Identifiability, the KL property in metric spaces, and subgradient curves

A.S. Lewis [*] Tonghua Tian [†]

May 5, 2022

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 Kurdya-Łojasiewicz property typically governs convergence in both discrete and continuous time: we explore its interplay with identifiability.

Key words: variational analysis, subgradient descent, partly smooth, active manifold, identification, KL property.

AMS Subject Classification: 49J52, 49Q12, 90C56, 65K10, 37C10

1 Introduction

Contemporary large-scale optimization problems involve 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 [17, 19, 29, 42] motivated the idea of an *identifiable set* for an objective function f over Euclidean space [26, Definition 3.10].

[*]ORIE, Cornell University, Ithaca, NY. `people.orie.cornell.edu/aslewis`
Research supported in part by National Science Foundation Grant DMS-2006990.
[†]ORIE, Cornell University, Ithaca, NY. `tt543@cornell.edu`

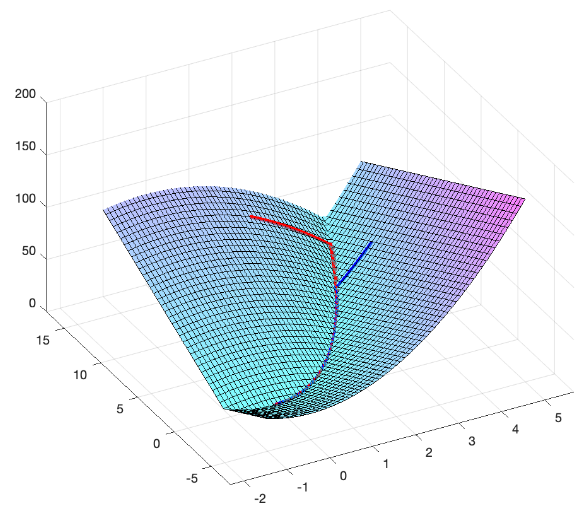

Figure 1: Subgradient curves for the objective $5|x_2 - x_1^2| + x_1^2$.

As a simple illustration, consider the nonsmooth nonconvex function on $\mathbf{R}^2$ defined by

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

We describe the set

$$\mathcal{M} \;=\; \{x : x_2 = x_1^2\}.$$

as *identifiable* at the minimizer $\bar{x} = 0$ because, clearly, no sequence of points x^k approaching $\bar{x}$ from outside $\mathcal{M}$ can have subgradients $y^k \in \partial f(x^k)$ approaching zero. In great generality, proximal-type optimization algorithms generate convergent iterates with corresponding subgradients approaching zero, and hence identify some associated structure. Examples include [35, 36].

In the search for illuminating themes in our understanding of iterative optimization algorithms, one appealing approach, pioneered in works like [3, 5], focuses on simple continuous-time analogues. The same identification behavior illustrated above in discrete time also manifests itself in continuous time. In the example, locally absolutely continuous curves $x\colon \mathbf{R}_+ \to \mathbf{R}^n$ satisfying

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

(following standard variational-analytic terminology [41]) always converge to $\bar{x}$ and are eventually confined to $\mathcal{M}$. Figure 1 plots two random examples of these *subgradient curves*.

We pursue here three main themes around the idea of identifiability. First, we redefine the idea in purely metric terms, highlighting both its simplicity and its applicability to more general settings like optimization over manifolds [1, 14]. Secondly, in the Euclidean setting, where the usual subgradient-based definition is equivalent, and closely related to *partial smoothness* [32], we demonstrate quite generally that subgradient curves are eventually confined to identifiable manifolds. Thirdly, motivated by the crucial role of the "Kurdyka-Łojasiewicz property" in the convergence of optimization algorithms and dynamics [4, 8, 10, 12], including those in metric-space settings [30], we show how the KL property is controlled entirely by the behavior of the objective restricted to any identifiable set. In generic semi-algebraic and more general concrete optimization problems [9, 25], local minimizers lie on identifiable *manifolds*, on which the objective function is analytic, making the KL property particularly transparent.

2 Identifiability in metric spaces

Our setting is a complete metric space (X, d), a function $f\colon X \to (-\infty, +\infty]$ that we assume to be closed, or equivalently, lower semicontinuous, and a point $\bar{x}$ in the domain

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

We will primarily be interested in $\bar{x}$ that are local minimizers, but we first consider the following weaker property: for some constant $\delta > 0$,

$$\bar{x} \quad \text{locally minimizes} \quad f + \delta d(\cdot, \bar{x}).$$

Following [22], the infimum of those $\delta > 0$ for which the property holds is called the *slope*, and denoted $|\nabla f|(\bar{x})$. The notation is suggestive of the case of a smooth function on a Euclidean space, when $|\nabla f|(x) = |\nabla f(x)|$. At points outside the domain, the slope is $+\infty$.

Our development relies on the following key property of the slope, which in this form may be found in [43, Theorem 1.4.1], for example.

Theorem 2.1 (Ekeland Variational Principle [28]) *If the closed function f is bounded below, then for any value $\epsilon > 0$ and point $x \in \operatorname{dom} f$, there exists a point v satisfying*

$$f(v) + \epsilon d(v, x) \;\le\; f(x) \qquad \text{and} \qquad |\nabla f|(v) \le \epsilon.$$

If the point $\bar{x}$ is a local minimizer, then the slope there is zero: $|\nabla f|(\bar{x}) = 0$. The latter property is clearly weaker: for example, on $\mathbf{R}$, the function $f(x) = -x^2$ has slope zero at the point zero, which is not a local minimizer. Weaker still is the property of being a critical point, which we define next.

Definition 2.2 A *critical sequence* for a function f is a sequence of points (x_r) converging to a point $\bar{x} \in \operatorname{dom} f$ such that $f(x_r) \to f(\bar{x})$ and $|\nabla f|(x_r) \to 0$. In that case $\bar{x}$ is called the corresponding *critical point*.

For example, on $\mathbf{R}^2$, the function $f(u,v) = \min\{|u| - v, 0\}$ has slope 1 at the point $(0,0)$, but that point is critical, because the local minimizers $(\frac{1}{r}, 0)$ form a critical sequence.

The structure of concrete objective functions around a critical point often restricts the possible corresponding critical sequences. Many optimization algorithms generate critical sequences, thereby "identifying" this structure. This observation motivates our key idea and terminology.

Definition 2.3 An *identifiable set* for a point $\bar{x} \in \operatorname{dom} f$ is a closed set $\mathcal{M}$ containing $\bar{x}$ whose complement contains no critical sequence, or equivalently, such that the *modulus of identifiability*

$$\liminf_{\substack{x \to \bar{x},\ x \notin \mathcal{M} \\ f(x) \to f(\bar{x})}} |\nabla f|(x) \tag{2.4}$$

is strictly positive.

If the point $\bar{x}$ is not critical, then every closed set containing $\bar{x}$ is identifiable. Our interest in identifiable sets therefore focuses on critical points. To take an example, 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 identifiable for the minimizer zero, and the modulus of identifiability is 1. More generally, we have the following classical example.

Example 2.5 (Max functions) Consider smooth functions $f_i \colon \mathbf{R}^n \to \mathbf{R}$, for $i = 1, 2, \ldots, k$, and define

$$f(x) \;=\; \max_i f_i(x) \qquad (x \in \mathbf{R}^n).$$

Denote by $\mathcal{M}$ the set of points x where the values $f_i(x)$ are all equal. Suppose that, at some point $\bar{x} \in \mathcal{M}$, the gradients $\nabla f_i(\bar{x})$ are affine-independent, and contain zero in the relative interior of their convex hull G. Then f has zero slope at $\bar{x}$. Furthermore, the set $\mathcal{M}$ is identifiable for $\bar{x}$, and the modulus of identifiability is the distance from zero to the relative boundary of G.

We can now state our central tool for using identifiability to study growth conditions. Loosely speaking, near a local minimizer $\bar{x}$ (or more generally a slope-zero point), as we leave an identifiable set $\mathcal{M}$, the function f grows at a linear rate governed by the modulus of identifiability. Precursors include [32, Proposition 2.10] [21, Theorem D.2]. Striking here, however, is that this result does not rely on manifolds, subgradients, Euclidean structure, or even nearest-point projections. This simple metric space setting instead reveals the close connection to the Ekeland principle. A precise statement follows.

Theorem 2.6 (Linear growth) *Suppose that a closed function f has slope zero at a point $\bar{x}$, and consider any identifiable set $\mathcal{M}$ and any nonnegative constant ϵ strictly less than the modulus of identifiability. Then for any sequence of points $x_r \to \bar{x}$ satisfying $f(x_r) \to f(\bar{x})$, there exists a sequence of points $v_r \to \bar{x}$ in $\mathcal{M}$ satisfying $f(v_r) \to f(\bar{x})$ and*

$$f(v_r) + \epsilon d(v_r, x_r) \;\le\; f(x_r) \qquad \textit{for all large } r.$$

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 $v_r \in X$ satisfying $|\nabla f|(v_r) \le \epsilon$ and

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

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 $r \le \bar{r}$ gives the desired sequence.

Now consider the more general case where $|\nabla f|(\bar{x}) = 0$. Fix any constant $\delta > 0$ satisfying $2\delta < \sigma - \epsilon$. The point $\bar{x}$ is a local minimizer of the function $\tilde{f} = f + \delta d(., \bar{x})$, and furthermore the following inequalities hold:

$$|\nabla \tilde{f}|(x) \;\ge\; |\nabla f|(x) - \delta \;\ge\; \sigma - \delta$$

for all points x near $\bar{x}$, outside the set $\mathcal{M}$, and with $f(x)$ near $f(\bar{x})$, or equivalently, with $\tilde{f}(x)$ near $\tilde{f}(\bar{x})$. Consequently, the set $\mathcal{M}$ is also identifiable for the function $\tilde{f}$, with modulus $\sigma - \delta > \epsilon + \delta$.

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

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

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

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

Our conclusion follows by the triangle inequality. □

The generality of the metric space framework for Theorem 2.6 results only in the existence of the "shadow" sequence v_r. If X is in fact a Euclidean space and $\mathcal{M}$ is a $C^{(2)}$-smooth manifold on which f is $C^{(2)}$-smooth, then we can be more descriptive: according to [21, Theorem D.2], for sufficiently small ϵ we can take v_r to be the nearest-point projection $\mathrm{proj}_{\mathcal{M}}(x_r)$.

The zero slope assumption in Theorem 2.6 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 $x_r = \frac{1}{r}$.

Optimality conditions

We next derive some simple consequences of the linear growth result for optimality conditions. In the case of Example 2.5, these reduce to classical active-set results.

Corollary 2.7 (Sufficient condition for optimality) *Suppose that a closed function f has slope zero at a point $\bar{x}$, and consider any identifiable set $\mathcal{M}$. Then $\bar{x}$ is a local minimizer if and only if it is a local minimizer relative to $\mathcal{M}$.*

Proof Suppose that $\bar{x}$ is not a local minimizer, so there exists a sequence $x_r \to \bar{x}$ such that $f(x_r) < f(\bar{x})$. Lower semicontinuity implies $f(x_r) \to f(\bar{x})$, so using Theorem 2.6, 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 next result is very similar.

Corollary 2.8 (Sufficient condition for strict optimality) *Suppose that a closed function f has slope zero at a point $\bar{x}$, and consider any identifiable set $\mathcal{M}$. Then $\bar{x}$ is a strict local minimizer if and only if it is a strict local minimizer relative to $\mathcal{M}$.*

Proof Suppose that $\bar{x}$ is not a strict local minimizer, so there exists a sequence $x_r \to \bar{x}$ such that $x_r \neq \bar{x}$ and $f(x_r) \leq f(\bar{x})$. Lower semicontinuity implies $f(x_r) \to f(\bar{x})$, so using Theorem 2.6, 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,

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

For any such f, if $v_r = \bar{x}$, we deduce

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

giving the contradiction $x_r = \bar{x}$. The converse is trivial. □

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

Corollary 2.9 (Quadratic growth) *Suppose that a closed function f has slope zero at a point $\bar{x}$, and consider any identifiable set $\mathcal{M}$. 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

$$f(x_r) - f(\bar{x}) \;<\; \gamma d(x_r, \bar{x})^2.$$

In particular, we deduce $f(x_r) \to f(\bar{x})$. By our linear growth result, Theorem 2.6, there exists a constant $\epsilon > 0$ and a sequence of points $v_r \to \bar{x}$ in $\mathcal{M}$ satisfying

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

This guarantees $v_r \neq \bar{x}$ for all large r, since otherwise we arrive at the contradiction

$$f(\bar{x}) + \epsilon d(\bar{x}, x_r) \;\leq\; f(\bar{x}) + \gamma d(x_r, \bar{x})^2.$$

Since $\gamma < \beta$, for all large r we must have

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

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 Section 3, we prove an analogous result for a growth condition fundamental to convergence analysis for optimization algorithms.

Proximal points

A canonical identification procedure is the classical proximal operator associated with any convex (or prox-regular) function f on $\mathbf{R}^n$:

$$x \;\mapsto\; \operatorname{argmin}\{f + \alpha|\cdot - x|^2\}.$$

See [34, Proposition 4.5], for example. We next discuss this relationship from a purely metric perspective.

To set the stage, given a closed function f on a metric space (X, d), we consider a property somewhat stronger than the slope-zero condition.

For some constant $\bar{\alpha}$, the point $\bar{x}$ minimizes the function $f + \bar{\alpha} d(\cdot, \bar{x})^2$.

This property clearly holds if $\bar{x}$ is a minimizer, and implies $|\nabla f|(\bar{x}) = 0$. In fact, when the space X is Euclidean and f is "prox-regular" (convex or $C^{(2)}$-smooth, for example) and bounded below by a quadratic, this property is equivalent to $|\nabla f|(\bar{x}) = 0$. In general, however, the property is stronger than slope-zero: for example, the real function $f(x) = -|x|^{\frac{3}{2}}$ has slope zero at the point $\bar{x} = 0$ but the property fails there.

The following simple tool is useful.

Lemma 2.10 *For a closed function f, given any point $x \in \operatorname{dom} f$ and constant $\alpha \geq 0$, if y minimizes $f + \alpha d(\cdot, x)^2$, then $|\nabla f|(y) \leq 2\alpha d(y, x)$.*

Proof There is nothing to prove if y minimizes f. Otherwise, we have

$$\begin{aligned}
|\nabla f|(y) &= \limsup_{y \neq z \to y} \frac{f(y) - f(z)}{d(y, z)} \;\leq\; \limsup_{y \neq z \to y} \frac{\alpha d(z, x)^2 - \alpha d(y, x)^2}{d(y, z)} \\
&= \limsup_{y \neq z \to y} \frac{\alpha\big(d(z, x) - d(y, x)\big)\big(d(z, x) + d(y, x)\big)}{d(y, z)} \\
&\leq \limsup_{y \neq z \to y} \alpha\big(d(z, x) + d(y, x)\big) \;=\; 2\alpha d(y, x),
\end{aligned}$$

as required. □

We now show how the proximal operator, when defined, allows us to construct critical sequences in identifiable sets.

Proposition 2.11 (Proximal sequences are critical) *For a closed function f and a constant $\bar{\alpha}$, suppose that the point $\bar{x}$ minimizes the function $f + \bar{\alpha} d(\cdot, \bar{x})^2$. Suppose that a sequence of points $x_r \to \bar{x}$ satisfies $f(x_r) \to f(\bar{x})$. Consider any constant $\alpha > \bar{\alpha}$. Then any sequence of points (y_r) satisfying*

$$y_r \quad \textit{minimizes} \quad f + \alpha d(\cdot, x_r)^2 \qquad \textit{for } r = 1, 2, 3, \ldots$$

is critical:

$$y_r \to \bar{x}, \qquad f(y_r) \to f(\bar{x}), \qquad \textit{and} \qquad |\nabla f|(y_r) \to 0.$$

Consequently, if the set $\mathcal{M}$ is identifiable for $\bar{x}$, then $y_r \in \mathcal{M}$ for all large r.

Proof We have

$$\begin{aligned}
f(x_r) - f(\bar{x}) &\geq f(y_r) + \alpha d(y_r, x_r)^2 - f(\bar{x}) \\
&\geq \alpha d(y_r, x_r)^2 - \bar{\alpha} d(y_r, \bar{x})^2 \\
&\geq \alpha d(y_r, x_r)^2 - \bar{\alpha}\big(d(y_r, x_r) + d(x_r, \bar{x})\big)^2 \\
&\geq (\alpha - \bar{\alpha}) d(y_r, x_r)^2 - 2\bar{\alpha} d(y_r, x_r) d(x_r, \bar{x}) - \bar{\alpha} d(x_r, \bar{x})^2.
\end{aligned}$$

Rearranging gives

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

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

$$f(y_r) + \alpha d(y_r, x_r)^2 \;\leq f(x_r)$$

for each r, we also know $\limsup f(y_r) \leq f(\bar{x})$, so $f(y_r) \to f(\bar{x})$. Finally, by Lemma 2.10,

$$|\nabla f|(y_r) \;\leq\; 2\alpha d(y_r, x_r) \;\to\; 0,$$

completing the proof. □

We emphasize that the proximal points y_r assumed in the result are not guaranteed to exist in general, unless the metric space X is *proper*, meaning that closed bounded sets are compact.

Minimal identifiable sets

Identifiable sets for a point $\bar{x}$ are not usually unique: any superset of an identifiable set is also identifiable. Identifiable sets always exist, since in particular any closed neighborhood of $\bar{x}$ is identifiable. Furthermore, finite intersections of identifiable sets are identifiable. Consequently, identifiability at $\bar{x}$ is a local property: if a set is identifiable, so is its intersection with any closed neighborhood of $\bar{x}$. Since smaller identifiable sets are more informative tools than larger ones, most interesting are identifiable sets $\mathcal{M}$ that are *locally minimal*, in the sense that any other identifiable set must contain $\mathcal{M}$ locally.

Unfortunately, in general there may exist no locally minimal identifiable set. For example, the continuous convex function $f(u,v) = \sqrt{u^2 + v^4}$ has slope zero at its minimizer $(0,0)$, and elsewhere we have

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

Consider any constant $\alpha > 0$. Close to zero and outside the set

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

the slope is bounded below by $\frac{\alpha^2}{1+\alpha^2} > 0$, so the sets $\mathcal{M}_\alpha$ are all identifiable. These sets shrink to the set $\mathcal{M}_0$ as $\alpha \downarrow 0$. However, $\mathcal{M}_0$ is not identifiable, because the sequence (k^{-3}, k^{-1}) for $k = 1, 2, 3, \ldots$ approaches zero from outside $\mathcal{M}_0$ and yet the slope at these points converges to zero. Thus no locally minimal identifiable set exists. In the Euclidean setting, $X = \mathbf{R}^n$, we can guarantee the existence of a locally minimal identifiable set using the property of *partial smoothness*, a topic to which we return later. For a related discussion, see [26].

Identifiability via subgradients

More robust than the slope is the *limiting slope* of f at a point $x \in X$:

$$\overline{|\nabla f|}(x) \;=\; \liminf_{\substack{z \to x \\ f(z) \to f(x)}} |\nabla f|(z).$$

With this terminology, critical points are exactly those where the limiting slope is zero. The limiting slope also leads to an equivalent definition.

Proposition 2.12 *A closed set $\mathcal{M}$ is identifiable for a point $\bar{x} \in \operatorname{dom} f$ if and only if $\bar{x} \in \mathcal{M}$ and no sequence (x_r) outside $\mathcal{M}$ can satisfy $x_r \to \bar{x}$, $f(x_r) \to f(\bar{x})$, and $\overline{|\nabla f|}(x_r) \to 0$.*

Proof If $x_r \notin \mathcal{M}$ satisfies $x_r \to \bar{x}$, $f(x_r) \to f(\bar{x})$ and $|\nabla f|(x_r) \to 0$, then the limiting slopes $\overline{|\nabla f|}(x_r) \leq |\nabla f|(x_r)$ also converge to zero. Conversely, if $x_r \notin \mathcal{M}$ satisfies $x_r \to \bar{x}$, $f(x_r) \to f(\bar{x})$ and $\overline{|\nabla f|}(x_r) \to 0$, then using the definition of the limiting slope and the fact that $\mathcal{M}$ is closed, for each $r = 1, 2, 3, \ldots$ we can choose $x'_r \notin \mathcal{M}$ satisfying $d(x_r, x'_r) < \frac{1}{r}$ (implying $x'_r \to \bar{x}$), $|f(x_r) - f(x'_r)| < \frac{1}{r}$ (implying $f(x'_r) \to f(\bar{x})$), and $|\nabla f|(x'_r) \leq 2\overline{|\nabla f|}(x_r) \to 0$. □

In the Euclidean case, we deduce the equivalence of this new definition of identifiability with the original version [26, Definition 3.10]. We follow standard variational-analytic terminology [41] throughout: in particular, ∂f denotes the usual subdifferential.

Theorem 2.13 (Identifiability via subgradients) *Consider a closed proper function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a point $\bar{x} \in \operatorname{dom} f$. Then a closed set $\mathcal{M}$ containing $\bar{x}$ is identifiable there if and only if there exists no sequence (x_r) outside $\mathcal{M}$ converging to $\bar{x}$ with $f(x_r) \to f(\bar{x})$ and with subgradients $y_r \in \partial f(x_r)$ converging to zero.*

Proof We simply apply Proposition 2.12 along with the following formula for the limiting slope of a closed function f at any point $x \in \operatorname{dom} f$:

$$\overline{|\nabla f|}(x) \;=\; d\big(0, \partial f(x)\big).$$

For a proof, see [24, Proposition 4.6]. □

3 The Kurdya-Łojasiewicz property

We begin this section with a discussion of sharpness. Our setting, as usual, is a complete metric space (X, d), a closed function $f\colon X \to (-\infty, +\infty]$, and a point $\bar{x}$ where the value $\bar{f} = f(\bar{x})$ is finite. We call $\bar{x}$ a *sharp local minimizer* when there exists a constant $\epsilon > 0$ such that all points x near $\bar{x}$ satisfy $f(x) \geq \bar{f} + \epsilon d(x, \bar{x})$. An immediate consequence of our linear growth tool, Theorem 2.6, is the following result.

Corollary 3.1 (Sharp minimizers) *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.*

For any value $\alpha \in \mathbf{R}$, we define the level set

$$\mathcal{L}_\alpha \;=\; \{x \in X : f(x) \leq \alpha\}.$$

Local minimizers $\bar{x}$ that are sharp are in particular strict: the level set $\mathcal{L}_{\bar{f}}$ is locally just $\{\bar{x}\}$. A weaker property than sharpness, originating with [18], is that $\bar{x}$ is a *weak sharp minimizer* (locally), meaning that there exists a constant $\epsilon > 0$ such that all points x near $\bar{x}$ satisfy

$$f(x) \;\geq\; \bar{f} + \epsilon d(x, \mathcal{L}_{\bar{f}}).$$

Identifiability helps us understand this property too.

Corollary 3.2 (Weak sharp minimizers) *For a closed function f and a local minimizer $\bar{x}$, if the level set $\mathcal{L}_{\bar{f}}$ is identifiable at $\bar{x}$ for f, then $\bar{x}$ is a weak sharp minimizer.*

Proof Suppose $\epsilon > 0$ is strictly less than the modulus of identifiability of the function f for the set $\mathcal{L}_{\bar{f}}$ at the point $\bar{x}$, namely

$$\liminf_{\substack{x \to \bar{x} \\ \bar{f} < f(x) \to \bar{f}}} |\nabla f|(x) \;>\; 0.$$

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

$$f(x_r) \;<\; \bar{f} + \epsilon d(x_r, \mathcal{L}_{\bar{f}}). \tag{3.3}$$

Since $\bar{x}$ is a local minimizer, for all large r we have $f(x_r) \geq f(\bar{x})$ and so $x_r \notin \mathcal{L}_{\bar{f}}$. The right-hand side converges to $f(\bar{x})$, and hence so does $f(x_r)$. By Theorem 2.6, there exists a sequence of points $v_r \to \bar{x}$ in $\mathcal{L}_{\bar{f}}$ satisfying, for all large r,

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

or in other words

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

contradicting inequality (3.3). □

The Kurdya-Łojasiewicz property then simply modifies the assumption of Corollary 3.2 by truncation and rescaling. Define the function $f_{\bar{x}} \colon X \to [0, +\infty]$ by

$$f_{\bar{x}}(x) \;=\; \max\{f(x) - \bar{f}, 0\} \qquad (x \in X).$$

Definition 3.4 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 space [37], and hence also for real-analytic functions on analytic manifolds.

Definition 3.5 A closed function f satisfies the *KL property* at a point $\bar{x} \in \operatorname{dom} f$ if there exists a desingularizer ϕ such that the level set $\mathcal{L}_{\bar{f}}$ is identifiable at $\bar{x}$ for the composite function $\phi \circ f_{\bar{x}}$.

The definition amounts to the existence of a constant $\delta > 0$ such that

$$\big|\nabla\big(\phi(f(\cdot) - \bar{f})\big)\big|(x) \;\geq\; \delta$$

for all points x near $\bar{x}$ with value $f(x)$ near and strictly larger than $\bar{f}$. We can rewrite the left-hand side using a simple chain rule [6, Lemma 4.1]:

$$\big|\nabla\big(\phi(f(\cdot) - \bar{f})\big)\big|(x) \;=\; \phi'\big(f(x)\big) \cdot |\nabla f|(x).$$

Our definition of the KL property essentially coincides with that given in [11], a relationship we discuss further at the end of Section 4.

The following observation, one of our main results, shows that, around a slope-zero point, a function inherits the KL property from the corresponding property restricted to any identifiable set.

Theorem 3.6 (Identifiability and the KL property) *Suppose that a closed function f has slope zero at a point $\bar{x}$ and consider any identifiable set $\mathcal{M}$ for $\bar{x}$. Consider 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) *The restriction* $f|_{\mathcal{M}}$ *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) are distinct 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 $\bar{f}$. Without loss of generality assume $\bar{f} = 0$.

Suppose that property (b) holds, so there exists a constant $\delta > 0$ such that

$$\big|\nabla\big(\phi \circ f|_{\mathcal{M}}\big)\big|(x) \;\geq\; \delta$$

for all points $x \in \mathcal{M}$ near $\bar{x}$ satisfying $0 < f(x) < \delta$. The definition of slope then implies

$$\big|\nabla\big(\phi \circ f\big)\big|(x) \;\geq\; \delta, \tag{3.7}$$

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 $\phi'\big(f(x)\big) \geq \delta$ and $|\nabla f|(x) > 2\epsilon$. The chain rule implies

$$\big|\nabla\big(\phi \circ f\big)\big|(x) \;\geq\; \delta \cdot 2\epsilon \;>\; 0.$$

Together with inequality (3.7), this proves property (a).

Conversely, suppose that property (a) holds, so there exists a constant $\delta > 0$ such that

$$\big|\nabla\big(\phi \circ f\big)\big|(x) \;>\; \delta \quad \text{and} \quad \phi'\big(f(x)\big) \geq \delta$$

for all points x satisfying $d(x, \bar{x}) < \delta$ and $0 < f(x) < \delta$. Shrinking δ if necessary, the identifiability property also ensures that all such points outside $\mathcal{M}$ also satisfy $|\nabla f|(x) > 2\epsilon$.

Now consider any such point $x \in \mathcal{M}$. There exists a sequence of points $x_r \to x$, each satisfying

$$\phi\big(f(x)\big) - \phi\big(f(x_r)\big) \;>\; \delta \cdot d(x, x_r).$$

By Theorem 2.6, there exist corresponding 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

$$\begin{aligned} \phi\big(f(x)\big) - \phi\big(f(v_r)\big) \;&>\; \delta \cdot d(x, x_r) + \phi\big(f(x_r)\big) - \phi\big(f(v_r)\big) \\ &=\; \delta \cdot d(x, x_r) + \phi'(t_r)\big(f(x_r) - f(v_r)\big) \end{aligned}$$

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

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

where $\gamma = \delta \cdot \min\{1, \epsilon\} > 0$, so

$$|\nabla(\phi \circ f|_{\mathcal{M}})|(x) \;\geq\; \gamma.$$

Property (b) follows. $\square$

Remark 3.8 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.

A particular case of interest is when the metric space (X, d) is Euclidean with the usual distance, and the set $\mathcal{M} \subset X$ is an embedded submanifold [13]. In the result above, we are considering $\mathcal{M}$ as a space whose metric is just the inherited Euclidean distance: $d(x, y) = |x - y|$. More natural may be to consider the *intrinsic* metric d' for the manifold $\mathcal{M}$, defined as the infimum of the length of paths in $\mathcal{M}$ from x to y. However, this results in no change to the KL property, as a consequence of the following fact [23, Proposition 3.1]:

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

The following consequence describes the generic situation for semi-algebraic and more general concrete optimization problems [9, 25]. However, this result applies to functions that may not be semi-algebraic (or tame) outside the identifiable set.

Corollary 3.9 (Partly analytic functions) *Suppose that $f \colon \mathbf{R}^n \to (-\infty, +\infty]$ is a closed function and consider a local minimizer $\bar{x}$ where $f(\bar{x})$ is finite. If $\bar{x}$ lies in an identifiable set $\mathcal{M}$ that is an analytic manifold such that the restriction $f|_{\mathcal{M}}$ is analytic, then the KL property holds at $\bar{x}$.*

Proof Any analytic function on an analytic manifold in $\mathbf{R}^n$ must satisfy the KL property, since it can be viewed as the restriction of an analytic function on $\mathbf{R}^n$. The conclusion then follows immediately from the preceding result. $\square$

4 Proximal sequences in metric spaces

To illustrate the power of the KL property in convergence analysis, we briefly summarize an argument from [11]. Beyond its intrinsic interest, in line with our current theme, we emphasize its purely metric essence.

As usual, we consider a complete metric space (X, d), and a closed function $f\colon X \to (-\infty, +\infty]$. We assume that f is bounded below, with $\inf f = \bar{\rho}$. We consider a proximal sequence, generated recursively from each current iterate x by selecting the next iterate x_{new} to satisfy

$$x_{\text{new}} \quad \text{minimizes} \quad f + \alpha d(\cdot, x)^2$$

for some constant $\alpha > 0$.

In quite general settings, there exists a proximal sequence (x_k) starting from any initial point $x_0 \in \operatorname{dom} f$. Consider in particular the case when the metric space X is *proper*, meaning that closed bounded sets are compact. For each k, the function $f + \alpha d(\cdot, x_k)^2$ is closed, so has a minimizer x_{k+1}, since the level set

$$\{x : f(x) + \alpha d(x, x_k)^2 \leq f(x_k)\}$$

is closed, contains x_k, and is contained in the ball

$$\Big\{x : d(x, x_k)^2 \leq \frac{1}{\alpha}\big(f(x_k) - \bar{\rho}\big)\Big\}.$$

Moreover, using Lemma 2.10, we have

$$|\nabla f|(x_{k+1}) \ \leq \ 2\alpha d(x_{k+1}, x_k).$$

In fact, proximal sequences exist for certain functions, even in metric spaces that are not proper. The classical proximal point iteration is well defined for closed convex functions on infinite-dimensional Hilbert spaces, for example, and also extends beyond Euclidean settings [7,31].

Next, we make the further common assumption [2] that f *is continuous on slope-bounded sets*, by which we mean that any sequence of points $x_k \to x$ with values $f(x_k)$ and slopes $|\nabla f|(x_k)$ uniformly bounded must satisfy $f(x_k) \to f(x)$. Closed convex functions on Euclidean spaces have this property, for example, as do many more general classes of objectives.

We also assume a *global* KL property: there exists a value $\rho > \inf f$ and a desingularizer ϕ such that

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

If the set of minimizers of f is compact and $\liminf_{x\to\infty} f(x) > \inf f$, for example, then this global property is equivalent to the KL property holding at every minimizer.

Under these conditions, assuming $f(x_0) < \rho$, we claim that the iterates x_k must converge to a critical point x^*. To see this, assume without loss of generality that $f(x_k) > \inf f$ for all k. For each $k = 1, 2, 3, \ldots$, we note

$$\begin{aligned} d(x_k, x_{k+1}) \ &= \ \operatorname{dist}\big(x_k, \{y : f(y) \leq f(x_{k+1})\}\big) \\ &\leq \ \operatorname{Dist}\big(\{y : f(y) \leq f(x_k)\}, \{y : f(y) \leq f(x_{k+1})\}\big), \end{aligned}$$

where $\mathrm{Dist}(S_1, S_2)$ denotes the Hausdorff distance between two sets S_1 and S_2. Hence it follows from [11, Corollary 4] that

$$d(x_k, x_{k+1}) \;\leq\; \phi\big(f(x_k)\big) - \phi\big(f(x_{k+1})\big).$$

Thus the length of the trajectory $\sum_k d(x_k, x_{k+1})$ is finite, so by completeness, x_k converges to some point x^* and furthermore $|\nabla f|(x_k) \to 0$. The values $f(x_k)$ are nonincreasing and bounded below, so continuity on slope bounded sets implies $f(x_k) \to f(x^*)$. Hence the limit x^* is a critical point.

Remark 4.1 In [11], the authors explore characterizations of the KL property via the finite length of *talweg curves*. For a suitably isolated critical point $\bar{x}$ with value 0, given any constant $R > 1$, they define such a curve to be a selection from the mapping defined for small $r > 0$ by

$$r \;\mapsto\; \Big\{x \in D : f(x) = r,\; \overline{|\nabla f|}(x) \leq R \inf_{\substack{y \in D \\ f(y)=r}} \overline{|\nabla f|}(y)\Big\}$$

where D is a closed bounded neighborhood of $\bar{x}$. If $\mathcal{M}$ is an identifiable set for $\bar{x}$, then talweg curves $x(r)$ lie in $\mathcal{M}$ for small r. This correlates with our observation that the KL property is only determined by the behavior of f on $\mathcal{M}$.

5 Subgradient curves and identification

We turn now from iterations in discrete time to the continuous-time setting. From an optimization perspective [3, 5], the canonical such dynamical systems, for objective functions $f\colon \mathbf{R}^n \to (-\infty, +\infty]$, are *subgradient curves*: locally absolutely continuous maps $x\colon \mathbf{R}_+ \to \operatorname{dom} f = \{z : f(z) < +\infty\}$ satisfying

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

As we shortly outline, existing literature covers a wide class of nonsmooth and nonconvex functions f for which subgradient curves must converge.

Our goal is the existence, around some given point x^*, of a small set $\mathcal{M}$ (in a suitable sense) with the property that any subgradient curve converging to x^* must eventually remain in $\mathcal{M}$. This behavior is the continuous-time 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. Our main continuous-time result, Theorem 5.5, proves the same behavior for curves. Figure 1 illustrates this behavior for the simple nonsmooth nonconvex example (1.1)

$$f(x) \;=\; 5|x_2 - x_1^2| + x_1^2,$$

with associated identifiable set $\mathcal{M}$ around the minimizer $x^* = 0$ defined by $x_2 = x_1^2$.

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 [15]. In fact this result holds more generally, under the assumption that f is *primal lower nice* [39].

Theorem 5.1 (Existence and uniqueness of subgradient curves [38]) *If the proper closed function $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ 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*

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

We defer until the next section the formal definition of primal lower nice functions. For now we simply note that they comprise a large class. For example, the function (1.1) is primal lower nice at zero. More generally, the following versatile example [41], composing more familiar smooth and convex ingredients, also belongs to this class [39, Theorem 5.1]. It covers in particular the case of *weakly convex* functions of the form $g - \rho|\cdot|^2$, for proper closed convex functions g and constant ρ. Unless otherwise stated, the term **"smooth" always means $C^{(2)}$-smooth**.

Example 5.2 (Strongly amenable functions) At a point $x \in \mathbf{R}^n$, we describe a function f as *strongly amenable* if it has the local representation $f = g \circ F$ for some proper closed convex function $g \colon \mathbf{R}^m \to \overline{\mathbf{R}}$ and smooth map $F \colon \mathbf{R}^n \to \mathbf{R}^m$ such that $F(x)$ lies in $\operatorname{dom} g$ and the normal cone to $\operatorname{cl}(\operatorname{dom} g)$ there intersects the null space of the adjoint map $\nabla F(x)^*$ trivially. In that case both the subdifferential $\partial f(x)$ and the regular subdifferential $\hat{\partial} f(x)$ are given by the chain rule formula $\nabla F(x)^* \partial g\big(F(x)\big)$.

The function (1.1) is strongly amenable at zero, because, for example, we can represent it using the smooth map $F(x) = (x_2 - x_1^2, x_1)$ and the convex function $g(u,v) = 5|u| + v$. All proper closed convex functions are strongly amenable, as are indicator functions of smooth manifolds. Furthermore, the property is preserved under addition of smooth functions [41, Example 10.24], so in particular "weakly convex" functions (see for example [20]) are strongly amenable.

For closed convex functions with minimizers, following the Brézis result, Bruck proved that subgradient curves converge to minimizers [16]. In fact, subgradient curves are known to converge more generally, assuming a suitable version of the KL property, holding in particular for closed semi-algebraic (or subanalytic) functions on bounded domains [8, 10]. 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 [38, 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.

Before presenting our main continuous-time result, we briefly review smooth manifolds and smooth functions defined on them. We take an elementary approach here.

A set $\mathcal{M} \subset \mathbf{R}^n$ is a *manifold* around a point $\bar{x} \in \mathbf{R}^n$ if there exists a smooth map $G\colon \mathbf{R}^n \to \mathbf{R}^m$, for some integer m, such that $G(\bar{x}) = 0$, the derivative $\nabla G(\bar{x})\colon \mathbf{R}^n \to \mathbf{R}^m$ is surjective, and all points $x \in \mathbf{R}^n$ near $\bar{x}$ lie in $\mathcal{M}$ if and only if $G(x) = 0$. We then say that 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 $\bar{x}$. In that case, the *Riemannian gradient* of h at points $x \in \mathcal{M}$ near $\bar{x}$ is given by the following orthogonal projection onto the tangent space to $\mathcal{M}$ at x:

$$\nabla_{\mathcal{M}} h(x) \;=\; \mathrm{Proj}_{T_{\mathcal{M}}(x)}\big(\nabla \tilde{h}(x)\big).$$

An easy exercise shows that the right-hand side is independent of the choice of $\tilde{h}$.

For example, the set $\mathcal{M} \subset \mathbf{R}^2$ defined by the equation $x_2 = x_1^2$ is a manifold around the point $\bar{x} = 0$, because we can choose $G(x) = x_2 - x_1^2$, and the restriction $f|_{\mathcal{M}}$ of the function f defined by equation (1.1) is smooth on $\mathcal{M}$ because it agrees with the smooth function $\tilde{f}(x) = 5x_2 - 4x_1^2$ on $\mathcal{M}$. Furthermore, the Riemannian gradient $\nabla_{\mathcal{M}}(f|_{\mathcal{M}})$, which we abbreviate to $\nabla_{\mathcal{M}} f$, is zero at zero.

Our main continuous-time result concerns identifiable sets that are smooth manifolds. Before stating it, we consider a special case.

Proposition 5.3 (Smooth restrictions) *Suppose that a set $\mathcal{M} \subset \mathbf{R}^n$ is a smooth manifold $\mathcal{M}$ around the point $\bar{x}$, and that the function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ 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*

$$x'(t) = -\nabla_{\mathcal{M}} f\big(x(t)\big). \tag{5.4}$$

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

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

At such times t we must have $x'(t) \in T_{\mathcal{M}}\big(x(t)\big)$, from which we deduce equation (5.4). The result now follows from classical smooth initial value theory. □

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

Theorem 5.5 (Identification for subgradient curves)
Consider a proper closed function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *and a subgradient curve* $x(\cdot)$ *converging to a point* $\bar{x}$ *at which* f *is primal lower nice and has an identifiable manifold* $\mathcal{M}$*. Suppose that* $f|_{\mathcal{M}}$ *is smooth around* $\bar{x}$*. Then, after a finite time,* $x(\cdot)$ *becomes a smooth curve on* $\mathcal{M}$*, satisfying*

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

The proof follows in the next section, until after a formal discussion of the primal lower nice property.

6 Primal lower nice functions

A function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is *primal lower nice* at a point $\bar{x} \in \mathbf{R}^n$ if the following weaker version of the standard subgradient inequality holds: there exist constants α and β such that all points x, x' near $\bar{x}$, and subgradients $y \in \partial f(x)$ satisfy

$$f(x') \;\geq\; f(x) \;+\; \langle y, x'-x\rangle \;-\; (\alpha + \beta|y|)|x'-x|^2. \tag{6.1}$$

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

Readers familiar with the notion of "prox-regularity" [40] will notice a similarity. In the definition of that property, we only require the inequality to hold under further restrictions on x, x' and y: the functions values $f(x)$ and $f(x')$ must both be close to $f(\bar{x})$, and the subgradient y must be close to some given subgradient $\bar{y} \in \partial f(\bar{x})$; furthermore, $\beta = 0$, but α may depend on $\bar{y}$. Clearly the primal lower nice property thus implies prox-regularity.

Conversely, as we have remarked, the primal lower nice property at a point is implied by strong amenability there. Hence, for locally Lipschitz functions f, the property coincides with various well-known variational analytic properties: prox regularity, strong amenability, the "lower $\mathcal{C}^2$" property [41, Exercise 10.36, Proposition 13.33], and weak convexity.

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

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

This gives a contradiction for t sufficiently small.

As we have seen in Theorem 5.1, the subgradient curves of functions that are primal lower nice and bounded below have desirable properties. Next we show that under the same condition, the velocity vector $x'(\cdot)$ for a bounded subgradient curve $x(\cdot)$ *essentially converges 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 6.2 *If a proper closed function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is primal lower nice and bounded below, then any bounded subgradient curve $x(\cdot)$ of f satisfies that the velocity $x'(\cdot)$ essentially converges to zero.*

Proof Suppose $x([0,\infty)) \subset rB$, where B denotes the closed unit ball in $\mathbf{R}^n$. Using a standard compactness argument, we can show there exist constants α, β and $\delta > 0$ such that the primal lower nice inequality (6.1) 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

$$x'(t) \in -\partial f\big(x(t)\big) \qquad \text{for all } t \in \Omega.$$

By [38, Lemma 2.1], the primal lower nice property ensures the existence of constants $\mu, \nu > 0$ such that

$$|x'(s)| \;\leq\; |x'(t)| \exp\Big(\mu(s-t) + \nu \int_t^s |x'(\tau)|\, d\tau\Big)$$

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

By way of contradiction, suppose 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 $j = 1, 2, 3, \ldots$. Taking a subsequence, we can suppose, for each j, the inequality $s_{j+1} - s_j > 1$, from which we deduce either

$$|x'(s_j)| \;\leq\; |x'(t)| \exp\Big(\mu + \nu \int_t^{s_j} |x'(\tau)|\, d\tau\Big) \text{ for all } t \in [s_j - 1, s_j] \cap \Omega, \tag{6.3}$$

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

$$|x(t) - x(s_j)| \;>\; \delta. \tag{6.4}$$

Assume (6.3) holds, then

$$|x'(t)| \exp\Big(\nu \int_t^{s_j} |x'(\tau)|\, d\tau\Big) \;>\; \epsilon e^{-\mu} \text{ for all } t \in [s_j - 1, s_j] \cap \Omega.$$

Hence either some time t satisfies

$$\nu \int_t^{s_j} |x'(\tau)|\, d\tau \;\geq\; 1,$$

in which case

$$\int_{s_j - 1}^{s_j} |x'(\tau)|\, d\tau \;\geq\; \frac{1}{\nu},$$

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

$$\int_{s_j-1}^{s_j} |x'(\tau)|\, d\tau \;\geq\; \epsilon e^{-\mu-1}.$$

On the other hand, if (6.4) holds, we easily deduce

$$\int_{s_j-1}^{s_j} |x'(\tau)|\, d\tau \;\geq\; \int_{t}^{s_j} |x'(\tau)|\, d\tau \;\geq\; |x(t) - x(s_j)| \;>\; \delta.$$

We have thus shown the existence of a constant $\rho > 0$ such that, for each j,

$$\int_{s_j-1}^{s_j} |x'(\tau)|\, d\tau \;\geq\; \rho$$

and hence, by Hölder's inequality,

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

But this contradicts the conclusion of Theorem 5.1. □

Like both the strong amenability and prox-regularity properties, the primal lower nice property is preserved by addition of smooth functions.

Proposition 6.5 (Primal lower nice preservation) *Consider a function $g\colon \mathbf{R}^n \to \mathbf{R}$ that is smooth around the point $\bar{x} \in \mathbf{R}^n$. If the proper closed function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is primal lower nice at $\bar{x}$, then so is the sum $f + g$.*

Proof Using the notation above, we can assume that inequality (6.1) holds. Since g is $C^{(2)}$-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

$$g(x') \;\geq\; g(x) \;+\; \langle \nabla g(x), x' - x \rangle \;-\; \gamma |x' - x|^2 \tag{6.6}$$

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 [41, Corollary 10.9] ensures the existence of a subgradient $y \in \partial f(x)$ satisfying $w = y + \nabla g(x)$. Adding the inequalities (6.1) and (6.6), 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. □

Both Theorem 5.1 and Theorem 6.2 concern functions that are both primal lower nice and bounded below. However, in the local analysis that we pursue here, the boundedness assumption involves no loss of generality, as we show next.

Lemma 6.7 (Localization) *If a proper closed function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ 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 $\bar{x} = 0$. From the definition, f is primal lower nice throughout the ball $2\delta B$ for some $\delta > 0$ (where B denotes the closed unit ball). 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

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

is smooth on the interior of $2\delta B$, so, by Proposition 6.5, the function $f+g$ is primal lower nice. But $f+g$ is also bounded below and agrees identically with f on the ball δB, as required. □

We next derive a localized version of Theorem 6.2.

Theorem 6.8 (Essential convergence) *If a proper closed function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ 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 $f\big(x(\cdot)\big)$ converges to $f(\bar{x})$, and the velocity $x'(\cdot)$ essentially converges to zero.*

Proof The first two conclusions essentially appear in [38], and the third conclusion appears in Theorem 6.2, all under the stronger conditions that f is both bounded below and also primal lower nice throughout its domain. Here, we lose no generality in assuming those conditions. More precisely, we could first use Lemma 6.7 to replace f by a function that is primal lower nice and bounded below, and that is unchanged near the point $\bar{x}$. Since the original subgradient curve converges to $\bar{x}$, 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 5.5. Denote the critical point of interest by $\bar{x}$. By Theorem 6.8 (Essential convergence), the function value $f\big(x(\cdot)\big)$ converges to $f(\bar{x})$, and there exists a full-measure set $\Omega \subset \mathbf{R}_+$ such that

$$\begin{aligned} x'(t) &\in -\partial f\big(x(t)\big) \qquad \text{for all } t \in \Omega \\ x'(t) &\to 0 \qquad \text{as } t \to +\infty \text{ in } \Omega. \end{aligned}$$

For every point $x \in \mathcal{M}$ around $\bar{x}$, the primal lower nice property ensures

$$\partial f(x) \;\subseteq\; \partial(f + \delta_{\mathcal{M}})(x).$$

Identifiability now implies

$$x(t) \in \mathcal{M} \qquad \text{and} \qquad x'(t) \in -\partial(f + \delta_{\mathcal{M}})\big(x(t)\big) \qquad \text{for all large } t \in \Omega$$

Since the set $\mathcal{M}$ is locally closed around $\bar{x}$, the subgradient curve $x(\cdot)$, being continuous, eventually lies in $\mathcal{M}$, and furthermore coincides with a subgradient curve for the function $f + \delta_{\mathcal{M}}$. We conclude by appealing to Proposition 5.3 (Smooth restrictions). □

An alternative approach to Theorem 5.5 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 [27, 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

$$\tilde{x}'(t) = -\nabla_{\mathcal{M}} f\big(\tilde{x}(t)\big), \qquad \tilde{x}(0) = x(t_k).$$

7 Identifiable manifolds and partial smoothness

We end by discussing briefly the relationship between identifiability and *partial smoothness*, and its evolution. An earlier form of identifiability, introduced in [26, Definition 3.10] and inspired by earlier terminology in [17, 19, 29, 42], is defined as follows. Given a function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$, we call a set $\mathcal{M} \subset \mathbf{R}^n$ *identifiable* at a point x in $\mathbf{R}^n$ for a subgradient $y \in \partial f(x)$ when all sequences of points $x_i \to x$ with $f(x_i) \to f(x)$ and sequences of subgradients $y_i \in \partial f(x_i)$ with $y_i \to y$ have the property that x_i must eventually lie in $\mathcal{M}$. In the special case when $y = 0$ is a regular subgradient at x, this coincides precisely with our previous terminology, by Theorem 2.13.

To discuss the relationship with partial smoothness, it is easiest to focus on the special case when y lies in the relative interior of the convex set of regular subgradients $\hat{\partial} f(x)$. In that case, when $\mathcal{M}$ is a manifold around x that is identifiable there for y, and the restriction $f|_{\mathcal{M}}$ is smooth around x, we call f *partly smooth* at x for y relative to $\mathcal{M}$.

The relative interior assumption is a type of nondegeneracy condition common in optimization. Recall a point x is critical for a function f if zero is a subgradient there. We call such points *nondegenerate* when zero in fact lies in the relative interior of $\hat{\partial} f(x)$. Rephrasing our main result using partial smoothness, we have the following theorem.

Theorem 7.1 *Consider a proper closed function* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *and a subgradient curve* $x(\cdot)$ *converging to a nondegenerate critical point at which* f *is both primal lower nice and partly smooth for zero, relative to a manifold* $\mathcal{M}$*. Then, after a finite time,* $x(\cdot)$ *becomes a smooth curve on* $\mathcal{M}$*: for all large* t*,*

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

The idea of partial smoothness was introduced in [32], using a definition relying more heavily on the variational-analytic properties of the function f. The special case of the definition we presented here, more natural for our current focus on identification, follows from [27, Proposition 10.12]. A third version [33, Theorem 6.5] instead emphasizes basic differential geometry: the graph of the subdifferential operator ∂f should be a smooth manifold around the point (x, y), and for nearby points (u, v) on this manifold, the projection $(u, v) \mapsto u$ is constant-rank.

The original variational-analytic definition of partial smoothness for a function f relative to a manifold $\mathcal{M}$ at a point x is most transparent when f is closed and convex. Along with the condition that the restriction $f|_{\mathcal{M}}$ is smooth around x, it requires two more conditions:

- The affine span of the subdifferential $\partial f(x)$ and the normal space $N_{\mathcal{M}}(x)$ are translates of each other.
- The subdifferential mapping ∂f, restricted to $\mathcal{M}$, is continuous at x.

More generally, consider a strongly amenable function f with representation $g \circ F$ as in Example 5.2. Suppose that the convex function g is partly smooth relative to a manifold $\mathcal{M}$ at the point $F(x)$. Assuming the transversality condition that the normal space $N_{\mathcal{M}}\big(F(x)\big)$ intersects the null space of the adjoint map $\nabla F(x)^*$ trivially, $F^{-1}(M)$ is a manifold around x, relative to which f is partly smooth at x [32, Theorem 4.2]. For example, our illustration (1.1), the function $5|x_2 - x_1^2| + x_1^2$, is partly smooth at the nondegenerate critical point zero relative to the manifold defined by $x_2 = x_1^2$.

As a variational-analytic property, partial smoothness seems rather involved. However, it holds commonly for concrete objective functions f. In particular, it is a *generic* property when f is semi-algebraic [25]. 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 concrete optimization problems, we can reasonably expect to see identifiable manifolds around the solutions.

Acknowledgement The authors thank D. Drusvyatskiy for helpful suggestions, in particular leading to the alternative argument for Theorem 5.5.

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, 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.

[5] H. Attouch and M. Teboulle. Regularized Lotka-Volterra dynamical system as continuous proximal-like method in optimization. *J. Optim. Theory Appl.*, 121:541–570, 2004.

[6] D. Azé and J.-N. Corvellec. Nonlinear error bounds via a change of function. *Journal of Optimization Theory and Applications*, 172:9–32, 2017.

[7] M. Bacak. The proximal point algorithm in metric spaces. *Isr. J. Math.*, 194:689–701, 2013.

[8] 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.

[9] J. Bolte, A. Daniilidis, and A.S. Lewis. Generic optimality conditions for semialgebraic convex programs. *Math. Oper. Res.*, 36:55–70, 2011.

[10] J. Bolte, A. Daniilidis, A.S. Lewis, and M. Shiota. Clarke subgradients of stratifiable functions. *SIAM J. Optimization*, 18(2):556–572, 2007.

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

[12] 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.

[13] W.M. Boothby. *An Introduction to Differential Manifolds and Riemannian Geometry*. Academic Press, Orlando, 1988.

[14] N. Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, Cambridge, 2022.

[15] H. Brézis. *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*. North-Holland ,New York, 1973.

[16] R.E. Bruck. Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. *J. Functional Analysis*, 18:15–26, 1975.

[17] J.V. Burke. On the identification of active constraints. II. The nonconvex case. *SIAM J. Numer. Anal.*, 27(4):1081–1103, 1990.

[18] J.V. Burke and M.C. Ferris. Weak sharp minima in mathematical programming. *SIAM J. Control Optim.*, 31:1340–1359, 1993.

[19] J.V. Burke and J.J. Moré. On the identification of active constraints. *SIAM J. Numer. Anal.*, 25(5):1197–1211, 1988.

[20] D. Davis and D. Drusvyatskiy. Stochastic model-based minimization of wealy convex functions. *SIAM J. Optim.*, 29:207–239, 2019.

[21] D. Davis, D. Drusvyatskiy, and V. Charisopoulos. Stochastic algorithms with geometric step decay converge linearly on sharp functions. `arXiv:1907.09547`, 2019.

[22] 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.

[23] A.L. Dontchev and A.S. Lewis. Perturbations and metric regularity. *Set-Valued Anal.*, 13:417–438, 2005.

[24] D. Drusvyatskiy, A.D. Ioffe, and A.S. Lewis. Curves of descent. *SIAM J. Control Optim.*, 53(1):114–138, 2015.

[25] D. Drusvyatskiy, A.D. Ioffe, and A.S. Lewis. Generic minimizing behavior in semialgebraic optimization. *SIAM J. Optim.*, 26:513–534, 2016.

[26] D. Drusvyatskiy and A.S. Lewis. Optimality, identifiability, and sensitivity. *Math. Program.*, 147:467–498, 2014.

[27] D. Drusvyatskiy and A.S. Lewis. Optimality, identifiability, and sensitivity. arXiv:1207.6628, 2014.

[28] I. Ekeland. On the variational principle. *J. Math. Anal. Appl.*, 47:324–353, 1974.

[29] S.D. Flåm. On finite convergence and constraint identification of subgradient projection methods. *Math. Program.*, 57:427–437, 1992.

[30] D. Hauer and J.M. Mazón. Kurdyka–Lojasiewicz–Simon inequality for gradient flows in metric spaces. *Transactions of the American Mathematical Society*, 372:4917–4976, 2019.

[31] J. Jost. Convex functionals and generalized harmonic maps into spaces of non positive curvature. *Comment. Math. Helvetici*, 70:659–673, 1995.

[32] A.S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM J. Optim.*, 13:702–725, 2002.

[33] A.S. Lewis, Jingwei Liang, and Tonghua Tian. Partial smoothness and constant rank. *SIAM J. Optim.*, 2022.

[34] A.S. Lewis and S. Zhang. Partial smoothness, tilt stability, and generalized Hessians. *SIAM J. Optim.*, 23(1):74–94, 2013.

[35] 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, *Advances in Neural Information Processing Systems 27*, pages 1970–1978. Curran Associates, Inc., 2014.

[36] J. Liang, J. Fadili, G. Peyré, and R. Luke. Activity identification and local linear convergence of Douglas-Rachford/ADMM under partial smoothness. In J.-F. Aujol, M. Nikolova, and N. Papadakis, editors, *Scale Space and Variational Methods in Computer Vision*, pages 642–653. Springer International Publishing, 2015.

[37] S. Łojasiewicz. Ensembles semi-analytiques. Preprint, IHES, 1965.

[38] S. Marcellin and L. Thibault. Evolution problems associated with primal lower nice functions. *J. Convex Anal.*, 13:385–421, 2006.

[39] R.A. Poliquin. Integration of subdifferentials of nonconvex functions. *Nonlinear Anal.*, 17:385–398, 1991.

[40] R.A. Poliquin and R.T. Rockafellar. Prox-regular functions in variational analysis. *Trans. Amer. Math. Soc.*, 348:1805–1838, 1996.

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

[42] S.J. Wright. Identifiable surfaces in constrained optimization. *SIAM J. Control Optim.*, 31:1063–1079, July 1993.

[43] C. Zalinescu. *Convex Analysis in General Vector Spaces*. World Scientific, New Jersey, 2002.