider the partition of $\operatorname{dom} f$ into finitely many $\mathbf{C}^2$-manifolds $\{M_i\}$, guaranteed to exist by Theorem 6.2, and let $x\colon [0,\eta) \to \operatorname{dom} f$ be any absolutely continuous path. Since there are only finitely many manifolds M_i, applying Lemma 6.1, we deduce that for any index i and for almost very $t \in [0,\eta)$, the implication

$$x(t) \in M_i \Longrightarrow \dot{x}(t) \in T_{M_i}(x(t)),$$

holds. On the other hand by Theorem 6.2, for any point x lying in a stratum M_i we have $\partial_c f(x) \subset v + N_M(x)$, for some vector $v \in \mathbf{R}^n$. The result follows. □

Next obtain a semi-algebraic counterpart of Proposition 5.3, where we completely eliminate the subdifferential regularity requirement.

PROPOSITION 6.4 (Equivalence for semi-algebraic Lipschitz functions). *Consider a semi-algebraic function* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *that is Lipschitz continuous relative to its domain. Then* $x\colon [0,T) \to \operatorname{dom} f$ *is a curve of near-maximal slope if and only if it satisfies*

$$\dot{x}(t) \in -\partial f(x(t)), \qquad \textit{for almost every } t \in [0,T).$$

Proof. Consider the partition of $\operatorname{dom} f$ into finitely many $\mathbf{C}^2$-manifolds $\{M_i\}$, guaranteed to exist by Theorem 6.2. We first record some preliminary observations. Clearly both x and $f \circ x$ are differentiable at a.e. $t \in [0,T)$. Furthermore, in light of Lemma 6.1, for any index i and for a.e. $t \in [0,\eta)$ the implication

$$x(t) \in M_i \Longrightarrow \dot{x}(t) \in T_{M_i}(x(t)), \qquad \text{holds.}$$

Now suppose that for such t, the point $x(t)$ lies in a stratum M_i and let $g\colon \mathbf{R}^n \to \mathbf{R}$ be a $\mathbf{C}^1$-smooth function agreeing with f on a neighborhood of $x(t)$ in M_i. Lipschitzness of f on its domain then easily implies

$$\begin{aligned}
\frac{d(f\circ x)}{dt}(t) &= \lim_{\epsilon \downarrow 0} \frac{f(x(t+\epsilon)) - f(x(t))}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{f(P_{M_i}(x(t+\epsilon))) - f(x(t))}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{g(P_{M_i}(x(t+\epsilon))) - g(x(t))}{\epsilon} \\
&= \frac{d}{dt}\, g \circ P_{M_i} \circ x(t) = \langle \nabla g(x(t)), \dot{x}(t) \rangle \\
&= \langle \nabla g(x(t)) + N_{M_i}(x(t)), \dot{x}(t) \rangle = -\|\dot{x}(t)\|^2.
\end{aligned}$$

Suppose now that $x\colon [0,T) \to \mathbf{R}^n$ satisfies the subdifferential inclusion for a.e. $t \in [0,T)$. Then we have $-\dot{x}(t) \in \partial f(x(t))$ and $-\dot{x}(t) \in T_{M_i}(x(t))$. Hence Lemma 2.13 implies $\|\dot{x}(t)\| = \overline{|\nabla f|}(x(t))$ for a.e. $t \in [0,T)$. Using the computation above, we conclude that $x(t)$ is curve of near-maximal slope.

Conversely, suppose that $x(t)$ is a curve of near-maximal slope. Then we have

$$-(|\nabla f|(x(t)))^2 = \frac{d(f \circ x)}{dt}(t) = \langle \nabla g(x(t)) + N_{M_i}(x(t)), \dot{x}(t)\rangle.$$

Now let $v(t) \in \partial f(x(t))$ be a vector achieving $d(0, \partial f(x(t)))$. Then from the equation above we conclude $(|\nabla f|(x(t)))^2 \leq \|v(t)\|\|\dot{x}(t)\|$, with equality if and only if $\dot{x}(t)$ and $v(t)$ are collinear. On the other hand we clearly have equality in this expression, and consequently we deduce $-\dot{x}(t) = v(t) \in \partial f(x(t))$ for a.e. $t \in [0,T)$. □

We end this section by showing that for semi-algebraic functions, bounded curves of near-maximal slope necessarily have bounded length. None of the arguments we present are new; rather we include this discussion with the purpose of painting a more complete picture for the reader. We begin with the celebrated Kurdyka-Łojasiewicz inequality.

Definition 6.5 (Kurdyka-Łojasiewicz inequality). A function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is said to satisfy the *Kurdyka-Łojasiewicz inequality* if for any bounded open set $U \subset \mathbf{R}^n$ and any real τ, there exists $\rho > 0$ and a non-negative continuous function $\psi\colon [\tau, \tau + \rho) \to \mathbf{R}$, which is $\mathbf{C}^1$-smooth and strictly increasing on $(\tau, \tau+\rho)$, and such that the inequality

$$\|\nabla(\psi \circ f)\|(x) \geq 1,$$

holds for all $x \in U$ with $\tau < f(x) < \tau + \rho$.

In particular, all semi-algebraic functions satisfy the Kurdyka-Łojasiewicz inequality [19]. For an extensive study the Kurdyka-Łojasiewicz inequality and a description of its historical significance, see for example [5]. The proof of the following theorem is almost identical to the proof of [19, Theorem 7.1]; hence we only provide a sketch. In fact, the theorem remains valid if rather than assuming semi-algebraicity, we only assume that the Kurdyka-Łojasiewicz inequality is satisfied.

Theorem 6.6 (Lengths of curves of near-maximal slope).
Consider a lsc, semi-algebraic function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, and let U be a bounded subset of $\mathbf{R}^n$. Then there exists a number $N > 0$ such that the length of any curve of near-maximal slope for f lying in U does not exceed N.

Proof. Let $x\colon [0,T)$ be a curve of near-maximal slope for f and let ψ be any strictly increasing $\mathbf{C}^1$-smooth function on an interval containing the image of $f \circ x$. It is then easy to see then that, up to a reparametrization, x is a curve of near-maximal slope for the composite function $\psi \circ f$. In particular, we may assume that f is bounded on U, since otherwise we may for example replace f by $\psi \circ f$ where $\psi(t) = \frac{t}{\sqrt{1+t^2}}$.

Define the function

$$\xi(s) = \inf\{|\nabla f|(x) : x \in U,\, f(x) = s\}.$$

Standard arguments show that ξ is semi-algebraic. Consequently, with an exception of finitely many points, the domain of ξ is a union of finitely many open intervals (α_i, β_i), with ξ continuous and either strictly monotone or constant on each such interval. Define for each index i, the quantity

$$c_i = \inf\{\xi(s) : s \in (\alpha_i, \beta_i)\}.$$

We first claim that ξ is strictly positive on each interval (α_i, β_i). This is clear for indices i with $c_i > 0$. On the other hand if we have $c_i = 0$, then by Sard's theorem [18] the function ξ is strictly positive on (α_i, β_i) as well.

Define ζ_i and η_i by

$$\zeta = \inf\{t : f(x(t)) = \alpha_i\} \quad \text{and} \quad \eta = \sup\{t : f(x(t)) = \beta_i\},$$

and let l_i be the length of $x(t)$ between ζ_i and η_i.

Then we have

$$l_i = \int_{\zeta_i}^{\eta_i} \|\dot{x}(t)\|dt = \int_{\zeta_i}^{\eta_i} \overline{|\nabla f|}(x(t))dt \le \Big((\eta_i - \zeta_i) \int_{\zeta_i}^{\eta_i} \overline{|\nabla f|}(x(t))^2 dt\Big)^{\frac{1}{2}}.$$

On the other hand, observe

$$\int_{\zeta_i}^{\eta_i} \overline{|\nabla f|}(x(t))^2 dt = f(x(\eta_i)) - f(x(\zeta_i)) = \beta_i - \alpha_i.$$

Finally in the case $c_i > 0$ we have $l_i \ge c_i(\eta_i - \zeta_i)$, which combined with the two equations above yields the bound

$$l_i \le \frac{\beta_i - \alpha_i}{c_i}.$$

If the equation $c_i = 0$ holds, then by the Kurdyka-Łojasiewicz inequality we can find a continuous function $\xi_i \colon [\alpha_i, \alpha_i + \rho) \to \mathbf{R}$, for some $\rho > 0$, where ξ is strictly positive and $\mathbf{C}^1$-smooth on $(\alpha_i, \alpha_i + \rho)$ and satisfying $|\nabla(\xi_i \circ f)|(y) \ge 1$ for any $y \in U$ with $\alpha_i < f(y) < \alpha_i + \rho$. Since ξ_i is strictly increasing on $(\alpha_i, \alpha_i + \rho)$, it is not difficult to check that we may extend ξ_i to a continuous function on $[\alpha_i, \beta_i]$ and so that this extension is $\mathbf{C}^1$-smooth and strictly increasing on (α_i, β_i) with the inequality $|\nabla(\xi_i \circ f)|(y) \ge 1$ being valid for any $y \in U$ with $\alpha_i < f(y) < \beta_i$.

Then as we have seen before, up to a reparametrization, the curve $x(t)$ for $t \in [\zeta_i, \eta_i]$ is a curve of near maximal slope for the function $\xi_i \circ f$. Then as above, we obtain the bound $l_i \le \xi_i(\beta_i) - \xi_i(\alpha_i)$.

We conclude that the length of the curve $x(t)$ is bounded by a constant that depends only on f and on U, thereby completing the proof. □

The following consequence is now immediate.

Corollary 6.7 (Convergence of curves of near-maximal slope). *Consider a lsc, semi-algebraic function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$. Then any curve of near-maximal slope for f that is bounded and has a maximal domain of definition converges to a generalized critical point of f (a point $\bar{x}$ satisfying $0 \in \partial f(\bar{x})$).*

REFERENCES

[1] L. Ambrosio. Minimizing movements. *Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5)*, 19:191–246, 1995.

[2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[3] J.-P. Aubin and A. Cellina. *Differential inclusions*, volume 264 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1984. Set-valued maps and viability theory.

[4] S. Basu, R. Pollack, and M. Roy. *Algorithms in Real Algebraic Geometry (Algorithms and Computation in Mathematics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[5] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet. Characterizations of łojasiewicz inequalities: subgradient flows, talweg, convexity. *Trans. Amer. Math. Soc.*, 362(6):3319–3363, 2010.

[6] Jérôme Bolte, Aris Daniilidis, and Adrian Lewis. The łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.*, 17(4):1205–1223 (electronic), 2006.

[7] J.M. Borwein and Q.J. Zhu. *Techniques of Variational Analysis*. Springer Verlag, New York, 2005.

[8] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contraction dans des espaces de Hilbert*. North-Holland Math. Stud. 5, North-Holland, Amsterdam, 1973.

[9] F.H. Clarke, Yu. Ledyaev, R.I. Stern, and P.R. Wolenski. *Nonsmooth Analysis and Control Theory*. Texts in Math. 178, Springer, New York, 1998.

[10] M. Coste. *An Introduction to o-minimal Geometry*. RAAG Notes, 81 pages, Institut de Recherche Mathématiques de Rennes, November 1999.

[11] M. Coste. *An Introduction to Semialgebraic Geometry*. RAAG Notes, 78 pages, Institut de Recherche Mathématiques de Rennes, October 2002.

[12] E. De Giorgi. New problems on minimizing movements. In *Boundary value problems for partial differential equations and applications*, volume 29 of *RMA Res. Notes Appl. Math.*, pages 81–98. Masson, Paris, 1993.

[13] E. De Giorgi, A. Marino, and M. Tosques. Problems of evolution in metric spaces and maximal decreasing curve. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8)*, 68(3):180–187, 1980.

[14] M. Degiovanni, A. Marino, and M. Tosques. Evolution equations with lack of convexity. *Nonlinear Anal.*, 9(12):1401–1443, 1985.

[15] E. De Giorgi, A. Marino, and M. Tosques. Problems of evolution in metric spaces and maximal decreasing curve. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8)*, 68(3):180–187, 1980.

[16] J.-B. Hiriart-Urruty. Extension of Lipschitz functions. *J. Math. Anal. Appl.*, 77(2):539–554, 1980.

[17] A.D. Ioffe. Metric regularity and subdifferential calculus. *Uspekhi Mat. Nauk*, 55(3(333)):103–162, 2000.

[18] A.D. Ioffe. Critical values of set-valued maps with stratifiable graphs. Extensions of Sard and Smale-Sard theorems. *Proc. Amer. Math. Soc.*, 136(9):3111–3119, 2008.

[19] A.D. Ioffe. An invitation to tame optimization. *SIAM Journal on Optimization*, 19(4):1894–1917, 2009.

[20] A.D. Ioffe and J.-P. Penot. Subdifferentials of performance functions and calculus of coderivatives of set-valued mappings. *Serdica Math. J.*, 22(3):257–282, 1996. Well-posedness and stability of variational problems.

[21] K. Kuratowski. Les fonctions semi-continues dans l'espace des ensembles fermés. *Fundamenta Mathematicae*, 18:148–159, 1932.

[22] A. Marino, C. Saccon, and M. Tosques. Curves of maximal slope and parabolic variational inequalities on nonconvex constraints. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 16(2):281–330, 1989.

[23] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation I: Basic Theory*. Grundlehren der mathematischen Wissenschaften, Vol 330, Springer, Berlin, 2005.

[24] R.T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[25] R.T. Rockafellar and A.L. Dontchev. *Implicit functions and solution mappings*. Monographs in Mathematics, Springer-Verlag, 2009.

[26] R.T. Rockafellar and R.J-B. Wets. *Variational Analysis*. Grundlehren der mathematischen Wissenschaften, Vol 317, Springer, Berlin, 1998.

[27] M. Shiota. *Geometry of subanalytic and semialgebraic sets*. Birkhäuser Boston Inc., Cambridge, MA, USA, 1997.

[28] L. van den Dries. *Tame topology and o-minimal structures*, volume 248 of *LMS Lecture Note Series*. Cambridge University Press, Cambridge, 1998.

[29] L. van den Dries and C. Miller. Geometric categories and o-minimal structures. *Duke Mathematical Journal*, 84:497–540, 1996.