SIAM J. CONTROL OPTIM.
Vol. 53, No. 1, pp. 114–138

CURVES OF DESCENT*

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

Abstract. Steepest descent is central in variational mathematics. We present a new transparent existence proof for curves of near-maximal slope—an influential notion of steepest descent in a nonsmooth setting. The existence theory is further amplified for semialgebraic functions, prototypical nonpathological functions in nonsmooth optimization: such functions always admit nontrivial descent curves emanating from any (even critical) nonminimizing point. We moreover show that curves of near-maximal slope of semialgebraic functions have a more classical description as solutions of subgradient dynamical systems.

Key words. descent, slope, subdifferential, subgradient dynamical system, semialgebraic

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

DOI. 10.1137/130920216

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? To facilitate the discussion, consider "the fastest instantaneous rate of decrease" of a function f on a complete metric space $(\mathcal{X}, d)$, namely, the *slope*

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

Here, we use the convention $r^+ := \max\{0, r\}$. The slope of a smooth function on $\mathbf{R}^n$ simply coincides with the norm of the gradient, and hence the notation. For more details on slope see, for example, [16]. Even though the definition is deceptively simple, slope plays a decisive role in regularity theory and sensitivity analysis; see [3, 19].

One can readily verify for any 1-Lipschitz curve $\gamma\colon (a, b) \to \mathcal{X}$ the upper bound on the speed of descent:

$$|\nabla(f \circ \gamma)|(t) \leq |\nabla f|(\gamma(t)) \qquad \text{for a.e. } t \in (a, b). \tag{1.1}$$

Supposing that f is continuous (for technical reasons), it is then natural to call γ a *steepest descent curve* if $f \circ \gamma$ is nonincreasing and the reverse inequality holds in (1.1). Such curves, up to a reparametrization and an integrability condition, are the curves of maximal slope studied in [1, 15, 17, 28]. Evidently, the slope is not a lower-semicontinuous function (lsc) of its argument and hence is highly unstable. Replacing the slope $|\nabla f|$ with its lower-semicontinuous envelope in the reverse inequality of (1.1) defines *near-steepest descent curves*. See Definition 2.13 for a more precise statement.

*Received by the editors May 8, 2013; accepted for publication (in revised form) August 18, 2014; published electronically January 13, 2015.
http://www.siam.org/journals/sicon/53-1/92021.html

†Department of Mathematics, University of Washington, Seattle, WA 98195 (ddrusv@uw.edu). The work of this author was partially supported by an 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 (ioffe@math.technion.ac.il). The work of this author was supported in part by US–Israel Binational Science Foundation grant 2008261.

§School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853 (aslewis@orie.cornell.edu). This author's research was supported in part by National Science Foundation grant DMS-0806057 and by US–Israel Binational Scientific Foundation grant 2008261.

The question concerning existence of near-steepest descent curves is at the core of the subject. Roughly speaking, there are two strategies in the literature for constructing such curves for a function f. The first one revolves around minimizing f on an increasing sequence of balls around a point until the radius hits a certain threshold; then 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 and a reparametrization, yields a near-steepest descent curve [28, section 4]. The second approach is based on De Georgi's minimizing movements [15], having origins in the work of Yosida, Brézis, and Moreau (see, for example, [9] and references therein). There, 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 [1, Chapter 2]. The analysis, in both cases, is highly nontrivial and moreover does not give an intuitive meaning to the parametrization of the curve used in the construction. We note in passing that there is yet another approach based on the techniques of differential inclusions [2]. This approach, however, requires strong conditions (Lipschitzness and convexity-like properties) on f and anyway does not extend to the metric setting. We will comment further on the relationship between near-steepest descent curves and differential inclusions shortly.

In the current work, we propose an alternate transparent strategy for constructing near-steepest descent curves. The key idea of our construction is to discretize the range of f and then build a piecewise linear curve by "projecting" iterates onto successive sublevel sets. Passing to the limit as the mesh of the partition tends to zero, under reasonable conditions and a reparametrization, yields a near-steepest descent curve. Moreover, the parametrization of the curve used in the construction is entirely intuitive: the values of the function parametrize the curve. From a technical viewpoint, this type of a parametrization opens the door to the deep theory of metric regularity [19, 31], yielding a simple and geometrically transparent proof. This construction allows us to further amplify the existence theory for *semialgebraic functions*—those functions on $\mathbf{R}^n$ whose epigraph can be written as a finite union of sets, each defined by finitely many polynomial inequalities [13, 34]. We prove that semialgebraic functions always admit near-steepest descent curves bypassing all nonminimizing critical points—a result that is decisively false even for $\mathbf{C}^\infty$-smooth functions (see Example 3.4). It is worthwhile, in passing, to compare our construction with De Georgi's minimizing movements. The latter, being an implicit Euler scheme, generates iterates that are also obtained through a projection, but implicitly so, in contrast to our construction, where we explicitly choose onto which sublevel sets to project. This seemingly minor distinction is key.

The question concerning when solutions of subgradient dynamical systems and curves of near-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 an lsc function [1, 28] and hence it coincides with the limiting slope—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 and show that for semialgebraic functions that are locally Lipschitz continuous on their domains, solutions of subgradient dynamical systems are one and the same as curves of near-maximal slope. Moreover, using an argument based on the Kurdyka–Łojasiewicz inequality (KŁ-inequality), in the spirit of [6, 7, 21, 24], we show that bounded curves of near-maximal slope for semialgebraic functions have finite length. Consequently, such curves defined on maximal domains converge to critical points.

Rather than striving for maximal generality, we have tried to make the basic ideas and techniques as clear as possible. In particular, all results pertaining to semialgebraic functions remain true more generally for definable (e.g., globally subanalytic) and tame functions; see [21, 34, 35] for the relevant definitions. The outline of the manuscript is as follows. Section 2 is a short self-contained treatment of variational analysis in metric spaces. In this section, we emphasize that the slope provides a very precise way of quantifying error bounds (Lemma 2.5). In section 3 we prove that curves of near-steepest descent exist under reasonable conditions. In section 4 we consider the relationship between curves of near-steepest descent and solutions to subgradient dynamical systems in Euclidean spaces. Most of the results in this section are not new; rather, our purpose is to paint a more complete picture and to set the groundwork for section 5, where we consider descent curves for semialgebraic functions.

2. Preliminaries: Variational analysis in metric spaces. Throughout this section, we will let $(\mathcal{X}, d)$ be a complete metric space. We stress that completeness of the metric space will be essential throughout. Consider the extended real line $\overline{\mathbf{R}} := \mathbf{R} \cup \{-\infty\} \cup \{+\infty\}$. We will always assume that extended-real-valued functions are proper, meaning they are never $\{-\infty\}$ and are not always $\{+\infty\}$. For a function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$, the *domain* of f is

$$\operatorname{dom} f := \{x \in \mathcal{X} : f(x) < +\infty\},$$

and the *epigraph* of f is

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

A function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$ is lsc at $\bar{x}$ if the inequality $\liminf_{x \to \bar{x}} f(x) \geq f(\bar{x})$ holds. For a set $Q \subset \mathcal{X}$ and a point $x \in \mathcal{X}$, the *distance* of x from Q is

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

and the *metric projection* of x onto Q is

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

2.1. Slope and error bounds. A fundamental notion in local variational analysis is that of *slope*—the "fastest instantaneous rate of decrease" of a function. For more details about slope and its relevance to the theory of metric regularity, see [3,19].

DEFINITION 2.1 (slope). *Consider a function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$ and a point $\bar{x} \in \mathcal{X}$ with $f(\bar{x})$ finite. The slope of f at $\bar{x}$ is*

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

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 $(x, f(x)) \to (\bar{x}, f(\bar{x}))$.

The two slopes can easily differ. For example, for the function of one variable $f(x) = \min\{0, x\}$, we clearly have $|\nabla f|(0) = 1$, while on the other hand $\overline{|\nabla f|}(0) = 0$.

Slope allows us to define generalized critical points.

DEFINITION 2.2 (lower-critical points). *Consider a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$. We will call any point $\bar{x}$ satisfying $\overline{|\nabla f|}(\bar{x}) = 0$ a* lower-critical point *of f.*

For $\mathbf{C}^1$-smooth functions f on a Hilbert space, both $\overline{|\nabla f|}(\bar{x})$ and $|\nabla f|(\bar{x})$ simply coincide with the norm of the gradient of f at $\bar{x}$, and hence the notation. In particular, lower-critical points of such functions are critical points in the classical sense

PROPOSITION 2.3 (slope of a composition). *Consider an lsc function $f\colon [a, b] \to \overline{\mathbf{R}}$ and a nondecreasing continuous function $s\colon [c, d] \to [a, b]$. Suppose that s is differentiable at a point $t \in (c, d)$ with $s'(t) \neq 0$. Then the equality*

$$|\nabla (f \circ s)|(t) = |\nabla f|(s(t)) \cdot |s'(t)| \qquad \textit{holds.}$$

Proof. First, since s is nondecreasing and continuous and satisfies $s'(t) \neq 0$, we deduce that s is locally open near t. Taking this into account, we deduce the chain of equalities

$$\begin{aligned} |\nabla (f \circ s)|(t) &= \limsup_{\tau \to t} \frac{(f(s(t)) - f(s(\tau)))^+}{|t - \tau|} \\ &= \limsup_{\tau \to t} \frac{(f(s(t)) - f(s(\tau)))^+}{|s(t) - s(\tau)|} \cdot \frac{|s(t) - s(\tau)|}{|t - \tau|} = |\nabla f|(s(t)) \cdot |s'(t)|, \end{aligned}$$

thereby establishing the claimed result. ☐

We record below the celebrated Ekeland's variational principle.

THEOREM 2.4 (Ekeland's variational principle). *Consider an lsc function $g\colon \mathcal{X} \to \overline{\mathbf{R}}$ that is bounded from below. Suppose that for some $\epsilon > 0$ and $x \in \mathbf{R}^n$, we have $g(x) \leq \inf g + \epsilon$. Then for any $\rho > 0$, there exists a point $\bar{u}$ satisfying*

- $g(\bar{u}) \leq g(x)$,
- $d(\bar{u}, x) \leq \rho^{-1}\epsilon$, *and*
- $g(u) + \rho\, d(u, \bar{u}) > g(\bar{u})$ *for all* $u \in \mathcal{X} \setminus \{\bar{u}\}$.

The following consequence of Ekeland's variational principle will play a crucial role in our work [19, Basic Lemma, Chapter 1]. We provide a proof for completeness.

LEMMA 2.5 (error bound). *Consider an lsc function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$. Assume that for some point $x \in \operatorname{dom} f$, there are constants $\alpha < f(x)$ and $r, K > 0$ so that the implication*

$$\alpha < f(u) \leq f(x) \quad \textit{and} \quad d(u, x) \leq K \quad \Longrightarrow \quad |\nabla f|(u) \geq r \ \textit{holds.}$$

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

Proof. Define an lsc function $g\colon \mathcal{X} \to \overline{\mathbf{R}}$ 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 2.4), there exists a point $\bar{u}$ satisfying

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

and

$$g(u) + \rho\, d(u, \bar{u}) \geq g(\bar{u}) \text{ for all } u.$$

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

2.2. Absolute continuity and the metric derivative. In this section, we adhere closely to the notation and the development in [1].

Definition 2.6 (absolutely continuous curves). *Consider a curve $\gamma\colon (a,b) \to \mathcal{X}$. We will say that γ is* absolutely continuous*, denoted $\gamma \in AC(a,b,\mathcal{X})$, provided that there exists an integrable function $m\colon (a,b) \to \mathbf{R}$ satisfying*

$$d(\gamma(s), \gamma(t)) \leq \int_s^t m(\tau)\, d\tau \quad \text{whenever } a < s \leq t < b. \tag{2.1}$$

Every curve $\gamma \in AC(a,b,\mathcal{X})$ is uniformly continuous. Moreover, the right and left limits of γ, denoted respectively by $\gamma(a)$ and $\gamma(b)$, exist. There is a canonical choice for the integrand appearing in the definition of absolute continuity, namely, the "metric derivative."

Definition 2.7 (metric derivative). *For any curve $\gamma\colon [a,b] \to \mathcal{X}$ and any $t \in (a,b)$, the quantity*

$$\|\dot{\gamma}(t)\| := \lim_{s \to t} \frac{d(\gamma(s), \gamma(t))}{|s-t|},$$

if it exists, is the metric derivative *of γ at t. If this limit does exist at t, then we will say that γ is* metrically differentiable *at t.*

Some comments concerning our notation are in order, since it deviates slightly from that used in the standard monograph on the subject [1]. The notation $\|\dot{\gamma}(t)\|$ is natural, since whenever γ is a differentiable curve into a Hilbert space, the metric derivative is simply the norm of its derivative. This abuse of notation should not cause confusion in what follows.

For any curve $\gamma \in AC(a,b,\mathcal{X})$, the metric derivative exists almost everywhere on (a,b). Moreover the function $t \mapsto \|\dot{\gamma}(t)\|$ is integrable on (a,b) and is an admissible integrand in inequality (2.1). In fact, as far as such integrands are concerned, the metric derivative is in a sense minimal. Namely, for any admissible integrand $m\colon (a,b) \to \mathbf{R}$ for the right-hand side of (2.1), the inequality

$$\|\dot{\gamma}(t)\| \leq m(t) \quad \text{holds for a.e. } t \in (a,b).$$

See [1, Theorem 1.1.2] for more details. We can now define the *length* of any absolutely continuous curve $\gamma \in AC(a,b,\mathcal{X})$ by the formula

$$\operatorname{length}(\gamma) := \int_a^b \|\dot{\gamma}(\tau)\|\, d\tau.$$

We adopt the following convention with respect to curve reparametrizations.

Definition 2.8 (curve reparametrization). *Consider a curve $\gamma\colon [a,b] \to \mathcal{X}$. Then any curve $\omega\colon [c,d] \to \mathcal{X}$ is a* reparametrization *of γ whenever there exists a nondecreasing absolutely continuous function $s\colon [c,d] \to [a,b]$ with $s(c) = a$, $s(d) = b$, and satisfying $\omega = \gamma \circ s$.*

Absolutely continuous curves can always be parametrized by arclength. See, for example, [1, Lemma 1.1.4] or [10, Proposition 2.5.9].

Theorem 2.9 (arclength parametrization). *Consider an absolutely continuous curve $\gamma \in AC(a,b,\mathcal{X})$, and denote its length by $L = \operatorname{length}(\gamma)$. Then there exists a* *nondecreasing absolutely continuous map* $s\colon [a,b] \to [0,L]$ *with* $s(a)=0$ *and* $s(b)=L$ *and a* 1*-Lipschitz curve* $v\colon [0,L] \to \mathcal{X}$ *satisfying*

$$\gamma = v \circ s \quad \text{ and } \quad \|\dot{v}\| = 1 \text{ a.e. in } [0,L].$$

Proposition 2.10 (metric derivative of a composition). *Consider a curve* $\gamma\colon [0,L] \to \mathcal{X}$ *and a continuous function* $s\colon [a,b] \to [0,L]$. *Consider a point* $t \in (a,b)$, *so that* s *is differentiable at* t *and* γ *is metrically differentiable at* $s(t)$. *Then the curve* $\gamma \circ s$ *is metrically differentiable at* t *with metric derivative* $\|\dot{\gamma}(s(t))\| \cdot |s'(t)|$.

Proof. Observe that for any sequence of points $t_i \to t$ with $t_i \neq t$ and $s(t_i) = s(t)$ for each index i, we have

$$\lim_{i\to\infty} \frac{d(\gamma(s(t_i)), \gamma(s(t)))}{|t_i - t|} = \|\dot{\gamma}(s(t))\| \cdot |s'(t)|.$$

On the other hand, for any sequence $t_i \to t$ with $t_i \neq t$ and $s(t_i) \neq s(t)$ for each index i, we have

$$\lim_{i\to\infty} \frac{d(\gamma(s(t_i)), \gamma(s(t)))}{|t_i - t|} = \lim_{i\to\infty} \frac{d(\gamma(s(t_i)), \gamma(s(t)))}{|s(t_i) - s(t)|} \cdot \frac{|s(t_i) - s(t)|}{|t_i - t|} = \|\dot{\gamma}(s(t))\| \cdot |s'(t)|.$$

The result follows. □

The following is a Sard type theorem for real-valued functions of one variable. See, for example, [36, Fundamental Lemma].

Theorem 2.11 (Sard theorem for functions of one variable). *For any function* $s\colon [a,b] \to \mathbf{R}$, *the set*

$$\{t \in \mathbf{R} : \exists \tau \in s^{-1}(t) \text{ with } s'(\tau) = 0\}$$

is Lebesgue null.

The following theorem provides a convenient way of determining when strictly monotone, continuous functions are absolutely continuous [25].

Theorem 2.12 (inverses of absolutely continuous functions). *Consider a continuous, strictly increasing function* $s\colon [a,b] \to \mathbf{R}$. *Then the inverse* $s^{-1}\colon [s(a), s(b)] \to [a,b]$ *is absolutely continuous if and only if the set*

$$E := \{t \in [a,b] : s'(t) = 0\}$$

has Lebesgue measure zero.

2.3. Steepest descent curves. In this section we consider steepest descent curves in a purely metric setting. To this end, consider an lsc function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$ and a 1-Lipschitz continuous curve $\gamma\colon [a,b] \to \mathcal{X}$. There are two intuitive requirements that we would like γ to satisfy in order to be called a steepest descent curve:

1. The composition $f \circ \gamma$ is nonincreasing on a full-measure subset of $[a,b]$.
2. The instantaneous rate of decrease of $f \circ \gamma$ is almost always as great as possible.

To elaborate on the latter requirement, suppose that the composition $f \circ \gamma$ is indeed nonincreasing on a full-measure subset of $[a,b]$. Then there exists a nonincreasing function $\phi\colon [a,b] \to \overline{\mathbf{R}}$ coinciding almost everywhere with $f \circ \gamma$. Note that in particular, whenever $f \circ \gamma$ is continuous, we may simply take $\phi := f \circ \gamma$. Now taking into account that γ is 1-Lipschitz continuous and that monotone functions are differentiable a.e., one can readily verify

$$|\phi'(t)| \leq |\nabla f|(\gamma(t)) \qquad \text{for a.e. } t \in [a,b]. \tag{2.2}$$

Requiring the reverse inequality to hold amounts to forcing the curve to achieve fastest instantaneous rate of decrease. The discussion above motivates the following definition.

Definition 2.13 (near-steepest descent curves). *Consider an lsc function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$. Then a 1-Lipschitz curve $\gamma\colon [a,b] \to \mathcal{X}$ is a* steepest descent curve *if $f \circ \gamma$ coincides a.e. with some nonincreasing function $\phi\colon [a,b] \to \mathcal{X}$ and the inequality*

$$|\phi'(t)| \geq |\nabla f|(\gamma(t)) \qquad \textit{holds a.e. on } [a,b].$$

If instead the weaker inequality

$$|\phi'(t)| \geq \overline{|\nabla f|}(\gamma(t)) \qquad \textit{holds a.e. on } [a,b],$$

then we will say that γ is a near-steepest descent curve.

Remark 2.14. It is easy to see that the defining inequalities of steepest and near-steepest descent curves are independent of the particular function $\phi\colon [a,b] \to \mathcal{X}$. Namely, if those inequalities hold for some nondecreasing function agreeing a.e. with $f \circ \gamma$, then they hold for any other function agreeing a.e with $f \circ \gamma$.

In principle, near-steepest descent curves may fall short of achieving true "steepest descent" since the analogue of inequality (2.2) for the limiting slope may fail to hold in general. Our work, however, will revolve around near-steepest descent curves since the limiting slope is a much better behaved object, and anyway this is common practice in the literature (see, for example, [1, 17, 28]). The following example illustrates the difference between the two notions.

Example 2.15 (steepest descent versus near-steepest descent). Consider the function $f\colon \mathbf{R}^2 \to \mathbf{R}$ defined by $f(x,y) := -x + \min(y,0)$. Then the curve $x(t) = (t,0)$ is a near-steepest descent curve but is not a steepest descent curve, as one can easily verify.

It is often convenient to reparametrize near-steepest descent curves so that their speed is given by the limiting slope. This motivates the following companion notion; related concepts appear in [1, section 1.3], [14, 28].

Definition 2.16 (curve of near-maximal slope). *Consider an lsc function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$. A curve $\gamma\colon [a,b] \to \mathcal{X}$ is a* curve of near-maximal slope *if the following conditions hold:*

(a) *γ is absolutely continuous,*

(b) *$\|\dot{\gamma}(t)\| = \overline{|\nabla f|}(\gamma(t))$ a.e. on $[a,b]$,*

(c) *$f \circ \gamma$ coincides a.e. with some nonincreasing function $\phi\colon [a,b] \to \mathcal{X}$ with*

$$\phi'(t) \leq -\big(\overline{|\nabla f|}(\gamma(t))\big)^2 \qquad \textit{a.e. on } [a,b].$$

The following proposition shows that, as alluded to above, near-steepest descent curves and curves of near-maximal slope are the same up to reparametrization, provided that a minor integrability condition is satisfied.

Proposition 2.17 (curves of near-steepest descent and near-maximal slope). *Let $f\colon \mathcal{X} \to \overline{\mathbf{R}}$ be an lsc function.*

Consider a near-steepest descent curve $\gamma\colon [a,b] \to \mathcal{X}$ satisfying $\|\dot{\gamma}(t)\| = 1$ a.e. on $[a,b]$. If $(\overline{|\nabla f|} \circ \gamma)^{-1}$ is integrable, then there exists a reparametrization of γ that is a curve of near-maximal slope.

Conversely, consider a curve of near-maximal slope $\gamma\colon [a,b] \to \mathcal{X}$, having finite length. Then there exists a reparametrization of γ that is a near-steepest descent curve satisfying $\|\dot{\gamma}(t)\| = 1$ a.e. on $[a,b]$.

Proof. To see the validity of the first claim, let $\eta := \int_a^b \frac{1}{\overline{|\nabla f|}(\gamma(r))}\, dr$ and define the function $s\colon [a,b] \to [0,\eta]$ by setting

$$s(t) := \int_a^t \frac{1}{\overline{|\nabla f|}(\gamma(r))}\, dr.$$

Then s is a strictly increasing, absolutely continuous function. Moreover, by Theorem 2.12, the inverse $s^{-1} : [0,\eta] \to [a,b]$ is absolutely continuous as well. Define now the function $\omega\colon [0,\eta] \to [a,b]$ by setting $\omega(\tau) := \gamma(s^{-1}(\tau))$. Clearly ω is absolutely continuous, and using Propositions 2.10 and 2.3 we immediately deduce that ω is a curve of near-maximal slope. The converse claim follows by similar means. ∎

Our focus in the current work is on existence of curves of near-steepest descent. As a byproduct of our arguments, the curves that we construct will in addition have certain regularity properties. To facilitate the exposition, we introduce the following notation.

DEFINITION 2.18 (reliable descent). *Given a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, we will say that there exists* reliable descent *from a point $\bar{x} \in \operatorname{dom} f$ if there exists a near-steepest descent curve $\gamma\colon [0,L] \to \mathcal{X}$, for some $L > 0$, emanating from $\bar{x}$ such that on a full-measure subset of $[0,L]$, the curve γ is moving at unit speed, the slope $\overline{|\nabla f|}(\gamma)$ is finite, and $f \circ \gamma$ is strictly decreasing. Such a near-steepest descent curve will be called a* reliable descent curve*.*

3. Existence of descent curves. In this section, we provide a natural and transparent existence proof for near-steepest descent curves in complete locally convex metric spaces. We begin with a few relevant definitions, adhering closely to the notation of [23].

DEFINITION 3.1 (metric segments). *A subset S of a metric space $\mathcal{X}$ is a* metric segment *between two points x and y in $\mathcal{X}$ if there exists a closed interval $[a,b]$ and an isometry $\omega\colon [a,b] \to \mathcal{X}$ satisfying $\omega([a,b]) = S$, $\omega(a) = x$, and $\omega(b) = y$.*

DEFINITION 3.2 (convex metric spaces). *We will say that $\mathcal{X}$ is a* convex metric space *if for any distinct points $x, y \in \mathcal{X}$ there exists a metric segment between them. We will call $\mathcal{X}$ a* locally convex metric space *if each point in $\mathcal{X}$ admits a neighborhood that is a convex metric space in the induced metric.*

Some notable examples of locally convex metric spaces are complete Riemannian manifolds and, more generally, length spaces that are complete and locally compact (see the Hopf–Rinow theorem). For more examples, we refer the reader to [18].

We now introduce the following very weak continuity condition, which has been essential in the study of descent curves in metric spaces; see, e.g., [1, Theorem 2.3.1].

DEFINITION 3.3 (continuity on slope-bounded sets). *Consider a function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$. We will say that f is* continuous on slope-bounded sets *provided that for any point $\bar{x} \in \operatorname{dom} f$ we have the implication*

$$x_i \to \bar{x} \quad \textit{with} \quad \sup_{i\in\mathbb{N}}\{|\nabla f|(x_i), f(x_i)\} < \infty \qquad \Longrightarrow \qquad f(x_i) \to f(\bar{x}).$$

We now arrive at the main result of this section. To make the reading easier, some comments are in order. To this end, consider a function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$ and a point $\bar{x}$. If the point $\bar{x}$ is lower-critical for f, then the constant curve $x(t) \equiv \bar{x}$ is trivially a near-steepest descent curve. On the other hand, if $\bar{x}$ is not a local minimizer, we can hope for nontrivial near-steepest descent curves emanating from $\bar{x}$. The situation is interesting even when f is a $\mathbf{C}^\infty$-smooth function on $\mathbf{R}^n$. One special case stands out.

If f is a Morse function at $\bar{x}$, that is, the Hessian $\nabla^2 f(\bar{x})$ is nonsingular, then it is easy to see that nontrivial steepest descent curves do exist. Indeed, due to the celebrated Morse lemma one can obtain nontrivial descent by following (in the local coordinate system) an eigenvector corresponding to a negative eigenvalue of the Hessian $\nabla^2 f(\bar{x})$. For general $\mathbf{C}^\infty$-smooth functions, possibly having a degenerate Hessian at the point of criticality, the answer is decisively false!

Example 3.4 (nontrivial descent curves of smooth functions may fail to exist). Consider the function $f\colon \mathbf{R} \to \mathbf{R}$ defined by $f(x) := e^{-\frac{1}{|x|}} \sin(\frac{1}{|x|})$. It is easy to see that f is $\mathbf{C}^\infty$-smooth and that it does not admit any nontrivial near-steepest descent curve emanating from $\bar{x} = 0$, even though $\bar{x}$ is not a local minimizer.

The following theorem shows that a nontrivial near-steepest descent curve emanating from $\bar{x}$ will exist provided that $\bar{x}$ is not a local minimizer of f and that the slope $|\nabla f|$ is uniformly bounded away from zero on the set $U \cap [f < f(\bar{x})]$, where U is some neighborhood of $\bar{x}$. Clearly the latter condition holds in the case $\overline{|\nabla f|}(\bar{x}) > 0$. The seemingly slight generality we obtain, on the other hand, will pay important dividends in subsection 5.2 in connection with the KŁ-inequality.

We should note that in the following theorem we will suppose tough compactness assumptions relative to the metric topology. As is now standard, such compactness assumptions can be sidestepped by instead introducing weaker topologies [1, section 2.1]. On the other hand, following this route would take us far afield and would lead to technical details that may obscure the main proof ideas for the reader. Hence we do not dwell on this issue further. We have, however, designed our proof so as to make such an extension as easy as possible for interested readers.

THEOREM 3.5 (existence of near-steepest descent curves). *Consider an lsc function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$ on a complete locally convex metric space $\mathcal{X}$, along with a point $\bar{x} \in \operatorname{dom} f$ which is not a local minimizer of f. Suppose that the following are true:*

1. *f is continuous on slope-bounded sets and bounded closed subsets of sublevel sets of f are compact.*
2. *There exists a neighborhood U of $\bar{x}$ so that the slope $\overline{|\nabla f|}$ is uniformly bounded away from zero on the set $U \cap [f < f(\bar{x})]$.*

Then there exists reliable descent from $\bar{x}$.

Proof. First, by restricting attention to a sufficiently small neighborhood of $\bar{x}$ we can clearly assume that $\mathcal{X}$ is a convex metric space. Using property 2 and lsc of f, we deduce that there exist constants $\alpha < f(\bar{x})$ and $r, R > 0$, so that the implication

$$\alpha < f(u) < f(\bar{x}) \quad \text{and} \quad d(u, \bar{x}) < R \quad \Longrightarrow \quad |\nabla f|(u) \geq r$$

holds for any point $u \in \mathcal{X}$. Define $\eta := f(\bar{x}) - \alpha$ and $C > 0$ to be slightly smaller than R. Increasing α we may enforce the inequality $\eta < rC$. Let $0 = \tau_0 < \tau_1 < \cdots < \tau_k = \eta$ be a partition of $[0, \eta]$ into k equal parts. We will adopt the notation

$$\lambda := \frac{\tau_{i+1} - \tau_i}{\eta}, \qquad \alpha_i = f(\bar{x}) - \tau_i, \qquad L_i := [f \leq \alpha_i].$$

With this partition, we will associate a certain curve $u_k(\tau)$ for $\tau \in [0, \eta]$, naturally obtained by concatenating metric segments between points x_i, x_{i+1} lying on consecutive sublevel sets. See Figure 1 for an illustration. For notational convenience, we will often suppress the index k in $u_k(\tau)$. The construction is as follows. Set $x_0 := \bar{x}$ and let x_1 be any point of L_1 satisfying $d(\bar{x}, L_1) \leq d(x_1, \bar{x}) + \frac{1}{k}$. Observe that since $\bar{x}$ is not a local minimizer of f, the point x_1 is well-defined for all sufficiently large indices k. We will assume throughout the proof that k is chosen sufficiently large for

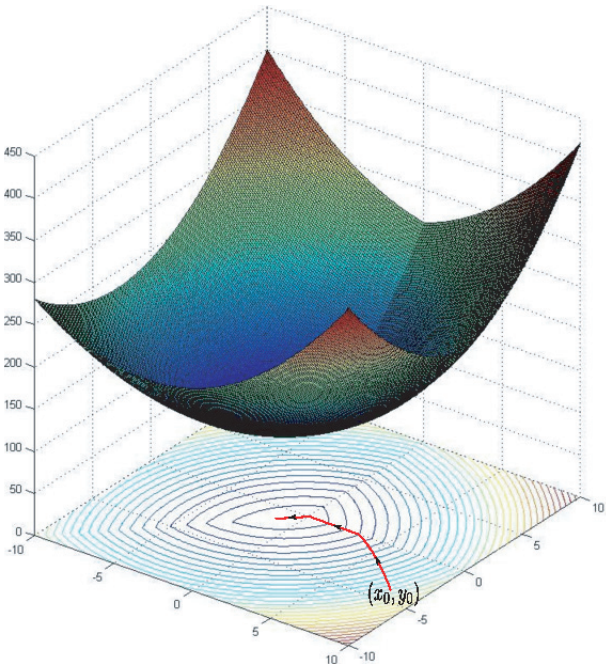

FIG. 1. $f(x, y) = \max\{x + y,\, |x - y|\} + x(x + 1) + y(y + 1) + 100$.

this to be the case. We will now inductively define x_{j+1}. It is important, however, to keep in mind that the point x_1 is rather exceptional, since we assume nothing about the slope of f exactly at $\bar{x}$. Nevertheless, define $\rho := d(x_0, x_1)$ and observe that since $\bar{x}$ is not a local minimizer of f, the distance ρ tends to zero monotonically as k tends to infinity. We will assume throughout that k is sufficiently large to ensure $\rho < C - r^{-1}\eta$, $C + \rho < R$, and $\rho < (1 - \lambda)C$.

We now proceed with the inductive definition of x_{j+1}. To this end, suppose that we have defined points x_i for $i = 1, \ldots, j$. Consider the quantity

$$r_j := \inf\{|\nabla f|(y) : \ \alpha_{j+1} < f(y) \le f(x_j),\ d(y, x_j) < \lambda C\}$$

and let x_{j+1} be any point satisfying

$$x_{j+1} \in L_{j+1} \qquad \text{and} \qquad d(x_{j+1}, x_j) \le r_j^{-1}(f(x_j) - \alpha_{j+1})^+.$$

(In our setting, due to the compactness of bounded closed subsets of sublevel sets of f, we may simply define x_{j+1} to be any closest point of L_{j+1} to x_j.)

CLAIM 3.6 (well-definedness). *For all indices $i = 1, \ldots, k$, the points x_i are well-defined and satisfy*

$$d(x_{i+1}, x_i) \le r_i^{-1}(\tau_{i+1} - \tau_i) \tag{3.1}$$

and

$$r_i \ge r, \quad d(x_i, \bar{x}) \le r^{-1}\tau_i + \rho. \tag{3.2}$$

Proof. The proof proceeds by induction. First, observe that due to our choice of how large to make k, the point x_1 is well-defined and inequalities (3.2) hold for

$i = 1$. Proceeding to the inductive step, suppose that the points x_i are well-defined for indices $i = 1, \ldots, j$, the inequalities (3.2) are valid for $i = 1, \ldots, j$, and the inequality (3.1) is valid for indices i satisfying $1 \le i \le j - 1$.

Observe if the inequality $f(x_j) \le \alpha_{j+1}$ were true, then we may set $x_{j+1} := x_j$ and the inductive step would be true trivially. Hence suppose otherwise. We claim that the conditions of Lemma 2.5 are satisfied with $x = x_j$, $\alpha = \alpha_{j+1}$, $K = \lambda C$, and with r_j in place of r.

To this end, we show the following:

- $f(x_j) - \alpha_{j+1} \le \lambda r_j C$;
- $\alpha_{j+1} < f(y) \le f(x_j)$ and $d(y, x_j) \le \lambda C \quad \Longrightarrow \quad |\nabla f|(y) \ge r_j$.

Observe

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

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

Applying Lemma 2.5, we conclude that the point x_{j+1} is well-defined and the inequality $d(x_{j+1}, x_j) \le r_j^{-1}(f(x_j) - \alpha_{j+1})^+ \le r_j^{-1}(\tau_{j+1} - \tau_j)$ holds. Consequently, we obtain

$$d(x_{j+1}, \bar{x}) \le d(x_{j+1}, x_j) + d(x_j, \bar{x}) \le r_j^{-1}(\tau_{j+1} - \tau_j) + r^{-1}\tau_j + \rho \le r^{-1}\tau_{j+1} + \rho.$$

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

$$d(y, \bar{x}) \le d(y, x_{j+1}) + d(x_{j+1}, \bar{x}) \le \frac{\tau_{j+2} - \tau_{j+1}}{\eta}C + \frac{\tau_{j+1}}{r} + \rho < \frac{\tau_{j+2}}{\eta}C + \rho < R.$$

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

For each index $i = 0, \ldots, k - 1$, let $\omega_i \colon [0, d(x_i, x_{i+1})] \to \mathcal{X}$ be the isometry parameterizing the metric segment between x_i and x_{i+1}. For reasons which will become apparent momentarily, we now rescale the domain of ω_i by instead declaring

$$\omega_i \colon [\tau_i, \tau_{i+1}] \to \mathcal{X} \quad \text{ to be } \quad \omega_i(t) = \omega_i\Big(\frac{d(x_{i+1}, x_i)}{\tau_{i+1} - \tau_i}(t - \tau_i)\Big).$$

Observe now that for any index $i = 1, \ldots, k - 1$ and any $s, t \in [\tau_i, \tau_{i+1}]$ with $s < t$, we have

$$d(\omega_i(t), \omega_i(s)) = \frac{d(x_{i+1}, x_i)}{\tau_{i+1} - \tau_i}(t - s) \le r_i^{-1}(t - s). \tag{3.3}$$

It follows that all the curves ω_i, for $i = 1, \ldots, k - 1$, are Lipschitz continuous with a uniform modulus r^{-1}. We may now define a curve $u_k \colon [0, \eta] \to \mathcal{X}$ by simply concatenating the domains of ω_i for each index $i = 0, \ldots, k - 1$. Each such curve u_k is Lipschitz continuous on $[\rho_k, \eta]$ with modulus r^{-1} and the sequence of curves $\{u_k\}$ is uniformly bounded. Moreover, since f is lsc and bounded subsets of sublevel sets of f are compact, one can readily verify that the sequence $\{u_k\}$ is pointwise relatively compact. The well-known theorem of Arzelà and Ascoli [22, section 7] then guarantees that a certain subsequence of u_k, which we continue to denote by u_k, converges uniformly on compact subsets of $(0,\eta]$ to some mapping $x\colon (0,\eta] \to \mathcal{X}$. We can then clearly extend the domain of x to the closed interval $[0,\eta]$ by declaring $x(0) := \bar{x}$. It follows that $x\colon [0,\eta] \to \mathcal{X}$ is Lipschitz continuous with modulus r^{-1}.

Observe that the metric derivative functions $\|\dot{u}_k(\cdot)\|$ are bounded in $L^2(a,\eta)$ for any $a \in (0,\eta)$. It follows that, up to a subsequence, the mappings $\|\dot{u}_k(\cdot)\|$ converge weakly in $L^2(0,\eta)$ to some integrable mapping $m\colon [0,\eta] \to \mathbf{R}$ satisfying

$$d(x(s),x(t)) \le \int_s^t m(\tau)\, d\tau \quad \text{ whenever } 0 < s \le t < \eta. \tag{3.4}$$

For what follows now, define the set of breakpoints

$$E := \bigcup_{k\in\mathbb{N}} \bigcup_{i\in\mathbb{N}\cap[0,k]} \left\{ \frac{i\lambda_k}{\eta} \right\}$$

and observe that it has zero measure in $[0,\eta]$. In addition, let D be the full-measure subset of $(0,\eta)$ on which all the curves u_k and x admit a metric derivative.

Claim 3.7. *For almost every $\tau \in [0,\eta]$ with $\|\dot{x}(\tau)\| \neq 0$, the following are true:*

- $f(x(\tau)) = f(\bar{x}) - \tau$,
- $\|\dot{x}(\tau)\| \le \frac{1}{\overline{|\nabla f|}(x(\tau))}$.

Proof. Fix a real $\tau \in D \setminus E$ with $\|\dot{x}(\tau)\| \neq 0$. Then using (3.3) we deduce that for sufficiently large k, we have

$$\|\dot{u}_k(\tau)\| \le \frac{1}{r_{i_k}^{(k)}} \tag{3.5}$$

for some $i_k \in \{0,\dots,k\}$, where the superscript (k) refers to partition of the interval $[0,\eta]$ into k equal pieces. Noting that weak convergence does not increase the norm and using minimality of the metric derivative, we deduce

$$\liminf_{k\to\infty} \|\dot{u}_k(\tau)\| \ge m(\tau) \ge \|\dot{x}(\tau)\| \quad \text{ for a.e. } \tau \in [0,\eta]. \tag{3.6}$$

Consequently there exists a subsequence of $\|\dot{u}_k(\tau)\|$, which we continue to denote by $\|\dot{u}_k(\tau)\|$, satisfying $\lim_{k\to\infty} \|\dot{u}_k(\tau)\| \neq 0$. Taking into account (3.5), we deduce that $r_{i_k}^{(k)}$ remain bounded. We may then choose points $x_{i_k}^{(k)}$, y_k, and reals λ_k, $\tau_{i_k}^{(k)}$ with $\tau \in (\tau_{i_k}^{(k)}, \tau_{i_k+1}^{(k)})$ satisfying

$$d(y_k, x_{i_k}^{(k)}) < \lambda_k C, \qquad f(\bar{x}) - \tau_{i_k+1}^{(k)} < f(y_k) \le f(x_{i_k}^{(k)}), \quad x_{i_k}^{(k)} \to x(\tau), \quad \tau_{i_k}^{(k)} \to \tau,$$

and $|\nabla f|(y_k) \le r_{i_k}^{(k)} + \frac{1}{k}$. Then since f is continuous on slope-bounded sets and the quantity $f(x_{i_k}^{(k)}) - (f(\bar{x}) - \tau_{i_k+1}^{(k)})$ tends to zero, we deduce

$$f(x(\tau)) = \lim_{k\to\infty} f(y_k) = \lim_{k\to\infty} f(\bar{x}) - \tau_{i_k+1}^{(k)} = f(\bar{x}) - \tau,$$

as claimed. Moreover

$$\liminf_{k\to\infty} r_{i_k}^{(k)} \ge \liminf_{k\to\infty} \left\{ |\nabla f|(y_k) - \frac{1}{k} \right\} \ge \overline{|\nabla f|}(x(\tau)).$$

Combining this with (3.6) and taking the limit in (3.5), we obtain

$$\|\dot{x}(\tau)\| \le \frac{1}{\overline{|\nabla f|}(x(\tau))},$$

as claimed. ∎

Define now a strictly decreasing function $\phi\colon [0,\eta] \to \mathbf{R}$ by the formula $\phi(\tau) = f(\bar{x}) - \tau$. In particular, it follows from Claim 3.7 that ϕ coincides with $f \circ x$ almost everywhere on the set $\{\tau \in (0,\eta) : \|\dot{x}(\tau)\| \neq 0\}$. Moreover, for almost every $\tau \in (0,\eta)$ the implication

$$\|\dot{x}(\tau)\| \neq 0 \quad \Longrightarrow \quad \overline{|\nabla f|}(x(\tau)) < \infty \quad \text{and} \quad |\phi'(\tau)| \geq \overline{|\nabla f|}(x(\tau)) \cdot \|\dot{x}(\tau)\|$$

holds.

Now in light of Theorem 2.9, there exists a nondecreasing absolutely continuous map $s\colon [a,b] \to [0,L]$ with $s(a) = 0$ and $s(b) = L$ and a 1-Lipschitz curve $\gamma\colon [0,L] \to \mathcal{X}$ satisfying

$$x(\tau) = (\gamma \circ s)(\tau) \quad \text{and} \quad \|\dot{\gamma}(t)\| = 1 \text{ for a.e. } t \in [0,L].$$

Define also the nondecreasing function $\tau\colon [0,L] \to [0,\eta]$ by setting

$$\tau(t) := \min\{\tau : s(\tau) = t\},$$

and observe the equality $(s \circ \tau)(t) = t$. Define now the nonincreasing function $\psi\colon [0,L] \to [0,\eta]$ by setting $\psi(t) := \phi(\tau(t))$.

It is easy to check that if s is differentiable at $\tau(t)$ with $s'(\tau(t)) \neq 0$, the function $\tau(\cdot)$ is continuous at t, and γ is metrically differentiable at t with $\|\dot{\gamma}(t)\| = 1$, then we have

$$\|\dot{x}(\tau(t))\| = s'(\tau(t)) \qquad \text{and} \qquad \tau'(t) = \frac{1}{s'(\tau(t))},$$

and consequently

$$\overline{|\nabla f|}(\gamma(t)) = \overline{|\nabla f|}(x(\tau(t))) \leq |\phi'(\tau(t))| \cdot \frac{1}{s'(\tau(t))} = |\psi'(t)|.$$

It easily follows (in part using Theorem 2.11) that the collection of such real numbers t has full measure in $[0,L]$ and that ψ and $f \circ \gamma$ coincide a.e. on $[0,L]$. This completes the proof. $\square$

Remark 3.8 (strict descent under continuity). It is easy to see that if the function f of Theorem 3.5 is continuous on its domain, then the near-steepest descent curve γ constructed in that theorem has the property that the composition $f \circ \gamma$ is strictly decreasing.

The following is now an easy consequence.

COROLLARY 3.9 (existence of curves of near-maximal slope). *Consider an lsc function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$ on a complete locally convex metric space $\mathcal{X}$ and a point $\bar{x} \in \mathcal{X}$, with f finite at $\bar{x}$. Suppose that f is continuous on slope-bounded sets and that bounded closed subsets of sublevel sets of f are compact. Then there exists a curve of near-maximal slope $\gamma\colon [0,T] \to \mathcal{X}$, for some $T > 0$, starting at $\bar{x}$.*

Proof. If $\bar{x}$ is a lower-critical point of f, then the constant curve $x(t) \equiv \bar{x}$ is a curve of near-maximal slope. Hence we may suppose that $\bar{x}$ is not a lower-critical point of f. The result is now immediate from Proposition 2.17 and Theorem 3.5. $\square$

Remark 3.10 (steepest descent versus near-steepest descent). In Example 2.15 we illustrated a near-steepest descent curve that fails to be a steepest descent curve. On the other hand, the construction used in Theorem 3.5 would not produce such a curve. This leads us to conjecture that our construction would always yield a true steepest descent curve at least when applied to semialgebraic functions. This class of functions is the focal point of section 5.

4. Descent curves and subgradient dynamical systems. In this section, we compare curves of near maximal slope to a more classical idea—solutions of subgradient dynamical systems. To do so we recall a notion of a generalized gradient, which in principle makes sense in Hilbert spaces. However, for the sake of simplicity we stay within the setting of Euclidean spaces throughout the section. Some of the arguments presented are not new. Rather we include them to paint a more complete picture for the reader. Moreover, this section will set up some of the groundwork and motivation for parts of section 5, which in contrast are entirely new.

4.1. Some elements of variational analysis. 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 and Zhu [8], Clarke et al. [11], Mordukhovich [29], Penot [30], and Rockafellar and Wets [32], and to the survey of Ioffe [19], for more details. Unless otherwise stated, we follow the terminology and notation of [19] and [32].

Throughout this section, we will consider a real Euclidean space $\mathbf{R}^n$ with inner product $\langle \cdot, \cdot \rangle$. The symbol $\|\cdot\|$ will denote the corresponding norm on $\mathbf{R}^n$. 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 $\operatorname{cl} Q$, $\operatorname{conv} Q$, $\operatorname{cone} Q$, and $\operatorname{aff} Q$ will denote the topological closure, the convex hull, the (nonconvex) conical hull, and the affine span of Q, respectively. The symbol $\operatorname{par} Q$ will denote the parallel subspace of Q, namely, the set $\operatorname{par} Q := \operatorname{aff} Q - \operatorname{aff} Q$. 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 4.1 (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 $(x_i, f(x_i), v_i)$ converging 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 $v_i \in \hat{\partial} f(x_i)$, so that $(x_i, f(x_i), \tau_i v_i)$ converge 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 $\bar{x}$ such that $f(\bar{x})$ is not finite, we follow the convention that $\hat{\partial} f(\bar{x}) = \partial f(\bar{x}) = \partial^\infty f(\bar{x}) = \partial_c f(\bar{x}) = \emptyset$.

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, an lsc function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is locally Lipschitz continuous around $\bar{x}$ if and only if we have $\partial^\infty f(\bar{x}) = \{0\}$.

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 4.2 (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 4.3 (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 4.4 (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* 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 [32, Corollary 8.11]. Shortly, we will need the following result describing normals to sublevel sets [32, 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 4.5 (normals to sublevel sets). *Consider an 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*

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

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

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

We first show the implication

$$(x^*, 0) \in N_{Q_{\bar{\alpha}}}(\bar{x}, \bar{\alpha}) \implies x^* \in (\operatorname{cone} \partial f(\bar{x})) \bigcup \partial^\infty f(\bar{x}). \tag{4.1}$$

Indeed, consider a vector $(x^*, 0) \in N_{Q_{\bar{\alpha}}}(\bar{x}, \bar{\alpha})$. Then fuzzy calculus [19, Chapter 2.1] implies that there are sequences $(x_{1k}, \alpha_{1k}) \in \operatorname{epi} f$, $(x_{1k}^*, \beta_{1k}) \in \hat{N}_{\operatorname{epi} f}(x_{1k}, \alpha_{1k})$, $(x_{2k}, \alpha_{2k}) \in \mathcal{L}_{\bar{\alpha}}$, 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 the nature of epigraphs we have $\beta_{1k} \le 0$. If up to a subsequence we had $\beta_{1k} = 0$, then (4.1) 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 (4.1) follows.

Now consider a vector $u^* \in \hat{N}_{[f \le \bar{\alpha}]}(u)$ for some $u \in [f \le \bar{\alpha}]$. Consequently the inequality $\langle u^*, h\rangle \le o(\|h\|)$ holds whenever h satisfies $f(u+h) \le \bar{\alpha}$. The latter in turn implies $(u^*, 0) \in \hat{N}_{Q_{\bar{\alpha}}}(u, \bar{\alpha})$. Together with (4.1), taking limits of Fréchet subgradients and applying (4.1) completes the proof. $\square$

The following result, which follows from the proofs of [19, Propositions 1 and 2, Chapter 3], establishes an elegant relationship between the slope and subdifferentials.

Proposition 4.6 (slope and subdifferentials). *Consider an 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 $|\nabla f|(\bar{x}) \le d(0, \hat{\partial} f(\bar{x}))$, and furthermore equality*

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

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

Proof. The inequality $|\nabla f|(\bar{x}) \le d(0, \hat{\partial} f(\bar{x}))$ 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 $g(u) \ge f(x)$. We deduce (see, e.g., [32, Exercise 10.10])

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

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

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 $\|v_i\| \ge |\nabla f|(x_i)$. Letting i tend to infinity, the result follows. $\square$

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. We conclude this subsection with the following standard result of linear algebra.

Lemma 4.7 (result in linear algebra). *Consider a subspace V of $\mathbf{R}^n$. Then for any vector $b \in \mathbf{R}^n$, the equations*

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

Proof. Observe $b = P_V(b) + P_{V^{\perp}}(b)$, and consequently the inclusion

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

The reverse inclusion follows from the trivial computation

$$z \in (b + V^{\perp}) \cap V \Longrightarrow z - P_V(b) \in V \cap V^{\perp} \Longrightarrow z = P_V(b).$$

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

$$\min_{z \in b + V^{\perp}} \|z\|^2$$

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

4.2. Main results. In this section, we consider curves of near-maximal slope in Euclidean spaces. In this context, it is interesting to compare such curves to solutions $x\colon [0,\eta] \to \mathbf{R}^n$ of subgradient dynamical systems

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

It turns out that the same construction as in the proof of Theorem 3.5 shows that there exist near-steepest descent curves x so that essentially, up to rescaling, the vector $\dot{x}(t)$ lies in $-\partial f(x(t))$ for a.e. $t \in [0,\eta]$.

THEOREM 4.8 (existence of near-steepest descent curves II). *Consider an lsc function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, along with a point $\bar{x}$ in the domain of f. Suppose that f is continuous on slope bounded sets. Then there exist $L > 0$ and a curve of near-maximal slope $x\colon [0,L] \to \mathcal{X}$ emanating from $\bar{x}$ and satisfying*

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

Proof. We can clearly assume that zero is not a subgradient of f at $\bar{x}$. We now specialize the construction of Theorem 3.5 to the Euclidean setting. Namely, let u_k be defined as in that theorem, except in this case define $u_k(0) = \bar{x}$, 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. In particular, up to a subsequence, u_k converge uniformly on compact subsets to a Lipschitz continuous curve x.

Observe that in light of Proposition 4.5, for any index k and any $\tau \in [\tau_i, \tau_{i+1}]$ (for $i = 1,\ldots,k$) we have $\dot{u}_k(\tau) \in -(\mathrm{cone}\,\partial f(u_k(\tau_{i+1}))) \cup \partial^\infty f(u_k(\tau_{i+1}))$. 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 sequence 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

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

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. After the reparametrization performed in the proof of Theorem 3.5, the curve γ is subdifferentiable almost everywhere on $[0,L]$ and consequently satisfies

$$\dot{\gamma}(t) \in -\mathrm{cl}\,\mathrm{cone}\,\partial_c f(\gamma(t)) \quad \text{for a.e. } t \in [0,L],$$

as we needed to show. $\square$

The above theorem motivates the question of when curves of near-maximal slope and solutions of subgradient dynamical systems are one and the same, that is, when the rescaling of the gradient $\dot{x}$ in the previous theorem is not needed. The following property turns out to be crucial.

Definition 4.9 (chain rule). *Consider an lsc function* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$. *We say that* $\overline{|\nabla f|}$ admits a chain rule *if for every curve* $x \in AC(a,b,\mathbf{R}^n)$ *for which the composition* $f\circ x$ *is nonincreasing and* f *is subdifferentiable almost everywhere along* x, *the equation*

$$(f\circ x)'(t) = \langle \partial f(x(t)), \dot{x}(t)\rangle \qquad \textit{holds for a.e. } t\in(a,b).$$

The following simple proposition shows that whenever $\overline{|\nabla f|}$ admits a chain rule, solutions to subgradient dynamical systems and curves of near-maximal slope coincide.

Proposition 4.10 (subgradient systems and curves of near-maximal slope). *Consider an lsc function* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *and suppose that* $\overline{|\nabla f|}$ *admits a chain rule. Then for any curve* $x\in AC(a,b,\mathbf{R}^n)$ *the following are equivalent:*

1. *x is a curve of near-maximal slope.*
2. *$f\circ x$ is nonincreasing and we have*

$$\dot{x}(t) \in -\partial f(x(t)) \quad \textit{a.e. on } [a,b]. \tag{4.2}$$

3. *$f\circ x$ is nonincreasing and we have*

$$\dot{x}(t)\in -\partial f(x(t)) \quad \textit{and} \quad \|\dot{x}(t)\| = d(0,\partial f(x(t))) \quad \textit{a.e. on } [a,b]. \tag{4.3}$$

Proof. We first prove the implication $1 \Rightarrow 3$. To this end, suppose that x is a curve of near-maximal slope. Then clearly $f\circ x$ is nonincreasing and we have

$$\langle \partial f(x(t)), \dot{x}(t)\rangle = (f\circ x)'(t) \le -\big(\overline{|\nabla f|}(x(t))\big)^2 \quad \text{a.e. on } [a,b].$$

Let $v(t)\in\partial f(x(t))$ be a vector of minimal norm. Then we have

$$\langle \partial f(x(t)), \dot{x}(t)\rangle \ge -\|v(t)\|\cdot\|\dot{x}(t)\| = -\big(\overline{|\nabla f|}(x(t))\big)^2$$

with equality if and only if $\dot{x}(t)$ and $v(t)$ are collinear. We deduce $\dot{x}(t) = -v(t)$, as claimed.

The implication $3\Rightarrow 2$ is trivial. Hence we focus now on $2\Rightarrow 1$. To this end suppose that 2 holds and observe

$$\langle \partial f(x(t)), \dot{x}(t)\rangle = -\|\dot{x}(t)\|^2 \quad \text{for a.e. } t\in(a,b). \tag{4.4}$$

Given such t consider the affine subspaces

$$V = \operatorname{par}\partial f(x(t)).$$

Then we have

$$\operatorname{aff}\partial f(x(t)) = -\dot{x}(t) + V.$$

We claim now that the inclusion $\dot{x}(t)\in V^{\perp}$ holds. To see this, observe that for any real λ_i and for vectors $v_i\in\partial f(x(t))$, we have

$$\left\langle \dot{x}(t), \sum_{i=1}^{k}\lambda_i(v_i+\dot{x}(t))\right\rangle = \sum_{i=1}^{k}\lambda_i\big[\langle \dot{x}(t), v_i\rangle + \|\dot{x}(t)\|^2\big] = 0,$$

where the latter equality follows from (4.4). Hence the inclusion

$$-\dot{x}(t) \in (-\dot{x}(t) + V) \cap V^{\perp}$$

holds. Consequently, using Lemma 4.7, we deduce that $-\dot{x}(t)$ achieves the distance of the affine space, $\operatorname{aff} \partial f(x(t))$, to the origin. On the other hand, the inclusion $-\dot{x}(t) \in \partial f(x(t))$ holds, and hence $-\dot{x}(t)$ actually achieves the distance of $\partial f(x(t))$ to the origin. The result follows. □

In light of the theorem above, it is interesting to understand for which functions f the slope $\overline{|\nabla f|}$ admits a chain rule. Subdifferentially regular (in particular, all lsc convex) functions furnish a simple example. The convex case can be found in [9, Lemma 3.3, p. 73] (chain rule).

LEMMA 4.11 (chain rule under subdifferential regularity). *Consider a subdifferentially regular function* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$. *Then* $\overline{|\nabla f|}$ *admits a chain rule.*

Proof. Consider a curve $x\colon (a,b) \to \mathbf{R}^n$. Suppose that for some real $t \in (a,b)$ both x and $f \circ x$ are differentiable at t and $\partial f(x(t))$ is nonempty. We then deduce

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

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 the equation

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

and the result follows. □

Subdifferentially regular functions are very special, however. In particular, many nonpathological functions such as $-\|\cdot\|$ are not subdifferentially regular. So it is natural to consider prototypical nonpathological functions appearing often in practice—those that are semialgebraic. This is the focus of the following section.

5. Near-steepest descent curves of semialgebraic functions. A *semialgebraic* set $S \subset \mathbf{R}^n$ is a finite union of sets of the form

$$\{x \in \mathbf{R}^n : P_1(x) = 0, \ldots, P_k(x) = 0, J_1(x) < 0, \ldots, J_l(x) < 0\},$$

where $P_1, \ldots, P_k$ and $J_1, \ldots, J_l$ 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 *semialgebraic* if $\operatorname{epi} f \subset \mathbf{R}^{n+1}$ is a semialgebraic set. For an extensive discussion on semialgebraic geometry, see the monographs of Basu, Pollack, and Roy [4], van den Dries [35], and Shiota [33]. For a quick survey, see the article of van den Dries and Miller [34] and the surveys of Coste [12, 13]. Unless otherwise stated, we follow the notation of [13] and [34].

5.1. Semialgebraic descent and subgradient dynamical systems. The main goal of this subsection is to establish an equivalence between curves of near-maximal slope and solution of subgradient dynamical systems for lsc semialgebraic functions that are locally Lipschitz continuous on their domains. To this end, we analyze the chain rule for the slope in the context of semialgebraic functions. Before we proceed, we need to recall the notion of tangent cones.

DEFINITION 5.1 (tangent cone). *Consider a set $Q \subset \mathbf{R}^n$ and a point $\bar{x} \in Q$. Then the* tangent cone *to Q at $\bar{x}$ is simply the set*

$$T_Q(\bar{x}) := \Big\{ \lim_{i\to\infty} \lambda_i(x_i - \bar{x}) : \lambda_i \uparrow \infty \text{ and } x_i \to \bar{x} \text{ in } Q \Big\}.$$

We now record the following simple lemma, whose importance in the context of semialgebraic geometry will become apparent shortly. We omit the proof since it is rather standard.

LEMMA 5.2 (generic tangency). *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*

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

The following is a key property of semialgebraic functions that we will exploit [5, Proposition 4].

THEOREM 5.3 (projection formula). *Consider an lsc semialgebraic function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$. Then there exists a partition of* $\operatorname{dom} f$ *into finitely many $\mathbf{C}^1$-manifolds $\{M_i\}$ so that f restricted to each manifold M_i is $\mathbf{C}^1$-smooth. Moreover for any point x lying in a manifold M_i, the inclusion*

$$\partial_c f(x) \subset \nabla g(x) + N_{M_i}(x) \quad \textit{holds,}$$

where $g\colon \mathbf{R}^n \to \mathbf{R}$ is any $\mathbf{C}^1$-smooth function agreeing with f on a neighborhood of x in M_i.

COROLLARY 5.4 (semialgebraic chain rule for the slope). *Consider an lsc semi-algebraic function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ that is locally Lipschitz continuous on its domain. Consider also a curve $\gamma \in AC(a,b,\mathbf{R}^n)$ whose image is contained in the domain of f. Then equality*

$$(f\circ\gamma)'(t) = \langle \partial f(\gamma(t)), \dot{x}(t)\rangle = \langle \partial_c f(\gamma(t)), \dot{x}(t)\rangle$$

holds for almost every $t \in [a,b]$. In particular, the slope $\overline{|\nabla f|}$ admits a chain rule.

Proof. Consider the partition of $\operatorname{dom} f$ into finitely many $\mathbf{C}^1$-manifolds $\{M_i\}$, guaranteed to exist by Theorem 5.3. 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 5.2, for any index i and for a.e. $t \in (0,\eta)$ the implication

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

Now suppose that for such t, the point $x(t)$ lies in a manifold 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

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

The result follows. ∎

A noteworthy point about the corollary above is the appearance of the Clarke subdifferential in the chain rule. As a result, we can even strengthen Proposition 4.10 in the context of lsc semialgebraic functions $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ that are locally Lipschitz continuous on their domains. The proof is analogous to that of Proposition 4.10.

PROPOSITION 5.5 (semialgebraic equivalence). *Consider an lsc semialgebraic function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ that is locally Lipschitz continuous on its domain. Then for any curve $x \in AC(a, b, \mathbf{R}^n)$ the following are equivalent:*

1. *x is a curve of near-maximal slope.*
2. *$f \circ x$ is nonincreasing and we have*

$$\dot{x} \in -\partial f(x) \qquad a.e.\ on\ [a, b].$$

3. *$f \circ x$ is nonincreasing and we have*

$$\dot{x} \in -\partial f(x), \quad \|\dot{x}\| = d(0, \partial f(x)), \quad and \quad \|\dot{x}\| = d(0, \partial_c f(x)) \quad a.e.\ on\ [a, b].$$

5.2. Semialgebraic descent: Existence, length, and convergence. In this subsection, we will be interested in existence of near-steepest descent curves emanating from nonminimizing points. This question was already addressed in Theorem 3.5 under a seemingly stringent condition, which will now come into focus. To this end, we first observe that the pathological function f in Example 3.4 is not analytic, and this is no accident. We will see shortly that analytic (and semi-algebraic) functions $f\colon \mathbf{R}^n \to \mathbf{R}$ do admit nontrivial near-steepest descent curves emanating from any point $\bar{x}$ that is not a local minimizer. The key to our development is the KŁ-inequality [5, 7, 24], generalizing the Łojasiewicz inequality for analytic functions [26, 27].

DEFINITION 5.6 (KŁ-inequality).

- *A function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is said to satisfy the* upper KŁ-inequality *if for any bounded open set $U \subset \mathbf{R}^n$ and any real τ, there exists $\rho > 0$ and a nonnegative 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 $\tau < f(x) < \tau + \rho$.
- *A function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is said to satisfy the* lower KŁ-inequality *if for any bounded open set $U \subset \mathbf{R}^n$ and any real τ, there exists $\rho > 0$ and a nonnegative continuous function $\psi\colon (\tau-\rho, \tau] \to \mathbf{R}$, which is $\mathbf{C}^1$-smooth and strictly*

increasing on $(\tau-\rho,\tau)$*, and such that the inequality*

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

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

In particular, all analytic and all semialgebraic functions satisfy both the upper and lower KŁ-inequalities [5, Theorem 14]. The following existence result is now immediate.

THEOREM 5.7 (KŁ-inequality and existence of near-steepest descent curves). *Consider an lsc function* $f\colon \mathbf{R}^n\to\overline{\mathbf{R}}$ *satisfying the lower KŁ-inequality, along with a point* $\bar{x}\in\operatorname{dom} f$ *that is not a local minimizer of* f*. Suppose moreover that* f *is continuous on slope-bounded sets. Then there exists reliable descent from* $\bar{x}$*.*

Proof. Let U be any bounded open neighborhood of $\bar{x}$. Since f satisfies the lower KŁ-inequality, we deduce that there exists a real $\rho>0$ and a nonnegative continuous function $\psi\colon (f(\bar{x})-\rho, f(\bar{x})]\to\mathbf{R}$, which is $\mathbf{C}^1$-smooth and strictly increasing on $(f(\bar{x})-\rho, f(\bar{x}))$, and such that the inequality

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

holds for all $x\in U$ with $f(\bar{x})-\rho<f(x)<f(\bar{x})$. Since f is lsc, there exists $\epsilon>0$ so that for each $x\in\overline{\mathbf{B}}_\epsilon(\bar{x})$ we have $f(x)>f(\bar{x})-\rho$. Define a function $h\colon\mathbf{R}^n\to\overline{\mathbf{R}}$ by setting $h(x)=(\psi\circ f)(x)+\delta_{\overline{\mathbf{B}}_\epsilon(\bar{x})}(x)+\delta_{[f\leq f(\bar{x})]}(x)$. Theorem 3.5 then immediately implies that there exists a reliable descent curve $\gamma\colon [0,L]\to\mathbf{R}^n$ emanating from $\bar{x}$

By decreasing L, we may assume without loss of generality that the image of γ is contained in $\mathbf{B}_\epsilon(\bar{x})$. Let $\theta\colon [0,L]\to\mathbf{R}$ be a nondecreasing function coinciding a.e. with $h\circ\gamma$. We then deduce $\psi^{-1}\circ\theta$ is a nondecreasing function coinciding a.e. with $f\circ\gamma$. Moreover for a.e. $t\in[0,L]$ we have

$$|(\psi^{-1}\circ\theta)'(t)|=\frac{1}{\psi'(f(\gamma(t)))}|\theta'(t)|\geq\frac{1}{\psi'(f(\gamma(t)))}\overline{|\nabla(\psi\circ f)|}(\gamma(t))=\overline{|\nabla f|}(\gamma(t)).$$

The result follows immediately. ∎

THEOREM 5.8 (lengths of near-steepest descent curves). *Consider an lsc function* $f\colon\mathbf{R}^n\to\overline{\mathbf{R}}$ *satisfying the upper KŁ-inequality. Then any bounded near-steepest descent curve* $\gamma\colon[0,\infty)\to\mathbf{R}^n$*, satisfying* $\overline{|\nabla f|}(\gamma(t))\neq 0$ *a.e. on* $[0,\infty)$*, has finite length.*

Proof. Consider a near-steepest descent curve $\gamma\colon[0,\infty)\to\mathbf{R}^n$ and let U be an open bounded set containing the image of γ. Then there exists a nonincreasing function $\phi\colon[0,\infty)\to\mathcal{X}$ coinciding with $f\circ\gamma$ on a full-measure subset $M\subset[0,\infty)$ and so that the inequality $|\phi'(t)|\geq\overline{|\nabla f|}(\gamma(t))$ holds on M. Let τ be the infimum of ϕ on M. Clearly then since f is lsc and proper, we have $\tau>-\infty$. Moreover τ is never attained on M, since $\overline{|\nabla f|}(\gamma(t))\neq 0$ almost everywhere. By the upper KŁ-inequality, there exists $\rho>0$ and a nonnegative 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 $\tau<f(x)<\tau+\rho$. We deduce that there exists $T\in[0,\infty)$ so that the image $\gamma[T,\infty)$ lies in the sublevel set $[f<\tau+\rho]$. We deduce

$$\begin{aligned}\int_0^\infty\|\dot\gamma(r)\|\,dr&\leq T+\int_T^\infty\overline{|\nabla(\psi\circ f)|}(\gamma(r))\,dr\leq T-\int_T^\infty(\psi\circ\phi)'(r)\,dr\\&\leq T+\psi(\phi(T))-\psi(\tau).\end{aligned}$$

The result follows. ∎

The following is now immediate

COROLLARY 5.9 (convergence to local minimizers). *Consider an lsc function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ satisfying the upper and lower KL-inequalities, along with a point $\bar{x} \in \operatorname{dom} f$ that is not a local minimizer of f. Suppose moreover that f is continuous on slope-bounded sets. Then there exists a reliable descent curve $\gamma\colon [0,L] \to \mathbf{R}^n$, for some $L > 0$, emanating from $\bar{x}$, so that either γ is unbounded or L is finite and γ converges to a local minimizer of f.*

Proof. This follows immediately by extending a reliable descent curve $\gamma\colon [0,L] \to \mathbf{R}^n$ emanating from $\bar{x}$ to have a maximal domain of definition using Zorn's lemma, along with Theorems 5.7 and 5.8. We omit the details for the sake of brevity. ∎

In conclusion, we show that curves of near-maximal slope of semialgebraic functions have finite length. The proof is almost identical to the proof of [21, Theorem 7.1]; hence we only provide a sketch.

THEOREM 5.10 (lengths of curves of near-maximal slope). *Consider an lsc, semialgebraic 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 $x\colon [0,T) \to \mathbf{R}^n$ 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 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 semialgebraic. 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 [20] 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

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

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 upper KL-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 $|\nabla(\xi_i \circ f)|(y) \geq 1$ 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 $|\nabla(\xi_i \circ f)|(y) \geq 1$ 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 $l_i \leq \xi_i(\beta_i) - \xi_i(\alpha_i)$. 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 5.11 (convergence of curves of near-maximal slope). *Consider an lsc, semialgebraic 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 lower-critical point of f.*

Acknowledgments. We thank Aris Daniilidis for insightful discussions and we thank the anonymous referees for their careful reading and useful suggestions.

REFERENCES

[1] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.

[2] J.-P. Aubin and A. Cellina, *Differential Inclusions: Set-Valued Maps and Viability Theory*, Grundlehren Math. Wiss. 264, Springer-Verlag, Berlin, 1984.

[3] D. Azé, *A unified theory for metric regularity of multifunctions*, J. Convex Anal., 13 (2006), pp. 225–252.

[4] S. Basu, R. Pollack, and M. Roy, *Algorithms in Real Algebraic Geometry*, Algorithms Comput. Math. 10, Springer-Verlag, New York, 2006.

[5] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM J. Optim., 18 (2007), pp. 556–572.

[6] J. Bolte, A. Daniilidis, and A. Lewis, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim., 17 (2006), pp. 1205–1223.

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

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

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

[10] D. Burago, Y. Burago, and S. Ivanov, *A Course in Metric Geometry*, Grad. Stud. Math. 33, AMS, Providence, RI, 2001.

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

[12] M. Coste, *An Introduction to o-Minimal Geometry*, RAAG Notes, Institut de Recherche Mathématiques de Rennes, 1999.

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

[14] S. Daneri and G. Savaré, *Lecture Notes on Gradient Flows and Optimal Transport*, arXiv:1009.3737 [math.CA], 2010.

[15] E. De Giorgi, *New problems on minimizing movements*, in Boundary Value Problems for Partial Differential Equations and Applications, RMA Res. Notes Appl. Math. 29, Masson, Paris, 1993, pp. 81–98.

[16] 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 (1980), pp. 180–187.
[17] M. Degiovanni, A. Marino, and M. Tosques, *Evolution equations with lack of convexity*, Nonlinear Anal., 9 (1985), pp. 1401–1443.
[18] M. Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Modern Birkhäuser Classics, Birkhäuser, Boston, 2007.
[19] A. Ioffe, *Metric regularity and subdifferential calculus*, Uspekhi Mat. Nauk, 55 (2000), pp. 103–162.
[20] A. Ioffe, *Critical values of set-valued maps with stratifiable graphs. Extensions of Sard and Smale-Sard theorems*, Proc. Amer. Math. Soc., 136 (2008), pp. 3111–3119.
[21] A. Ioffe, *An invitation to tame optimization*, SIAM J. Optim., 19 (2009), pp. 1894–1917.
[22] J. Kelley, *General Topology*, Grad. Texts Math. 27, Springer-Verlag, New York, 1975.
[23] M. Khamsi and W. Kirk, *An Introduction to Metric Spaces and Fixed Point Theory*, Pure and Appl. Math., Wiley-Interscience, New York, 2001.
[24] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Ann. Inst. Fourier (Grenoble), 48 (1998), pp. 769–783.
[25] G. Leoni, *A First Course in Sobolev Spaces*, Grad. Stud. Math. 105, AMS, Providence, RI, 2009.
[26] S. Łojasiewicz, *Une propriété topologique des sous-ensembles analytiques réels*, in Les Équations aux Dérivées Partielles (Paris, 1962), Éditions du Centre National de la Recherche Scientifique, Paris, 1963, pp. 87–89.
[27] S. Łojasiewicz, *Sur les ensembles semi-analytiques*, in Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 2, Gauthier-Villars, Paris, 1971, pp. 237–241.
[28] 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 (1989), pp. 281–330.
[29] B. Mordukhovich, *Variational Analysis and Generalized Differentiation* I: *Basic Theory*, Grundlehren Math. Wissen. 330, Springer, Berlin, 2005.
[30] J.-P. Penot, *Calculus Without Derivatives*, Grad. Texts Math. 266, Springer, New York, 2013.
[31] R. Rockafellar and A. Dontchev, *Implicit Functions and Solution Mappings*, Monogr. Math., Springer-Verlag, Berlin, 2009.
[32] R. Rockafellar and R.-B. Wets, *Variational Analysis*, Grundlehren Math. Wiss. 317, Springer, Berlin, 1998.
[33] M. Shiota, *Geometry of Subanalytic and Semialgebraic Sets*, Birkhäuser, Cambridge, MA, 1997.
[34] L. van den Dries and C. Miller, *Geometric categories and o-minimal structures*, Duke Math. J., 84 (1996), pp. 497–540.
[35] L. van den Dries, *Tame Topology and o-Minimal Structures*, LMS Lecture Note Ser. 248, Cambridge University Press, Cambridge, UK, 1998.
[36] D. Varberg, *On absolutely continuous functions*, Amer. Math. Monthly, 72 (1965), pp. 831–841.