FOUNDATIONS OF
COMPUTATIONAL
MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

Identifiability, the KL Property in Metric Spaces, and Subgradient Curves

A. S. Lewis[1] · Tonghua Tian[1]

Abstract

Identifiability, and the closely related idea of partial smoothness, unify classical active set methods and more general notions of solution structure. Diverse optimization algorithms generate iterates in discrete time that are eventually confined to identifiable sets. We present two fresh perspectives on identifiability. The first distills the notion to a simple metric property, applicable not just in Euclidean settings but to optimization over manifolds and beyond; the second reveals analogous continuous-time behavior for subgradient descent curves. The Kurdyka–Łojasiewicz property typically governs convergence in both discrete and continuous time: we explore its interplay with identifiability.

1 Introduction

Contemporary optimization involves notions of structure, such as sparsity and rank, that have prompted a re-examination of classical active-set philosophy. Early generalizations and terminology such as [25, 27, 40, 67] motivated the idea of an *identifiable set* for an objective function f over Euclidean spaces [35, Definition 3.10].

∆ Springer

Consider, for example, the nonsmooth nonconvex function on $\mathbf{R}^2$ defined by

$$f(x) = 5|x_2 - x_1^2| + x_1^2. \tag{1.1}$$

The set $\mathcal{M} = \{x : x_2 = x_1^2\}$ is identifiable at the minimizer $\bar{x} = 0$ because, in the language of subgradients [66], a quick calculation shows

$$\text{dist}(0, \partial f(x^k)) \to 0 \quad \Rightarrow \quad x^k \in \mathcal{M} \text{ eventually.} \tag{1.2}$$

A wide variety of algorithms for nonsmooth or constrained optimization generate convergent sequences of iterates that behave well with respect to some merit function. Specifically, some corresponding sequence of subgradients of the merit function converges to zero. Consequently, the algorithm identifies some associated structure — in example (1.1), the set $\mathcal{M}$. Examples include classical projected gradient or subgradient methods [27, 40, 67], the proximal point method [43] and composite generalizations such as the proximal gradient method [56], alternating projection and Gauss-Seidel schemes [5, 19], some bundle methods [64], and the broad majorization-minimization framework [20], including sequential quadratic programming. Interesting contemporary examples include [58, 59].

Within optimization and its interface with machine learning, interest in non-Euclidean settings has grown rapidly, supported by expositions like [1, 21]. In particular, [69] proved complexity bounds for the projected subgradient approach to geodesically convex optimization on Hadamard manifolds. Earlier work had already extended the proximal point philosophy far beyond Euclidean settings, in particular to Hadamard manifolds: [12, 39, 57] study geodesically convex and nonconvex optimization, and alternating projection schemes. Extensions to more general geodesic spaces include [9, 10, 48] (in the case of nonpositive curvature) and [54]. A reappraisal of the fundamentals of identifiability from a purely metric perspective is therefore appealing.

A complementary perspective on identifiability arises through analogues of iterative algorithms like the proximal point method but in continuous time [22]. The same identification behavior illustrated above in discrete time also manifests itself in continuous time. In the example, as we shall see later, locally absolutely continuous curves $x : \mathbf{R}_+ \to \mathbf{R}^n$ with initial points $x(0)$ near zero and satisfying

$$x'(t) \in -\partial f(x(t)) \quad \text{for almost all times } t > 0$$

(following standard variational-analytic terminology [66]) always converge to zero and are eventually confined to $\mathcal{M}$. Figure 1 shows plots of two randomly initialized examples of these *subgradient curves*.

We pursue three main themes around the idea of identifiability for an objective f. First, we reframe the idea in purely *metric* terms, using the *slope* $|\nabla f|$, which measures instantaneous rate of decrease [31]. For example, property (1.2) becomes

$$|\nabla f|(x^k) \to 0 \quad \Rightarrow \quad x^k \in \mathcal{M} \text{ eventually.}$$
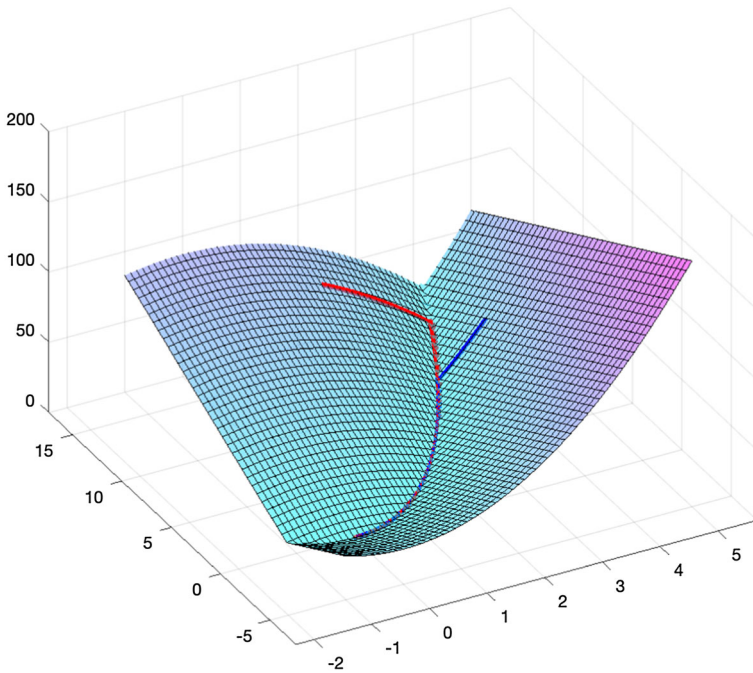
Fig. 1 Subgradient curves for the objective $5|x_2 - x_1^2| + x_1^2$

By switching from the sophisticated dual perspective of subgradient vectors to the simple primal scalar perspective of the slope, we highlight the intuitive nature of identifiability and extend its applicability to more general settings like optimization over manifolds [1, 21]. This setting also allows us to highlight a new linear growth tool (Theorem 4.2), on which our whole development fundamentally depends. Secondly, in the Euclidean setting, we demonstrate quite generally that subgradient curves are eventually confined to identifiable manifolds. In the Euclidean setting, our metric definition of identifiability is equivalent to the usual subgradient-based definition, and closely related to *partial smoothness* [52]. Thirdly, motivated by the crucial role of the "Kurdyka–Łojasiewicz property" in the convergence of optimization algorithms and dynamics [5, 15, 17, 19], including those in metric-space settings [44], we show how the KL property is controlled entirely by the KL property for the restriction of the objective to any identifiable set. Optimization problems we encounter in practice are semi-algebraic or, more generally, tame [46], and for generic problem instances in such settings, local minimizers lie on identifiable analytic manifolds on which the objective function is analytic [16, 34]. The KL property for analytic functions on analytic manifolds is classical [49, Section 9], so our results apply, shedding a fresh light on the KL property in nonsmooth optimization.

We present our development in two parts. Part I—Sects. 2 through 5—studies identifiability and its interplay with the KL property in metric spaces; Part II—Sects. 6 through 10—focuses on the special case of Euclidean space.

Part I begins in Sect. 2 by reviewing the notion of slope in metric spaces, and its relationship with subgradients in the Euclidean case. We introduce the idea of a critical sequence, noting the prevalence of such sequences in optimization as proximal images of minimizing sequences. Thus motivated, Sect. 3 presents our new metric definition of identifiability: sets in which all critical sequences are eventually contained. In particular, expanding on a Euclidean precedent [35], we study minimal identifiable sets. In Sect. 4, in striking contrast to previous subgradient-based developments, we deploy the Ekeland variational principle to explain the fundamental geometry of linear objective growth around identifiable sets in complete metric spaces. Our new linear growth result explains, in broad generality, how objective growth around minimizers is governed entirely by growth on the identifiable set. As Sect. 5 explains, a particularly important growth condition is the Kurdyka–Łojasiewicz property, which we newly reframe in terms of identifiability. We prove that the KL property is inherited from the analogous property on an identifiable set, a useful tool for understanding the property in practice.

Moving to the Euclidean setting of Part II, we begin in Sect. 6 with a quick review of some variational analysis. Section 7 develops the basic example of "max functions": pointwise maxima of finitely many continuously differentiable functions. Section 8 studies identifiability from a variational-analytic perspective, throwing a fresh and intuitive light on the equivalence between identifiable manifolds and partial smoothness [35]. We return in particular to the KL property, now relying on analytic settings. In Sect. 9 we move to the continuous time setting, presenting a result that explains for the first time the identification behavior illustrated in Fig. 1. We end in Sect. 10 with a proof of this result.

Part I

Identifiability in Metric Spaces

2 Critical Sequences in Metric Spaces

Our setting is a metric space (X, d). For any point $x \in X$, the function $d_x \colon X \to \mathbf{R}$ is defined by $d_x(y) = d(x, y)$ for points $y \in X$. On X, we consider a function f which, unless otherwise stated, takes values in $(-\infty, +\infty]$. We focus on a point $\bar{x}$ in the domain

$$\operatorname{dom} f = \{x : f(x) < +\infty\}.$$

We are mainly concerned with points $\bar{x}$ that are local minimizers for f. First, however, we recall a fundamental definition [31] leading to a weaker property.

Definition 2.1 For a function f on a metric space, the *slope* $|\nabla f|(\bar{x})$ at a point $\bar{x} \in \operatorname{dom} f$ is zero if $\bar{x}$ is a local minimizer, and otherwise is the quantity

$$\limsup_{\bar{x} \neq x \to \bar{x}} \frac{f(\bar{x}) - f(x)}{d(\bar{x}, x)}.$$

At points outside of the domain, the slope is $+\infty$.

The notation is suggestive of the easy case of a differentiable function on a Euclidean space, when $|\nabla f|(x) = |\nabla f(x)|$ [31, Observation 4.3]. Notice that the slope may be infinite at points in the domain: for example, $|\nabla \sqrt[3]{\cdot}|(0) = +\infty$. An easy exercise shows an equivalent definition:

$$|\nabla f|(\bar{x}) = \inf\{\delta > 0 : \bar{x} \text{ is a local minimizer for } f + \delta d_{\bar{x}}\}.$$

(The infimum of the empty set is defined to be $+\infty$.) By definition, the slope is zero at any local minimizer. However, the converse may fail: for example, on $\mathbf{R}$, the function $f(x) = -x^2$ has slope zero at zero but is not locally minimized there. The following result [47, Proposition 3.3] collects some simple tools for the slope.

Proposition 2.2 *Functions f and g on a metric space always satisfy*

$$|\nabla(f + g)|(x) \leq |\nabla f|(x) + |\nabla g|(x) \quad \text{and} \quad |\nabla(\delta f)|(x) = \delta|\nabla f|(x)$$

for all points x and scalars $\delta > 0$. Furthermore, if f is L-lipschitz, then $|\nabla f|(x) \leq L$.

By Proposition 2.2, all points x, y in the metric space X satisfy $|\nabla d_x|(y) \leq 1$. When X is a *length space*, meaning that the distance $d(x, y)$ is always the infimum of the lengths of curves joining x and y, we can say more.

Theorem 2.3 (Slope characterization of length spaces [7]) *A complete metric space (X, d) is a length space if and only if all distinct points $x, y \in X$ satisfy $|\nabla d_x|(y) = 1$.*

Any connected Riemannian manifold with its Riemannian distance is a length space [51, Theorem 2.55]. We also observe the following easy inequality.

Proposition 2.4 *On any metric space, the squared distance function satisfies*

$$|\nabla d_x^2|(y) \leq 2d(x, y) \quad \text{and} \quad |\nabla(-d_x^2)|(y) \leq 2d(x, y)$$

for all points $x, y \in X$.

Proof Since

$$\limsup_{y \neq z \to y} \frac{d_x^2(y) - d_x^2(z)}{d(y, z)} = \limsup_{y \neq z \to y} \frac{\big(d(y, x) - d(z, x)\big)\big(d(y, x) + d(z, x)\big)}{d(y, z)}$$

$$\leq \limsup_{y \neq z \to y} \big(d(y, x) + d(z, x)\big) = 2d(y, x),$$

the first inequality follows. The second inequality follows similarly. □

A property even weaker than the slope being zero is the criticality idea that we define next. Following [66], for any function f and any point $\bar{x} \in X$, we write $x_r \to_f \bar{x}$ to denote *f-attentive convergence*: $x_r \to \bar{x}$ and $f(x_r) \to f(\bar{x})$. Furthermore, for any function g, we define the quantity

$$\liminf_{x \to_f \bar{x}} g(x) = \inf\{\liminf_r g(x_r) : x_r \to_f \bar{x}\},$$

which by definition is no larger than $g(\bar{x})$.

Definition 2.5 Given a function f, a *critical sequence* for a point $\bar{x} \in \mathrm{dom}\, f$ is a sequence $x_r \to_f \bar{x}$ such that $|\nabla f|(x_r) \to 0$, and $\bar{x}$ is *critical* if such a sequence exists.

For example, any sequence $x_r \to x$ in X is critical for the function d_x^2, by Proposition 2.4. On $\mathbf{R}$, the function $f(x) = \min\{x, 0\}$ has slope 1 at the point 0, but that point is critical because the local minimizers $\frac{1}{r}$ for $r = 1, 2, 3, \ldots$ form a critical sequence. The notion of a critical point is related to the following more robust version of the slope [47, Definition 8.4].

Definition 2.6 The *limiting slope* of f at a point $\bar{x} \in \mathrm{dom}\, f$ is the quantity

$$\overline{|\nabla f|}(\bar{x}) \;=\; \liminf_{x \to_f \bar{x}} |\nabla f|(x).$$

With this terminology, critical points are exactly those where the limiting slope is zero. Notice that the limiting slope is never larger than the slope.

Our development of identifiability in metric spaces studiously avoids the use of subgradients. However, in the special case when the underlying space is Euclidean, readers familiar with standard variational analysis (for example [66]) will find the following formula [47, Proposition 8.5] illuminating. It relates the limiting slope of a closed function with the usual subdifferential ∂f in the terminology of [66].

Proposition 2.7 (Distance formula) *On $\mathbf{R}^n$, any closed function f and any point $x \in \mathrm{dom}\, f$ satisfy $\overline{|\nabla f|}(x) = \mathrm{dist}\big(0, \partial f(x)\big)$.*

In Euclidean space, critical points are therefore those at which zero is a subgradient.

A wide variety of optimization procedures generate critical sequences, because the procedures involve projection or proximal operations corresponding to closed functions $f \colon X \to (-\infty, +\infty]$:

$$\mathrm{prox}_f(x) \;=\; \mathrm{argmin}\Big\{ f + \frac{1}{2} d_x^2 \Big\}. \tag{2.8}$$

In Euclidean settings, this observation is very familiar [5, 19, 27, 40, 43, 56, 58, 59, 64, 67]. However, the phenomenon persists in non-Euclidean settings such as Hadamard manifolds [12, 39, 57, 69] and more general geodesic spaces [9, 10, 48, 54]. We next discuss this property from a purely metric perspective.

The proximal operator is single-valued for closed convex functions f on a Hilbert space (by Minty's Theorem [63]) and in some more general metric spaces [9, 48]. However, for general spaces and closed functions, the value $\inf\{ f + \frac{1}{2} d_x^2 \}$ can be $-\infty$, and even when finite can be unattained. We do, however, have the following routine result in *proper* metric spaces (meaning that all closed balls are compact).

Lemma 2.9 *On a proper metric space X, if a closed function $f \colon X \to (-\infty, +\infty]$ is not identically $+\infty$ and is bounded below, then the operator prox_f is everywhere nonempty-valued.*

Proof Consider any lower bound α for f and any point $z \in X$ for which $f(z)$ is finite. Given any point $x \in X$, the function $g = f + \frac{1}{2}d_x^2$ is closed, so the level set $L = \{y : g(y) \leq g(z)\}$ is closed. Notice that L contains z, and furthermore is bounded, since all points $y \in L$ satisfy $\frac{1}{2}d_x^2(y) \leq g(z) - \alpha$. Hence L is nonempty and compact, so the set $\text{prox}_f(x) = \text{argmin}_L g$ is nonempty. $\square$

The proximal operator has the obvious property that any minimizer of the function f is a *fixed point*: $x \in \text{prox}_f(x)$, and more generally, $x \in \text{prox}_{\frac{1}{\alpha}f}(x)$ for any constant $\alpha > 0$. In turn, f has slope zero at any such fixed point. In Euclidean space this follows from the property that $y \in \text{prox}_f(x)$ implies $x - y \in \partial f(y)$. More generally, it is a consequence of the following easy tool.

Lemma 2.10 *For any closed function f on a metric space, any points $x \in \text{dom } f$ and $y \in \text{prox}_f(x)$ satisfy $|\nabla f|(y) \leq d(x, y)$. Consequently, f has slope zero at any fixed point of prox_f.*

Proof Propositions 2.2 and 2.4 and the fact that $\left|\nabla\left(f + \frac{1}{2}d_x^2\right)\right|(y) = 0$ imply

$$|\nabla f|(y) \leq \left|\nabla\left(f + \frac{1}{2}d_x^2\right)\right|(y) + \left|\nabla\left(-\frac{1}{2}d_x^2\right)\right|(y) \leq d(x, y).$$

The final claim follows by setting $y = x$. $\square$

We next show that the proximal operator, when nonempty-valued, allows us to associate a critical sequence with any minimizing sequence. Special cases of this mechanism, which underlies a great deal of the identification literature that we discussed after equation (2.8), date back at least to [13], although this simple metric version seems a new perspective.

Proposition 2.11 (Proximal images of minimizing sequences) *For a closed function f on a metric space and constants $\alpha > \beta > 0$, consider any point $\bar{x}$ that minimizes f, or more generally, is a fixed point of the proximal operator $\text{prox}_{\frac{1}{\beta}f}$. Then, for any sequence of points $x_r \to_f \bar{x}$, any sequence of points $y_r \in \text{prox}_{\frac{1}{\alpha}f}(x_r)$ for $r = 1, 2, 3, \ldots$ is critical for $\bar{x}$.*

Proof We have

$$f(x_r) - f(\bar{x}) \geq f(y_r) + \frac{\alpha}{2}d(y_r, x_r)^2 - f(\bar{x}) \geq \frac{\alpha}{2}d(y_r, x_r)^2 - \frac{\beta}{2}d(y_r, \bar{x})^2$$

$$\geq \frac{\alpha}{2}d(y_r, x_r)^2 - \frac{\beta}{2}\left(d(y_r, x_r) + d(x_r, \bar{x})\right)^2$$

$$\geq \frac{\alpha - \beta}{2}d(y_r, x_r)^2 - \beta d(y_r, x_r)d(x_r, \bar{x}) - \frac{\beta}{2}d(x_r, \bar{x})^2.$$

Rearranging gives

$$f(x_r) - f(\bar{x}) + \frac{\alpha\beta}{2(\alpha - \beta)}d(x_r, \bar{x})^2 \geq \frac{\alpha - \beta}{2}\left(d(y_r, x_r) - \frac{\beta}{\alpha - \beta}d(x_r, \bar{x})\right)^2.$$

Since the left-hand side converges to zero, so does the right-hand side, so we deduce $d(y_r, x_r) \to 0$, and hence $y_r \to \bar{x}$. Since the function f is closed, we know $\liminf f(y_r) \geq f(\bar{x})$. But since $f(y_r) + \frac{\alpha}{2} d(y_r, x_r)^2 \leq f(x_r)$ for each r, we also know $\limsup f(y_r) \leq f(\bar{x})$, so $f(y_r) \to f(\bar{x})$. Finally, Proposition 2.2 and Lemma 2.10 complete the proof: $|\nabla f|(y_r) \leq \alpha d(y_r, x_r) \to 0$. $\qquad\square$

Although minimizers are always fixed points of the proximal operator, points where the slope is zero may not be. For example, the real function $f(x) = -|x|^{\frac{3}{2}}$ has slope zero at the point $\bar{x} = 0$, but $\bar{x}$ is not a fixed point of $\mathrm{prox}_{\frac{1}{\beta} f}$ for any value $\beta > 0$. However, when the space X is Euclidean, for well-behaved functions f, any critical point is a fixed point of $\mathrm{prox}_{\frac{1}{\beta} f}$ for all sufficiently large $\beta > 0$. Specifically, following the standard terminology of [66], this property holds if f is both "prox-regular" and "prox-bounded" [66, Proposition 13.37].

3 Identifiable Sets in Metric Spaces

In the previous section we saw how minimizing sequences are often associated with critical sequences. As we discussed in the introduction, many optimization problems possess inherent structure around critical points that restricts the possible corresponding critical sequences. To formalize this phenomenon, and unify many precedents in the literature (originating with [28]), [35] used the language of subgradients in Euclidean spaces to develop the notion of an "identifiable set", and to study when such sets are "locally minimal" [35, Definitions 3.10 and 3.11]. The terminology originated with [27]. Our contribution is a fresh view and extension of this theory from a simple metric space perspective.

Definition 3.1 Given a function f on a metric space X, consider a point $\bar{x} \in \mathrm{dom}\, f$ and a set $\mathcal{M} \subset X$ containing $\bar{x}$.

- $\mathcal{M}$ is *identifiable* at $\bar{x}$ if its complement contains no critical sequence for $\bar{x}$.
- $\mathcal{M}$ is *strongly critical* at $\bar{x}$ if all sequences (x_r) in $\mathcal{M}$ satisfying $x_r \to_f \bar{x}$ are critical for $\bar{x}$.
- The *modulus of identifiability* for $\mathcal{M}$ at $\bar{x}$ is

$$\liminf_{x \to_f \bar{x},\ x \notin \mathcal{M}} |\nabla f|(x). \tag{3.2}$$

Notice that $\mathcal{M}$ is identifiable if and only if the modulus of identifiability is strictly positive. Furthermore, if the point $\bar{x} \in \mathrm{dom}\, f$ is not critical, then every set containing $\bar{x}$ is identifiable, so our interest in identifiable sets focuses on critical points. Identifiable sets always exist. In particular, for any radius $\epsilon > 0$, setting $B_\epsilon(\bar{x}) = \{x \in X : d(x, \bar{x}) \leq \epsilon\}$, the f-attentive neighborhood

$$B_\epsilon^f(\bar{x}) = \{x \in B_\epsilon(\bar{x}) : |f(x) - f(\bar{x})| \leq \epsilon\} \tag{3.3}$$

is identifiable. The following tools are easy to verify.

Proposition 3.4 *On a metric space, at any point in the domain of a function, finite intersections of identifiable sets are identifiable, and finite unions of strongly critical sets are strongly critical.*

Proposition 3.5 *On any metric space, consider functions $f \geq g$, a point $\bar{x} \in \mathrm{dom}\, f$ and a set $\mathcal{M}$ containing $\bar{x}$. Suppose that f and g are equal on $\mathcal{M}$ around $\bar{x}$. At $\bar{x}$, if $\mathcal{M}$ is strongly critical for g, then $\mathcal{M}$ is also strongly critical for f.*

Proof For any sequence $x_r \to_f \bar{x}$ in $\mathcal{M}$ since $g \leq f$ and $g(x_r) = f(x_r)$, strong criticality for g implies $0 \leq |\nabla f|(x_r) \leq |\nabla g|(x_r) \to 0$, so (x_r) is critical for f. $\quad\square$

Remark 3.6 If a set $\mathcal{M}$ is identifiable at a point $\bar{x}$, then it is easily seen to be identifiable at every point in the intersection of $\mathcal{M}$ with some f-attentive neighborhood of $\bar{x}$. Moreover, the modulus of identifiability σ_x, as a function of $x \in \mathcal{M}$, is lower semicontinuous with respect to f-attentive convergence:

$$\liminf_{x \to_f \bar{x},\, x \in \mathcal{M}} \sigma_x \geq \sigma_{\bar{x}}.$$

Identifiable sets are not unique: any superset of an identifiable set at a point $\bar{x}$ is also identifiable. Identifiability is a local property (f-attentively): if a set is identifiable, so is its intersection with any f-attentive neighborhood. Our interest in strongly critical sets stems from the fact that smaller identifiable sets are more informative tools than larger ones. To be precise, any strongly critical identifiable set at $\bar{x}$ must be locally minimal (in an f-attentive sense) among identifiable sets. A subgradient-based version of the following result appeared in [35, Proposition 3.12].

Proposition 3.7 (Locally minimal identifiable sets) *Given a function f on a metric space, suppose that a set $\mathcal{M}$ is identifiable at a point $\bar{x}$. Then $\mathcal{M}$ is strongly critical at $\bar{x}$ if and only if the following property holds:*

$$\begin{cases} \textit{For any identifiable set } \mathcal{M}' \textit{ at } \bar{x}, \textit{ there exists an } f\textit{-attentive} \\ \textit{neighborhood } B_\epsilon^f(\bar{x}) \textit{ with } \epsilon > 0 \textit{ satisfying } \mathcal{M} \cap B_\epsilon^f(\bar{x}) \subset \mathcal{M}'. \end{cases} \quad (3.8)$$

Proof Suppose that the set $\mathcal{M}$ is strongly critical at $\bar{x}$ but that property (3.8) fails for some set $\mathcal{M}'$ that is identifiable at $\bar{x}$. Then, for each $r = 1, 2, 3, \ldots$, there exists a point x_r in the set $\mathcal{M} \cap B_{1/r}^f(\bar{x})$ but outside the set $\mathcal{M}'$. Consequently we have $x_r \to_f \bar{x}$ in $\mathcal{M}$, so by definition, the sequence (x_r) is critical, contradicting the identifiability of $\mathcal{M}'$.

Conversely, if the set $\mathcal{M}$ is identifiable but not strongly critical at $\bar{x}$, then there exists a sequence $x_r \to_f \bar{x}$ in $\mathcal{M}$ that is not critical. After taking a subsequence, we can suppose that there exists a value $\delta > 0$ such that $|\nabla f|(x_r) > \delta$ for all r. We claim that the set defined by $\mathcal{M}' = \mathcal{M} \setminus (x_r)$ is identifiable. If not, there exists a critical sequence (y_s) for $\bar{x}$ outside $\mathcal{M}'$. Since $\mathcal{M}$ is identifiable, $y_s \in \mathcal{M}$ for all large s, and hence $y_s \in (x_r)$, from which we deduce $|\nabla f|(y_s) > \delta$, contradicting the criticality of (y_s). But now we note that for all $\epsilon > 0$, the point x_r lies in the set $\mathcal{M} \cap B_\epsilon^f(\bar{x})$ for all large r but lies outside $\mathcal{M}'$, contradicting property (3.8). $\quad\square$

Identifiability and strong criticality play symmetric roles here: a result analogous to Proposition 3.7 is easy to derive for locally *maximal* strongly critical sets.

We next note an immediate consequence of Proposition 3.7: strongly critical identifiable sets (or equivalently, locally minimal identifiable sets), if they exist, must be locally unique. A subgradient-based version, with stronger assumptions, appeared in [42, Corollary 4.2].

Proposition 3.9 (Uniqueness of strongly critical identifiable sets) *Given a function f on a metric space, for any two strongly critical identifiable sets $\mathcal{M}$ and $\mathcal{M}'$ at a point $\bar{x}$, there exists an f-attentive neighborhood $B_\epsilon^f(\bar{x})$ with $\epsilon > 0$ such that*

$$\mathcal{M} \cap B_\epsilon^f(\bar{x}) = \mathcal{M}' \cap B_\epsilon^f(\bar{x}).$$

To illustrate the idea of a strongly critical identifiable set, we consider some simple examples. At any point $\bar{x} \in X$, any neighborhood of $\bar{x}$ is a strongly critical identifiable set for the squared distance function $d_{\bar{x}}^2$, by Proposition 2.4: the modulus of identifiability is $+\infty$. On the other hand, if X is a length space, then the set $\mathcal{M} = \{\bar{x}\}$ is a strongly critical identifiable set for the distance function $d_{\bar{x}}$, by Theorem 2.3: the modulus of identifiability is 1. A simple concrete example to keep in mind is the following.

Example 3.10 On $\mathbf{R}^2$, the function $f(u, v) = |u| + v^2$ has slope

$$|\nabla f|(u, v) = \begin{cases} \sqrt{1 + 4v^2} & (u \neq 0) \\ 2v & (u = 0). \end{cases}$$

The set $\mathcal{M} = \{(u, v) : u = 0\}$ is a strongly critical identifiable set at the minimizer zero, and the modulus of identifiability is 1.

Unfortunately, as the following example [35, Example 4.13] shows, even continuous convex functions do not always admit strongly critical identifiable sets at minimizers. In such cases, among the collection of sets that eventually capture every critical sequence, no one particular example stands out: more precisely, there is no locally minimal element.

Example 3.11 The function $f(u, v) = \sqrt{u^2 + v^4}$ has slope zero at its minimizer $(0, 0)$, and

$$|\nabla f|(u, v) = \sqrt{\frac{u^2 + 4v^6}{u^2 + v^4}} \quad \text{for } (u, v) \neq (0, 0).$$

Consider any constant $\alpha > 0$. Close to $(0, 0)$ and outside the set

$$\mathcal{M}_\alpha = \{(u, v) : |u| \leq \alpha v^2\},$$

the slope is bounded below by $\frac{\alpha}{\sqrt{1+\alpha^2}} > 0$, so the set $\mathcal{M}_\alpha$ is identifiable at $(0, 0)$. These sets shrink to the set $\mathcal{M}_0$ as $\alpha \downarrow 0$. However, $\mathcal{M}_0$ is not identifiable, because the

critical sequence (k^{-3}, k^{-1}) for $k = 1, 2, 3, \ldots$ approaches $(0, 0)$ from outside $\mathcal{M}_0$. Thus no locally minimal identifiable set exists at $(0, 0)$, and hence, by Proposition 3.7, neither does a strongly critical identifiable set.

The limiting slope gives an equivalent definition of the modulus of identifiability.

Proposition 3.12 (Identifiability via limiting slope) *In Definition 3.1, if the set $\mathcal{M}$ is closed, then the modulus of identifiability is*

$$\liminf_{x \to_f \bar{x}, \, x \notin \mathcal{M}} \overline{|\nabla f|}(x). \tag{3.13}$$

Proof We just need to show that the modulus (3.2) is no larger than the quantity (3.13), the reverse inequality being immediate. If the claim fails, then there exists a constant $\epsilon > 0$ such that the modulus is at least $\overline{|\nabla f|}(x_r) + \epsilon$ for some sequence $x_r \notin \mathcal{M}$ satisfying $x_r \to_f \bar{x}$. Since $\mathcal{M}$ is closed, each distance $\delta_r = \mathrm{dist}(x_r, \mathcal{M})$ is strictly positive, so there exist points $x'_r \in X$ satisfying

$$d(x_r, x'_r) < \min\left\{\delta_r, \frac{1}{r}\right\}, \qquad |f(x_r) - f(x'_r)| < \frac{1}{r}, \qquad |\nabla f|(x'_r) < \overline{|\nabla f|}(x_r) + \frac{\epsilon}{2}.$$

Since $x'_r \notin \mathcal{M}$ and $x'_r \to_f \bar{x}$, we have arrived at a contradiction. $\square$

This result may fail if the set $\mathcal{M}$ is not closed. For example, consider the function $f \colon \mathbf{R}^2 \to \mathbf{R}$ defined by $f(u, v) = \min\{u, 0\}$ at the point $\bar{x} = (0, 0)$ in the set $\mathcal{M}$ whose complement is $\{(0, v) : v \neq 0\}$. The modulus (3.2) is 1, so $\mathcal{M}$ is identifiable at $\bar{x}$, but the quantity (3.13) is zero.

As a consequence of Proposition 3.12, in the special case when the underlying space is Euclidean, the following result shows that our new metric notion of identifiability coincides with the original version [35, Definition 3.10].

Corollary 3.14 (Identifiability in Euclidean spaces) *Consider a closed function f on $\mathbf{R}^n$ and a point $\bar{x}$ contained in a closed set $\mathcal{M} \subset \mathbf{R}^n$. Then the set $\mathcal{M}$ is identifiable at $\bar{x}$ if and only if there exists no sequence (x_r) outside $\mathcal{M}$ satisfying $x_r \to_f \bar{x}$ and with subgradients $y_r \in \partial f(x_r)$ converging to zero. The modulus of identifiability is*

$$\liminf_{x \to_f \bar{x}, \, x \notin \mathcal{M}, \, y \in \partial f(x)} |y|.$$

Proof The formula for the modulus follows from Proposition 3.12 by applying Proposition 2.7. The characterization of identifiability follows. $\square$

We defer a more detailed investigation of the Euclidean case until Part II.

4 Linear Growth Around Identifiable Sets

A crucial geometric feature of objective functions around identifiable sets is often described as "sharpness". An early appearance of this idea was [26], and it was a

core ingredient of precursors of identifiability like $\mathcal{VU}$-decomposition [62] and partial smoothness [52]. Loosely speaking, near a local minimizer, the objective grows in a manner determined entirely by its behavior on any identifiable set, because as we leave this set, the objective grows at a linear rate.

The original approaches to this linear growth property, including [52, Proposition 2.10], [42, Theorem 6.2], and [30, Theorem D.2], rely on smoothness assumptions on the identifiable set $\mathcal{M}$ and the restriction of the objective function f to $\mathcal{M}$. These assumptions fail in even some simple examples like

$$f(x) = |x|^{\frac{3}{2}} \quad (x \in \mathbf{R}), \qquad \bar{x} = 0, \qquad \mathcal{M} = \mathbf{R}$$
$$f(x) = x_1^2 + \left||x_1|^{\frac{3}{2}} - x_2\right| \quad (x \in \mathbf{R}^2), \qquad \bar{x} = 0, \qquad \mathcal{M} = \{x : |x_1|^{\frac{3}{2}} = x_2\}.$$

By contrast, we develop a simple new linear growth tool for identifiable sets that does not rely on manifolds, subgradients, Euclidean structure, or even nearest-point projections. Our approach is applicable in any complete metric space, and reveals for the first time the close connection between identifiability and the Ekeland variational principle, which we use in the following form [68, Theorem 1.4.1].

Theorem 4.1 (Ekeland variational principle [37]) *On any complete metric space, if a closed function f is bounded below, then for any value $\epsilon > 0$ and any point $x \in \mathrm{dom}\, f$, there exists a point v satisfying $f(v) + \epsilon d(v, x) \leq f(x)$ and $|\nabla f|(v) \leq \epsilon$.*

Theorem 4.2 (Linear growth) *On a complete metric space, consider a closed function f with slope zero at a point $\bar{x}$, an identifiable set $\mathcal{M}$ at $\bar{x}$, and any nonnegative constant ϵ strictly less than the modulus of identifiability. Then, for any sequence of points $x_r \to_f \bar{x}$, there exists a sequence of points $v_r \to_f \bar{x}$ in $\mathcal{M}$ such that, for all large r,*

$$f(v_r) + \epsilon d(v_r, x_r) \leq f(x_r).$$

Proof Denote the modulus of identifiability by σ. It suffices to prove the result when $0 < \epsilon < \sigma$, since the case $\epsilon = 0$ then follows.

Suppose first that $\bar{x}$ is a local minimizer. By redefining f to take the value $+\infty$ outside a closed ball centered at the local minimizer $\bar{x}$, we can assume that $\bar{x}$ is a global minimizer, and so f is bounded below. Applying the Ekeland principle to each point x_r ensures the existence of points v_r satisfying $|\nabla f|(v_r) \leq \epsilon$ and

$$\epsilon d(v_r, x_r) \leq f(x_r) - f(v_r) \leq f(x_r) - f(\bar{x}) \to 0.$$

We deduce $v_r \to \bar{x}$ and $f(v_r) \to f(\bar{x})$, so our assumption about ϵ ensures $v_r \in \mathcal{M}$ for all r larger than some $\bar{r}$. Redefining $v_r = \bar{x}$ for all $r \leq \bar{r}$ gives the desired sequence.

Now consider the general case where $|\nabla f|(\bar{x}) = 0$. Fix any constant $\delta > 0$ satisfying $2\delta < \sigma - \epsilon$. The point $\bar{x}$ locally minimizes the function $\tilde{f} = f + \delta d_{\bar{x}}$. Furthermore, since $f = \tilde{f} + (-\delta d_{\bar{x}})$, and the function $-\delta d_{\bar{x}}$ is δ-Lipschitz, Proposition 2.2 implies, for all points x, the inequality $|\nabla \tilde{f}|(x) \geq |\nabla f|(x) - \delta$. Taking the lim inf of both sides as $x \to_f \bar{x}$ (or equivalently as $x \to_{\tilde{f}} \bar{x}$) outside the set $\mathcal{M}$ shows that $\mathcal{M}$ is also identifiable for the function $\tilde{f}$, with modulus at least $\sigma - \delta > \epsilon + \delta$.

We can now apply the result proved by the first argument, using the function $\tilde{f}$ in place of the function f and constant $\epsilon + \delta$ in place of ϵ. We deduce that there exists a sequence of points $v_r \to \bar{x}$ in $\mathcal{M}$ satisfying $\tilde{f}(v_r) \to \tilde{f}(\bar{x})$ and

$$\tilde{f}(v_r) + (\epsilon + \delta)d(v_r, x_r) \leq \tilde{f}(x_r) \quad \text{for all large } r.$$

Consequently $f(v_r) \to f(\bar{x})$, and for all large r we have

$$f(v_r) + \delta d(v_r, \bar{x}) + (\epsilon + \delta)d(v_r, x_r) \leq f(x_r) + \delta d(x_r, \bar{x}).$$

Our conclusion follows by the triangle inequality. $\qquad\square$

The slope zero assumption in Theorem 4.2 cannot be relaxed to criticality. For example, the function $f(x) = \min\{0, x\}$ for $x \in \mathbf{R}$ has a critical point $\bar{x} = 0$ for which the set $\mathcal{M} = \mathbf{R}_+$ is identifiable, but the result fails for the sequence $x_r = -\frac{1}{r}$.

We next derive simple optimality conditions from Theorem 4.2 (Linear growth). We model these on [42, Theorem 6.2] and [35, Proposition 7.2 and Corollary 7.3], but replacing the subdifferential arguments for those Euclidean results with our simpler Ekeland-based approach, valid in an arbitrary complete metric space. In the case of max functions (Sect. 7), the results reduce to classical active-set properties.

Corollary 4.3 (Sufficient condition for optimality) *On a complete metric space, suppose that a closed function f has slope zero at a point $\bar{x}$, and consider any identifiable set $\mathcal{M}$ at $\bar{x}$. Then $\bar{x}$ is a local minimizer if and only if it is a local minimizer relative to $\mathcal{M}$. Furthermore, $\bar{x}$ is a strict local minimizer if and only if it is a strict local minimizer relative to $\mathcal{M}$.*

Proof To prove the first equivalence, suppose that $\bar{x}$ is not a local minimizer, so there exists a sequence $x_r \to \bar{x}$ such that $f(x_r) < f(\bar{x})$. Closedness implies $f(x_r) \to f(\bar{x})$, so using Theorem 4.2 (Linear growth), we deduce the existence of a constant $\epsilon > 0$ and a sequence of points $v_r \to \bar{x}$ in $\mathcal{M}$ satisfying, for all large r,

$$f(v_r) \leq f(v_r) + \epsilon d(v_r, x_r) \leq f(x_r) < f(\bar{x}),$$

so $\bar{x}$ is not a local minimizer relative to $\mathcal{M}$. The converse is trivial. The second equivalence concerns strict local minimizers of f, namely points $\bar{x}$ such that $f(x) > f(\bar{x})$ for all points $x \neq \bar{x}$ near $\bar{x}$. The proof is very similar. $\qquad\square$

As a further illustration along classical lines, we show that quadratic growth rates are determined by the growth rate on any identifiable set.

Corollary 4.4 (Quadratic growth) *On a complete metric space, suppose that a closed function f has slope zero at a point $\bar{x}$, and consider any identifiable set $\mathcal{M}$ at $\bar{x}$. Then f has quadratic growth around $\bar{x}$ if and only if it has quadratic growth around $\bar{x}$ relative to $\mathcal{M}$. Indeed, the two growth rates are identical:*

$$\liminf_{\substack{x \to \bar{x} \\ x \neq \bar{x}}} \frac{f(x) - f(\bar{x})}{d(x, \bar{x})^2} = \liminf_{\substack{x \to \bar{x}, \, x \in \mathcal{M} \\ x \neq \bar{x}}} \frac{f(x) - f(\bar{x})}{d(x, \bar{x})^2}.$$

Proof Denote the left-hand and right-hand sides by α and β respectively. Clearly $\alpha \leq \beta$. Suppose in fact $\alpha < \beta$. Choose any value γ in the interval (α, β). Since $\gamma > \alpha$ there is a sequence of points $x_r \to \bar{x}$ satisfying $f(x_r) - f(\bar{x}) < \gamma d(x_r, \bar{x})^2$. Taking the lim sup of both sides shows $\limsup_r f(x_r) \leq f(\bar{x})$. Since f is closed, we also know $\liminf_r f(x_r) \geq f(\bar{x})$, so we deduce $f(x_r) \to f(\bar{x})$. By Theorem 4.2 (Linear growth), there exists a constant $\epsilon > 0$ and a sequence of points $v_r \to \bar{x}$ in $\mathcal{M}$ satisfying $f(v_r) + \epsilon d(v_r, x_r) \leq f(x_r)$ for all large r. Hence $v_r \neq \bar{x}$ for all large r, since otherwise we arrive at the contradiction $f(\bar{x}) + \epsilon d(\bar{x}, x_r) < f(\bar{x}) + \gamma d(x_r, \bar{x})^2$. Since $\gamma < \beta$, for all large r we must have

$$\gamma d(v_r, \bar{x})^2 < f(v_r) - f(\bar{x}) \leq f(x_r) - \epsilon d(v_r, x_r) - f(\bar{x})$$
$$< \gamma d(x_r, \bar{x})^2 - \epsilon d(v_r, x_r) \leq \gamma \big(d(v_r, \bar{x}) + d(v_r, x_r)\big)^2 - \epsilon d(v_r, x_r).$$

We deduce the inequality $0 < d(v_r, x_r)\big(2d(v_r, \bar{x}) + d(v_r, x_r) - \epsilon\big)$, which is impossible for r sufficiently large. □

In Sect. 5, we prove an analogous result for a growth condition fundamental to convergence analysis for optimization algorithms.

Theorem 4.2 asserts only the existence of a "shadow" sequence v_r. If the underlying space is Euclidean and $\mathcal{M}$ is a smooth manifold restricted to which f is smooth, then [30, Theorem D.2] is more geometrically descriptive: for small ϵ we can take as v_r the nearest-point projection $\mathrm{proj}_{\mathcal{M}}(x_r)$. However, we cannot expect this more generally. First, unless the metric space X is proper, the requisite nearest points may not exist. Secondly, the argument of [30, Theorem D.2] relies on smoothness of $\mathcal{M}$ and the corresponding restriction of f, and furthermore on a characterization of identifiable manifolds using subgradients [36] that is very Euclidean in spirit.

However, by relaxing our notion of projection, we can leverage Theorem 4.2 (Linear growth) to prove an analogous result in great generality. This new result uses *approximate* projections of points x onto $\mathcal{M}$: after fixing a constant $C \geq 1$, we consider points $v \in \mathcal{M}$ such that $d(v, x) \leq C \mathrm{dist}(x, \mathcal{M})$.

We need some standard metric space notions. As usual, we consider a metric space (X, d), a subset $\mathcal{M}$, and a function $f : X \to (-\infty, +\infty]$. By restricting the original metric d to $\mathcal{M}$, we arrive at the *induced* metric on $\mathcal{M}$, also denoted d for simplicity. On the resulting metric space $(\mathcal{M}, d)$, we denote the *restriction* of f by $f|_{\mathcal{M}}$. In $\mathcal{M}$, if any two points can be joined by a rectifiable curve in $\mathcal{M}$, then we can also define the *intrinsic* metric $d^{\mathcal{M}}$: the intrinsic distance between any two points is the infimum of the lengths of curves joining them. If there exists a constant $\mu > 0$ such that $d^{\mathcal{M}}(x, y) \leq \mu d(x, y)$ for all points $x, y \in \mathcal{M}$, then we say that $\mathcal{M}$ is *normally embedded* [14]. (By the triangle inequality, $d(x, y) \leq d^{\mathcal{M}}(x, y)$ always holds.) Clearly the property is preserved by bi-Lipschitz homeomorphisms. Examples include convex sets and compact submanifolds in $\mathbf{R}^n$. Curves in normally embedded sets have the same length with respect to both the induced and intrinsic metrics [24, Proposition 2.3.12], and furthermore, when rectifiable, can be reparametrized to have unit speed (with respect to both metrics) [24, Proposition 2.5.9].

Theorem 4.5 (Linear growth from approximate projection) *On a complete metric space, consider a closed function f with slope zero at a point $\bar{x}$, and a normally embedded identifiable set $\mathcal{M}$ at $\bar{x}$ such that the restriction $f|_{\mathcal{M}}$ is continuous.*

- *If $\mathcal{M}$ is strongly critical at $\bar{x}$, then the slope $|\nabla(f|_{\mathcal{M}})|$ is continuous at $\bar{x}$.*
- *Conversely, if $|\nabla(f|_{\mathcal{M}})|$ is continuous at $\bar{x}$, then for any constant $C \geq 1$, there exists a constant $\epsilon > 0$ such that all points x close to $\bar{x}$ with value $f(x)$ close to $f(\bar{x})$ satisfy the property*

$$v \in \mathcal{M} \quad \text{and} \quad d(v, x) \leq C \text{dist}(x, \mathcal{M}) \quad \Rightarrow \quad f(v) + \epsilon d(v, x) \leq f(x).$$

If, furthermore, $\mathcal{M}$ is locally closed, then $\mathcal{M}$ is strongly critical at $\bar{x}$.

Proof The first claim is clear, because $|\nabla f|(x) \geq |\nabla(f|_{\mathcal{M}})|(x)$ for all points $x \in \mathcal{M}$. Turning to the converse direction, by way of contradiction, consider sequences of points $x_r \to_f \bar{x}$ and v_r in $\mathcal{M}$ satisfying the inequalities $d(v_r, x_r) \leq C \text{dist}(x_r, \mathcal{M})$ and $f(v_r) + \frac{1}{r} d(v_r, x_r) > f(x_r)$ for all r. By Theorem 4.2 (Linear growth), there exists $\epsilon > 0$ and a sequence $w_r \to \bar{x}$ in $\mathcal{M}$ satisfying $f(w_r) + \epsilon d(w_r, x_r) \leq f(x_r)$ for all large r. From the inequalities $d(v_r, x_r) \leq C \text{dist}(x_r, \mathcal{M}) \leq C d(w_r, x_r)$, we deduce

$$d(w_r, v_r) \leq d(w_r, x_r) + d(x_r, v_r) \leq (1 + C) d(w_r, x_r).$$

Combining the inequalities shows, for all large r,

$$f(v_r) - f(w_r) > \epsilon d(w_r, x_r) - \frac{1}{r} d(v_r, x_r) \geq \left(\epsilon - \frac{C}{r}\right) d(w_r, x_r) \geq \frac{\epsilon}{2} d(w_r, x_r)$$

$$\geq \frac{\epsilon}{2(1 + C)} d(w_r, v_r).$$

Since the set $\mathcal{M}$ is normally embedded, there exists a number $\mu > 0$ such that the intrinsic metric $d^{\mathcal{M}}$ satisfies $d^{\mathcal{M}}(w, v) \leq \mu d(w, v)$ for all points $w, v \in \mathcal{M}$. In particular, since $w_r, v_r \in \mathcal{M}$, we have

$$d^{\mathcal{M}}(w_r, v_r) \leq \mu d(w_r, v_r) \leq \frac{1}{\delta}\big(f(v_r) - f(w_r)\big)$$

where $\delta = \frac{\epsilon}{2\mu(1+C)}$, so there exist unit-speed curves $\gamma_r : [0, l_r] \to \mathcal{M}$ from w_r to v_r with length $l_r \leq 2d^{\mathcal{M}}(w_r, v_r) \leq \frac{2}{\delta}\big(f(v_r) - f(w_r)\big)$. The corresponding continuous functions $g_r = f \circ \gamma_r$ satisfy $g_r(l_r) - g_r(0) > \frac{\delta}{2} l_r$. Let t_r be the smallest number $t \in [0, l_r]$ such that

$$g_r(s) \geq g_r(l_r) - \frac{\delta}{2}(l_r - s) \qquad \text{for all } s \in [t, l_r].$$

Continuity implies $t_r > 0$ and $g_r(t_r) = g_r(l_r) - \frac{\delta}{2}(l_r - t_r)$. Moreover, for each large r, there exists a sequence of points $s_k^r \nearrow t_r$ such that $g_r(s_k^r) < g_r(l_r) - \frac{\delta}{2}(l_r - s_k^r)$ for

all $k = 1, 2, 3, \ldots$. We deduce

$$\limsup_{t \nearrow t_r} \frac{g_r(t_r) - g_r(t)}{t_r - t} \geq \limsup_{k \to \infty} \frac{g_r(t_r) - g_r(s_k^r)}{t_r - s_k^r} \geq \frac{\delta}{2}.$$

Consider the points $y_r = \gamma_r(t_r) \in \mathcal{M}$. Since γ_r is unit-speed, we have

$$|\nabla(f|_{\mathcal{M}})|(y_r) = \limsup_{x \xrightarrow{\mathcal{M}} y_r, x \neq y_r} \frac{f(y_r) - f(x)}{d(y_r, x)} \geq \limsup_{x \xrightarrow{\mathcal{M}} y_r, x \neq y_r} \frac{f(y_r) - f(x)}{d^{\mathcal{M}}(y_r, x)}$$

$$\geq \limsup_{t \nearrow t_r} \frac{f(y_r) - f(\gamma_r(t))}{d^{\mathcal{M}}(\gamma_r(t_r), \gamma_r(t))} \geq \limsup_{t \nearrow t_r} \frac{g_r(t_r) - g_r(t)}{t_r - t} \geq \frac{\delta}{2}.$$

Furthermore we have

$$d(y_r, \bar{x}) \leq d^{\mathcal{M}}(y_r, w_r) + d(w_r, \bar{x}) \leq 2d^{\mathcal{M}}(v_r, w_r) + d(w_r, \bar{x}) \to 0,$$

contradicting the continuity of the slope $|\nabla(f|_{\mathcal{M}})|$ at $\bar{x}$.

It remains to prove that the set $\mathcal{M}$ is strongly critical at the point $\bar{x}$. Fix any constant $\delta > 0$. By the slope continuity assumption, there exists a constant $\epsilon > 0$ such that $|\nabla(f|_{\mathcal{M}})|(x) < \delta$ for all points x in the ball $B_{2\epsilon}(\bar{x})$. Since $\mathcal{M}$ is locally closed, we can furthermore assume that such x have approximate projections $v \in \mathcal{M}$ with $d(v, x) \leq 2\text{dist}(x, \mathcal{M})$. Our argument so far shows that, after shrinking ϵ, we can assume that these approximate projections satisfy $f(v) \leq f(x)$.

We claim $|\nabla f|(x) \leq 3\delta$ for all points $x \in \mathcal{M} \cap B_{\epsilon}(\bar{x})$. We can assume that x is not a local minimizer for f, since otherwise the claim is immediate. Choose a sequence $x_r \to x$ in $B_{\epsilon}(\bar{x})$ such that

$$q_r = \frac{f(x) - f(x_r)}{d(x, x_r)} \to |\nabla f|(x)$$

and choose approximate projections $v_r \in \mathcal{M}$ with $d(v_r, x_r) \leq 2\text{dist}(x_r, \mathcal{M})$. By definition, these points satisfy $d(v_r, x_r) \leq 2d(x, x_r)$ and hence

$$d(x, v_r) = \big(d(x, v_r) - d(v_r, x_r)\big) + d(v_r, x_r) \leq 3d(x, x_r),$$

implying in particular $v_r \to x$. Moreover, as we have seen, $f(v_r) \leq f(x_r)$ for all large r. Since $|\nabla(f|_{\mathcal{M}})|(x) < \delta$, we know $f(x) - f(v_r) < \delta d(x, v_r)$ for all large r, and hence

$$q_r < \frac{\delta d(x, v_r)}{d(x, x_r)} \leq 3\delta.$$

Letting $r \to \infty$ proves the claim. Since δ was arbitrary, the result follows. $\square$

While we have assumed, for simplicity that the set $\mathcal{M}$ is normally embedded, we only need that assumption to hold *locally* around the point $\bar{x}$, in the obvious sense.

In the most basic case, the underlying space is Euclidean and the identifiable set $\mathcal{M}$ is a smooth submanifold, so the nearest point projection mapping $\text{proj}_\mathcal{M}$ is single-valued near any point $\bar{x} \in \mathcal{M}$. Setting $v = \text{proj}_\mathcal{M}(x)$ in Theorem 4.5, we then recover [30, Theorem D.2] as a special case. However, this new result applies more broadly, including situations when $\mathcal{M}$ is not a manifold, as we illustrate later.

5 The Kurdyka–Łojasiewicz Property

Our initial motivation for the preceding section was the idea of sharpness, a notion highlighted in the optimization literature in [26]. We now return to that idea in order to discuss the nonsmooth Kurdyka–Łojasiewicz inequality [15, 17], a property that has had far-reaching impact on complexity analysis for optimization. We begin with the simple idea of sharpness at a point.

Definition 5.1 On a metric space (X, d), a point $\bar{x}$ is a *sharp local minimizer* of a function $f \colon X \to (-\infty, +\infty]$ if there exists a constant $\epsilon > 0$ such that $\bar{x}$ locally minimizes $f - \epsilon d_{\bar{x}}$.

For example, any point $\bar{x} \in X$ is a sharp local minimizer of the distance function $d_{\bar{x}}$. Notice that any sharp local minimizer is also a strict local minimizer. An immediate consequence of Theorem 4.2 (Linear growth) is the following result.

Corollary 5.2 (Sharp minimizers) *On a complete metric space, if a closed function f has slope zero at a point $\bar{x}$, and the set $\{\bar{x}\}$ is identifiable, then $\bar{x}$ is a sharp local minimizer.*

The following example shows that the converse may fail.

Example 5.3 (Sharp minimizers may not be identifiable) The locally Lipschitz function defined on $\mathbf{R}$ by $f(0) = 0$ and $f(x) = |x| + x^2 \sin \frac{1}{x}$ for $x \neq 0$ satisfies $f(x) \geq \frac{1}{2}|x|$ for all small x, and hence has a sharp minimizer at 0. However, f has derivative zero at each point $\frac{1}{2\pi r}$ for $r = 1, 2, 3, \ldots$. These points approach zero, so the set $\{0\}$ is not identifiable.

To handle the more practical situation where minimizers may not be strict, [26] introduced the following weaker property, which we use in a local form.

Definition 5.4 On a metric space X, a point $\bar{x}$ is a *weak sharp minimizer* (locally) for a function $f \colon X \to (-\infty, +\infty]$ if there exists a constant $\epsilon > 0$ such that, for the level set $\mathcal{L} = \{x : f(x) \leq f(\bar{x})\}$, the point $\bar{x}$ locally minimizes the function $f - \epsilon \text{dist}(\cdot, \mathcal{L})$.

By analogy with Corollary 5.2, we now draw a new relationship between weak sharpness and identifiability.

Corollary 5.5 (Weak sharp minimizers) *On a complete metric space, consider a closed function f and a local minimizer $\bar{x}$. If the level set of points x with value $f(x) \leq f(\bar{x})$ is identifiable at $\bar{x}$ for f, then $\bar{x}$ is a weak sharp minimizer.*

Proof Denote the level set by $\mathcal{L}$. Suppose $\epsilon > 0$ is strictly less than the modulus of identifiability of $\mathcal{L}$ for the function f at the point $\bar{x}$:

$$\liminf_{x \to_f \bar{x},\ f(x) > f(\bar{x})} |\nabla f|(x) > \epsilon.$$

If the result fails, then there exists a sequence of points $x_r \to \bar{x}$ in X satisfying

$$f(x_r) < f(\bar{x}) + \epsilon \cdot \text{dist}(x_r, \mathcal{L}). \tag{5.6}$$

Since $\bar{x}$ is a local minimizer, for large r we have $f(x_r) \geq f(\bar{x})$ and so inequality (5.6) implies $x_r \notin \mathcal{L}$. The right-hand side converges to $f(\bar{x})$, whence so does $f(x_r)$. By Theorem 4.2 (Linear growth), there exists a sequence of points $v_r \to \bar{x}$ in $\mathcal{L}$ satisfying, for large r, the inequality $f(v_r) + \epsilon d(v_r, x_r) \leq f(x_r)$. For large r we know $f(v_r) \geq f(\bar{x})$, so $\epsilon d(v_r, x_r) \leq f(x_r) - f(\bar{x})$, contradicting inequality (5.6). $\square$

Example 5.3 shows that the converse of Corollary 5.5 may fail.

The Kurdyka–Łojasiewicz property may be thought of as a modification of the assumption of Corollary 5.5, obtained by first shifting and truncating the function f, replacing each value $f(x)$ by $[f(x) - f(\bar{x})]^+$, where u^+ denotes the positive part of the value u, and then rescaling those new function values using the following standard tool.

Definition 5.7 A *desingularizer* is a function $\phi \colon [0, +\infty] \to [0, +\infty]$ that is continuous on $[0, +\infty)$, satisfies $\phi(0) = 0$ and $\phi(\tau) \to \phi(+\infty)$ as $\tau \to +\infty$, and has continuous strictly positive derivative on $(0, +\infty)$.

Typically we use desingularizers of the form $\phi(\tau) = \tau^{1-\alpha}$, where the *KL exponent* α lies in the interval $[0, 1)$. Such desingularizers, for example, suffice for the original Łojasiewicz proof of the KL property for real-analytic functions on Euclidean spaces [60], and hence also for real-analytic functions on analytic manifolds [49, Section 9]. Unlike some definitions, we do not explicitly require concavity.

We now present a definition of the Kurdyka–Łojasiewicz property in metric spaces. Our definition is inspired by a version of the KL property from [18, Corollary 4], reframed in the language of identifiability. In the Euclidean case, some earlier subgradient-based versions of the KL property [4] reduce precisely to this metric definition; the original nonsmooth property [15, 17] (again subgradient-based) is slightly stronger, as we shortly discuss. Researchers on Riemannian nonsmooth optimization have proposed generalizations to manifolds [11, 45], but by comparison with the subgradient approach in general, the metric definition we present is simple and direct. By contrast with [18, Section 2], which aimed to illuminate the KL inequality as a metric regularity property, the new connection to identifiability that we present allows us to deploy the linear growth results of Sect. 4.

Definition 5.8 On a metric space X, a function $f \colon X \to (-\infty, +\infty]$ satisfies the *KL property* at a point $\bar{x} \in \text{dom } f$ if there exists a desingularizer ϕ such that the level set of points x satisfying $f(x) \leq f(\bar{x})$ is identifiable at $\bar{x}$ for the composite function g defined on X by $g(x) = \phi([f(x) - f(\bar{x})]^+)$.

We make some immediate comments. The level set for f in the definition is also the level set defined by $g(x) \leq 0$, for the composite function g. Hence, by Corollary 5.5, when the metric space X is complete and f is closed, the KL property implies that the point $\bar{x}$ is a weak sharp minimizer of the composite function g.

For closed functions f, the KL property amounts to the existence of $\delta > 0$ such that $\big|\nabla\big(\phi(f(\cdot) - f(\bar{x}))\big)\big|(x) \geq \delta$ for all points x near $\bar{x}$ with value $f(x)$ near and strictly larger than $f(\bar{x})$, as stated in [18, Corollary 4]. Using a simple chain rule [8, Lemma 4.1], we can rewrite the left-hand side as $\phi'\big(f(x) - f(\bar{x})\big) \cdot |\nabla f|(x)$. When the underlying space X is Euclidean, Proposition 2.7 transforms our definition into an equivalent but more familiar inequality:

$$\phi'\big(f(x) - f(\bar{x})\big) \cdot \mathrm{dist}\big(0, \partial f(x)\big) \;\geq\; \delta. \tag{5.9}$$

The original Euclidean nonsmooth property [17] (sometimes called the *strong* KL inequality [64]), while theoretically stronger, since it replaces the subdifferential $\partial f(x)$ in inequality (5.9) by the larger *Clarke* subdifferential, in practice seems no stronger: for more discussion, see [55]. In Euclidean spaces, the strong KL property always holds for functions f that are semi-algebraic, or, more generally, tame [17].

Remarkably, by-now-standard KL-based convergence analyses like [18, Theorem 24] often transpire on inspection to be entirely metric in nature. This new perspective clarifies proofs and extends their reach to settings such as manifolds and more general metric spaces. Before illustrating this view, we require a standard definition.

Definition 5.10 A function f on a metric space is *continuous on slope-bounded sets* [2] if sequences of points $x_r \to x$ with values $f(x_r)$ and slopes $|\nabla f|(x_r)$ uniformly bounded must satisfy $f(x_r) \to f(x)$.

In particular, on Euclidean spaces, all closed convex functions are continuous on slope-bounded sets, as, more generally, are all closed functions that are subdifferentially continuous in the sense of [66].

We next illustrate KL-based analysis in metric spaces by proving convergence of the proximal point method. The argument follows that for [18, Theorem 24]. In that result, the setting is a Hilbert space and the objective must be weakly convex. However, on inspection, as we show, the result has an entirely metric version. For comparison, [9] also proves convergence of the proximal point method in metric spaces, but requires the space to be of nonpositive curvature and the objective to be geodesically convex; furthermore, the argument only guarantees a notion of convergence weaker than the convergence in distance that we prove next.

Example 5.11 (Proximal sequences in metric spaces) On a complete metric space, we consider a closed function f that is bounded below, and continuous on slope-bounded sets. We assume that there exists a value $\rho > \inf f$ and a desingularizer ϕ such that

$$\inf f < f(x) < \rho \quad \Rightarrow \quad \big|\nabla\big(\phi(f(\cdot) - \inf f)\big)\big|(x) \;\geq\; 1.$$

A simple open cover argument (cf. [3, Lemma 1(ii)]) shows that this uniform version of the KL property holds if the level set $\{x : f(x) \leq \rho\}$ is compact and the KL

property holds at every minimizer. We consider a sequence of points (x_k) arising from the proximal point iteration:

$$x_{k+1} \in \operatorname{prox}_f(x_k) \qquad (k = 0, 1, 2, \ldots).$$

As we discussed in Sect. 2, in quite general settings there exist such sequences starting from any initial point $x_0 \in \operatorname{dom} f$. The values $f(x_k)$ are nonincreasing and bounded below, and Lemma 2.10 implies $|\nabla f|(x_{k+1}) \leq d(x_{k+1}, x_k)$ for all k. We furthermore assume the initial value satisfies $f(x_0) < \rho$. We claim that the iterates x_k must converge to a critical point. We lose no generality in supposing $f(x_k) > \inf f$ for all k. Consider the level sets $L_k = \{x : f(x) \leq f(x_k)\}$. By definition, all points $x \in X$ satisfy

$$f(x_{k+1}) + \frac{1}{2}d^2(x_k, x_{k+1}) \leq f(x) + \frac{1}{2}d^2(x_k, x),$$

so all x in the level set L_{k+1} satisfy $d(x_k, x) \geq d(x_k, x_{k+1})$. We deduce the equation $d(x_k, x_{k+1}) = \operatorname{dist}(x_k, L_{k+1})$. The right-hand side is no larger than the Hausdorff distance between the sets L_k and L_{k+1}, which by [18, Corollary 4] is bounded above by $\phi(f(x_k)) - \phi(f(x_{k+1}))$. Thus trajectory $\sum_k d(x_k, x_{k+1})$ has finite length, so by completeness, x_k converges to some point x^* and furthermore $|\nabla f|(x_k) \to 0$. Continuity on slope bounded sets implies $f(x_k) \to f(x^*)$, so x^* is a critical point.

We are now ready for one of our main new results. In the previous section, we used Theorem 4.2 (Linear growth) to show how a function, around any point where its slope is zero, inherits various growth properties from the corresponding properties restricted to any identifiable set. We now extend that list to include the KL property. The possibility of such an extension is perhaps not surprising: in general metric spaces, the KL inequality is closely related to a growth property [18, Corollary 4], and in the convex Euclidean case the ideas are essentially equivalent [19, Section 3]. However, this powerful technique for recognizing the KL property through identifiability is new even in the Euclidean case. Furthermore, we show that the unrestricted property holds with the same desingularizer (and, in particular, KL exponent) as the restricted property.

Recall that on any subset $\mathcal{M}$ of the metric space (X, d), we can consider the induced metric, where the distance between points $x, y \in \mathcal{M}$ is just $d(x, y)$. The set $\mathcal{M}$ equipped with the induced metric then forms a metric space $(\mathcal{M}, d)$.

Theorem 5.12 (Identifiability and the KL property) *On a complete metric space (X, d), suppose that a closed function f has slope zero at a point $\bar{x}$. Consider any identifiable set $\mathcal{M}$ at $\bar{x}$, and a desingularizer ϕ that is concave (or, more generally, that satisfies $\liminf_{\tau \searrow 0} \phi'(\tau) > 0$). Then the following properties are equivalent.*

(a) *The function f has the KL property at $\bar{x}$ with desingularizer ϕ.*
(b) *On the metric space $(\mathcal{M}, d)$, the function $f|_{\mathcal{M}}$ has the KL property at $\bar{x}$ with desingularizer ϕ.*
(c) *The function $f + \delta_{\mathcal{M}}$ has the KL property at $\bar{x}$ with desingularizer ϕ.*

Proof Properties (b) and (c) differ only in notation. Observe that identifiability ensures the existence of a constant $\epsilon > 0$ such that $|\nabla f|(x) > 2\epsilon$ for all points $x \notin \mathcal{M}$ near $\bar{x}$ with value $f(x)$ near $f(\bar{x})$. Without loss of generality assume $f(\bar{x}) = 0$.

Suppose that property (b) holds. Then there exists a constant $\delta > 0$ such that $\left|\nabla(\phi \circ f|_{\mathcal{M}})\right|(x) \geq \delta$ for all points $x \in \mathcal{M}$ near $\bar{x}$ satisfying $0 < f(x) < \delta$. From this inequality, the definition of slope then implies

$$\left|\nabla(\phi \circ f)\right|(x) \geq \delta, \tag{5.13}$$

since the left-hand side is no smaller than the previous left-hand side.

On the other hand, consider points $x \notin \mathcal{M}$ near $\bar{x}$ satisfying $0 < f(x) < \delta$. Shrinking δ if necessary, we can ensure $\phi'(f(x)) \geq \delta$ and $|\nabla f|(x) > 2\epsilon$. The chain rule implies $\left|\nabla(\phi \circ f)\right|(x) \geq \delta \cdot 2\epsilon > 0$. Property (a) follows from inequality (5.13).

Conversely, if property (a) holds, then there exists a constant $\delta > 0$ such that both the inequalities $\left|\nabla(\phi \circ f)\right|(x) > \delta$ and $\phi'(t) \geq \delta$ hold for all points x satisfying $d(x, \bar{x}) < \delta$ and $0 < f(x) < \delta$, and all $t \in (0, \delta)$. Shrinking δ if necessary, identifiability ensures that all such points x outside $\mathcal{M}$ also satisfy $|\nabla f|(x) > 2\epsilon$. On the other hand, for any such point x in $\mathcal{M}$, there exists a sequence of points $x_r \to x$, each satisfying $\phi(f(x)) - \phi(f(x_r)) > \delta \cdot d(x, x_r)$. By Theorem 4.2 (Linear growth), there exist points $v_r \in \mathcal{M}$, satisfying $f(v_r) + \epsilon d(v_r, x_r) \leq f(x_r)$ for all large r, so

$$\phi(f(x)) - \phi(f(v_r)) > \delta \cdot d(x, x_r) + \phi(f(x_r)) - \phi(f(v_r))$$
$$= \delta \cdot d(x, x_r) + \phi'(t_r)(f(x_r) - f(v_r))$$

for some value $t_r \in [f(v_r), f(x_r)]$, by the mean value theorem. Hence we deduce

$$\phi(f(x)) - \phi(f(v_r)) > \delta \cdot d(x, x_r) + \delta \cdot \epsilon \cdot d(v_r, x_r) \geq \gamma \cdot d(x, v_r),$$

where $\gamma = \delta \cdot \min\{1, \epsilon\} > 0$, so $|\nabla(\phi \circ f|_{\mathcal{M}})|(x) \geq \gamma$. Property (b) follows. □

Remark 5.14 By restricting attention to desingularizers ϕ satisfying the condition $\liminf_{\tau \searrow 0} \phi'(\tau) > 0$, we lose no essential generality. If a desingularizer ϕ fails this condition, then the function defined by $\varphi(t) = \int_0^t \max\{1, \phi'(\tau)\}d\tau, \forall t \geq 0$, is another desingularizer, and it maintains the KL property and satisfies the condition.

Talweg Curves and Identifiability

In [18], the authors explore characterizations of the KL property via the finite length of "talweg" curves (German for "valley paths"). For a suitably isolated critical point $\bar{x}$, given any constant $R > 1$, they consider the mapping $\chi: (f(\bar{x}), +\infty) \rightrightarrows X$ defined, for a closed bounded neighborhood D of $\bar{x}$, by

$$\chi(r) = \left\{x \in D : f(x) = r, \overline{|\nabla f|}(x) \leq R \inf_{\substack{y \in D \\ f(y) = r}} \overline{|\nabla f|}(y)\right\}.$$

A *talweg* is a curve $x \colon (f(\bar{x}), \rho) \to X$ satisfying $x(r) \in \chi(r)$ for all r.

Consider, for example, Example 3.10, where $f(u, v) = |u| + v^2$. The unique critical point $\bar{x} = (0, 0)$ is also the strict global minimizer, with value $f(\bar{x}) = 0$. Let D be the unit disk. Suppose $0 < r < \frac{1}{4R^2}$. Then we have

$$R \inf_{\substack{y \in D \\ f(y) = r}} \overline{|\nabla f|}(y) = 2R\sqrt{r} < 1,$$

so $\chi(r) = \{(0, \pm\sqrt{r})\}$. Thus any talweg curve $x \colon (0, r) \to \mathbf{R}^2$ lies in the identifiable set $\mathcal{M} = \{0\} \times \mathbf{R}$ for all small $r > 0$.

We can generalize this observation using the following assumption.

Assumption 5.15 At some strict local minimizer $\bar{x}$, the set $\mathcal{M}$ is strongly critical, identifiable, and locally connected. Furthermore, $\bar{x}$ is not an isolated point of $\mathcal{M}$, and the restriction $f|_{\mathcal{M}}$ is continuous.

With this assumption, we claim

$$\inf_{\substack{y \in D \\ f(y) = r}} \overline{|\nabla f|}(y) \to 0 \quad \text{as } r \downarrow f(\bar{x}). \tag{5.16}$$

If not, then there exists a value $\delta > 0$ and a sequence $f(\bar{x}) < r_i \downarrow f(\bar{x})$ such that

$$\overline{|\nabla f|}(y) > \delta \quad \text{for all } y \in D \text{ such that } f(y) = r_i \text{ for some } i. \tag{5.17}$$

We can assume the existence of points $y_i \in \mathcal{M}$ approaching $\bar{x}$ and satisfying $f(y_i) = r_i$ for all i. Otherwise, taking a subsequence, there would exist an open neighborhood U of $\bar{x}$ such that $f(x) \neq r_i$ for all $x \in U$ and all i. We can assume, after shrinking, that U is connected and $f(x) > f(\bar{x})$ whenever $\bar{x} \neq x \in U$. But the image $f(U)$ must then be connected, and hence contained in the interval $(-\infty, f(\bar{x})]$. We have arrived at the contradiction $U = \{\bar{x}\}$. Since $\mathcal{M}$ is strongly critical, we know $\overline{|\nabla f|}(y_i) \to 0$, contradicting inequality (5.17). Property (5.16) follows.

We note that the identifiability of $\mathcal{M}$ ensures that any talweg $x \colon (f(\bar{x}), r) \to X$ lies in $\mathcal{M}$ for all r near $f(\bar{x})$. This correlates with our observation that the KL property is determined just by the behavior of f on $\mathcal{M}$.

Part II

Identifiability in Euclidean spaces

6 Four Classes of Nonsmooth Functions

The second part of this work focuses on identifiable sets in Euclidean space. We focus initially on four objective classes, consisting of functions that are, in increasing order of strength, prox-regular, primal lower nice, strongly amenable, or "smoothly embedded". The first three are broad classes, familiar from earlier literature [65, 66].

The last class, consisting just of smooth functions, indicator functions of manifolds, and their sums, is much narrower. Nonetheless, we shall see in Part II that, from the perspective of identifiability and its continuous-time analog, smoothly embedded functions often model the behavior of much more general functions.

We begin by reviewing standard variational analysis [66]. Throughout Part II, **the term "smooth" means** $\mathcal{C}^{(2)}$**-smooth**. For a function $f\colon \mathbf{R}^n \to (-\infty, +\infty]$ and a point $\bar{x} \in \mathrm{dom}\, f$, a vector $y \in \mathbf{R}^n$ is a *regular subgradient* of f at $\bar{x}$, written $y \in \hat{\partial} f(\bar{x})$, if $f(x) \geq f(\bar{x}) + \langle y, x - \bar{x}\rangle + o(|x - \bar{x}|)$ as $x \to \bar{x}$. If furthermore we can replace the term $o(|x - \bar{x}|)$ by the term $O(|x - \bar{x}|^2)$, then we instead call y a *proximal subgradient*, written $y \in \partial_P f(\bar{x})$. We say that y is a *subgradient* of f at $\bar{x}$, written $y \in \partial f(\bar{x})$, if there are sequences $x_r \to_f \bar{x}$ and $y_r \in \hat{\partial} f(x_r)$ with $y_r \to y$. As we observed after Proposition 2.7, $\bar{x}$ is critical if and only if $0 \in \partial f(\bar{x})$. By contrast, it is easy to check that f has slope zero at $\bar{x}$ if and only if $0 \in \hat{\partial} f(\bar{x})$. We define the *regular normal cone* of a set $\mathcal{M} \subset \mathbf{R}^n$ at a point $\bar{x} \in \mathcal{M}$ by $\widehat{N}_{\mathcal{M}}(\bar{x}) = \hat{\partial} \delta_{\mathcal{M}}(\bar{x})$, and the *normal cone* by $N_{\mathcal{M}}(\bar{x}) = \partial \delta_{\mathcal{M}}(\bar{x})$. When $\mathcal{M}$ is convex, $\mathrm{ri}\, \mathcal{M}$ denotes its relative interior, and $\mathrm{par}\, \mathcal{M}$ denotes the unique linear subspace that is a translate of its affine span. We denote the closed unit ball by B.

We next summarize some basic ideas about manifolds: more details may be found in standard references such as [50]. In parallel with our terminology for functions, throughout Part II, **the term "manifold" means a** $\mathcal{C}^{(2)}$**-smooth manifold**. Given a manifold $\mathcal{M} \subset \mathbf{R}^n$ and a point $\bar{x} \in \mathcal{M}$, there exists a smooth map $F\colon \mathbf{R}^n \to \mathbf{R}^m$ with $F(\bar{x}) = 0$ and surjective Jacobian $\nabla F(\bar{x})$ such that

$$x \in \mathcal{M} \quad \Leftrightarrow \quad F(x) = 0 \qquad \text{for all } x \in \mathbf{R}^n \text{ near } \bar{x}. \tag{6.1}$$

Consequently [66, Example 6.8] both the normal cone $N_{\mathcal{M}}(\bar{x})$ and the regular normal cone $\widehat{N}_{\mathcal{M}}(\bar{x})$ coincide with the classical normal space to $\mathcal{M}$ at $\bar{x}$. The nearest-point projection $\mathrm{Proj}_{N_{\mathcal{M}}(x)}(y)$ is well-defined and depends continuously on points x, y near $\bar{x}$ with $x \in \mathcal{M}$. A function $h\colon \mathcal{M} \to \mathbf{R}$ is *smooth* around $\bar{x}$ if there exists a smooth function $\tilde{h}\colon \mathbf{R}^n \to \mathbf{R}$ such that $h(x) = \tilde{h}(x)$ for all points $x \in \mathcal{M}$ near $\bar{x}$. We then call $\tilde{h}$ a *smooth extension* of h at $\bar{x}$, and the *Riemannian gradient* of h at $\bar{x}$, denoted $\nabla_{\mathcal{M}} h(\bar{x})$, is the orthogonal projection of the gradient $\nabla \tilde{h}(\bar{x})$ onto the tangent space $T_{\mathcal{M}}(\bar{x})$. The projected vector is in fact independent of the choice of extension $\tilde{h}$. For example, the set $\mathcal{M} \subset \mathbf{R}^2$ defined by the equation $x_2 = x_1^2$ is a manifold, and the restriction $f|_{\mathcal{M}}$ of the function f defined by equation (1.1) is smooth because it agrees with the smooth function $\tilde{f}(x) = 5x_2 - 4x_1^2$ on $\mathcal{M}$. At the point $0 \in \mathbf{R}^2$, the Riemannian gradient $\nabla_{\mathcal{M}}(f|_{\mathcal{M}})$, which we abbreviate to $\nabla_{\mathcal{M}} f$, is zero.

We consider four objective classes of interest, beginning with prox-regularity [66].

Definition 6.2 A closed function $f\colon \mathbf{R}^n \to (-\infty, +\infty]$ is *prox-regular* at a point $\bar{x}$ for a subgradient $\bar{y} \in \partial f(\bar{x})$ if there exists a constant $\rho \geq 0$ such that all points x, x' near $\bar{x}$ with $f(x)$ near $f(\bar{x})$ and subgradients $y \in \partial f(x)$ near $\bar{y}$ satisfy the inequality $f(x') \geq f(x) + \langle y, x' - x\rangle - \rho|x' - x|^2$.

The second class [65] is clearly more restrictive than prox-regularity.

Definition 6.3 A closed function $f : \mathbf{R}^n \to (\infty, +\infty]$ is *primal lower nice* at a point $\bar{x} \in \mathbf{R}^n$ if there exist constants α and β such that all points x, x' near $\bar{x}$, and subgradients $y \in \partial f(x)$ satisfy

$$f(x') \geq f(x) + \langle y, x' - x \rangle - (\alpha + \beta|y|)|x' - x|^2. \tag{6.4}$$

If this property holds for all $\bar{x} \in \operatorname{dom} f$, then we simply call f *primal lower nice*.

The third class, discussed in [66, Definition 10.23 and Proposition 13.32], while versatile, is more restrictive still [65, Theorem 5.1].

Definition 6.5 At a point $\bar{x} \in \mathbf{R}^n$, a function f is *strongly amenable* if it has the local representation $f = g \circ F$ for some closed convex function $g : \mathbf{R}^m \to (-\infty, +\infty]$ and smooth map $F : \mathbf{R}^n \to \mathbf{R}^m$ such that $F(\bar{x})$ lies in $\operatorname{dom} g$ and the normal cone to $\operatorname{cl}(\operatorname{dom} g)$ there intersects the null space of the adjoint map $\nabla F(\bar{x})^*$ trivially.

In the scenario of Definition 6.5, a standard chain rule [66] implies

$$\partial f(\bar{x}) = \hat{\partial} f(\bar{x}) = \nabla F(\bar{x})^* \partial g(F(\bar{x})).$$

The final class can be seen, via property (6.1), to consist of particularly simple strongly amenable functions.

Definition 6.6 At a point $\bar{x} \in \mathbf{R}^n$, a function f is *smoothly embedded* if it has the local representation $f = g + \delta_{\mathcal{M}}$ for some smooth function $g : \mathbf{R}^n \to \mathbf{R}$ and some manifold $\mathcal{M}$ containing $\bar{x}$.

In the scenario of Definition 6.6, a standard sum rule [66, Corollary 10.9] implies

$$\partial f(\bar{x}) = \hat{\partial} f(\bar{x}) = \nabla g(\bar{x}) + N_{\mathcal{M}}(\bar{x}). \tag{6.7}$$

The function (1.1) is strongly amenable at zero, because we can represent it using the smooth map $F(x) = (x_2 - x_1^2, x_1)$ and the convex function $g(u, v) = 5|u| + v$. *Weakly convex* functions — those of the form $g - \rho| \cdot |^2$, for proper closed convex functions g and constant ρ (see for example [29]) — are easily seen to be strongly amenable. For locally Lipschitz functions, prox-regularity, the primal lower nice property, and strong amenability are equivalent, each coinciding with the "lower $\mathcal{C}^2$ property" [66]. Indeed, a locally Lipschitz f has each of these properties at a point if and only if it equals a weakly convex function locally [66, Exercise 10.36, Proposition 13.33].

In general, however, prox-regular functions may not be primal lower nice, even if they are continuous. For example, the function $f(x) = \sqrt{|x|}$ is prox-regular at zero, but not primal lower nice there. To see this, assume that inequality (6.4) holds. For any small $t > 0$, we can set $x = t^2$, $x' = 4t^2$, $y = \frac{1}{2t}$, to deduce

$$2t \geq t + \frac{1}{2t}3t^2 - \left(\alpha + \beta\frac{1}{2t}\right)(3t^2)^2.$$

This gives a contradiction for t sufficiently small. Also worth noting is that primal lower nice functions may not be weakly convex: a smoothly embedded function as described in Definition 6.6 is weakly convex if and only if the set $\mathcal{M}$ is locally affine.

We prove two simple tools for later use. First, the primal lower nice property is preserved by addition of smooth functions.

Proposition 6.8 (Primal lower nice preservation) *For a function $g\colon \mathbf{R}^n \to \mathbf{R}$ that is smooth around the point $\bar{x} \in \mathbf{R}^n$, if a closed function $f\colon \mathbf{R}^n \to (-\infty, +\infty]$ is primal lower nice at $\bar{x}$, then so is the sum $f + g$.*

Proof Using the notation above, we can assume that inequality (6.4) holds. Since g is smooth, its gradient is Lipschitz around $\bar{x}$. (In fact this property suffices for our proof.) Hence, for some constants $\gamma, \lambda > 0$, all points x and x' near $\bar{x}$ satisfy

$$g(x') \;\geq\; g(x) \;+\; \langle \nabla g(x), x' - x\rangle \;-\; \gamma |x' - x|^2 \tag{6.9}$$

and $|\nabla g(x)| \leq \lambda$.

Now consider the sum $h = f + g$. For any points x and x' near $\bar{x}$ and subgradient $w \in \partial h(x)$, the sum rule [66, Corollary 10.9] ensures the existence of a subgradient $y \in \partial f(x)$ satisfying $w = y + \nabla g(x)$. Adding the inequalities (6.4) and (6.9), we deduce

$$h(x') \;\geq\; h(x) \;+\; \langle w, x' - x\rangle \;-\; \big(\gamma + (\alpha + \beta|y|)\big)|x' - x|^2.$$

Since $|y| \leq \lambda + |w|$, the primal lower nice property for h follows. □

Secondly, we show that when considering primal lower nice functions, we lose no generality in assuming that they are bounded below.

Lemma 6.10 (Localization) *If a closed function $f\colon \mathbf{R}^n \to (-\infty, +\infty]$ is primal lower nice at a point $\bar{x} \in \mathbf{R}^n$, then there is another proper closed function that is primal lower nice and bounded below, and that agrees identically with f near $\bar{x}$.*

Proof For simplicity, suppose $\bar{x} = 0$. By definition, f is primal lower nice throughout the ball $2\delta B$ for some $\delta > 0$. Since f is closed, we can shrink $\delta > 0$ if necessary to ensure f is bounded below on $2\delta B$. The function $g\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ defined by

$$g(x) \;=\; \begin{cases} 0 & (|x| \leq \delta) \\ \frac{(|x|^2 - \delta^2)^3}{4\delta^2 - |x|^2} & (\delta < |x| < 2\delta) \\ +\infty & (|x| \geq 2\delta) \end{cases}$$

is smooth on the interior of $2\delta B$, so, by Proposition 6.8, the function $f + g$ is primal lower nice throughout its domain. But $f + g$ is also bounded below and agrees identically with f on the ball δB, as required. □

A particularly simple class of strongly amenable functions f arise when the convex function g in Definition 6.5 is given by $g(y) = \max_i y_i$. The resulting functions f illustrate well the theory of identifiability: we discuss them in the next section.

7 Max Functions

We consider the important basic example of *max functions*: those functions of the form $x \mapsto \max_i f_i(x)$ for $x \in \mathbf{R}^n$, where the index i ranges over some finite index set and each function f_i is continuously differentiable. Identifiable sets for max functions were studied in [52, Corollary 4.8], but under nondegeneracy conditions that often fail in practice [38]. By contrast, [35, Corollary 4.11] dispenses with nondegeneracy, but omits important details in the result and proof. Here we present a short and direct argument. We start with the smooth case.

Proposition 7.1 *If a function f on $\mathbf{R}^n$ is continuously differentiable around a point $\bar{x}$, with $\nabla f(\bar{x}) = 0$, then any set containing $\bar{x}$ is strongly critical at $\bar{x}$.*

Proof As noted, $|\nabla f|(x) = |\nabla f(x)|$ at all points x near $\bar{x}$. The result follows. □

The next result uses Proposition 2.7 and standard variational analysis: [66, Theorem 10.31] reveals the property $G(x) = \partial f(x)$. A direct proof is also short.

Proposition 7.2 (Slope for max functions) *Consider a nonempty finite index set I and continuously differentiable functions $f_i : \mathbf{R}^n \to \mathbf{R}$ for $i \in I$. For each point $x \in \mathbf{R}^n$, define the value $f(x) = \max_i f_i(x)$, and sets*

$$I(x) = \{i : f_i(x) = f(x)\} \quad and \quad G(x) = \mathrm{conv}\{\nabla f_i(x) : i \in I(x)\}.$$

Then the slope and limiting slope of f at x both equal $\mathrm{dist}\big(0, G(x)\big)$.

Proof Elementary calculus shows

$$|\nabla f|(x) = \max_{|y|=1} \min_{i \in I(x)} \langle \nabla f_i(x), y \rangle.$$

A simple limiting argument then shows that the slope and limiting slope agree everywhere, and the result then follows from the separating hyperplane theorem. □

For any index set I, the *support set* of a vector $\lambda \in \mathbf{R}^I$ is the set $\{i \in I : \lambda_i \neq 0\}$.

Theorem 7.3 (Identifiable sets for max functions) *With the assumptions of Proposition 7.2, consider a critical point $\bar{x}$ for the max function f. Let $\bar{I} = I(\bar{x})$, and denote by Ω the finite set consisting of support sets of vectors $\lambda \in \mathbf{R}_+^{\bar{I}}$ satisfying*

$$\sum_{i \in \bar{I}} \lambda_i = 1, \quad \sum_{i \in \bar{I}} \lambda_i \nabla f_i(\bar{x}) = 0. \tag{7.4}$$

Then the union of the sets $\mathcal{M}_J = \{x : I(x) \supset J\}$ as the support set J ranges over Ω is strongly critical and identifiable at $\bar{x}$.

Proof We first prove identifiability. Consider a sequence $x_r \to \bar{x}$ satisfying $|\nabla f|(x_r) \to 0$ and with $x_r \notin \mathcal{M}_J$ for $r = 1, 2, \ldots$ and all $J \in \Omega$. For large r, continuity implies $I(x_r) \subset \bar{I}$, and Proposition 7.2 then implies the existence of a vector

$\lambda^r \in \mathbf{R}_+^{\bar{I}}$ satisfying $\sum_i \lambda_i^r = 1$ and $\lambda_i^r = 0$ for all $i \notin I(x_r)$, with $\sum_i \lambda_i^r \nabla f_i(x_r) \to 0$. After taking a subsequence, λ^r converges to some vector $\bar{\lambda} \in \mathbf{R}_+^{\bar{I}}$ with support set $\bar{J} \subset \bar{I}$. Since $\bar{\lambda}$ satisfies equation (7.4), by continuous differentiability, $\bar{J} \in \Omega$. Consequently $I(x_r) \not\supset \bar{J}$ for each r, and hence there exists an index $i(r) \notin I(x_r)$ with $\bar{\lambda}_{i(r)} > 0$. Taking a further subsequence, some index $i \in \bar{I}$ satisfies $\bar{\lambda}_i > 0$ and $i \notin I(x_r)$ for all r. This gives the desired contradiction: $0 = \lambda_i^r \to \bar{\lambda}_i > 0$.

Turning to strong criticality, by Proposition 3.4, it suffices to prove, for each support set $J \in \Omega$, that the set $\mathcal{M}_J$ is strongly critical. Some corresponding solution $\lambda \in \mathbf{R}_+^{\bar{I}}$ of equation (7.4) has support set J. The continuously differentiable function $g = \sum_i \lambda_i f_i$ satisfies the inequality $g \leq f$, with equality throughout $\mathcal{M}_J$. Strong criticality follows by Propositions 3.5 and 7.1. $\qquad\qquad\square$

To illustrate, consider the max function on $\mathbf{R}^2$ with the two ingredients defined by $(u, v) \mapsto v^2 \pm u$ at the critical point $(0, 0)$. Theorem 7.3 then recovers Example 3.10. Although in this example the strongly critical identifiable set is a manifold (defined by $u = 0$), that might not be true more generally. For example, the function on $\mathbf{R}$ with ingredients the zero function and the function $x \mapsto -x$, at the critical point 0, has $\mathbf{R}_+$ as a strongly critical identifiable set. Furthermore, in general, the set Ω in Theorem 7.3 may contain several distinct supports. Consider, for example, the function on $\mathbf{R}$ with ingredients the zero function and the function $x \mapsto x^3$, at the critical point 0, in which case $\mathbf{R}$ is a strongly critical identifiable set. When the gradients $\nabla f_i(\bar{x})$ for $i \in \bar{I}$ are affinely independent, however, equation (7.4) has a unique solution $\bar{\lambda}$. If we fix some index i in the index set $I_> = \{i \in \bar{I} : \bar{\lambda}_i > 0\}$, then M is defined around $\bar{x}$ by the constraints

$$\begin{aligned} f_j(x) - f_i(x) = 0 \qquad & \text{for } j \in I_> \setminus \{i\} \\ f_j(x) - f_i(x) \leq 0 \qquad & \text{for } j \in \bar{I} \setminus I_>. \end{aligned}$$

The functions on the left-hand side are continuously differentiable, with linearly independent gradients, by the affine independence assumption. The inverse function theorem therefore guarantees that, locally, the set $\mathcal{M}$ is diffeomorphic to a polyhedral cone. Consequently, around $\bar{x}$, the strongly critical identifiable set $\mathcal{M}$ is normally embedded, so Theorem 4.5 applies.

8 Identifiability and Partial Smoothness

In Euclidean space, identifiability, an idea inspired by explorations in [25, 27, 40, 67], was introduced in [35, Definition 3.10]: by Corollary 3.14, that original definition coincides with ours. A primary aim of [35] was to relate identifiability and the idea of *partial smoothness* introduced in [52]. We now revisit that relationship using a fresh proof scheme, replacing the intricate proximal epigraphical analysis suggested in [35] by a direct and natural argument based on Theorem 4.2 (Linear growth). The next result, following a pattern observed in [36, Section 10], is the key tool.

Proposition 8.1 (Identifiable sets and subgradients) *Suppose that the closed function* $f \colon \mathbf{R}^n \to (-\infty, +\infty]$ *has slope zero at the point* $\bar{x}$. *Consider a closed, normally embedded set* $\mathcal{M} \subset \mathbf{R}^n$ *that is strongly critical and identifiable at* $\bar{x}$, *and has continuous restriction* $f|_{\mathcal{M}}$. *Then for all points* x *near* $\bar{x}$ *with value* $f(x)$ *near* $f(\bar{x})$ *and small vectors* $y \in \mathbf{R}^n$, *the following two properties are equivalent:*

(a) $y \in \partial f(x)$
(b) $x \in \mathcal{M}$ *and* $y \in \partial(f + \delta_{\mathcal{M}})(x)$.

The same equivalence holds for both regular and proximal subgradients. Furthermore, prox-regularity at x *for* y *holds for* f *if and only if it holds for* $f + \delta_{\mathcal{M}}$.

Proof We first prove the regular subdifferential result. In that case, property (a) implies $x \in \mathcal{M}$, by identifiability, and then property (b) follows from the inclusion $\hat{\partial} f(x) \subset \hat{\partial}(f + \delta_{\mathcal{M}})(x)$. Hence, in the regular case, we just need to prove (b) implies (a). By Theorem 4.5, there exists $\epsilon > 0$ such that all points $x \in B_{2\epsilon}^f(\bar{x})$ and nearest points $v \in \mathrm{Proj}_{\mathcal{M}}(x)$ satisfy $f(v) + \epsilon|v - x| \leq f(x)$. To complete the proof, we claim

$$x \in \mathcal{M} \cap B_\epsilon^f(\bar{x}), \quad y \in \hat{\partial}(f + \delta_{\mathcal{M}})(x), \quad \text{and} \quad |y| < \epsilon \quad \Rightarrow \quad y \in \hat{\partial} f(x).$$

Assume x and y satisfy the left-hand side properties. Then points $v \in \mathcal{M}$ satisfy

$$f(v) \geq f(x) + \langle y, v - x \rangle + o(|v - x|) \quad \text{as } v \to x.$$

If $y \notin \hat{\partial} f(x)$, then there exists $\delta > 0$ and a sequence $z_s \to x$ satisfying

$$f(z_s) - f(x) - \langle y, z_s - x \rangle < -\delta|z_s - x|.$$

For all large s, since f is closed, $f(z_s) > f(x) - \epsilon$, and we also have $f(z_s) < f(x) + \epsilon$, so in fact $z_s \in B_\epsilon^f(x) \subset B_{2\epsilon}^f(\bar{x})$. Choosing $v_s \in \mathrm{Proj}_{\mathcal{M}}(z_s)$, we have

$$\begin{aligned} f(z_s) &\geq f(v_s) + \epsilon|v_s - z_s| \\ &\geq f(x) + \langle y, v_s - x \rangle + \epsilon|v_s - z_s| + o(|v_s - x|) \\ &= f(x) + \langle y, z_s - x \rangle + \langle y, v_s - z_s \rangle + \epsilon|v_s - z_s| + o(|v_s - x|) \\ &\geq f(x) + \langle y, z_s - x \rangle + o(|v_s - x|), \end{aligned}$$

so $-\delta|z_s - x| > o(|v_s - x|)$, and hence in fact $|z_s - x| = o(|v_s - x|)$, contradicting

$$|v_s - x| \leq |v_s - z_s| + |z_s - x| \leq 2|z_s - x|.$$

This completes the proof in the regular case. The analogous proof for proximal subgradients is almost identical, and the result for subgradients follows from the regular case by a simple limiting argument. Finally, prox-regularity of f at x for a proximal subgradient y depends only on the behavior of f-attentive localizations of the mapping ∂f around (x, y) [66, Theorem 13.36], so the final claim follows. $\qquad\square$

Definition 8.2 Given a closed function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$, consider a point $\bar{x} \in \mathrm{dom}\, f$ and a manifold $\mathcal{M}$ containing $\bar{x}$. Then f is *partly smooth* at $\bar{x}$ for a subgradient $\bar{y} \in \partial f(\bar{x})$ relative to $\mathcal{M}$ if it satisfies the following properties:

Prox-regularity:	the function f is prox-regular at $\bar{x}$ for $\bar{y}$.
Restricted smoothness:	the restriction $f|_{\mathcal{M}}$ is smooth around $\bar{x}$.
Sharpness:	the subspace $\mathrm{par}\, \hat{\partial} f(\bar{x})$ is just the normal space $N_{\mathcal{M}}(\bar{x})$.
Inner semicontinuity:	for all $y \in \partial f(\bar{x})$ near $\bar{y}$ and sequences $x_r \to \bar{x}$ in $\mathcal{M}$, there exist $y_r \in \partial f(x_r)$ converging to y.

This definition, from [36], slightly refined the original [52]. We note an easy example.

Proposition 8.3 *Any smoothly embedded function is partly smooth at every point in its domain $\mathcal{M}$ relative to $\mathcal{M}$.*

Proof As we have observed, any smoothly embedded function $f = g + \delta_{\mathcal{M}}$, for a smooth function g and a manifold $\mathcal{M}$, must be prox-regular. The restricted smoothness condition is immediate. The sharpness condition follows from formula (6.7). Finally, given any subgradient $y \in \partial f(\bar{x})$, define a vector $\tilde{y} = y - \nabla g(\bar{x}) \in N_{\mathcal{M}}(\bar{x})$. Given any sequence $x_r \to \bar{x}$ in $\mathcal{M}$, we have

$$\nabla g(x_r) + \mathrm{Proj}_{N_{\mathcal{M}}(x_r)}(\tilde{y}) \;\to\; \nabla g(\bar{x}) + \mathrm{Proj}_{N_{\mathcal{M}}(\bar{x})}(\tilde{y}) \;=\; \nabla g(\bar{x}) + \tilde{y} \;=\; y.$$

The left-hand side vector lies in $\partial f(x_r)$, so inner semicontinuity follows. □

We focus on partial smoothness for the subgradient $\bar{y} = 0$. By Theorem 2.7, the following result follows immediately from the inner semicontinuity condition.

Proposition 8.4 *If a closed function $f \colon \mathbf{R}^n \to (-\infty, +\infty]$ is partly smooth for zero at a point $\bar{x}$ relative to a manifold $\mathcal{M}$, then $\mathcal{M}$ is strongly critical at $\bar{x}$.*

Although the properties required for partial smoothness appear at first to comprise a complex list, remarkably, in semi-algebraic optimization, partial smoothness holds generically [34], along with the *nondegeneracy condition* $0 \in \mathrm{ri}\, \hat{\partial} f(\bar{x})$. More precisely, for all vectors y in a dense open semi-algebraic subset of $\mathbf{R}^n$, the perturbed function $f + \langle y, \cdot \rangle$ is partly smooth at all critical points x, each of which furthermore must be nondegenerate. Consequently, the perturbed function has an identifiable manifold at each critical point, for zero. Thus, for generic semi-algebraic optimization problems, we expect to see identifiable manifolds around the solutions. Partial smoothness and nondegeneracy also imply an important constant-rank property for the subdifferential operator ∂f [53, Theorem 6.5].

Most striking from our current perspective is the following result [36, Proposition 10.12]: partial smoothness and nondegeneracy together are equivalent to a seemingly simpler identifiability property with respect to a manifold. The argument is now a straightforward application of Proposition 8.1.

Theorem 8.5 (Identifiability and partial smoothness) *Consider a closed function $f \colon \mathbf{R}^n \to (-\infty, +\infty]$ with slope zero at a point $\bar{x}$ lying in a manifold $\mathcal{M}$. Suppose that $f|_{\mathcal{M}}$ is smooth around $\bar{x}$. Then $\mathcal{M}$ is identifiable at $\bar{x}$ if and only if f is partly smooth at $\bar{x}$ for zero relative to $\mathcal{M}$ with $0 \in \mathrm{ri}\, \hat{\partial} f(\bar{x})$.*

Proof Suppose that the manifold $\mathcal{M}$ is identifiable at the point $\bar{x}$. By assumption, the restriction $f|_{\mathcal{M}}$ has a smooth extension $\tilde{f}$ at $\bar{x}$. Like the Euclidean case, it is easy to verify that the slope $|\nabla(f|_{\mathcal{M}})|(x)$ is just given by $|\nabla(f|_{\mathcal{M}})(x)|$, at all points x near $\bar{x}$, and hence is continuous at $\bar{x}$. Applying Theorem 4.5, we know $\mathcal{M}$ is strongly critical at $\bar{x}$. Consider the function $h = f + \delta_{\mathcal{M}} = \tilde{f} + \delta_{\mathcal{M}}$. We have

$$0 \in \hat{\partial} f(\bar{x}) \subset \hat{\partial} h(\bar{x}) = \nabla \tilde{f}(\bar{x}) + N_{\mathcal{M}}(\bar{x}),$$

so $\nabla \tilde{f}(\bar{x}) \in N_{\mathcal{M}}(\bar{x})$ and $\hat{\partial} h(\bar{x}) = N_{\mathcal{M}}(\bar{x})$. Using Proposition 8.1, the convex sets $\hat{\partial} h(\bar{x})$ and $\hat{\partial} f(\bar{x})$ are identical near zero so the nondegeneracy condition $0 \in \mathrm{ri}\, \hat{\partial} f(\bar{x})$ and the sharpness condition follow, and since h is smoothly embedded, it is prox-regular at $\bar{x}$ for 0, whence so is f. Restricted smoothness holds by assumption.

Inner semicontinuity also holds by Proposition 8.1. To see this claim, note that any small subgradient $y \in \partial f(\bar{x})$ also lies in $\partial h(\bar{x})$. Given any sequence $x_r \to \bar{x}$ in $\mathcal{M}$, since the smoothly embedded function h is partly smooth by Proposition 8.3, there exist subgradients $y_r \in \partial h(x_r)$ approaching y. For all large r we have $y_r \in \partial f(x_r)$, so the claim follows.

Our proof in the converse direction is more standard, following [42] but shorter. If the result fails, then by Corollary 3.14 there exists a sequence of points $x_r \to_f \bar{x}$ in $\mathcal{M}^c$ with subgradients $y_r \in \partial f(x_r) \to 0$. For large r, the point x_r has a unique nearest point x_r' in $\mathcal{M}$. Normalizing the vector $x_r - x_r'$ gives a unit vector $w_r \in N_{\mathcal{M}}(x_r')$, which, taking a subsequence, we can suppose converges to a unit vector $w \in N_{\mathcal{M}}(\bar{x})$. The sharpness and nondegeneracy conditions imply the existence of a constant $\epsilon > 0$ such that $\epsilon w \in \partial f(\bar{x})$, and the continuity condition then implies the existence of a sequence of subgradients $y_r' \in \partial f(x_r') \to \epsilon w$. Since f is prox-regular at $\bar{x}$ for 0, there exists a constant $C > 0$ such that, for all large r we have

$$f(x_r') \geq f(x_r) + \langle y_r, x_r' - x_r \rangle - C|x_r' - x_r|^2 \quad \text{and}$$
$$f(x_r) \geq f(x_r') + \langle y_r', x_r - x_r' \rangle - C|x_r' - x_r|^2,$$

and hence $\langle y_r' - y_r, w_r \rangle \leq 2C|x_r - x_r'|$. But in this inequality, the left-hand side converges to ϵ, contradicting the convergence of the right-hand side to zero. $\square$

The original result, [36, Proposition 10.12], concerns partial smoothness for an arbitrary subgradient, rather than zero specifically. This more general result follows immediately by adding a linear perturbation to the objective f.

Partly Analytic Functions

We end this section by discussing an important special case of Theorem 8.5. We present a simple tool that often suffices for verifying the KL property of Sect. 5 and helps find desingularizers. We describe how the KL property can decouple, becoming, like partial smoothness, a simple and intuitive hybrid of smooth and sharp behavior. In general, semi-algebraic (or tame) functions always have the KL property, but the proof [17] is not straightforward, revealing little about desingularizers. For *generic* semi-algebraic objectives, on the other hand — specifically for the sum of any specific

semi-algebraic function and a generic linear perturbation — the situation is simpler: [16, 34] prove the existence of an identifiable analytic manifold on which the objective is analytic. We call this situation the *partly analytic* case.

We can equip any manifold $\mathcal{M}$ in Euclidean space with two natural metrics: the metric d induced by the Euclidean distance, and the Riemannian distance d_{Rie}. For any points $x, y \in \mathcal{M}$, $d_{\mathrm{Rie}}(x, y)$ is the infimum of the length of paths in $\mathcal{M}$ from x to y [51]. We can move easily between these metrics using the following technique.

Proposition 8.6 *Given a manifold $\mathcal{M}$ in Euclidean space, consider a function $g \colon \mathcal{M} \to (-\infty, +\infty]$, a point $\bar{x} \in \mathrm{dom}\, g$, and a desingularizer ϕ. Then, with respect to the Riemannian and induced Euclidean metrics d_{Rie} and d, the KL properties for g at $\bar{x}$ with desingularizer ϕ are equivalent.*

Proof This is a consequence of the following fact, proved in [32, Proposition 3.1]:

$$\lim_{\substack{x, y \to \bar{x} \\ x \neq y}} \frac{d_{\mathrm{Rie}}(x, y)}{d(x, y)} = 1.$$

$\square$

In the partly analytic case, the KL property follows from the following result. The result does not in fact require tameness. For example, it applies to the objective $x \mapsto f(x) + |x|$, for the function f defined in Example 5.3. This objective is not tame, but does satisfy the KL property at zero, since $\{0\}$ is an identifiable set.

Corollary 8.7 (Partly analytic functions) *Suppose that $f \colon \mathbf{R}^n \to (-\infty, +\infty]$ is closed and has slope zero at a point $\bar{x}$. If some analytic manifold $\mathcal{M}$ is identifiable at $\bar{x}$ and the restriction $f|_{\mathcal{M}}$ is analytic, then the KL property for f holds at $\bar{x}$, with KL exponent equal to that of $f|_{\mathcal{M}}$ at $\bar{x}$.*

Proof Analytic functions on analytic manifolds satisfy the KL property [49, Section 9]. The conclusion then follows from Theorem 5.12 and Proposition 8.6. $\square$

9 Subgradient Curves and Identification

The property of identifiability fundamentally concerns critical sequences. We now move on from that discrete-time perspective to study continuous-time trajectories. Consider once again the question of minimizing an objective function f on a metric space. The central continuous-time analogue of the sequences considered thus far is the family of curves of "maximal slope" [2, 31]. Around some given point x^*, we pose the question of the existence of a small set $\mathcal{M}$ such that curves of maximal slope converging to x^* always remain in $\mathcal{M}$ eventually.

Figure 1 illustrates this behavior for the simple nonsmooth nonconvex objective $5|x_2 - x_1^2| + x_1^2$, around the minimizer $x^* = 0$: the associated small set $\mathcal{M}$ is defined by $x_2 = x_1^2$. The behavior is the precise analogue of the phenomenon we observed in preceding sections, where various iterative procedures for minimizing f identify an associated set $\mathcal{M}$ around x^*: sequences of iterates converging to x^* must eventually lie in $\mathcal{M}$, thereby revealing some solution structure like constraint activity, sparsity, or

matrix rank. For objectives f on general metric spaces, the question of identification in continuous time remains open. As a first step, therefore, we focus on the Euclidean case, $f : \mathbf{R}^n \to (-\infty, +\infty]$. Following the arguments of [33], for interesting functions f the question then typically reduces to the behavior of *subgradient curves*: locally absolutely continuous maps $x : \mathbf{R}_+ \to \operatorname{dom} f$ satisfying

$$x'(t) \in -\partial f(x(t)) \qquad \text{for almost all times } t > 0.$$

Our main continuous-time result, Theorem 9.2, demonstrates for the first time identification behavior for subgradient curves.

The significance of subgradient curves from an optimization perspective has been highlighted in [3, 6]. Existing literature covers a wide class of nonsmooth and non-convex functions f for which they must converge. In particular, the existence and uniqueness of subgradient curves, given any initial point $x(0)$ in $\operatorname{dom} f$, was shown for convex functions f in a well-known 1973 monograph of Brézis [22]. In fact this result holds more generally, under the assumption that f is primal lower nice, as shown in the following theorem from [61].

Theorem 9.1 (Existence and uniqueness of subgradient curves) *If the closed function $f : \mathbf{R}^n \to (-\infty, +\infty]$ is primal lower nice and bounded below, then there exists a unique subgradient curve $x(\cdot)$ corresponding to any initial point $x(0) \in \operatorname{dom} f$. The curve furthermore satisfies*

$$\int_0^\infty |x'(t)|^2 \, dt \; < \; +\infty.$$

For closed convex functions with minimizers, following [22], Bruck proved that subgradient curves converge to minimizers [23]. Subgradient curves are known to converge more generally, assuming some version of the KL property, holding in particular for closed semi-algebraic (or subanalytic) functions on bounded domains [15, 17]. Such convergence results hold for all closed convex functions, "lower $\mathcal{C}^2$" functions, and weakly convex functions that are bounded below [15, 18]. We note that these results generalize: for a primal lower nice function f that is bounded below, if the KL property holds throughout some subgradient curve, then a simple argument using [61, Theorem 3.2] shows that the curve must have finite length and hence converge. We deduce, for example, that all subgradient curves for the function (1.1) converge to the minimizer at zero. To summarize, we can reasonably focus our current study on subgradient curves that converge.

The following is our main continuous-time result.

Theorem 9.2 (Identification for subgradient curves) *Consider a closed function $f : \mathbf{R}^n \to (-\infty, +\infty]$ and a subgradient curve $x(\cdot)$ converging to a point $\bar{x}$ at which f is primal lower nice and has a closed identifiable set $\mathcal{M}$. Then, after a finite time, $x(\cdot)$ lies on $\mathcal{M}$, and becomes a subgradient curve for the function $f + \delta_{\mathcal{M}}$.*

To derive continuous-time identification results of this general type, we must prove that the speed of any subgradient curve always converges (essentially) to zero. In the

convex case, this essential convergence property is well known. Indeed, the convex special case of Theorem 9.2 is part of the identification folklore, although we do not know a reference. In the next section, we prove essential convergence in the primal lower nice case, and hence deduce Theorem 9.2.

In Theorem 9.2, if the identifiable set $\mathcal{M}$ is a manifold, then we deduce that the dynamic of the subgradient curves is eventually governed by a smooth system. Before stating the general result, we consider a special case.

Proposition 9.3 (Smooth restrictions) *Suppose that the set $\mathcal{M} \subset \mathbf{R}^n$ is a manifold around the point $\bar{x}$, and that the function $f : \mathbf{R}^n \to (-\infty, +\infty]$ is smooth around $\bar{x}$. Then, near $\bar{x}$, subgradient curves $x(\cdot)$ for the function $f + \delta_{\mathcal{M}}$ are just smooth curves in $\mathcal{M}$ solving the classical differential equation*

$$x'(t) = -\nabla_{\mathcal{M}} f(x(t)). \tag{9.4}$$

Proof By definition, at all times t for which $x(t)$ is near $\bar{x}$, we must also have $x(t) \in \mathcal{M}$, and furthermore, by the sum rule for subdifferentials [66],

$$x'(t) \in -\nabla f(x(t)) + N_{\mathcal{M}}(x(t)) \quad \text{almost surely.}$$

At such times t we must have $x'(t) \in T_{\mathcal{M}}(x(t))$ since $x(s) \in \mathcal{M}$ for all time s near t. Therefore we deduce equation (9.4). The smoothness of $x(\cdot)$ now follows from classical smooth initial value theory [41]. □

The general case asserts the same eventual behavior for a broader class of functions.

Theorem 9.5 (Subgradient curves on manifolds) *Consider a closed function $f : \mathbf{R}^n \to (-\infty, +\infty]$ and a subgradient curve $x(\cdot)$ converging to a point $\bar{x}$ at which f is primal lower nice and has an identifiable manifold $\mathcal{M}$. If $f|_{\mathcal{M}}$ is smooth around $\bar{x}$, then, after a finite time, $x(\cdot)$ becomes a smooth curve on $\mathcal{M}$, satisfying*

$$x(t) \in \mathcal{M} \quad \text{and} \quad x'(t) = -\nabla_{\mathcal{M}} f(x(t)).$$

Proof Manifolds are locally closed. By Theorem 9.2, the curve $x(\cdot)$ eventually lies in $\mathcal{M}$ and becomes a subgradient curve of the function $f + \delta_{\mathcal{M}}$. Let $\tilde{f}$ be a smooth extension of f. Then $f + \delta_{\mathcal{M}} = \tilde{f} + \delta_{\mathcal{M}}$. The result follows by Proposition 9.3. □

10 Trajectories for Primal Lower Nice Functions

Theorem 9.1 listed several important properties of subgradient curves for primal lower nice functions. For our main result, Theorem 9.2, we rely crucially on one more property. We prove that velocity vectors $x'(\cdot)$ for bounded subgradient curves $x(\cdot)$ *essentially converge to zero*: in other words, for all $\epsilon > 0$, there exists a time T such that for almost all times $t \geq T$ we have $|x'(t)| < \epsilon$.

Theorem 10.1 *If a closed function $f : \mathbf{R}^n \to (-\infty, +\infty]$ is primal lower nice and bounded below, then the velocity of any bounded subgradient curve for f essentially converges to zero.*

Proof Consider a bounded subgradient curve $x(\cdot)$. Suppose $x([0, \infty)) \subset rB$. A standard compactness argument guarantees constants α, β and $\delta > 0$ such that the primal lower nice inequality (6.4) holds for all points x, $x' \in rB$ with $|x - x'| < 2\delta$.

By assumption, there exists a full-measure subset Ω of the interval $[0, +\infty)$ such that $x'(t) \in -\partial f(x(t))$ for all $t \in \Omega$. Consider the measure ρ on $\mathbf{R}_+$ with density $|x'(\cdot)|$. By [61, Lemma 2.1], the primal lower nice property ensures the existence of constants μ, $\nu > 0$ such that

$$|x'(s)| \leq |x'(t)| \exp(\mu(s - t) + \nu\rho[t, s])$$

for all times $t < s$ in Ω satisfying $x([t, s]) \subset x(t) + \delta B$.

If the result fails, there is a constant $\epsilon > 0$ and a sequence of times $s_j \to +\infty$ in the set Ω satisfying $|x'(s_j)| > \epsilon$ for all j. Taking a subsequence, we can suppose, for each j, the inequality $s_{j+1} - s_j > 1$, from which we deduce either

$$|x'(s_j)| \leq |x'(t)| \exp(\mu + \nu\rho[t, s_j]) \quad \text{for all } t \in [s_j - 1, s_j] \cap \Omega, \quad (10.2)$$

or there exists $t \in [s_j - 1, s_j]$ such that

$$|x(t) - x(s_j)| > \delta. \tag{10.3}$$

Assume (10.2) holds. Then $|x'(t)| \exp(\nu\rho[t, s_j]) > \epsilon e^{-\mu}$ for all $t \in [s_j - 1, s_j] \cap \Omega$. Hence either some time t satisfies $\nu\rho[t, s_j] \geq 1$, in which case $\rho[s_j - 1, s_j] \geq \frac{1}{\nu}$, or there is no such time t, in which case all $t \in [s_j - 1, s_j] \cap \Omega$ satisfy the inequality $|x'(t)| > \epsilon e^{-\mu-1}$, implying $\rho[s_j - 1, s_j] \geq \epsilon e^{-\mu-1}$. On the other hand, if (10.3) holds, we easily deduce

$$\rho[s_j - 1, s_j] \geq \rho[t, s_j] \geq |x(t) - x(s_j)| > \delta.$$

We have thus shown the existence of a constant $\epsilon > 0$ such that $\rho[s_{j-1}, s_j] \geq \epsilon$ for each j, and hence, by Hölder's inequality,

$$\int_{s_j-1}^{s_j} |x'(\tau)|^2 d\tau \geq \epsilon^2.$$

But this contradicts the conclusion of Theorem 9.1. $\qquad\square$

Theorems 9.1 and 10.1 concern functions that are bounded below. However, our analysis here is local, so Lemma 6.10 shows that the boundedness assumption involves no loss of generality. The following result is a local version of Theorem 10.1.

Theorem 10.4 (Essential convergence)
If a closed function $f : \mathbf{R}^n \to (-\infty, +\infty]$ is primal lower nice at a point $\bar{x} \in \mathbf{R}^n$, and a subgradient curve $x(\cdot)$ for f converges to $\bar{x}$, then $\bar{x}$ is critical, the function value $f(x(\cdot))$ converges to $f(\bar{x})$, and the velocity $x'(\cdot)$ essentially converges to zero.

Proof The first two conclusions essentially appear in [61], and the third appears in Theorem 10.1, all under the additional assumption that f is bounded below and primal lower nice throughout its domain. More generally, we can apply Proposition 6.10 to replace f by a function that is primal lower nice and bounded below, and that is unchanged near the point $\bar{x}$. The original subgradient curve converges to $\bar{x}$, so it coincides eventually with a subgradient curve for the new function. □

The proof of our main continuous-time result is now straightforward.

Proof of Theorem 9.2 By Theorem 10.4, the function value $f(x(\cdot))$ converges to $f(\bar{x})$, and there exists a full-measure set $\Omega \subset \mathbf{R}_+$ such that $x'(t) \in -\partial f(x(t))$ for all $t \in \Omega$ and $x'(t) \to 0$ as $t \to +\infty$ in Ω. Hence by Corollary 3.14, we know $x(t) \in \mathcal{M}$ for all large $t \in \Omega$. Note that since f is primal lower nice at the critical point $\bar{x}$, it has slope zero there. We then deduce by Proposition 8.1 that $x'(t) \in -\partial(f + \delta_{\mathcal{M}})(x(t))$ for all large $t \in \Omega$. Since the set $\mathcal{M}$ is closed, the subgradient curve $x(\cdot)$, being continuous, eventually lies in $\mathcal{M}$. □

An alternative approach to Theorem 9.5 bypasses the essential convergence result. We proceed as follows. By the square integrability of the velocity x', there exists a sequence of times $t_k \to \infty$ for which $x(t_k) \in -\partial f(x(t_k))$ for all k and $x'(t_k) \to 0$. We deduce $x(t_k) \in \mathcal{M}$ for all large k, and hence, by [36, Proposition 10.12], $\nabla_{\mathcal{M}} f(y) \in \partial f(y)$ for all points $y \in \mathcal{M}$ near $\bar{x}$. Consequently, for sufficiently large k, by the uniqueness of the subgradient curve starting at the point $x(t_k)$, we know that the trajectory $x(t_k + \cdot)$ coincides with the solution $\tilde{x} : \mathbf{R}_+ \to \mathcal{M}$ of the initial value problem $\tilde{x}'(t) = -\nabla_{\mathcal{M}} f(\tilde{x}(t))$ with $\tilde{x}(0) = x(t_k)$.

References

1. P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008.
2. L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser-Verlag, Basel, 2nd edition, 2008.
3. H. Attouch and J. Bolte. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Math. Program.*, 116:5–16, 2009.
4. H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality. *Math. Oper. Res.*, 35:438–457, 2010.
5. H. Attouch, J. Bolte, and B.F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. *Math. Program.*, 137(1-2, Ser. A):91–129, 2013.

6. H. Attouch and M. Teboulle. Regularized Lotka-Volterra dynamical system as continuous proximal-like method in optimization. *JOTA*, 121:541–570, 2004.
7. D. Azé and J.-N. Corvellec. On some variational properties of metric spaces. *J. Fixed Point Theory Appl.*, 5:185–200, 2009.
8. D. Azé and J.-N. Corvellec. Nonlinear error bounds via a change of function. *Journal of Optimization Theory and Applications*, 172:9–32, 2017.
9. M. Bačák. The proximal point algorithm in metric spaces. *Isr. J. Math.*, 194:689–701, 2013.
10. M. Bačák, I. Searston, and B. Sims. Alternating projections in cat(0) spaces. *Journal of Mathematical Analysis and Applications*, 385:599–607, 2012.
11. C.G. Bento, J.X. Cruz Neto, and P.R. Oliveira. Convergence of inexact descent methods for nonconvex optimization on Riemannian manifolds. arXiv:1103.4828, 2011.
12. G. Bento, O. Ferreira, and P. Oliveira. Local convergence of the proximal point method for a special class of nonconvex functions on Hadamard manifolds. *Nonlinear Analysis*, 73:564–572, 2010.
13. D.P. Bertsekas. On the Goldstein-Levitin-Polyak gradient projection method. *IEEE Transactions on Automatic Control*, 21:174 – 184, 1976.
14. L. Birbrair and T. Mostowski. Normal embeddings of semialgebraic sets. *Michigan Math. J.*, 47:125–132, 2000.
15. J. Bolte, A. Daniilidis, and A.S. Lewis. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.*, 17:1205–1223, 2006.
16. J. Bolte, A. Daniilidis, and A.S. Lewis. Generic optimality conditions for semialgebraic convex programs. *Math. Oper. Res.*, 36:55–70, 2011.
17. J. Bolte, A. Daniilidis, A.S. Lewis, and M. Shiota. Clarke subgradients of stratifiable functions. *SIAM J. Optimization*, 18(2):556–572, 2007.
18. J. Bolte, A. Daniilidis, O. Ley, and L. Mazet. Characterizations of łojasiewicz inequalities: subgradient flows, talweg, convexity. *TAMS*, 362:3319–3363, 2010.
19. J. Bolte, T.P. Nguyen, J. Peypouquet, and B. Suter. From error bounds to the complexity of first order descent methods for convex functions. *Math. Programming*, 165:471–507, 2017.
20. J. Bolte and E. Pauwels. Majorization-minimization procedures and convergence of SQP methods for semi-algebraic and tame programs. *Mathematics of Operations Research*, 41:442–465, 2016.
21. N. Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, Cambridge, 2023.
22. H. Brézis. *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*. North-Holland ,New York, 1973.
23. R.E. Bruck. Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. *J. Functional Analysis*, 18:15–26, 1975.
24. D. Burago, Yu Burago, and S. Ivanov. *A Course in Metric Geometry*. Amer. Math. Soc., Providence, RI, 2001.
25. J.V. Burke. On the identification of active constraints. II. The nonconvex case. *SIAM J. Numer. Anal.*, 27(4):1081–1103, 1990.
26. J.V. Burke and M.C. Ferris. Weak sharp minima in mathematical programming. *SIAM J. Control Optim.*, 31:1340–1359, 1993.
27. J.V. Burke and J.J. Moré. On the identification of active constraints. *SIAM J. Numer. Anal.*, 25(5):1197–1211, 1988.
28. P.H. Calamai and J.J. Moré. Projected gradient methods for linearly constrained problems. *Math. Program.*, 39(1):93–116, 1987.
29. D. Davis and D. Drusvyatskiy. Stochastic model-based minimization of weakly convex functions. *SIAM J. Optim.*, 29:207–239, 2019.
30. D. Davis, D. Drusvyatskiy, and V. Charisopoulos. Stochastic algorithms with geometric step decay converge linearly on sharp functions. *Mathematical Programming*, 2023.
31. 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. Nat. (8)*, 68:180–187, 1980.
32. A.L. Dontchev and A.S. Lewis. Perturbations and metric regularity. *Set-Valued Anal.*, 13:417–438, 2005.
33. D. Drusvyatskiy, A.D. Ioffe, and A.S. Lewis. Curves of descent. *SIAM J. Control Optim.*, 53(1):114–138, 2015.
34. D. Drusvyatskiy, A.D. Ioffe, and A.S. Lewis. Generic minimizing behavior in semi-algebraic optimization. *SIAM J. Optim.*, 26:513–534, 2016.

35. D. Drusvyatskiy and A.S. Lewis. Optimality, identifiability, and sensitivity. *Math. Program.*, 147:467–498, 2014.
36. D. Drusvyatskiy and A.S. Lewis. Optimality, identifiability, and sensitivity. arXiv:1207.6628, 2014.
37. I. Ekeland. On the variational principle. *J. Math. Anal. Appl.*, 47:324–353, 1974.
38. J. Fadili, J. Malick, and G. Peyré. Sensitivity analysis for mirror-stratifiable convex functions. *SIAM Journal on Optimization*, 28:2975–3000, 2018.
39. O. Ferreira and P. Oliveira. Proximal point algorithm on Riemannian manifolds. *Optimization*, 51:257–270, 2002.
40. S.D. Flåm. On finite convergence and constraint identification of subgradient projection methods. *Math. Program.*, 57:427–437, 1992.
41. E. Hairer. Solving differential equations on manifolds. https://www.unige.ch/~hairer/poly-sde-mani.pdf, 2011.
42. W.L. Hare and A.S. Lewis. Identifying active constraints via partial smoothness and prox-regularity. *J. Convex Anal.*, 11(2):251–266, 2004.
43. W.L. Hare and A.S. Lewis. Identifying active manifolds. *Algorithmic Oper. Res.*, 2(2):75–82, 2007.
44. D. Hauer and J.M. Mazón. Kurdyka–Lojasiewicz–Simon inequality for gradient flows in metric spaces. *TAMS*, 372:4917–4976, 2019.
45. S. Hosseini. Convergence of nonsmooth descent methods via Kurdyka–Łojasiewicz inequality on Riemannian manifolds. INS Preprint No. 1523., 2015.
46. A.D. Ioffe. An invitation to tame optimization. *SIAM Journal on Optimization*, 19(4):1894–1917, 2009.
47. A.D Ioffe. *Variational Analysis of Regular Mappings*. Springer US, 2017.
48. J. Jost. Convex functionals and generalized harmonic maps into spaces of non positive curvature. *Comment. Math. Helvetici*, 70:659–673, 1995.
49. K. Kurdyka, T. Mostowski, and A. Parursiński. Proof of the gradient conjecture of R. Thom. *Annals of Mathematics*, 152:763–792, 2000.
50. J.M. Lee. *Introduction to Smooth Manifolds*. Springer, New York, 2003.
51. J.M. Lee. *Introduction to Riemannian Manifolds*. Springer, 2nd edition, 2018.
52. A.S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM J. Optim.*, 13:702–725, 2002.
53. A.S. Lewis, Jingwei Liang, and Tonghua Tian. Partial smoothness and constant rank. *SIAM J. Optim.*, 2022.
54. A.S. Lewis, G. López, and A. Nicolae. Local linear convergence of alternating projections in metric spaces with bounded curvature. *SIAM Journal on Optimization*, 32:1094–1119, 2022.
55. A.S. Lewis and Tonghua Tian. The complexity of first-order optimization methods from a metric perspective. arXiv:2305.03208, 2023.
56. A.S. Lewis and S.J. Wright. A proximal method for composite minimization. *Math. Program.*, pages 1–46, 2015.
57. C. Li, G. López, and V. Martín-Márquez. Monotone vector fields and the proximal point algorithm on Hadamard manifolds,. *Journal of the London Mathematical Society*, 79:663–683, 2009.
58. J. Liang, J. Fadili, and G. Peyré. Local linear convergence of forward–backward under partial smoothness. In Z. Ghahramani, M. Welling, C. Cortes, N.d. Lawrence, and K.q. Weinberger, editors, *NeurIPS*, pages 1970–1978. Curran Associates, Inc., 2014.
59. J. Liang, J. Fadili, G. Peyré, and R. Luke. Activity identification and local linear convergence of Douglas-Rachford/ADMM under partial smoothness. In *Scale Space and Variational Methods in Computer Vision*, pages 642–653. Springer US, 2015.
60. S. Łojasiewicz. Ensembles semi-analytiques. Preprint, IHES, 1965.
61. S. Marcellin and L. Thibault. Evolution problems associated with primal lower nice functions. *J. Convex Anal.*, 13:385–421, 2006.
62. R. Mifflin and C. Sagastizábal. $\mathcal{VU}$-smoothness and proximal point results for some nonconvex functions. *Optim. Methods Softw.*, 19(5):463–478, 2004.
63. G.J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Mathematical Journal*, 29:341–346, 1962.
64. D. Noll. Convergence of non-smooth descent methods using the Kurdyka–Łojasiewicz inequality. *Journal of Optimization Theory and Applications*, 160:553–572, 2014.
65. R.A. Poliquin. Integration of subdifferentials of nonconvex functions. *Nonlinear Anal.*, 17:385–398, 1991.
66. R.T. Rockafellar and R.J-B. Wets. *Variational Analysis*. Grundlehren der mathematischen Wissenschaften, Vol 317, Springer, Berlin, 1998.

67. S.J. Wright. Identifiable surfaces in constrained optimization. *SIAM J. Control Optim.*, 31:1063–1079, July 1993.
68. C. Zalinescu. *Convex Analysis in General Vector Spaces*. World Scientific, NJ, 2002.
69. Hongyi Zhang and S. Sra. First-order methods for geodesically convex optimization. *Journal of Machine Learning Research*, 49:1–22, 2016.