TRANSFERABILITY OF ERROR BOUNDS AND THE KURDYKA-ŁOJASIEWICZ PROPERTY UNDER C^1 PARTIAL SMOOTHNESS

HANJU WU*, CHING-PEI LEE†, AND YUE XIE‡

Abstract. This work studies how a generalized error bound (GEB) property in the ambient Euclidean space transfers to the active manifold under $\mathcal{C}^1$ partial smoothness. The GEB generalizes the standard error bound, which upper bounds the distance to a subset of critical points using first-order stationarity residuals, by allowing the distance to be composed with a gauge function. In the ambient space, these residuals are measured by the subdifferential or the slope of the original function, while on the manifold they are measured by the Riemannian gradient or the slope of the restricted function. We develop complementary geometric and metric analyses to show that the ambient and manifold GEBs are equivalent with the same gauge function. The geometric argument shows that the projection of the subdifferential onto the tangent space coincides with the Riemannian gradient, while the metric approach combines slope estimates, identifiability, and a new uniform linear-growth result for $\mathcal{C}^1$ manifolds. We further extend our analysis to establish the analogous equivalence for the Kurdyka–Łojasiewicz (KL) property, with the same desingularizer, and separately characterize their connections with the GEB. Finally, for regularized optimization, we relate subdifferential and proximal error bounds, thereby placing the previously established equivalence between proximal and manifold error bounds for ℓ_1-regularization within a broader framework.

1. Introduction. Active manifolds capture the local structure of many nonsmooth optimization problems, such as the support of a sparse solution and the rank of a matrix solution. On an active manifold, the nonsmooth function inducing the structure often becomes smooth. This viewpoint is formalized through *partial smoothness* [27, 15] and is closely related to *identifiability* [44, 20, 21], which characterizes when a sequence approaching a critical point eventually remains on an active manifold. This framework supports sensitivity analysis and active-set algorithms, and enables local analysis of a nonsmooth problem through its smooth restriction.

To leverage this smooth reduction in convergence analysis, one needs to understand how quantitative regularity transfers between the ambient space and the active manifold. Two prominent regularity properties, the Kurdyka–Łojasiewicz (KL) property and the error bound (EB), have been central to such analyses; see, for example, [37, 42, 41, 47, 1, 6]. The KL property relates a first-order stationarity residual to the objective-value gap and uses a desingularizer function to quantify the relationship, whereas an EB upper bounds the distance to a solution or critical-point set by this residual. Motivated in part by the flexibility of the desingularizer in the KL property, we work with a generalized error bound (GEB) that replaces the linear dependence on the distance in the standard EB by a possibly nonlinear gauge function. The desingularizing and gauge functions allow us to compare the quantitative strength of the corresponding KL and GEB properties.

In [46], we established an error-bound equivalence for ℓ_1-regularized problems of the form $g(x) + \lambda\|x\|_1$ with g Lipschitz continuously differentiable, and used it to analyze a two-metric adaptive projection method. We showed that the standard proximal EB at a nondegenerate minimizer is equivalent to the EB on the active manifold determined by its support, with a linear gauge in both cases. This result motivates our question:

> *Does the same equivalence extend beyond ℓ_1 regularization to general partly smooth problems, and does restriction to the active manifold yield a sharper local error bound?*

To answer the second part of the question, we adopt the generalization of the standard EB with a gauge function for characterizing sharpness of the regularity condition and the resulting convergence rates.

Existing analyses provide mixed responses to this question. On one hand, one might expect restriction to sharpen regularity. Indeed, a smooth loss can have a singular positive semidefinite Hessian in the ambient space, while its restriction to the tangent space of the active manifold is positive definite. To illustrate this, consider

$$f(x) = g(x) + |x_1| + |x_2|, \qquad g(x) = 0.5(x_1 + 0.5x_2 - 2)^2.$$

At the nondegenerate minimizer $\bar{x} = (1, 0)$, the active manifold is $\mathcal{M} = \{x_2 = 0\}$, and

$$\nabla^2 g \equiv \begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.25 \end{pmatrix}, \qquad f(x_1, 0) = 1.5 + 0.5(x_1 - 1)^2, \quad (x_1 > 0).$$

The ambient Hessian of g has the null direction $(1, -2)$, whereas the second derivative of $f|_{\mathcal{M}}$ is 1 near $\bar{x}$. Such a regularity condition on the active manifold has been used to design algorithms with provably fast local convergence [3, 18], and to explain the empirically observed fast local convergence of existing methods [34]. Similarly, [25, 22, 11] assume on-manifold error-bound conditions to design their algorithms and establish local superlinear convergence.

On the other hand, existing transfer results for other regularity properties suggest otherwise. Under suitable assumptions, local minimality, strict local minimality, and quadratic growth are preserved upon restriction to an

*Department of Mathematics, The University of Hong Kong, Pokfulam, Hong Kong (wuhanju@connect.hku.hk).
†Institute of Statistical Mathematics and University of Tokyo, Tokyo, Japan (leechingpei@gmail.com).
‡Department of Mathematics and Institute of Data Science, The University of Hong Kong, Pokfulam, Hong Kong (yxie21@hku.hk), (corresponding author).

active manifold [27, 31, 15, 29]. In the other direction, [44, 20, 45] showed that quadratic growth propagates from the active manifold to the ambient space in various settings, and [33] established manifold-to-ambient transfer of the KL property under partial smoothness. Lewis and Tian [29] subsequently studied the equivalence of the KL property with general desingularizers on identifiable sets in complete metric spaces. As discussed in detail in Section 5, the ambient-to-identifiable-set implication requires additional care and fails in this generality, which provides an additional motivation for our Euclidean analysis under $\mathcal{C}^1$ partial smoothness. Nevertheless, the manifold-to-ambient transfers above already rule out sharpening under their respective hypotheses.

To resolve this tension, we aim at minimal assumptions and study the transferability of GEBs between the ambient space and the active manifold. We therefore only assume that the underlying active manifold is $\mathcal{C}^1$, the minimal regularity needed for the tangent space and the Riemannian gradient to be well defined. Such manifolds arise, for example, in the following ℓ_p-norm ball constrained optimization problems with $p \in (1,2)$.

$$\min_{x\in\mathbb{R}^n} g(x) \quad \text{s.t.} \quad \|x\|_p \le r. \tag{1.1}$$

The active manifold at any $\bar{x}$ with $\|\bar{x}\|_p = r$ is $\mathcal{M} = \{x \in \mathbb{R}^n : \|x\|_p = r\}$, and $F(x) = \sum_{i=1}^n |x_i|^p - r^p$ is a local defining map for $\mathcal{M}$ on $\mathbb{R}^n \setminus \{0\}$, where $\nabla F(x) = p(\operatorname{sgn}(x_i)|x_i|^{p-1})_{i=1}^n \neq 0$ is Hölder continuous with exponent $p-1$, so $\mathcal{M}$ is $\mathcal{C}^{1,p-1}$ but not $\mathcal{C}^2$, and this regularity is sharp at points of $\mathcal{M}$ with zero components.

Analyses in the literature regarding partly smooth functions often assume $\mathcal{C}^2$ smoothness so that the nearest-point projection onto the manifold is single valued and well defined. Without these properties, we develop new tools by constructing a retraction-like mapping through a local graph view of $\mathcal{C}^1$ manifolds. Our results show that, under such minimal assumptions, the ambient and manifold GEBs hold with the same gauge function, and we further confirm the same for the desingularizer of the KL property. Consequently, the answer to the question raised above is affirmative in the first part, while negative in the second part: *Restriction to the active manifold does not sharpen the local EB or KL property* in this sense, although the constants and neighborhoods may differ. This negative result also suggests that the fast local convergence results observed in existing works might hold regardless of identifying the active manifold, which is consistent with and provides a complementary explanation to the recent discovery of [12] for degenerate situations.

More detailed contributions of this work are summarized below.

- We derive a uniform linear-growth estimate for nearby points off the active manifold under $\mathcal{C}^1$ partial smoothness and nondegeneracy. The estimate requires only a $\mathcal{C}^1$ active manifold and a $\mathcal{C}^1$ restriction, and supplies the local control needed for the transfer arguments without relying on a smooth nearest-point projection available for $\mathcal{C}^2$ manifolds (Theorem 3.6).
- We establish the equivalence between ambient and manifold GEBs with the same gauge function under $\mathcal{C}^1$ partial smoothness and nondegeneracy, relative to subsets of critical points on which f is locally constant. We give two complementary proofs: an intuitive geometric proof relating subdifferential residuals to the Riemannian gradient (Theorem 4.10), and a slope- and identifiability-based metric proof that has the potential to extend to more general settings (Theorem 4.11). For ℓ_1-regularized problems, this result recovers the known transferability of the error bound as a special case (Section 6.2).
- We show ambient-to-manifold transfer of the KL property under $\mathcal{C}^1$ partial smoothness and nondegeneracy for desingularizers locally concave near 0 (Theorem 5.3). Together with the known converse, this yields equivalence with the same desingularizer in the Euclidean setting. We also provide counterexamples (Section A) showing that the ambient-to-restricted implication may not hold when the identifiable set fails to be a $\mathcal{C}^1$ manifold.
- We give sufficient conditions for a GEB to imply a KL inequality and quantify the relationship between its gauge function and the resulting desingularizer (Proposition 5.2). This implication does not require partial smoothness and is therefore of independent interest.

This paper is organized as follows. Section 2 presents the preliminaries, and Section 3 develops the tools based on partial smoothness, identifiability, and linear growth. The GEB and KL transfer results are established in Sections 4 and 5. In Section 6, we discuss regularized optimization and apply our results to the motivating examples of ℓ_1-regularization and ℓ_p-norm constraints. Some final comments appear in Section 7. The appendices contain the metric-space counterexample and supplementary proofs.

2. Notation and Preliminaries.

In this paper, $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product on $\mathbb{R}^n$, $\|\cdot\|$ the induced Euclidean norm, and d the Euclidean distance between two points, namely $d(x,y) = \|x-y\|$. For a set $C \subset \mathbb{R}^n$, $\operatorname{ri} C$ denotes its relative interior, $\operatorname{aff} C$ its affine hull, and $\operatorname{par} C$ the subspace of $\mathbb{R}^n$ parallel to $\operatorname{aff} C$. We write $\operatorname{dist}(x, C) \coloneqq \inf_{y\in C} \|x-y\|$ as the distance between x and C, and use $P_C(x)$ for the Euclidean projection onto C whenever it is well defined and single valued. We follow the convention to set $\operatorname{dist}(x, C) = +\infty$ when $C = \emptyset$. For any given point x, the Euclidean open ball centered at x with radius ϵ is denoted by $B_\epsilon(x)$. The unit sphere is denoted by $\mathbb{S}^{n-1} \coloneqq \{v \in \mathbb{R}^n : \|v\| = 1\}$. The extended real number is defined as $\overline{\mathbb{R}} \coloneqq \mathbb{R} \cup \{+\infty\}$, and the nonnegative real number is denoted by $\mathbb{R}_+$. For an extended-real-valued function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, its domain is denoted by $\operatorname{dom} f \coloneqq \{x \in \mathbb{R}^n : f(x) < +\infty\}$. The sublevel set of f at level α is denoted by

$[f \le \alpha] \coloneqq \{x \in \mathbb{R}^n : f(x) \le \alpha\}$, and we also use $[\alpha \le f \le \beta]$ to denote the set $\{x \in \mathbb{R}^n : \alpha \le f(x) \le \beta\}$. We write $x_k \to_f x$ for $(x_k, f(x_k)) \to (x, f(x))$. We let Id to denote the identity map.

2.1. Manifolds. We follow the standard definition and notation for manifold-related concepts in [7]. All manifolds $\mathcal{M}$ are assumed throughout to be embedded in $\mathbb{R}^n$ and we consider the tangent and normal spaces to $\mathcal{M}$ as subspaces of $\mathbb{R}^n$. A non-empty set $\mathcal{M} \subset \mathbb{R}^n$ is a $\mathcal{C}^p$-smooth embedded submanifold of dimension m if around any $x \in \mathcal{M}$ there exists an open neighborhood $U \subset \mathbb{R}^n$ of x and a $\mathcal{C}^p$-smooth map $F : U \to \mathbb{R}^{n-m}$ such that $\mathcal{M} \cap U = F^{-1}(0)$, and for all $y \in U$, the Jacobian matrix $\nabla F(y)$ has full rank $n - m$. We call such an F a *local defining map* for $\mathcal{M}$. The tangent space of $\mathcal{M}$ at x is defined as $T_x\mathcal{M} \coloneqq \ker \nabla F(x)$. Another definition of the tangent space is that $T_x\mathcal{M}$ consists of all tangent vectors to smooth curves on $\mathcal{M}$ passing through x, i.e., $\dot\gamma(0)$ for all $\mathcal{C}^1$-smooth curves $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ with $\gamma(0) = x$. The normal space $N_x\mathcal{M}$ of $\mathcal{M}$ at x is defined as the orthogonal complement of $T_x\mathcal{M}$, i.e., $N_x\mathcal{M} \coloneqq (T_x\mathcal{M})^\perp = \operatorname{Range}\left(\nabla F(x)^\top\right)$.

When $\mathcal{M}$ is $\mathcal{C}^2$-smooth, the nearest-point projection $P_\mathcal{M}$ is well defined, single valued, and $\mathcal{C}^1$-smooth in a neighborhood of $\mathcal{M}$, and thus $\mathcal{M}$ being $\mathcal{C}^2$-smooth is the standard assumption in the literature for partial smoothness or identifiability (c.f. [15, 28, 29]). On the contrary, this work tries to relax the $\mathcal{C}^2$-smoothness condition to $\mathcal{C}^1$ for accommodating more general applications like those we discussed in Section 1. In this scenario, projection may not be well defined or smooth, and we therefore develop new tools that do not resort to the projection.

For a $\mathcal{C}^p$-smooth manifold $\mathcal{M}$, a function $f : \mathcal{M} \to \mathbb{R}$ is said to be $\mathcal{C}^p$-smooth if for each $x \in \mathcal{M}$ there exists an open neighborhood $U \subset \mathbb{R}^n$ of x and a $\mathcal{C}^p$-smooth function $\tilde f : U \to \mathbb{R}$ such that $\tilde f|_{\mathcal{M}\cap U} = f|_{\mathcal{M}\cap U}$. The function $\tilde f$ is called a $\mathcal{C}^p$-smooth extension of f around x.

It is convenient to equip each tangent space of the manifold $\mathcal{M}$ with an inner product $\langle\cdot,\cdot\rangle_x : T_x\mathcal{M}\times T_x\mathcal{M} \to \mathbb{R}$. If this inner product varies smoothly with x, then $\mathcal{M}$ is called a Riemannian manifold and the inner product is called a Riemannian metric. When $\mathcal{M}$ is an embedded submanifold of $\mathbb{R}^n$, the tangent space $T_x\mathcal{M}$ is a subspace of $\mathbb{R}^n$. It is therefore natural to choose the restriction of the Euclidean inner product on $\mathbb{R}^n$ to $T_x\mathcal{M}$ as the Riemannian metric, i.e., $\langle \eta, \xi\rangle_x \coloneqq \langle \eta, \xi\rangle$ for all $\eta, \xi \in T_x\mathcal{M}$. In this case, we call $\mathcal{M}$ a Riemannian submanifold of $\mathbb{R}^n$, and consequently, the distance measure in this work is also the ambient Euclidean distance.

For a Riemannian manifold $\mathcal{M}$, we can define the Riemannian gradient of a smooth function $f : \mathcal{M} \to \mathbb{R}$ as follows. The Riemannian gradient $\nabla_\mathcal{M} f(x) \in T_x\mathcal{M}$ is the unique tangent vector satisfying

$$\langle \nabla_\mathcal{M} f(x), \xi\rangle_x = \mathrm{D}f(x)[\xi] = \frac{d}{dt} f(\gamma(t))\big|_{t=0}, \quad \forall \xi \in T_x\mathcal{M},$$

where $\mathrm{D}f(x)[\xi]$ is the directional derivative of f at x along ξ and $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ is any smooth curve satisfying $\gamma(0) = x$ and $\dot\gamma(0) = \xi$. If $\mathcal{M}$ is a Riemannian submanifold of $\mathbb{R}^n$, then the Riemannian gradient can be computed via the projection of the Euclidean gradient onto the tangent space, i.e.,

$$\nabla_\mathcal{M} f(x) = P_{T_x\mathcal{M}} \nabla \tilde f(x)$$

where $\tilde f : U \to \mathbb{R}$ is any smooth extension of f around x. We slightly abuse the notation and write $\nabla_\mathcal{M} f(x)$ for both $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $f|_\mathcal{M}$ as well when f is smooth on $\mathcal{M}$, even if f is not smooth on $\mathbb{R}^n$.

The following result from [38, Thm. 2.2, Cor. 2.3] is critical for our analysis. Its proof is a simple application of the inverse function theorem, and we include it in Section B for completeness. More details about the inverse function theorem for manifolds can be found in [26, Thm. 4.5].

Lemma 2.1 (Manifold locally as a graph). *Let $\mathcal{M} \subset \mathbb{R}^n$ be a $\mathcal{C}^p$-smooth embedded submanifold around $\bar x \in \mathcal{M}$ for some $p \ge 1$. Then, there exist a neighborhood U of $\bar x$, $\rho > 0$ and a $\mathcal{C}^p$ map*

$$h : T_{\bar x}\mathcal{M} \cap B_\rho(0) \to N_{\bar x}\mathcal{M}$$

such that $h(0) = 0$, $\nabla h(0) = 0$, and

$$\mathcal{M} \cap U = \{\bar x + u + h(u) : u \in T_{\bar x}\mathcal{M} \cap B_\rho(0)\}.$$

Consequently, for all y sufficiently close to $\bar x$ with decomposition $y - \bar x = u + v$ for some $u \in T_{\bar x}\mathcal{M}$ and $v \in N_{\bar x}\mathcal{M}$,

$$R_{\bar x}(y) \coloneqq \bar x + u + h(u) \in \mathcal{M}$$

is well defined, and $R_{\bar x} = \mathrm{Id}$ on $\mathcal{M}$ around $\bar x$.

We use the notation $R_{\bar x}(y)$ due to its retraction-like behavior locally. A point y near $\bar x$ admits a unique decomposition $y = \bar x + u + w$ with $u \in T_{\bar x}\mathcal{M}$ and $w \in N_{\bar x}\mathcal{M}$, and $R_{\bar x}$ maps y to $\bar x + u + h(u) \in \mathcal{M}$. Moreover, the curve $c(t) \coloneqq \bar x + tu + h(tu)$ satisfies $c(0) = \bar x$ and $c'(0) = u$.

2.2. Slope and Variational Analysis. We use standard definitions and notation of variational analysis following [40]. For an extended-real-valued function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $\hat{\partial} f(\bar{x})$ denotes its *Fréchet subdifferential* at $\bar{x} \in \operatorname{dom} f$, $\partial f(\bar{x})$ the *limiting subdifferential* and $\partial_P f(\bar{x})$ the *proximal subdifferential*, and it is well-known that $\partial_P f(\bar{x}) \subseteq \hat{\partial} f(\bar{x}) \subseteq \partial f(\bar{x})$. The limiting critical-point set is defined as

$$\operatorname{crit} f := \{x \in \operatorname{dom} f : 0 \in \partial f(x)\}.$$

We say f is *subdifferentially regular* at $\bar{x}$ if $\hat{\partial} f(\bar{x}) = \partial f(\bar{x})$, and *prox-regular* at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ if there exist $\epsilon > 0$ and $\rho \geq 0$ such that

$$\begin{aligned} f(x') \geq f(x) + \langle v, x' - x \rangle - \frac{\rho}{2}\|x' - x\|^2,& \\ \forall x' \in B_\epsilon(\bar{x}), \quad x \in B_\epsilon(\bar{x}) \cap [f < f(\bar{x}) + \epsilon], \quad v \in \partial f(x) \cap B_\epsilon(\bar{v}).& \end{aligned}$$

If f is prox-regular at $\bar{x}$ for all $\bar{v} \in \partial f(\bar{x})$, we simply say it is prox-regular at $\bar{x}$. Prox-regularity implies that for all $(x, v) \in \operatorname{gph} \partial f$ close enough to $(\bar{x}, \bar{v})$ with $f(x)$ close enough to $f(\bar{x})$, v is a proximal subgradient of f at x.

The following proposition shows that the Riemannian gradient is the projection of any Fréchet subgradient onto the tangent space of the manifold. Similar results for $\mathcal{C}^2$ manifolds can be found in [27, Prop. 2.2] and [8, Prop. 12]. We provide the proof of this result in Section B for completeness.

Proposition 2.2. *Suppose that $\mathcal{M}$ is a $\mathcal{C}^1$-smooth manifold containing x, and that for a given function f, the restriction $f|_{\mathcal{M}}$ is $\mathcal{C}^1$-smooth around x. Then,*

$$\operatorname{par} \hat{\partial} f(x) \subseteq N_x\mathcal{M}, \tag{2.1}$$

and for every $g \in \hat{\partial} f(x)$, the Riemannian gradient of $f|_{\mathcal{M}}$ satisfies

$$\nabla_{\mathcal{M}} f(x) = P_{T_x\mathcal{M}}(g). \tag{2.2}$$

We now introduce the *slope* $|\nabla f|(x)$, a central tool in this paper. We follow the definition in [29, Def. 2.1], and it is easy to see that this definition is equivalent to the classical one in, for example, [13, Def. 2.2.1], [24, Def. 3.2] and [14, Sec. 2].

Definition 2.3 (Slope; [29, Def. 2.1]). *For a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the slope of f at $\bar{x} \in \mathbb{R}^n$ is defined as*

$$|\nabla f|(\bar{x}) := \begin{cases} \infty & \text{if } \bar{x} \notin \operatorname{dom} f, \\ 0 & \text{if } \bar{x} \text{ is a local minimizer of } f, \\ \limsup_{x \to \bar{x},\ x \neq \bar{x}} \frac{f(\bar{x}) - f(x)}{d(\bar{x}, x)} & \text{otherwise.} \end{cases}$$

We recall several basic properties of the slope from [24, Prop. 8.3, 8.5], [13, Prop. 2.3.3], and [10, Observation 4.3].

Proposition 2.4. *For any closed function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, we always have*
- $0 \leq \operatorname{dist}(0, \partial f(x)) \leq |\nabla f|(x) \leq \operatorname{dist}(0, \hat{\partial} f(x))$ *for all x.*
- *If f is subdifferentially regular at x, then $|\nabla f|(x) = \operatorname{dist}(0, \partial f(x))$.*

The following lemma from [30, Ex. 2.1] that relates the quantities $|\nabla(f|_{\mathcal{M}})|(x)$ and $\|\nabla_{\mathcal{M}} f(x)\|$ will be useful in our analysis. For completeness, we provide a concise proof in Section B.

Lemma 2.5. *Suppose that $\mathcal{M}$ is a $\mathcal{C}^1$-smooth embedded submanifold around x and that $f|_{\mathcal{M}}$ is $\mathcal{C}^1$-smooth around x, then*

$$|\nabla(f|_{\mathcal{M}})|(x) = \|\nabla_{\mathcal{M}} f(x)\|.$$

3. Partial Smoothness, Identifiability and Linear Growth. This section develops tools essential to the main results of this paper. We first introduce the definitions of partial smoothness and identifiability. The following is the definition of partial smoothness from [15, 29].[1]

Definition 3.1 ($\mathcal{C}^p$-Partial smoothness). *Given a closed function $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, a point $\bar{x} \in \operatorname{dom} f$, and a $\mathcal{C}^p$-smooth manifold $\mathcal{M}$ containing $\bar{x}$ for some $p \geq 1$. f is $\mathcal{C}^p$* 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 $\bar{y}$.*[2]

Restricted smoothness: *the restriction $f|_{\mathcal{M}}$ is $\mathcal{C}^p$-smooth around $\bar{x}$.*

[1] For the part of inner semicontinuity, our results only need it to hold for the Fréchet subdifferential $\hat{\partial} f$, instead of the limiting subdifferential ∂f in the standard definition of partial smoothness stated here.

[2] This definition differs slightly from the original one in [27], which uses subdifferential regularity instead of prox-regularity (and treats prox-regularity as an additional, independent assumption).

Sharpness: *the subspace* par $\hat{\partial} f(\bar{x})$ *is just the normal space* $N_{\bar{x}}\mathcal{M}$.

Inner semicontinuity: *for all* $y \in \partial f(\bar{x})$ *near* $\bar{y}$ *and every sequence* $x_r \to \bar{x}$ *in* $\mathcal{M}$, *there exists a sequence* $y_r \in \partial f(x_r)$ *converging to* y.

This manifold $\mathcal{M}$ is often called the *active manifold* of f at $\bar{x}$.

Notice that $\hat{\partial} f(x)$ is nonempty for every $x \in \mathcal{M}$ sufficiently close to $\bar{x}$ if the function f is $\mathcal{C}^p$-partly smooth at the point $\bar{x}$ for 0 relative to $\mathcal{M}$. Otherwise, there would be a sequence $x_r \to \bar{x}$ in $\mathcal{M}$ with $\hat{\partial} f(x_r) = \emptyset$. Inner semicontinuity at the subgradient 0 then provides $v_r \in \partial f(x_r)$ with $v_r \to 0$. Since $f(x_r) \to f(\bar{x})$, prox-regularity at $\bar{x}$ for 0 would imply $v_r \in \partial_P f(x_r) \subseteq \hat{\partial} f(x_r)$ for all large r, a contradiction.

For a partly smooth function, the sharpness condition in Definition 3.1 is stable: it also holds at all nearby points on the *active manifold*. A similar result has been established in [27, Prop. 2.10] for $\mathcal{C}^2$-partly smooth functions with a slightly different definition.

PROPOSITION 3.2. *If the function* f *is* $\mathcal{C}^p$*-partly smooth for some* $p \geq 1$ *at the point* $\bar{x}$ *for* 0 *relative to* $\mathcal{M}$, *then all* $x \in \mathcal{M}$ *sufficiently close to* $\bar{x}$ *satisfy*

$$\operatorname{par} \hat{\partial} f(x) = N_x\mathcal{M} \tag{3.1}$$

Proof. From (2.1), we always have par $\hat{\partial} f(x) \subseteq N_x\mathcal{M}$. Suppose now that the reverse inclusion fails, then there exist points $x_r \in \mathcal{M}$ converging to $\bar{x}$ and unit vectors

$$z_r \in N_{x_r}\mathcal{M} \cap \left(\operatorname{par} \hat{\partial} f(x_r)\right)^{\perp}.$$

Passing to a subsequence and using the continuity of the normal spaces of a $\mathcal{C}^1$ manifold, we may assume that $z_r \to z \in N_{\bar{x}}\mathcal{M}$ with $\|z\| = 1$.

Prox-regularity at $\bar{x}$ for 0 implies $0 \in \partial_P f(\bar{x}) \subseteq \hat{\partial} f(\bar{x})$. Let $u, v \in \hat{\partial} f(\bar{x})$ be sufficiently close to 0. By the inner semicontinuity condition in Definition 3.1, there exist $u_r, v_r \in \partial f(x_r)$ such that $u_r \to u$ and $v_r \to v$. Since $f|_{\mathcal{M}}$ is continuous, we also have $f(x_r) \to f(\bar{x})$. Therefore, by prox-regularity at $\bar{x}$ for 0, we get that for all sufficiently large r, u_r, v_r are proximal subgradients, and hence belong to $\hat{\partial} f(x_r)$. Consequently,

$$\langle z_r, u_r - v_r \rangle = 0.$$

Letting $r \to \infty$ then yields $\langle z, u - v \rangle = 0$.

Finally, because $\hat{\partial} f(\bar{x})$ is convex and contains 0, its intersection with any neighborhood of 0 has the same parallel space as $\hat{\partial} f(\bar{x})$. Thus the preceding identity shows that

$$z \in \left(\operatorname{par} \hat{\partial} f(\bar{x})\right)^{\perp}.$$

The sharpness condition at $\bar{x}$ gives par $\hat{\partial} f(\bar{x}) = N_{\bar{x}}\mathcal{M}$. This contradicts the facts that $z \in N_{\bar{x}}\mathcal{M}$ and $\|z\| = 1$. Hence (3.1) holds for all $x \in \mathcal{M}$ sufficiently close to $\bar{x}$. ◻

Partial smoothness is closely related to the concept of identifiability. We follow the definition from [16, 29].

DEFINITION 3.3 (Identifiability; Modulus of Identifiability). *For a closed function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, *consider a point* $\bar{x} \in \operatorname{dom} f$ *with* $0 \in \hat{\partial} f(\bar{x})$ *and a set* $\mathcal{M} \subset \mathbb{R}^n$ *containing* $\bar{x}$. $\mathcal{M}$ *is identifiable at* $\bar{x}$ *if*

$$\eta := \liminf_{x \to_f \bar{x}, x \notin \mathcal{M}} \operatorname{dist}(0, \partial f(x)) > 0.$$

The value η *is called the* modulus of identifiability.

Remark 3.4. When $\mathcal{M}$ is a $\mathcal{C}^p$ manifold for some $p \geq 1$, since a $\mathcal{C}^p$ manifold $\mathcal{M}$ is locally closed, [29, Prop. 3.12, Cor. 3.14] indicated that

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

It is shown in [15, Thm. 10.12] and [29, Thm. 8.5] that if f is $\mathcal{C}^2$-partly smooth at $\bar{x}$ for zero relative to $\mathcal{M}$ with the nondegeneracy condition

$$0 \in \operatorname{ri} \hat{\partial} f(\bar{x}) \tag{3.2}$$

holds, then $\mathcal{M}$ is identifiable at $\bar{x}$. We relax the condition from $\mathcal{C}^2$ to $\mathcal{C}^1$-partial smoothness in Lemma 3.5 below. Throughout the remainder of this section, we impose the following assumption.

Assumption 1. The function f is proper, closed, and $\mathcal{C}^1$-partly smooth at $\bar{x}$ for 0 relative to $\mathcal{M} \ni \bar{x}$, with $0 \in \operatorname{ri} \hat{\partial} f(\bar{x})$.

LEMMA 3.5. *For a function* f *under Assumption* 1, $\mathcal{M}$ *is identifiable at* $\bar{x}$ *for* 0 *in the sense of Definition* 3.3.

Proof. Assume that the conclusion does not hold, then there exists a sequence $x_r \to_f \bar{x}$ with $x_r \notin \mathcal{M}$ such that $y_r \in \partial f(x_r) \to 0 \in \partial f(\bar{x})$ as $r \to \infty$. Since $\mathcal{M}$ is $\mathcal{C}^1$, we can use Lemma 2.1 to write $\mathcal{M}$ around $\bar{x}$ as a $\mathcal{C}^1$ graph over $T_{\bar{x}}\mathcal{M}$ with graph map $h : T_{\bar{x}}\mathcal{M} \cap B_\epsilon(0) \to N_{\bar{x}}\mathcal{M}$ satisfying $h(0) = 0$ and $\nabla h(0) = 0$. For each x_r, we have the decomposition $x_r - \bar{x} = p_r + q_r$ with $p_r \in T_{\bar{x}}\mathcal{M}$ and $q_r \in N_{\bar{x}}\mathcal{M}$, and we define

$$x'_r := \bar{x} + p_r + h(p_r) \in \mathcal{M}.$$

Then $x'_r \to \bar{x}$, $x_r \neq x'_r$ for all large r, and for all such r we have

$$u_r := \frac{x_r - x'_r}{\|x_r - x'_r\|} = \frac{q_r - h(p_r)}{\|q_r - h(p_r)\|} \in N_{\bar{x}}\mathcal{M}.$$

By taking a subsequence if necessary, we can assume that $u_r \to u \in N_{\bar{x}}\mathcal{M}$ with $\|u\| = 1$. Sharpness and nondegeneracy conditions imply that there exists $\delta > 0$ small enough such that $\delta u \in \hat{\partial} f(\bar{x}) \subseteq \partial f(\bar{x})$ and δu lies in the neighborhood of 0 where the inner semicontinuity condition holds. Then, inner semicontinuity of ∂f at $\bar{x}$ for 0 on $\mathcal{M}$ implies that there exists $z_r \in \partial f(x'_r)$ such that $z_r \to \delta u$ as $r \to \infty$. By prox-regularity of f at $\bar{x}$ for 0, there exists $\rho \geq 0$ such that for all sufficiently large r, we have

$$\begin{cases} f(x_r) & \geq f(x'_r) + \langle z_r, x_r - x'_r \rangle - \frac{\rho}{2}\|x_r - x'_r\|^2, \\ f(x'_r) & \geq f(x_r) + \langle y_r, x'_r - x_r \rangle - \frac{\rho}{2}\|x_r - x'_r\|^2. \end{cases}$$

Summing the two inequalities up gives

$$0 \geq \langle z_r - y_r, x_r - x'_r \rangle - \rho\|x_r - x'_r\|^2.$$

Therefore, for all r large enough, since $\|x_r - x'_r\| > 0$, we have

$$\rho\|x_r - x'_r\| \geq \langle z_r - y_r, u_r \rangle.$$

Taking limits on both sides as $r \to \infty$ gives the following contradiction.

$$0 \geq \langle \delta u - 0, u \rangle = \delta\|u\|^2 = \delta > 0,$$ □

By the lower semicontinuity of f, Lemma 3.5 implies there exist $\epsilon, \nu, \eta > 0$ such that

$$\operatorname{dist}(0, \partial f(x)) > \eta, \qquad \forall x \in B_\epsilon(\bar{x}) \cap [f \leq f(\bar{x}) + \nu] \setminus \mathcal{M}.$$

We close this section by providing a linear growth for identifiable manifolds. When an identifiable set $\mathcal{M}$ is a $\mathcal{C}^2$-smooth manifold, its nearest-point projection is locally well defined and single valued. [9, Thm. D.2] states that in this case, f grows at least linearly around $\bar{x}$ in the normal direction of $\mathcal{M}$, i.e., there exist constants $\delta, \epsilon > 0$ such that

$$f(x) \geq f(P_{\mathcal{M}}(x)) + \delta d(x, P_{\mathcal{M}}(x)), \quad \forall x \in B_\epsilon(\bar{x}).$$

Their result leverages the nearest-point projection and therefore does not extend easily to the $\mathcal{C}^1$ setting, where the nearest-point projection needs not be locally well defined. Motivated by the sharp-minimizer result on $\mathcal{C}^2$ manifolds in [27, Thm. 6.1], we develop in Theorem 3.6 a linear-growth result for $\mathcal{C}^1$ manifolds.

THEOREM 3.6 (Linear growth under $\mathcal{C}^1$-partial smoothness). *Consider $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ satisfying Assumption 1 and set*

$$U := T_{\bar{x}}\mathcal{M}, \qquad V := N_{\bar{x}}\mathcal{M},$$

then there are neighborhoods $U_0 \subseteq U$ and $V_0 \subseteq V$ of 0 and a $\mathcal{C}^1$ map $v : U_0 \to V$ such that $v(0) = 0$, $\nabla v(0) = 0$, and

$$\bar{x} + u + w \in \mathcal{M} \quad \Longleftrightarrow \quad w = v(u) \tag{3.3}$$

for all $u \in U_0$ and all $w \in V_0$. Moreover, there exists a constant $\kappa > 0$ such that

$$f(\bar{x} + u + w) \geq f(\bar{x} + u + v(u)) + \kappa\|w - v(u)\| \tag{3.4}$$

for all $u \in U_0$ and all $w \in V_0$.

Consequently, every point y sufficiently close to $\bar{x}$ can be written uniquely as $y = \bar{x} + u + w$ with $u \in U$ and $w \in V$, and there exists a point

$$R_{\bar{x}}(y) := \bar{x} + u + v(u) \in \mathcal{M}$$

that satisfies

$$f(y) \geq f(R_{\bar{x}}(y)) + \kappa\|y - R_{\bar{x}}(y)\| \geq f(R_{\bar{x}}(y)) + \kappa \operatorname{dist}(y, \mathcal{M}).$$

Proof. The part of (3.3) follows directly from Lemma 2.1, and we thus focus on proving (3.4). If $V = \{0\}$, then $w = v(u) = 0$ in the local representation and the conclusion is immediate. We therefore suppose $V \neq \{0\}$. Similarly, we also assume $w \neq v(u)$ below.

Since $0 \in \operatorname{ri} \hat{\partial} f(\bar{x})$ and $\operatorname{par} \hat{\partial} f(\bar{x}) = V$, there exists $\delta > 0$ such that

$$B_\delta(0) \cap V \subseteq \hat{\partial} f(\bar{x}) \subseteq \partial f(\bar{x}). \tag{3.5}$$

By prox-regularity of f at $\bar{x}$ for 0, there exist $\tau > 0$ and $\rho \geq 0$ such that

$$f(z) \geq f(y) + \langle s, z - y\rangle - \frac{\rho}{2}\|z - y\|^2 \tag{3.6}$$

whenever $y, z \in B_\tau(\bar{x})$, $f(y) < f(\bar{x}) + \tau$, and $s \in \partial f(y) \cap B_\tau(0)$. For any $\alpha \in (0, \delta)$, we claim that there is a neighborhood W of $\bar{x}$ such that

$$\operatorname{dist}(\alpha d, \partial f(y)) \leq \frac{\alpha}{4}, \quad \forall y \in W \cap \mathcal{M},\ \forall d \in \mathbb{S}^{n-1} \cap V. \tag{3.7}$$

Indeed, if this fails, there must exist sequences $y_k \to \bar{x}$ in $\mathcal{M}$ and $d_k \in \mathbb{S}^{n-1} \cap V$ and $d \in \mathbb{S}^{n-1} \cap V$ such that

$$\operatorname{dist}(\alpha d_k, \partial f(y_k)) > \frac{\alpha}{4}, \quad \lim_{k\to\infty} d_k = d. \tag{3.8}$$

Since $\alpha d \in \partial f(\bar{x})$ by (3.5) and that $\alpha < \delta$, the inner semicontinuity condition gives points $s_k \in \partial f(y_k)$ with $s_k \to \alpha d$, which contradicts (3.8) and thus (3.7) holds.

Now we select the neighborhood U_0 and $\epsilon > 0$ small enough so that: (a) $y := \bar{x} + u + v(u) \in \mathcal{M} \cap W$ for the W defined in the previous paragraph and satisfies $f(y) < f(\bar{x}) + \tau$ for all $u \in U_0$ (such U_0 exists by Lemma 2.1 and the continuity of $f|_{\mathcal{M}}$), (b) $\bar{x} + u + w \in B_\tau(\bar{x})$ and $y \in B_\tau(\bar{x})$ whenever $u \in U_0$ and $w \in B_\epsilon(0) \cap V$, and (c) $\|v(u)\| < \epsilon$ for all $u \in U_0$. When $\rho > 0$, we also require

$$2\epsilon \leq \alpha/(2\rho) \tag{3.9}$$

We then fix $u \in U_0$ and $w \in B_\epsilon(0) \cap V$ with $w \neq v(u)$, set

$$d = \frac{w - v(u)}{\|w - v(u)\|} \in \mathbb{S}^{n-1} \cap V, \quad y = \bar{x} + u + v(u) \in \mathcal{M},$$

and choose $\alpha \in (0, \delta)$ small enough and $s \in \partial f(y)$ satisfying (3.7) so that $\|s\| \leq \|s - \alpha d\| + \|\alpha d\| < 2\alpha < \tau$. We can thus apply (3.6) with $z := \bar{x} + u + w$ and y and obtain

$$f(\bar{x} + u + w) - f(\bar{x} + u + v(u)) \geq \langle s, w - v(u)\rangle - \frac{\rho}{2}\|w - v(u)\|^2. \tag{3.10}$$

Since $\|d\| = 1$,

$$\frac{\langle s, w - v(u)\rangle}{\|w - v(u)\|} = \langle s, d\rangle \geq (\langle s - \alpha d, d\rangle + \alpha) \geq (\alpha - \|s - \alpha d\|) \geq \frac{3\alpha}{4}.$$

When $\rho > 0$, our choice for U_0 and ϵ yields $\|w - v(u)\| \leq 2\epsilon$, so we get from (3.10), the inequality above, and (3.9) that

$$f(\bar{x} + u + w) - f(\bar{x} + u + v(u)) \geq \frac{\alpha}{2}\|w - v(u)\|.$$

Thus, (3.4) holds with $\kappa = \alpha/2$. Finally, every point near $\bar{x}$ has a unique decomposition in the fixed orthogonal splitting $\mathbb{R}^n = U \oplus V$, and the last assertion follows because $y - R_{\bar{x}}(y) = w - v(u)$ and $\operatorname{dist}(y, \mathcal{M}) \leq \|y - R_{\bar{x}}(y)\|$. ☐

Instead of the nearest-point projection, Theorem 3.6 uses the map $R_{\bar{x}}$ constructed from a local graph representation of $\mathcal{M}$ in Lemma 2.1, which needs not be a nearest-point projection.

A related abstract linear-growth result is given in [29, Thm. 4.5] under normal embedding of the identifiable set, continuity of the restricted function and its slope, and the existence of an approximate projection. Under Assumption 1, these conditions can be verified using our $R_{\bar{x}}$ constructed using the local graph view. Their result, however, is localized in both distance and objective value, whereas Theorem 3.6 gives a uniform estimate throughout a sufficiently small spatial neighborhood under the direct geometric assumptions of $\mathcal{C}^1$ partial smoothness and nondegeneracy.

4. Transferability of Error Bound. In this section, we present our main result of whether an error bound in the ambient space $\mathbb{R}^n$ is equivalent to the corresponding error bound on an active manifold $\mathcal{M}$ under $\mathcal{C}^1$ partial smoothness. We investigate this question from two complementary perspectives. Throughout this section, we assume that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper and closed.

Our first perspective is a geometric one that deals with a formulation of the error bound based on subdifferential and Riemannian-gradient residuals. These residuals are commonly used to measure first-order stationarity respectively in nonsmooth and manifold optimization. We first give the definition of such error bounds in the ambient space and on the manifold. Our definitions use a gauge function ϕ to describe the relation. We call a function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ a *gauge function* if it is continuous, strictly increasing, and satisfies $\phi(0) = 0$ and $\lim_{t\to\infty} \phi(t) = \infty$.[3] The definitions of local generalized error bounds on $(\mathbb{R}^n, d)$ and $(\mathcal{M}, d)$ are given below.

Definition 4.1 (Generalized Error Bound (GEB) on $(\mathbb{R}^n, d)$). *Let $S \subseteq \mathbb{R}^n$ be a set and let $\bar{x} \in S$. We say that f satisfies the local (subdifferential) generalized error bound (GEB) property around $\bar{x}$ (on $(\mathbb{R}^n, d)$) relative to S if there exist $\epsilon, \nu, \theta > 0$ and a gauge function ϕ such that*

$$\phi(\operatorname{dist}(x, S)) \le \theta \operatorname{dist}(0, \hat{\partial} f(x)), \quad \forall x \in B_\epsilon(\bar{x}) \cap [f \le f(\bar{x}) + \nu]. \tag{4.1}$$

Definition 4.2 (Generalized Error Bound (GEB) on $(\mathcal{M}, d)$). *Let $S \subseteq \mathbb{R}^n$ be a set, $\mathcal{M}$ be a $\mathcal{C}^1$ manifold, and let $\bar{x} \in S \cap \mathcal{M}$. We say that f satisfies the local (subdifferential) generalized error bound (GEB) property around $\bar{x}$ on $(\mathcal{M}, d)$ relative to S if there exist $\epsilon, \nu, \theta > 0$ and a gauge function ϕ such that*

$$\phi(\operatorname{dist}(x, S \cap \mathcal{M})) \le \theta \left\| \nabla_{\mathcal{M}} f(x) \right\|, \quad \forall x \in (B_\epsilon(\bar{x}) \cap [f \le f(\bar{x}) + \nu]) \cap \mathcal{M}. \tag{4.2}$$

The complementary perspective considers a metric approach starting from a formulation of the GEB based on the slopes $|\nabla f|$ and $|\nabla(f|_{\mathcal{M}})|$ respectively for the ambient space and the manifold $\mathcal{M}$.

Definition 4.3 (Generalized Error Bound (GEB) on $(\mathbb{R}^n, d)$). *Let $S \subseteq \mathbb{R}^n$ be a set and let $\bar{x} \in S$. We say that f satisfies the local (slope) generalized error bound (GEB) property around $\bar{x}$ relative to S if there exist $\epsilon, \nu, \theta > 0$ and a gauge function ϕ such that*

$$\phi(\operatorname{dist}(x, S)) \le \theta |\nabla f|(x), \quad \forall x \in B_\epsilon(\bar{x}) \cap [f \le f(\bar{x}) + \nu]. \tag{4.3}$$

Definition 4.4 (Generalized Error Bound (GEB) on $(\mathcal{M}, d)$). *Let $S, \mathcal{M} \subseteq \mathbb{R}^n$ be two sets and let $\bar{x} \in S \cap \mathcal{M}$. We say that f satisfies the local (slope) generalized error bound (GEB) property around $\bar{x}$ on $(\mathcal{M}, d)$ relative to S if there exist $\epsilon, \nu, \theta > 0$ and a gauge function ϕ such that*

$$\phi(\operatorname{dist}(x, S \cap \mathcal{M})) \le \theta |\nabla(f|_{\mathcal{M}})|(x), \quad \forall x \in (B_\epsilon(\bar{x}) \cap [f \le f(\bar{x}) + \nu]) \cap \mathcal{M}. \tag{4.4}$$

More concretely, the set S above is often taken as either $\operatorname{crit} f$ or $\operatorname{argmin} f$ in the literature as well as in our analysis.

Remark 4.5. In the literature, error bounds and related conditions are often stated without our sublevel restriction $[f \le f(\bar{x}) + \nu]$. In other cases, they are usually assumed in a neighborhood of S, or in a region where a stationarity residual is small (as in [37, 42, 33, 43]). A main reason is that we need such restrictions to invoke prox-regularity and identifiability, both of which are stated only at points whose objective value is close to $f(\bar{x})$. When f is (weakly) convex, as is often assumed alongside such error bounds, these properties already hold at all nearby points and the restriction can be dropped.

These two formulations are tightly connected. In particular, on the manifold, Lemma 2.5 gives $|\nabla(f|_{\mathcal{M}})|(x) = \|\nabla_{\mathcal{M}} f(x)\|$. In the ambient space, equivalence between (4.1) and (4.3) for the standard linear error bound ($\phi = \mathrm{Id}$) is proven in [14, Remark 3.6]; the same approximation argument, together with the continuity of ϕ, extends directly to our generalized formulation after possibly shrinking the neighborhood and the sublevel parameter. We thus refer to both as the GEB in most situations, but when we need to distinguish them, we will refer to the first one as the *subdifferential GEB* and the second one as the *slope GEB*.

Despite this connection, the two approaches reveal different mechanisms behind the transfer of error bounds. The geometric approach identifies the common tangential projection of nearby Fréchet subgradients with the Riemannian gradient on $\mathcal{M}$. The metric approach instead relies on identifiability, an off-manifold lower bound for the slope, and the linear-growth result of Theorem 3.6. We develop the two approaches separately to make these mechanisms explicit. The four formulations and the relations we develop in this section are summarized in Figure 1.

Our ambient error bounds are closely related to the *generalized metric subregularity* recently introduced by Li, Mordukhovich, and Zhu [32, Def. 3.1], which states that for $S \subseteq \operatorname{crit} f$, there exist $\epsilon, \theta > 0$ and an admissible function $\hat{\phi} : \mathbb{R}_+ \to \mathbb{R}_+$ (meaning $\hat{\phi}(0) = 0$ and $\hat{\phi}(t) \to 0$ only if $t \to 0$) such that

$$\hat{\phi}(\operatorname{dist}(x, S)) \le \theta \operatorname{dist}(0, \partial f(x)), \quad \forall x \in B_\epsilon(\bar{x}). \tag{4.5}$$

[3] The part of $\phi(t) \to \infty$ is not required in this section as the error bound is a local property. We only need it for simplifying the relation between the error bound and the KL property in Section 5, and this condition can be omitted otherwise.

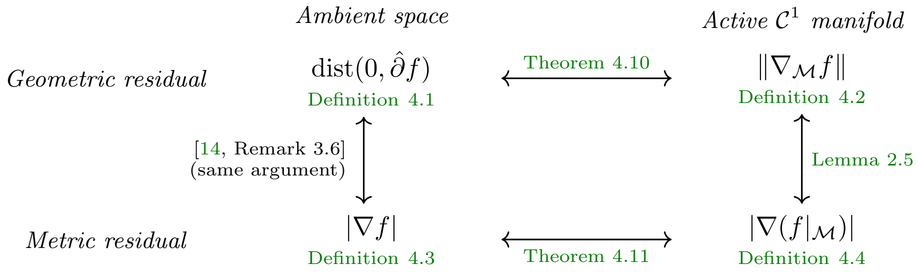

Fig. 1. *Relation between the four formulations of generalized error bounds.*

From Proposition 2.4, we have $\operatorname{dist}(0, \partial f(x)) \leq |\nabla f|(x)$, and hence (4.5) implies both (4.1) and (4.3) with $\nu = \infty$, that is, without the sublevel restriction discussed in Remark 4.5.

In our analysis, we often make the following assumption of proper separation of isocost surfaces, which is commonly assumed in nonconvex settings for analyses involving the Luo–Tseng error bound [37] for the critical set in the literature [42, 33, 43]. We will see in Section 6 that our generalized error bound condition is equivalent to a generalization of the Luo–Tseng error bound in the context of regularized optimization.

Assumption 2. We say that f has the proper separation of isocost surfaces property on S at $\bar{x}$ if there exists $\sigma > 0$ such that

$$f(y) = f(\bar{x}) \qquad \text{whenever } y \in S \text{ and } \|y - \bar{x}\| \leq \sigma. \tag{4.6}$$

Assumption 2 holds automatically when $S = \operatorname{argmin} f$ and also holds true for $S = \operatorname{crit} f$ of a convex function f. This mild condition, or rather its conclusion that f is constant on S near $\bar{x}$, is also reminiscent of the assumption needed for the uniformized KL property in [6, Lemma 6]. A sufficient condition for Assumption 2 is that f has only finitely many critical values and is continuous on $\operatorname{crit} f$ at $\bar{x}$. The former condition is satisfied by the broad class of tame functions (see [23, Thm. 5.2] and [4, Cor. 9]), which is the canonical example on which the Kurdyka–Łojasiewicz property holds.

The following lemma will be useful in both perspectives.

Lemma 4.6. *Consider f satisfying Assumptions 1 and 2 with $\bar{x} \in S \cap \mathcal{M}$ and $S \subseteq \operatorname{crit} f$. Then, for all $x \in \mathcal{M}$ near $\bar{x}$,*

$$\operatorname{dist}(x, S) = \operatorname{dist}(x, S \cap \mathcal{M}). \tag{4.7}$$

Proof. The inequality "$\leq$" is immediate. If it is strict, there exists $y \in S \backslash \mathcal{M}$ with $d(x, y) < \operatorname{dist}(x, S \cap \mathcal{M}) \leq d(x, \bar{x})$, and thus $d(y, \bar{x}) \leq d(y, x) + d(x, \bar{x}) \leq 2d(x, \bar{x})$. Assumption 2 then gives $f(y) = f(\bar{x})$ for x close enough to $\bar{x}$. However, Lemma 3.5 and the lower semicontinuity of f imply that there exists $\eta > 0$ such that $\operatorname{dist}(0, \partial f(y)) > \eta$ for all $y \notin \mathcal{M}$ near $\bar{x}$ with $f(y)$ near $f(\bar{x})$, contradicting with that $y \in \operatorname{crit} f$. ∎

4.1. Main Result: *Geometric* Perspective. We start from the formulation of error bounds defined by the Fréchet subdifferential and the Riemannian gradient. We adapt results from [38, 8] in Lemmas 4.7 to 4.9 below. Although the definition of partial smoothness in [38, 8] is slightly different from Definition 3.1, their proofs are still valid with minor modifications. For completeness, we provide proofs in Section C. For a function f $\mathcal{C}^1$-partly smooth at $\bar{x}$ for 0 relative to $\mathcal{M}$, we have from Definition 3.1 and Proposition 3.2 that

$$\operatorname{par} \hat{\partial} f(x) = N_x \mathcal{M}, \quad (N_x \mathcal{M})^{\perp} = T_x \mathcal{M}$$

for all $x \in \mathcal{M}$ near $\bar{x}$.

Lemma 4.7. *If f is proper, closed, and $\mathcal{C}^1$-partly smooth at $\bar{x}$ for 0 relative to $\mathcal{M}$, then for any x in $\mathcal{M}$ near $\bar{x}$, the projection of a Fréchet subgradient onto $T_x \mathcal{M}$ is independent of its choice. More precisely, $P_{T_x \mathcal{M}}(g) = P_{\operatorname{aff} \hat{\partial} f(x)}(0)$ for every $g \in \hat{\partial} f(x)$.*

Our next lemma extends the results in [38, Lemma 2.11, Thm. 2.12] and [8, Lemma 20] to $\mathcal{C}^1$ partial smoothness.

Lemma 4.8. *If Assumption 1 holds for a function f, then $P_{T_x \mathcal{M}}(g) \in \operatorname{ri} \hat{\partial} f(x)$ for all $x \in \mathcal{M}$ near $\bar{x}$ and every $g \in \hat{\partial} f(x)$.*

Combining Lemmas 4.7 and 4.8 with (2.2) yields the following result.

Lemma 4.9. *Under the setting of Lemma 4.8, for all $x \in \mathcal{M}$ near $\bar{x}$, we have*

$$\nabla_{\mathcal{M}} f(x) = P_{\hat{\partial} f(x)}(0) \qquad \text{and} \qquad \operatorname{dist}(0, \hat{\partial} f(x)) = \|\nabla_{\mathcal{M}} f(x)\|.$$

We are now ready for our main result of the geometric proof of the error bound transferability.

THEOREM 4.10. *Consider a function f with Assumption 1 holds, $S \subseteq \operatorname{crit} f$, and $\bar{x} \in S \cap \mathcal{M}$, and assume that Assumption 2 holds. Then the subdifferential GEB property around $\bar{x}$ relative to S on $(\mathbb{R}^n, d)$ is equivalent to the subdifferential GEB property around $\bar{x}$ relative to S on $(\mathcal{M}, d)$ for the same gauge function ϕ.*

Proof. For $x \in \mathcal{M}$ near $\bar{x}$, by Lemma 4.9, we have

$$\operatorname{dist}(0, \hat{\partial} f(x)) = \|\nabla_{\mathcal{M}} f(x)\|. \tag{4.8}$$

Moreover, by Lemma 3.5 and lower semicontinuity of f, there exist $\epsilon_0, \nu_0, \eta > 0$ such that

$$\operatorname{dist}(0, \hat{\partial} f(x)) \geq \operatorname{dist}(0, \partial f(x)) > \eta, \quad \forall x \in (B_{\epsilon_0}(\bar{x}) \cap [f \leq f(\bar{x}) + \nu_0]) \setminus \mathcal{M}. \tag{4.9}$$

We now prove the two implications. Suppose first that the ambient GEB holds with parameters $(\epsilon, \nu, \theta, \phi)$. Shrink ϵ so that $\epsilon \leq \epsilon_0$ and $B_\epsilon(\bar{x}) \cap \mathcal{M} \subseteq [f \leq f(\bar{x}) + \nu]$. Then, for every $x \in B_\epsilon(\bar{x}) \cap \mathcal{M}$,

$$\begin{aligned}\phi(\operatorname{dist}(x, S \cap \mathcal{M})) &= \phi(\operatorname{dist}(x, S)) \\ &\leq \theta \operatorname{dist}(0, \hat{\partial} f(x)) = \theta \|\nabla_{\mathcal{M}} f(x)\|,\end{aligned}$$

where the first equality follows from (4.7) and the last (4.8). This is exactly the GEB on $(\mathcal{M}, d)$.

Conversely, suppose now that the GEB on $(\mathcal{M}, d)$ holds with parameters $(\epsilon, \nu, \theta, \phi)$. We shrink ϵ small enough so that $\epsilon \leq \epsilon_0$ and (by the continuity of ϕ)

$$\phi(\|x - \bar{x}\|) \leq \theta\eta, \quad \forall x \in B_\epsilon(\bar{x}).$$

Set $\hat{\nu} := \min\{\nu, \nu_0\}$ and consider any $x \in B_\epsilon(\bar{x}) \cap [f \leq f(\bar{x}) + \hat{\nu}]$. If $x \in \mathcal{M}$, then (4.7) and the manifold GEB give

$$\phi(\operatorname{dist}(x, S)) = \phi(\operatorname{dist}(x, S \cap \mathcal{M})) \leq \theta \|\nabla_{\mathcal{M}} f(x)\| = \theta \operatorname{dist}(0, \hat{\partial} f(x)).$$

When $x \notin \mathcal{M}$, then since $\operatorname{dist}(x, S) \leq \|x - \bar{x}\|$, we have from (4.9) that

$$\phi(\operatorname{dist}(x, S)) \leq \phi(\|x - \bar{x}\|) \leq \theta\eta < \theta \operatorname{dist}(0, \hat{\partial} f(x)).$$

Combining the two cases proves the ambient GEB with parameters $(\epsilon, \hat{\nu}, \theta, \phi)$. ◻

4.2. Main Result: *Metric* Perspective. We now turn to the metric formulation of error bounds formulated via slope. In Theorem 4.11 below, we establish the equivalence between the slope GEB property on $(\mathbb{R}^n, d)$ and its restriction to the active manifold $(\mathcal{M}, d)$ under $\mathcal{C}^1$-partial smoothness and nondegeneracy (3.2).

THEOREM 4.11. *Consider a function f with Assumption 1 holds, $S \subseteq \operatorname{crit} f$, and $\bar{x} \in S \cap \mathcal{M}$, and assume that Assumption 2 holds. Then the slope GEB property around $\bar{x}$ relative to S on $(\mathbb{R}^n, d)$ is equivalent to the slope GEB property around $\bar{x}$ relative to S on $(\mathcal{M}, d)$ for the same gauge function ϕ.*

Proof. By Lemma 3.5 and Remark 3.4, there exists $\eta > 0$ such that

$$\operatorname{dist}(0, \partial f(x)) > \eta \text{ and } |\nabla f|(x) > \eta \tag{4.10}$$

for all $x \notin \mathcal{M}$ near $\bar{x}$ with $f(x)$ near $f(\bar{x})$. We now prove the two directions respectively.

Case 1: GEB on $(\mathcal{M}, d)$ implies GEB on $(\mathbb{R}^n, d)$.
We first assume that GEB on $(\mathcal{M}, d)$ holds at $\bar{x}$ with constants $\epsilon, \nu, \theta > 0$ and function ϕ. Fix any $x \in B_\epsilon(\bar{x}) \cap [f \leq f(\bar{x}) + \nu]$. If $x \in \mathcal{M}$, by the definition of GEB on $(\mathcal{M}, d)$ and the definition of the respective slopes, we have

$$\phi(\operatorname{dist}(x, S)) \leq \phi(\operatorname{dist}(x, S \cap \mathcal{M})) \leq \theta |\nabla (f|_{\mathcal{M}})|(x) \leq \theta |\nabla f|(x).$$

On the other hand, if $x \notin \mathcal{M}$, (4.10) gives $|\nabla f|(x) > \eta$. After shrinking ϵ if necessary, we have $\phi(d(x, \bar{x})) \leq \theta\eta$, and thus

$$\phi(\operatorname{dist}(x, S)) \leq \phi(d(x, \bar{x})) \leq \theta\eta < \theta |\nabla f|(x).$$

Case 2: GEB on $(\mathbb{R}^n, d)$ implies GEB on $(\mathcal{M}, d)$.
Now suppose that GEB on $(\mathbb{R}^n, d)$ holds with constants $\epsilon, \nu, \theta > 0$ and function ϕ. By shrinking ϵ if necessary, we have $\operatorname{dist}(x, S) = \operatorname{dist}(x, S \cap \mathcal{M})$ for all $x \in \mathcal{M} \cap B_\epsilon(\bar{x})$ from (4.7), and $B_\epsilon(\bar{x}) \cap \mathcal{M} \subseteq [f \leq f(\bar{x}) + \nu]$ and $\phi(\operatorname{dist}(x, S)) \leq \theta$ for all $x \in B_\epsilon(\bar{x})$ from that $f|_{\mathcal{M}}$ is continuous around $\bar{x}$.

If $\operatorname{dist}(x, S) = 0$, the desired GEB inequality on $(\mathcal{M}, d)$ is trivially true by (4.7). Thus, we assume $\operatorname{dist}(x, S) > 0$ onward. By GEB on $(\mathbb{R}^n, d)$, $|\nabla f|(x) \geq \theta^{-1} \phi(\operatorname{dist}(x, S)) > 0$. Hence, the definition of slope indicates that there exists a sequence $x_r \to x$ with $x_r \neq x$ such that

$$\frac{1}{2} \phi(\operatorname{dist}(x, S)) \leq \theta \frac{f(x) - f(x_r)}{d(x, x_r)} \iff \frac{1}{2} \phi(\operatorname{dist}(x, S)) d(x, x_r) \leq \theta (f(x) - f(x_r)). \tag{4.11}$$

From Theorem 3.6, there is $\kappa > 0$ such that, for all sufficiently large r, with $v_r := R_{\bar{x}}(x_r) \in \mathcal{M}$

$$f(x_r) \geq f(v_r) + \kappa d(x_r, v_r). \tag{4.12}$$

Substituting (4.12) into (4.11) gives

$$\theta(f(x) - f(v_r)) \geq \frac{1}{2}\phi(\mathrm{dist}(x, S))d(x, x_r) + \theta\kappa d(x_r, v_r).$$

As $\phi(\mathrm{dist}(x, S)) \leq \theta$, we further obtain

$$\theta(f(x) - f(v_r)) \geq \left(\frac{1}{2}d(x, x_r) + \kappa d(x_r, v_r)\right)\phi(\mathrm{dist}(x, S)).$$

Since $x_r \neq x$ and $\mathrm{dist}(x, S) > 0$, the inequality above implies that $f(v_r) < f(x)$, and hence $v_r \neq x$ for all r, so division by $d(x, v_r)$ is always valid. Let $\gamma := \min\{1/2, \kappa\}$, then the triangle inequality gives

$$\theta\frac{f(x) - f(v_r)}{d(x, v_r)} \geq \gamma\phi(\mathrm{dist}(x, S)).$$

The local graph representation in Lemma 2.1 shows that near $\bar{x}$, $R_{\bar{x}}$ is continuous and $R_{\bar{x}} = \mathrm{Id}$ on $\mathcal{M}$, so

$$v_r = R_{\bar{x}}(x_r) \to R_{\bar{x}}(x) = x.$$

Therefore, we get

$$\phi(\mathrm{dist}(x, S)) \leq \gamma^{-1}\theta|\nabla(f|_{\mathcal{M}})|(x).$$

Since $\mathrm{dist}(x, S) = \mathrm{dist}(x, S \cap \mathcal{M})$, we conclude that GEB indeed holds on $(\mathcal{M}, d)$ around $\bar{x}$ with $\gamma^{-1}\theta$. ∎

Remark 4.12. The proof of Theorem 4.11 in the manifold-to-ambient direction does not utilize any property beyond identifiability. It is hence applicable to any identifiable set in a complete metric space. On the contrary, the geometric approach in Theorem 4.10 uses Riemannian gradient and thus does not extend beyond $\mathcal{C}^1$ active manifolds.

When $S = \operatorname{argmin} f$, Assumption 2 holds automatically, and we get the following corollary.

COROLLARY 4.13. *Consider the case that $S = \operatorname{argmin} f$ and let $\bar{x} \in S \cap \mathcal{M}$. Suppose that Assumption 1 holds. Then the GEB property around $\bar{x}$ relative to S on $(\mathbb{R}^n, d)$ is equivalent to the GEB property around $\bar{x}$ relative to S on $(\mathcal{M}, d)$, i.e., Definition 4.1 ⇔ Definition 4.2 and Definition 4.3 ⇔ Definition 4.4.*

5. The KL Property. We now turn to the Kurdyka–Łojasiewicz (KL) property. We first recall the slope formulation of the KL property from [5, 29]. A function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ is called a *desingularizer* if it is continuous on $\mathbb{R}_+$, is continuously differentiable with $\psi'(t) > 0$ on $(0, \infty)$, and satisfies $\psi(0) = 0$ and $\psi(t) \to \infty$ as $t \to \infty$.

DEFINITION 5.1 (KL property on a metric space [29, Def. 5.8]). *Let $(Y, \bar{d})$ be a metric space. We say that a closed function $g : Y \to \overline{\mathbb{R}}$ has the* KL property *at $\bar{y} \in \operatorname{dom} g$* with desingularizer *$\psi$ if there exist constants $\epsilon, \delta > 0$, such that*

$$|\nabla(\psi(g(\cdot) - g(\bar{y})))|(y) \geq \delta, \quad \forall y \in B_\epsilon(\bar{y}) \cap [g(\bar{y}) < g < g(\bar{y}) + \epsilon], \tag{5.1}$$

where both the slope and the ball radius are computed using the metric $\bar{d}$ on Y.

As noted in [29, Sec. 5], we also have the following equivalent form for (5.1).

$$\psi'(g(y) - g(\bar{y}))\,|\nabla g|(y) \geq \delta, \quad \forall y \in B_\epsilon(\bar{y}) \cap [g(\bar{y}) < g < g(\bar{y}) + \epsilon]. \tag{5.2}$$

In [33, Thm. 3.7], Li and Pong showed that in Euclidean spaces, if the KL property holds on the active manifold at a point for which f is partly smooth, it also holds in the ambient space with the same exponent $\alpha \in [0, 1)$. Lewis and Tian studied in [29, Thm. 5.12] the equivalence between the KL property in a complete metric space and the KL property with the same desingularizer on an identifiable set (not necessarily a manifold) using a linear-growth property established in [29, Thm. 4.2]. Their linear-growth property states that for a closed function f on a complete metric space $(Y, \bar{d})$, a point $\bar{x}$ with $|\nabla f|(\bar{x}) = 0$, and an identifiable set $\mathcal{M} \ni \bar{x}$, for any $\epsilon > 0$ smaller than the modulus of identifiability and any sequence $x_r \to_f \bar{x}$, there is a sequence $v_r \to_f \bar{x}$ in $\mathcal{M}$ such that

$$f(x_r) \geq f(v_r) + \epsilon\bar{d}(x_r, v_r), \quad \forall r \text{ large}. \tag{5.3}$$

Compared with Theorem 3.6, their estimate has more flexibility in the linear growth parameter and applies in a more general space to a more general identifiable set. It is, however, pointwise: it holds only at the reference point $\bar{x}$ rather than uniformly at nearby points of the identifiable set. Such a pointwise estimate is insufficient for transferring the ambient slope inequality to $f|_{\mathcal{M}}$ at an arbitrary nearby point $x \in \mathcal{M}$, and we therefore develop Theorem 3.6. In particular, in (4.11) and (4.12), we considered a point x near $\bar{x}$ but with a positive slope, and thus (5.3) does not apply. The proof of [29, Thm. 5.12] encounters precisely the same issue in the ambient-to-restricted direction. Indeed, as our counterexample in Section A shows, this implication does not hold in such generality (while the manifold-to-ambient direction remains valid), and additional regularity of $\mathcal{M}$ and $f|_{\mathcal{M}}$ is needed. Utilizing the results developed in this paper, we consider two ways to recover the ambient-to-manifold direction of [29, Thm. 5.12] under $\mathcal{C}^1$ partial smoothness in Euclidean space.

The first approach utilizes the developed equivalence of the GEB property in the ambient space and on the $\mathcal{C}^1$ active manifold and shows that GEB implies the KL property on both the ambient space and the active manifold. We in particular deduce the relation between the function ϕ in GEB and the desingularizer ψ in KL. The implication of GEB to KL does not rely on partial smoothness, and is therefore of independent interest. In our second approach, we use Theorem 3.6 to directly show that if f is $\mathcal{C}^1$-partly smooth at $\bar{x}$ for 0 relative to $\mathcal{M}$, then the KL property around $\bar{x}$ on $(\mathbb{R}^n, d)$ is equivalent to the KL property around $\bar{x}$ on $(\mathcal{M}, d)$.

5.1. GEB Implies KL. For the first approach, the following proposition shows that the GEB property for $\operatorname{crit} f$ together with the separation of isocost surfaces property and prox-regularity implies the KL property. For this proposition, we need to define prox-regularity for the restriction of a function to a manifold on which it is differentiable. We say $f|_{\mathcal{M}}$ is prox-regular at $\bar{x}$ if there are $\epsilon' > 0$ and $\rho \geq 0$ such that

$$f|_{\mathcal{M}}(y) \geq f|_{\mathcal{M}}(x) + \langle \nabla_{\mathcal{M}} f(x), y - x\rangle - \frac{\rho}{2}\|y - x\|^2, \quad \forall x, y \in B_{\epsilon'}(\bar{x}) \cap \mathcal{M}. \tag{5.4}$$

Proposition 5.2. *Consider a closed function $f : \mathbb{R}^n \to (-\infty, \infty]$ and a point $\bar{x} \in S \subseteq \operatorname{crit} f$.*

(a) Assume there are $\epsilon, \nu, \theta > 0$ and a gauge function ϕ such that the GEB property (4.3) holds around $\bar{x}$ on $(\mathbb{R}^n, d)$, Assumption 2 holds, and f is prox-regular at $\bar{x}$ for 0 with constant $\rho \geq 0$. Let

$$\Phi(t) := t\,\phi^{-1}(\theta t) + \frac{\rho}{2}\phi^{-1}(\theta t)^2, \quad \psi(t) := \int_0^t \frac{ds}{\Phi^{-1}(s)}. \tag{5.5}$$

If the improper integral above is finite for some $t > 0$, then f has the KL property around $\bar{x}$ with desingularizer ψ.

(b) Consider a $\mathcal{C}^1$-embedded submanifold $\mathcal{M} \subseteq \mathbb{R}^n$ around $\bar{x} \in \mathcal{M}$ such that $f|_{\mathcal{M}}$ is $\mathcal{C}^1$-smooth around $\bar{x}$. Assume there are $\epsilon, \nu, \theta > 0$ and a gauge function ϕ such that the GEB property (4.4) holds around $\bar{x}$ on $(\mathcal{M}, d)$, Assumption 2 holds, and $f|_{\mathcal{M}}$ is prox-regular at $\bar{x}$ with constant $\rho \geq 0$. Define Φ and ψ by (5.5) with the present ϕ, θ, ρ. If the improper integral defining ψ is finite for some $t > 0$, then $f|_{\mathcal{M}}$ has the KL property around $\bar{x}$ on $(\mathcal{M}, d)$ with desingularizer ψ.

Since ϕ is continuous, strictly increasing, and strictly positive on $(0, \infty)$, it is straightforward to verify that Φ is well defined and satisfies these properties of ϕ.

Proof. **(a)**. We first decrease ϵ, if necessary, so that $\epsilon \leq \nu$ in (4.3), $2\epsilon \leq \sigma$ for the σ in (4.6), and the prox-regularity inequality holds on $B_{2\epsilon}(\bar{x})$ with subgradients in $B_{2\epsilon}(0)$ with constant ρ. We now fix this ϵ and choose

$$\eta := \frac{1}{2}\min\{\epsilon, \Phi(\epsilon)\} > 0.$$

We will prove the KL property on

$$B_\eta(\bar{x}) \cap [f(\bar{x}) < f < f(\bar{x}) + \eta],$$

while retaining ϵ as the threshold separating the two slope cases.

Fix x in this set. If $|\nabla f|(x) = \infty$, the KL property is immediate, so we assume $|\nabla f|(x) < \infty$.

Suppose first that $|\nabla f|(x) < \epsilon$. Since $\partial f(x)$ is nonempty and closed, we can choose $v \in \operatorname{argmin}_{s \in \partial f(x)} \|s\|$. Then by Proposition 2.4, we have $\|v\| \leq |\nabla f|(x) < \epsilon$, so this subgradient lies in the range required by prox-regularity.

Consider a sequence $y_r \in S$ such that $d(x, y_r) \to d(x, S)$. Since $\bar{x} \in S$,

$$d(x, S) \leq d(x, \bar{x}) < \eta \leq \epsilon/2.$$

After discarding finitely many terms, we may assume $d(x, y_r) < \epsilon$ for all r. Consequently, $d(\bar{x}, y_r) \leq d(\bar{x}, x) + d(x, y_r) < \eta + \epsilon < 2\epsilon \leq \sigma$, and thus $f(y_r) = f(\bar{x})$ for all r by (4.6). Moreover, $x, y_r \in B_{2\epsilon}(\bar{x})$ and $f(x) < f(\bar{x}) + \eta < f(\bar{x}) + \epsilon$. The prox-regularity inequality thus gives

$$\begin{aligned} f(x) - f(\bar{x}) &= f(x) - f(y_r) \\ &\leq -\langle v, y_r - x\rangle + \frac{\rho}{2} d(x, y_r)^2 \\ &\leq \|v\| d(x, y_r) + \frac{\rho}{2} d(x, y_r)^2. \end{aligned} \tag{5.6}$$

Using Proposition 2.4 again and taking the limit as $r \to \infty$, we obtain

$$f(x) - f(\bar{x}) \le |\nabla f|(x) d(x, S) + \frac{\rho}{2} d(x, S)^2. \tag{5.7}$$

Since $\eta \le \epsilon \le \nu$, the error bound (4.3) applies at x and yields

$$d(x, S) \le \phi^{-1}(\theta |\nabla f|(x)).$$

Substituting this into (5.7) gives

$$f(x) - f(\bar{x}) \le |\nabla f|(x)\, \phi^{-1}(\theta |\nabla f|(x)) + \frac{\rho}{2} \phi^{-1}(\theta |\nabla f|(x))^2 = \Phi(|\nabla f|(x)). \tag{5.8}$$

Next, suppose that $|\nabla f|(x) \ge \epsilon$. Our choice of η gives

$$f(x) - f(\bar{x}) < \eta < \Phi(\epsilon) \le \Phi(|\nabla f|(x)). \tag{5.9}$$

Combining (5.8) and (5.9) and applying the increasing map Φ^{-1}, we conclude that

$$\Phi^{-1}(f(x) - f(\bar{x})) \le |\nabla f|(x), \quad \forall x \in B_\eta(\bar{x}) \cap [f(\bar{x}) < f < f(\bar{x}) + \eta].$$

Therefore,

$$\psi'(f(x) - f(\bar{x}))\, |\nabla f|(x) = \frac{|\nabla f|(x)}{\Phi^{-1}(f(x) - f(\bar{x}))} \ge 1.$$

This proves (5.2) with neighborhood parameter η and $\delta = 1$.

It remains to verify that ψ is a desingularizer. The assumptions on ϕ clearly imply that Φ^{-1} is well defined, continuous and strictly increasing on $\mathbb{R}_+$, and is positive on $(0, \infty)$. Finiteness of the improper integral defining ψ at some positive endpoint implies its finiteness at every finite endpoint and $\lim_{t \downarrow 0} \psi(t) = 0$. Consequently, ψ is continuous at 0 with $\psi(0) = 0$, and $\psi'(t) = 1/\Phi^{-1}(t) > 0$ for all $t > 0$. Finally, since $\phi^{-1}(\theta t) \to \infty$, there exists $T > 0$ such that $\phi^{-1}(\theta t) \ge 1$ and thus $\Phi(t) \ge t$ for $t \ge T$, or equivalently $\Phi^{-1}(s) \le s$ for $s \ge \Phi(T)$. It follows that, for $t \ge \Phi(T)$,

$$\psi(t) \ge \psi(\Phi(T)) + \int_{\Phi(T)}^{t} \frac{ds}{s} \overset{t \to \infty}{\longrightarrow} \infty.$$

Thus ψ satisfies all the requirements of a desingularizer.

(b). The proof is similar to part (**a**), and since the Riemannian gradient is continuous, we do not need to consider the case that the slope is larger than the prox-regularity threshold. ∎

The condition (5.4) might seem unnatural at first glance. However, Lemma 4.9 shows that it holds automatically under Assumption 1 (prox-regularity is implied by partial smoothness).

Proposition 5.2 recovers the widely seen power-type KL property with desingularizers $\psi(t) = ct^{1-\alpha}$. For a gauge $\phi(t) = t^q$, $\rho > 0$ with $q \in (0, 2)$ yields the KL exponent $\alpha = \max\{q/(q+1), q/2\}$, covering $\alpha \in (0, 1)$ while $\rho = 0$ gives $\alpha = q/(q+1)$ for all $q > 0$. By using regular-variation techniques for inverse-integral constructions in [36], we can also obtain explicit asymptotic forms of desingularizers beyond the KL exponent type. For example, when $\rho > 0$ and $\phi(t) = t^2(\log(1/t))^\beta$ near 0 with $\beta > 1$, the construction in (5.5) yields a desingularizer of order $(\log(1/t))^{1-\beta}$ as $t \downarrow 0$.

Due to the term involving $\phi^{-1}(\theta t)^2$ (arising from prox-regularity) in Φ, the conversion from ϕ to ψ in Proposition 5.2 may lose some sharpness, and we can often get faster convergence rate guarantees from analyzing through the GEB property than through the KL property. This is consistent with the observation in [32, Sec. 5].

With Proposition 5.2, Theorem 4.11 implies that if f is $\mathcal{C}^1$-partly smooth at $\bar{x}$ for 0 relative to $\mathcal{M}$ and $f|_\mathcal{M}$ satisfies (5.4), then under (3.2) and Assumption 2, GEB for either f or $f|_\mathcal{M}$ on $S \subseteq \operatorname{crit} f$ around $\bar{x}$ implies the KL property for both f and $f|_\mathcal{M}$ around $\bar{x}$ with the desingularizer ψ defined in (5.5), provided the improper integral defining ψ is finite.

Given the result of Proposition 5.2, one might wonder if the reverse implication holds, i.e., whether the KL property implies the generalized error bound property. For this direction, [14, Thm. 3.7] shows that if the KL property (with $\psi(t) = t^\alpha$ for $\alpha \in (0, 1)$) holds on $U \cap [f^* < f < r]$ for some $f^* < r$ and some open set U, then it implies a generalized error bound condition in the form of (4.3), but relative to the set $S = [f \le f^*]$, which may not be a subset of $\operatorname{crit} f$. Rebjock and Boumal [39] established equivalence for $\mathcal{C}^2$ functions near local minima, with KL exponent 1/2. Except for these special cases, the reverse implication from KL to GEB is generally not true. The reason is clear: the KL property controls the behavior of only x with $f(x) > f(\bar{x})$, while the (generalized) error bound also includes those x with $f(x) < f(\bar{x})$ beyond the region governed by the KL property.

5.2. Ambient-to-Manifold for KL Under $\mathcal{C}^1$-Partial Smoothness. Our second approach takes a direct fix for the ambient-to-manifold direction of the KL equivalence of Lewis and Tian [29, Thm. 5.12] in the Euclidean setting under $\mathcal{C}^1$-partial smoothness and nondegeneracy through Theorem 3.6. For completeness and for contrasting the two directions, we also include in the theorem statement the correct manifold-to-ambient part from [29, Thm. 5.12].

THEOREM 5.3. *Let ψ be a desingularizer satisfying*

$$\liminf_{t\downarrow 0} \psi'(t) > 0. \tag{5.10}$$

(a) Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ satisfy Assumption 1. If f satisfies the KL property at $\bar{x}$ with desingularizer ψ on $(\mathbb{R}^n, d)$, then $f|_{\mathcal{M}}$ satisfies the KL property at $\bar{x}$ with the same desingularizer on $(\mathcal{M}, d)$.
(b) On a complete metric space $(Y, \bar{d})$, consider a closed function f that has slope zero at a point $\bar{x}$ and let $\mathcal{M}$ be an identifiable set at $\bar{x}$.[4] If $f|_{\mathcal{M}}$ has the KL property at $\bar{x}$ with desingularizer ψ on $(\mathcal{M}, \bar{d})$, then f also has the KL property at $\bar{x}$ with the same desingularizer on $(Y, \bar{d})$.

Proof. We only prove (a), as the proof in [29, Thm. 5.12] for (b) is valid. Without loss of generality, we may assume $f(\bar{x}) = 0$. Thus (5.1) becomes

$$|\nabla(\psi \circ f)|(x) \geq \delta, \quad \forall x \in B_\epsilon(\bar{x}) \cap [0 < f < \epsilon].$$

After shrinking ϵ and δ if necessary, (5.10) implies

$$\psi'(t) > \delta, \quad \forall t \in (0, \epsilon). \tag{5.11}$$

Theorem 3.6 gives a constant $\kappa > 0$, a neighborhood W of $\bar{x}$, and a continuous map $R_{\bar{x}} : W \to \mathcal{M}$ such that

$$f(z) \geq f\big(R_{\bar{x}}(z)\big) + \kappa\|z - R_{\bar{x}}(z)\|, \quad \forall z \in W. \tag{5.12}$$

Moreover, $R_{\bar{x}}(z) = z$ for all $z \in W \cap \mathcal{M}$. Therefore, through shrinking ϵ if necessary so that $B_\epsilon(\bar{x}) \subset W$, we have $R_{\bar{x}}(x) = x$ for any $x \in \mathcal{M} \cap B_\epsilon(\bar{x}) \cap [0 < f(x) < \epsilon]$.

Now we pick an arbitrary $x \in \mathcal{M} \cap B_\epsilon(\bar{x}) \cap [0 < f(x) < \epsilon]$. By the definition of slope, there exists a sequence $x_r \to x$ with $x_r \neq x$ such that

$$\psi(f(x)) - \psi(f(x_r)) > \frac{1}{2}\delta d(x, x_r) > 0, \quad \forall r \geq 0. \tag{5.13}$$

We first show that $0 < f(x_r) < \epsilon$ for all sufficiently large r. Since ψ is strictly increasing, we have $f(x_r) < f(x) < \epsilon$ for all r. By (5.12), let $v_r = R_{\bar{x}}(x_r)$, then $v_r \in \mathcal{M}$ and $v_r \to R_{\bar{x}}(x) = x$. Since $f|_{\mathcal{M}}$ is continuous from partial smoothness, we have $f(v_r) \to f(x)$. Hence, $0 < f(v_r) < \epsilon$ for all sufficiently large r. By (5.12), we have $0 < f(v_r) \leq f(x_r) < \epsilon$ for all sufficiently large r. By (5.13) and the mean value theorem, there exists $\xi_r \in [f(v_r), f(x_r)]$ such that

$$\begin{aligned}
\psi(f(x)) - \psi(f(v_r)) &> \frac{1}{2}\delta d(x, x_r) + \psi(f(x_r)) - \psi(f(v_r)) \\
&= \frac{1}{2}\delta d(x, x_r) + \psi'(\xi_r)\big(f(x_r) - f(v_r)\big).
\end{aligned}$$

As $\xi_r \in [f(v_r), f(x_r)] \subset (0, \epsilon)$ for all sufficiently large r, (5.11) holds. By inserting (5.11) and (5.12) into the inequality above and defining $\gamma := \delta \min\{1/2, \kappa\}$, we obtain

$$\begin{aligned}
\psi(f(x)) - \psi(f(v_r)) &> \frac{1}{2}\delta d(x, x_r) + \psi'(\xi_r)\kappa d(x_r, v_r) \\
&\geq \frac{1}{2}\delta d(x, x_r) + \delta\kappa d(x_r, v_r) \\
&\geq \gamma d(x, v_r).
\end{aligned} \tag{5.14}$$

From the first inequality in (5.14) and that $x_r \neq x$, we get $\psi(f(x)) - \psi(f(v_r)) > 0$, and thus $v_r \neq x$, for all sufficiently large r. Combining this with $v_r \in \mathcal{M}$ and $v_r \to x$, we conclude that

$$|\nabla(\psi \circ f|_{\mathcal{M}})|(x) \geq \gamma. \qquad \square$$

[4] The definition of identifiability in a complete metric space $(Y, \bar{d})$ replaces the distance $\operatorname{dist}(0, \partial f(x))$ with the slope $|\nabla f|(x)$ and the Euclidean distance with $\bar{d}$.

6. Discussion and Application. We now illustrate the application of our main results on error bounds with the two motivating examples discussed in Section 1. We first treat the ℓ_p-norm ball constrained problem in (1.1) whose active manifold is not $\mathcal{C}^2$, and then show that for the ℓ_1-regularized case, our result recovers the proximal error bound equivalence in [46] as a special case.

6.1. An ℓ_p-ball Constrained Example. Consider (1.1) with a differentiable g, and let $\bar{x}$ be the nondegenerate critical point of interest. Whenever $\bar{x}$ is sparse, the active manifold $\mathcal{M} = \{x : \|x\|_p = r\}$ is only $\mathcal{C}^{1,p-1}$ near $\bar{x}$ and not $\mathcal{C}^2$, so our Theorem 3.6 is indeed needed, and Theorems 4.10 and 4.11 provide the GEB equivalence in $\mathbb{R}^n$ and on $\mathcal{M}$. We consider the following simple example for demonstration.

Take $n = 2$, $p \in (1, 2)$, and $g(x) = -x_1$ in (1.1), and write $C := \{x : \|x\|_p \le r\}$ for its feasible set. We can then equivalently write this problem as $\min f(x)$ with $f = g + \delta_C$, where δ_C is the indicator function of C. Clearly, for this convex problem, $S = \operatorname{crit} f = \operatorname{argmin} f = \{\bar{x}\}$ with $\bar{x} := (r, 0)$. Moreover, Assumption 1 holds at $\bar{x}$. Indeed, f is convex and thus prox-regular and subdifferentially regular everywhere, and the restriction $f|_{\mathcal{M}} = g|_{\mathcal{M}}$ is $\mathcal{C}^1$. Writing $F(x) := |x_1|^p + |x_2|^p - r^p$, we have $\partial f(x) = \nabla g(x) + \mathbb{R}_+ \nabla F(x)$ for $x \in \mathcal{M}$. As ∇g and ∇F are continuous, this identity gives both inner semicontinuity along $\mathcal{M}$ and the sharpness condition $\operatorname{par} \hat{\partial} f(\bar{x}) = \operatorname{span} \nabla F(\bar{x}) = N_{\bar{x}}\mathcal{M}$. Finally, $\nabla g(\bar{x}) = -e_1$ and $\nabla F(\bar{x}) = pr^{p-1}e_1$ give $\hat{\partial} f(\bar{x}) = \{(t-1)e_1 : t \ge 0\}$, whose relative interior contains 0. We now compute both sides of the GEB on $(\mathcal{M}, d)$ for $x \in \mathcal{M}$ near $\bar{x}$, writing $x_1 = (r^p - |x_2|^p)^{1/p} > 0$. By Proposition 2.2, $\nabla_{\mathcal{M}} f(x) = P_{T_x\mathcal{M}}(\nabla g(x))$, and $T_x\mathcal{M} = (\nabla F(x))^\perp$ with $\nabla F(x) = p(a, b)$ for $a := x_1^{p-1}$ and $b := \operatorname{sgn}(x_2)|x_2|^{p-1}$. Since $\nabla g(x) = -e_1$, projecting gives $\nabla_{\mathcal{M}} f(x) = (-b^2, ab)/(a^2+b^2)$ and hence $\|\nabla_{\mathcal{M}} f(x)\| = |b|/\sqrt{a^2 + b^2} = |x_2|^{p-1}/(x_1^{2(p-1)} + |x_2|^{2(p-1)})^{1/2} = \Theta(|x_2|^{p-1})$ near $\bar{x}$. On the other side, $x_1 - r = -|x_2|^p/(pr^{p-1}) + O(|x_2|^{2p})$ is $o(|x_2|)$ as $p > 1$, so the displacement along x_1 is of higher order and $\operatorname{dist}(x, S) = \sqrt{(x_1 - r)^2 + x_2^2} = \Theta(|x_2|)$. This computation establishes a GEB on $(\mathcal{M}, d)$ around $\bar{x}$ with the gauge $\phi(t) = t^{p-1}$, and Theorem 4.10 or Theorem 4.11 transfers it to the GEB on $(\mathbb{R}^2, d)$ with the same gauge.

6.2. Recovering the ℓ_1 Case. We now assume that the objective function f has a regularized structure of $f = g + h$, with the following structural assumptions.

Assumption 3. ∇g is Lipschitz continuous with constant $L \ge 0$ and h is proper, closed and *ρ-weakly convex* for some $\rho \ge 0$ (meaning that $h + \rho\|\cdot\|^2/2$ is convex).

Some useful weakly convex regularizers include folded-concave penalties such as SCAD and MCP [35]. For such a regularized function, f is subdifferentially regular, hence the slope GEB and the subdifferential GEB coincide by Proposition 2.4. We refer to both as the *subdifferential GEB* in this section. The following definition is an alternative form of error bound (proximal EB) commonly used in the optimization-algorithms literature, and with the linear gauge, it is widely known as the Luo–Tseng error bound [37].

Definition 6.1. *Suppose that $f = g + h$ satisfies Assumption 3 and let $S \subset \operatorname{crit} f$. We say that f satisfies the proximal generalized error bound around $\bar{x} \in S$ on $(\mathbb{R}^n, d)$ if there exist constants $\hat{\epsilon}, \hat{\theta}, \gamma > 0$ and a gauge function $\hat{\phi}$ such that*

$$\hat{\phi}(\operatorname{dist}(x, S)) \le \hat{\theta}\|x - \operatorname{prox}_{\gamma h}(x - \gamma\nabla g(x))\|, \quad \forall x \in B_{\hat{\epsilon}}(\bar{x}),$$

where

$$\operatorname{prox}_{\gamma h}(x) := \operatorname{argmin}_{y \in \mathbb{R}^n} \ h(y) + \frac{1}{2\gamma}\|y - x\|^2. \tag{6.1}$$

Unlike Definitions 4.1 and 4.2, no sublevel restriction appears above. From Assumption 3, f is weakly convex, and Remark 4.5 shows that the restriction is unnecessary in this setting. We may therefore also drop it from Definitions 4.1 and 4.2 here, so that both error bounds on the same region.

We now show that, under Assumption 3, the subdifferential GEB and the proximal GEB are equivalent. A similar result for convex g and h with $\phi(t) = t$ was shown in [17, Thm.s 3.4 and 3.5].

Lemma 6.2 (Subdifferential EB and proximal EB). *Let Assumption 3 hold, $S \subseteq \operatorname{crit} f$ with $\bar{x} \in S$, and $\gamma\rho < 1$. Then proximal GEB (Definition 6.1) with gauge ϕ implies subdifferential GEB (Definition 4.1) with the same gauge. Conversely, subdifferential GEB with gauge ϕ and constant θ implies proximal GEB with*

$$\hat{\phi}(t) = \min\{t, \ \phi(t/2)\}, \qquad \hat{\theta} = \max\{2, c\}, \qquad c := \theta(1 + \gamma L)/\gamma. \tag{6.2}$$

Proof. We first show that $\operatorname{prox}_{\gamma h}$ is nearly nonexpansive. That is,

$$\frac{1}{1 - \rho\gamma}\|z_1 - z_2\| \ge \|\operatorname{prox}_{\gamma h}(z_1) - \operatorname{prox}_{\gamma h}(z_2)\|, \quad \forall z_1, z_2. \tag{6.3}$$

Let $h_\rho(x) := h(x) + \rho\|x\|^2/2$. Consider any $x_1, x_2 \in \operatorname{dom} \partial h$. For any $y_i \in \partial h(x_i), i = 1, 2$, we always have $y_i + \rho x_i \in \partial h_\rho(x_i)$. If h is ρ-weakly convex, we have that h_ρ is convex, and thus

$$\langle y_1 + \rho x_1 - (y_2 + \rho x_2), x_1 - x_2\rangle \ge 0 \quad \Rightarrow \quad \langle y_1 - y_2, x_1 - x_2\rangle \ge -\rho\|x_1 - x_2\|^2 \tag{6.4}$$

by the monotonicity of subdifferentials of proper and closed convex functions; see, for example, [2, Thm. 20.25]. Since $\rho\gamma < 1$, we have that $\operatorname{prox}_{\gamma h}(x)$ is single valued for any x as the objective function in (6.1) is strongly convex. Consider any z_1, z_2 and now let $x_i = \operatorname{prox}_{\gamma h}(z_i), i = 1, 2$, we see from the first-order optimality condition of (6.1) that there are $y_i \in \partial h(x_i)$ such that $z_i - x_i = \gamma y_i, i = 1, 2$. We therefore deduce from (6.4) and the Cauchy–Schwarz inequality that

$$\langle z_1 - x_1 - (z_2 - x_2), x_1 - x_2\rangle \geq -\gamma\rho\|x_1 - x_2\|^2 \quad \Rightarrow \quad \|z_1 - z_2\| \geq (1-\rho\gamma)\|x_1 - x_2\|,$$

which translates to (6.3).

Now assume subdifferential GEB holds with constants $\epsilon, \theta > 0$ and gauge ϕ, and let $\hat{x} := \operatorname{prox}_{\gamma h}(x - \gamma\nabla g(x))$. Since $\bar{x} \in \operatorname{crit} f$ gives $-\nabla g(\bar{x}) \in \partial h(\bar{x})$, which is the optimality condition of (6.1) at $\bar{x}$ with the input $\bar{x} - \gamma\nabla g(\bar{x})$, we have $\operatorname{prox}_{\gamma h}(\bar{x} - \gamma\nabla g(\bar{x})) = \bar{x}$, and thus (6.3) gives

$$\|\hat{x} - \bar{x}\| \leq \frac{1}{1-\rho\gamma}\|x - \bar{x} - \gamma\nabla g(x) + \gamma\nabla g(\bar{x})\| \leq \frac{1+\gamma L}{1-\rho\gamma}\|x - \bar{x}\|.$$

Hence, for any $x \in B_{\hat{\epsilon}}(\bar{x})$ with $\hat{\epsilon} := \epsilon(1-\rho\gamma)/(1+\gamma L)$, the point $\hat{x}$ lies in the region where the subdifferential GEB applies. The optimality condition of (6.1) at $\hat{x}$ reads $\gamma^{-1}(x - \hat{x}) - \nabla g(x) \in \partial h(\hat{x})$, so that $\gamma^{-1}(x-\hat{x}) + \nabla g(\hat{x}) - \nabla g(x) \in \nabla g(\hat{x}) + \partial h(\hat{x}) = \partial f(\hat{x})$, and therefore

$$\phi\big(\operatorname{dist}(\hat{x}, S)\big) \leq \theta \operatorname{dist}\big(0, \partial f(\hat{x})\big) \leq \frac{\theta(1+\gamma L)}{\gamma}\|x - \hat{x}\| = c\|x - \hat{x}\|. \tag{6.5}$$

It remains to transfer this estimate from $\hat{x}$ back to x. For a nonlinear gauge, the two terms in $\operatorname{dist}(x, S) \leq \|x - \hat{x}\| + \operatorname{dist}(\hat{x}, S) \leq 2\max\{\|x - \hat{x}\|, \operatorname{dist}(\hat{x}, S)\}$ cannot be combined under ϕ, so we keep the maximum. If it is attained by the first term, then $\operatorname{dist}(x, S) \leq 2\|x - \hat{x}\|$; otherwise $\operatorname{dist}(x, S)/2 \leq \operatorname{dist}(\hat{x}, S)$, and (6.5) together with the monotonicity of ϕ gives $\phi(\operatorname{dist}(x, S)/2) \leq c\|x - \hat{x}\|$. In both cases, $\hat{\phi}(\operatorname{dist}(x, S)) \leq \max\{2, c\}\|x - \hat{x}\|$ with $\hat{\phi}$ defined in (6.2), which is again a gauge function as the minimum of two gauge functions. This is exactly the proximal GEB.

The converse result also holds. Assume proximal GEB holds with constants $\hat{\epsilon}, \hat{\theta} > 0$ and gauge $\hat{\phi}$. Noting that $\operatorname{prox}_{\gamma h}(x + \gamma v - \gamma\nabla g(x)) = x$ for any $v \in \nabla g(x) + \partial h(x)$ and using (6.3), we have

$$\|x - \hat{x}\| = \|\operatorname{prox}_{\gamma h}(x + \gamma v - \gamma\nabla g(x)) - \operatorname{prox}_{\gamma h}(x - \gamma\nabla g(x))\| \leq \frac{\gamma}{1-\rho\gamma}\|v\|.$$

Since v is arbitrary, we have $(1-\rho\gamma)\|x - \hat{x}\| \leq \gamma \operatorname{dist}(0, \partial f(x))$, and hence,

$$\hat{\phi}(\operatorname{dist}(x, S)) \leq \hat{\theta}\|x - \hat{x}\| \leq \frac{\hat{\theta}\gamma}{1-\rho\gamma}\operatorname{dist}(0, \partial f(x)), \quad \forall x \in B_{\hat{\epsilon}}(\bar{x}),$$

where the inequality holds trivially at any x with $\partial f(x) = \emptyset$. This is exactly the subdifferential GEB. □

This confirms the claim in Section 4: with a general gauge, Definition 6.1 generalizes the Luo–Tseng error bound (which corresponds to $\phi(t) = t$), and Lemma 6.2 shows its relation to Definition 4.1.

As the error bounds concern only a neighborhood of $\bar{x}$ and weak convexity is essentially the globalization of prox-regularity, the assumption on h in Lemma 6.2 can be relaxed to prox-regularity at the single point $\bar{x}$, with proper modifications (but then we would need to add back the sublevel restriction). Likewise, the Lipschitz continuity of ∇g can be weakened to Hölder continuity with exponent $\alpha \in (0, 1]$, at the cost of raising the second branch of (6.2) to the power $1/\alpha$ and hence a less sharp gauge.

In the introduction, we recalled that the equivalence between EB on $(\mathbb{R}^n, d)$ and EB on $(\mathcal{M}, d)$ for ℓ_1-regularization was shown in [46] (for $S = \operatorname{argmin} f$). Next, we show that our main result (Theorems 4.10 and 4.11) can be applied to ℓ_1-regularization and recover this result with more generality on S. First, the ℓ_1-norm is $\mathcal{C}^1$-partly smooth at any point $\bar{x}$ relative to the manifold $\mathcal{M} = \{x : x_i = 0, \forall i \in I\}$ where $I = \{i : \bar{x}_i = 0\}$, see for example [19, Ex. 1.2.5]. Moreover, since g is $\mathcal{C}^{1,1}$ by Assumption 3, the sum $f = g + h$ is also $\mathcal{C}^1$-partly smooth at $\bar{x}$ relative to the same manifold $\mathcal{M}$.

As $h = \lambda\|\cdot\|_1$ is convex, we have $\rho = 0$, so Lemma 6.2 applies to every $\gamma > 0$ and we may simply take $\gamma = 1$. Moreover, if the subdifferential EB holds with the linear gauge $\phi(t) = t$, Lemma 6.2 gives the proximal EB with the same linear gauge. Since Theorems 4.10 and 4.11 preserve the gauge as well, they imply that if Assumption 2 and (3.2) hold at $\bar{x}$, then proximal EB (Definition 6.1) with $\gamma = 1$ is equivalent to EB on $\mathcal{M}$ (Definition 4.2), both with a linear gauge. The setting $S = \operatorname{argmin} f$ in [46, Thm. 4.3] makes Assumption 2 automatic, so the above recovers it (whose EB definition likewise carries no sublevel restriction), and extends to any $S \subseteq \operatorname{crit} f$ satisfying Assumption 2.

7. Conclusion. We have studied the transfer of local error bounds between an ambient Euclidean space and an identifiable $\mathcal{C}^1$ manifold, showing from both a geometric and a metric perspective that an error bound in the ambient space is equivalent to that on the manifold and sharpening is not possible. To establish this equivalence, we have developed tools to obtain a uniform linear-growth estimate for nearby points off the manifold without resorting to nearest-point projections that might not be available for $\mathcal{C}^1$ manifolds. We have further applied the developed tools to show ambient-to-manifold transfer of the Kurdyka–Łojasiewicz property under the same setting, and established that the generalized error bound with a wide range of gauge functions also implies this property, with an explicit conversion from the gauge to its desingularizer. Finally, for regularized optimization, we have related the generalized error bound to its proximal form, of which the Luo–Tseng error bound is the linear-gauge special case. A natural direction for future work is to investigate the implications of these generalized error bounds for local convergence rates of active-set and manifold-based algorithms.

Acknowledgments. The research of Hanju Wu and Yue Xie is funded by Hong Kong RGC General Research Fund (Project number: 17300824), and Guangdong Province Fundamental and Applied Fundamental Research Regional Joint Fund (Project number: 2022B1515130009). Ching-pei's research is supported in part by JSPS KAKENHI (24K20845), JST PRESTO (JPMJPR25KC), and JST BOOST (JPMJBY25F7). The authors thank Bruno Lourenço and Man-Chung Yue for helpful discussions on the improper integral in Proposition 5.2.

REFERENCES

[1] 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*, Mathematical programming, 137 (2013), pp. 91–129.
[2] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, second ed., 2017.
[3] D. P. Bertsekas, *Projected Newton methods for optimization problems with simple constraints*, SIAM Journal on control and Optimization, 20 (1982), pp. 221–246.
[4] J. Bolte, A. Daniilidis, A. S. Lewis, and M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM Journal on Optimization, 18 (2007), pp. 556–572.
[5] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet, *Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity*, Transactions of the American Mathematical Society, 362 (2010), pp. 3319–3363.
[6] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming, 146 (2014), pp. 459–494.
[7] N. Boumal, *An introduction to optimization on smooth manifolds*, Cambridge University Press, UK, 2023.
[8] A. Daniilidis, W. L. Hare, and J. Malick, *Geometrical interpretation of the predictor-corrector type algorithms in structured optimization problems*, Optimization, 55 (2006), pp. 481–503.
[9] D. Davis, D. Drusvyatskiy, and V. Charisopoulos, *Stochastic algorithms with geometric step decay converge linearly on sharp functions*, Mathematical Programming, 207 (2024), pp. 145–190.
[10] E. De Giorgi, A. Marino, and M. Tosques, *Problems of evolution in metric spaces and maximal decreasing curve*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.(8), 68 (1980), pp. 180–187.
[11] Z. Deng, J. Hu, K. Deng, and Z. Wen, *An efficient primal dual semismooth Newton method for semidefinite programming*, 2025, https://arxiv.org/abs/2504.14333.
[12] M. Díaz, P. I. Lehmann, H. Lu, and J. Yang, *Active set identification and rapid convergence for degenerate primal-dual problems*, 2026, https://arxiv.org/abs/2602.10436.
[13] D. Drusvyatskiy, *Slope and geometry in variational mathematics*, PhD thesis, Cornell University, 2013.
[14] D. Drusvyatskiy, A. D. Ioffe, and A. S. Lewis, *Nonsmooth optimization using Taylor-like models: error bounds, convergence, and termination criteria*, Mathematical Programming, 185 (2021), pp. 357–383.
[15] D. Drusvyatskiy and A. S. Lewis, *Optimality, identifiability, and sensitivity*, 2012, https://arxiv.org/abs/1207.6628.
[16] D. Drusvyatskiy and A. S. Lewis, *Optimality, identifiability, and sensitivity*, Mathematical Programming, 147 (2014), pp. 467–498.
[17] D. Drusvyatskiy and A. S. Lewis, *Error bounds, quadratic growth, and linear convergence of proximal methods*, Mathematics of operations research, 43 (2018), pp. 919–948.
[18] E. M. Gafni and D. P. Bertsekas, *Two-metric projection methods for constrained optimization*, SIAM Journal on Control and Optimization, 22 (1984), pp. 936–964.
[19] W. L. Hare, *Identifying active manifolds in regularization problems*, in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer, 2011, pp. 261–271.
[20] W. L. Hare and A. S. Lewis, *Identifying active constraints via partial smoothness and prox-regularity*, Journal of Convex Analysis, 11 (2004), pp. 251–266.
[21] W. L. Hare and A. S. Lewis, *Identifying active manifolds*, Algorithmic Operations Research, 2 (2007), pp. 75–82.
[22] J. Hu, T. Tian, S. Pan, and Z. Wen, *On the analysis of semismooth Newton-type methods for composite optimization*, Journal of Scientific Computing, 103 (2025), p. 59.
[23] A. D. Ioffe, *An invitation to tame optimization*, SIAM Journal on Optimization, 19 (2009), pp. 1894–1917.
[24] A. D. Ioffe, *Variational analysis of regular mappings*, Springer Monographs in Mathematics. Springer, Cham, (2017).
[25] C.-p. Lee, *Accelerating inexact successive quadratic approximation for regularized optimization through manifold identification*, Mathematical Programming, 201 (2023), pp. 599–633.
[26] J. M. Lee, *Smooth manifolds*, in Introduction to smooth manifolds, Springer, 2013, pp. 1–31.
[27] A. S. Lewis, *Active sets, nonsmoothness, and sensitivity*, SIAM Journal on Optimization, 13 (2002), pp. 702–725.
[28] A. S. Lewis, J. Liang, and T. Tian, *Partial smoothness and constant rank*, SIAM Journal on Optimization, 32 (2022), pp. 276–291.
[29] A. S. Lewis and T. Tian, *Identifiability, the KL property in metric spaces, and subgradient curves*, Foundations of Compu-

tational Mathematics, (2024), pp. 1–38.
[30] A. S. Lewis and T. Tian, *The complexity of first-order optimization methods from a metric perspective*, Mathematical Programming, 212 (2025), pp. 49–78.
[31] A. S. Lewis and S. Zhang, *Partial smoothness, tilt stability, and generalized Hessians*, SIAM Journal on Optimization, 23 (2013), pp. 74–94.
[32] G. Li, B. S. Mordukhovich, and J. Zhu, *Generalized metric subregularity with applications to high-order regularized Newton methods*, Mathematics of Operations Research, (2026). Articles in Advance.
[33] G. Li and T. K. Pong, *Calculus of the exponent of Kurdyka–Łojasiewicz inequality and its applications to linear convergence of first-order methods*, Foundations of computational mathematics, 18 (2018), pp. 1199–1232.
[34] J. Liang, J. Fadili, and G. Peyré, *Activity identification and local linear convergence of forward–backward-type methods*, SIAM Journal on Optimization, 27 (2017), pp. 408–437.
[35] H. Liu, T. Yao, R. Li, and Y. Ye, *Folded concave penalized sparse linear regression: sparsity, statistical performance, and algorithmic theory for local solutions*, Mathematical programming, 166 (2017), pp. 207–240.
[36] T. Liu and B. F. Lourenço, *Concrete convergence rates for common fixed point problems under Karamata regularity*, Mathematical Programming, (2026).
[37] Z.-Q. Luo and P. Tseng, *Error bounds and convergence analysis of feasible descent methods: a general approach*, Annals of Operations Research, 46 (1993), pp. 157–178.
[38] S. A. Miller and J. Malick, *Newton methods for nonsmooth convex minimization: connections among U-lagrangian, Riemannian Newton and SQP methods*, Mathematical programming, 104 (2005), pp. 609–633.
[39] Q. Rebjock and N. Boumal, *Fast convergence to non-isolated minima: four equivalent conditions for c^2 functions*, Mathematical Programming, 213 (2025), pp. 151–199.
[40] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.
[41] P. Tseng, *Approximation accuracy, gradient methods, and error bound for structured convex optimization*, Mathematical Programming, 125 (2010), pp. 263–295.
[42] P. Tseng and S. Yun, *A coordinate gradient descent method for nonsmooth separable minimization*, Mathematical Programming, 117 (2009), pp. 387–423.
[43] B. Wen, X. Chen, and T. K. Pong, *Linear convergence of proximal gradient algorithm with extrapolation for a class of nonconvex nonsmooth minimization problems*, SIAM Journal on Optimization, 27 (2017), pp. 124–145.
[44] S. J. Wright, *Identifiable surfaces in constrained optimization*, SIAM Journal on Control and Optimization, 31 (1993), pp. 1063–1079.
[45] S. J. Wright, *Accelerated block-coordinate relaxation for regularized optimization*, SIAM Journal on Optimization, 22 (2012), pp. 159–186.
[46] H. Wu and Y. Xie, *On resolution of L1-norm minimization via a two-metric adaptive projection method*, 2025, https://arxiv.org/abs/2504.12260.
[47] M.-C. Yue, Z. Zhou, and A. M.-C. So, *A family of inexact SQA methods for non-smooth convex minimization with provable convergence guarantees based on the Luo–Tseng error bound property*, Mathematical Programming, 174 (2019), pp. 327–358.

Appendix A. Counterexamples.

A.1. A disconnected metric space. Let $(a_n)_{n\geq 1}$ be any sequence of positive real numbers satisfying $a_n \to 0$, and define

$$p_n := (n^{-1}, 0), \qquad q_n := (n^{-1}, a_n), \qquad S_n := \{(n^{-1}, t) : 0 \leq t \leq a_n\}.$$

Set

$$X := \{(0,0)\} \cup \bigcup_{n=1}^{\infty} S_n,$$

equipped with the Euclidean metric inherited from $\mathbb{R}^2$. Since $a_n \to 0$, the only possible accumulation point of the segments S_n away from finitely many fixed segments is $(0,0)$. Hence X is closed in $\mathbb{R}^2$. Since $\mathbb{R}^2$ is complete and every closed subset of a complete metric space is complete, X is complete. Define $f : X \to \mathbb{R}$ by

$$f(0,0) = 0, \qquad f(n^{-1}, t) = a_n - t \quad (0 \leq t \leq a_n).$$

Since $a_n \to 0$, the function f is continuous at $(0,0)$; on each segment it is linear. Thus f is continuous, hence closed, and $f \geq 0$. Therefore

$$|\nabla f|(0,0) = 0.$$

Let

$$\mathcal{M} := \{(0,0)\} \cup \{p_n, q_n : n \geq 1\}.$$

We first check that $\mathcal{M}$ is identifiable at $\bar{x} = (0,0)$ for f. If

$$x = (n^{-1}, t) \in X \setminus \mathcal{M}, \qquad 0 < t < a_n,$$

then moving upward along the same segment decreases f at unit rate:

$$\lim_{s\downarrow 0} \frac{f(x) - f(n^{-1}, t+s)}{d(x, (n^{-1}, t+s))} = 1.$$

Thus $|\nabla f|(x) \geq 1$ for every $x \in X \setminus \mathcal{M}$. Hence

$$\liminf_{x \to_f \bar{x},\ x \notin \mathcal{M}} |\nabla f|(x) \geq 1,$$

so $\mathcal{M}$ is identifiable at $\bar{x}$.

Now take the desingularizer

$$\psi(t) = t.$$

The sublevel set $\{x \in X : f(x) \le f(\bar{x})\}$ is

$$\{(0,0)\} \cup \{q_n : n \ge 1\}.$$

At every point $x \in X$ with $0 < f(x) < \delta$, either $x = p_n$ for some n, or x lies in the interior of a segment S_n. In both cases, moving upward along S_n gives

$$|\nabla f|(x) \ge 1.$$

Consequently f has the KL property at $\bar{x}$ with desingularizer $\psi(t) = t$. Thus property (a) in [29, Thm. 5.12] holds.

However, on the metric subspace $\mathcal{M}$, each point p_n is isolated. Indeed,

$$d(p_n, q_n) = a_n > 0,$$

and the remaining points of $\mathcal{M}$ also have positive distance from p_n. Hence p_n is a local minimizer of $f|_{\mathcal{M}}$, so

$$|\nabla(f|_{\mathcal{M}})|(p_n) = 0.$$

But

$$p_n \to \bar{x}, \qquad f(p_n) = a_n \to 0, \qquad p_n \notin \{x \in \mathcal{M} : f(x) \le f(\bar{x})\}.$$

Therefore, $f|_{\mathcal{M}}$ does not have the KL property at $\bar{x}$ with desingularizer $\psi(t) = t$. In fact, for any desingularizer ψ, each p_n is a local minimizer of $\psi \circ (f|_{\mathcal{M}})$. Thus no desingularizer yields the KL property for the restriction.

A.2. A convex function on Euclidean space. Let $X = \mathbb{R}^2$ with the Euclidean metric, and define

$$f(u, v) = |v|, \qquad \bar{x} = (0,0), \qquad \mathcal{M} = (\mathbb{R} \times \{0\}) \cup \{(u, u^2) : u \in \mathbb{R}\}.$$

The function f is convex and Lipschitz continuous, and $\mathcal{M}$ is closed, connected, and semialgebraic. Since $|\nabla f|(\bar{x}) = 0$ and $|\nabla f|(u, v) = 1$ whenever $v \neq 0$, the function f has the KL property at $\bar{x}$ with desingularizer $\psi(t) = t$. Moreover, every point outside $\mathcal{M}$ has $v \neq 0$, so we get $\mathcal{M}$ is identifiable at $\bar{x}$ for f with modulus $\eta = 1$.

For $u \neq 0$, the set $\mathcal{M}$ coincides with the parabola near $x_u = (u, u^2)$. The parametrization $u \mapsto (u, u^2)$ has derivative $(1, 2u)$, and $f(x_u) = u^2$. Consequently, Lemma 2.5 gives

$$|\nabla(f|_{\mathcal{M}})|(x_u) = \frac{2|u|}{\sqrt{1+4u^2}} \longrightarrow 0 \qquad \text{as } u \to 0.$$

Since $x_u \to \bar{x}$ and $f(x_u) \to f(\bar{x}) = 0$ from above, the restriction $f|_{\mathcal{M}}$ fails the KL inequality with $\psi(t) = t$. This disproves the ambient-to-restricted implication in [29, Thm. 5.12] even in this Euclidean setting.

The restricted slope is continuous at $\bar{x}$: it vanishes on the horizontal axis and tends to zero along the parabola. However, $\mathcal{M}$ is not a manifold at $\bar{x}$. In this example, however, the restriction satisfies the KL property with a different desingularizer, namely $\psi(t) = \sqrt{t}$, because

$$\psi'(f(x_u))\, |\nabla(f|_{\mathcal{M}})|(x_u) = \frac{1}{\sqrt{1+4u^2}},$$

which is bounded away from zero for u near 0.

Appendix B. Proofs of Results in Section 2.

B.1. Proof of Lemma 2.1. Let $m = \dim \mathcal{M}$ and choose a $\mathcal{C}^p$ local defining map $F : U_0 \to \mathbb{R}^{n-m}$ around $\bar{x}$ such that $\mathcal{M} \cap U_0 = F^{-1}(0)$ and $\nabla F(\bar{x})$ has full rank. Set $T := T_{\bar{x}}\mathcal{M} = \ker \nabla F(\bar{x})$ and $N := N_{\bar{x}}\mathcal{M} = T^{\perp}$. Consider

$$G(u, w) := F(\bar{x} + u + w), \qquad u \in T, \quad w \in N.$$

The derivative of G with respect to w at $(0,0)$ is the linear map

$$\mathrm{D}_w G(0,0)w = \nabla F(\bar{x})w, \qquad w \in N.$$

This map from N to $\mathbb{R}^{n-m}$ is an isomorphism: it is injective because if $w \in N$ and $\nabla F(\bar{x})w = 0$, then $w \in T \cap N = \{0\}$, and the dimensions of N and $\mathbb{R}^{n-m}$ agree. By the implicit function theorem, there exist $\rho > 0$, a neighborhood O_N of 0 in N, and a unique $\mathcal{C}^p$ map $h : T \cap B_\rho(0) \to O_N$ such that $h(0) = 0$ and

$$F(\bar{x} + u + w) = 0 \quad \Longleftrightarrow \quad w = h(u)$$

for (u,w) sufficiently small. Hence $\mathcal{M}$ is locally represented as the stated graph. Differentiating the identity $F(\bar{x}+u+h(u))=0$ at $u=0$ gives

$$\nabla F(\bar{x})\big(u+\nabla h(0)u\big)=0, \qquad \forall u\in T.$$

Since $u\in\ker\nabla F(\bar{x})$, we have $\nabla F(\bar{x})\nabla h(0)u=0$. But $\nabla h(0)u\in N$ and $\nabla F(\bar{x})|_N$ is injective, so $\nabla h(0)u=0$ for all $u\in T$, and therefore $\nabla h(0)=0$. The final assertion follows by decomposing $y-\bar{x}$ according to $\mathbb{R}^n=T\oplus N$ and shrinking the neighborhood so that the tangential component lies in $B_\rho(0)$.

B.2. Proof of Proposition 2.2. Fix $g\in\hat{\partial}f(x)$ and $\xi\in T_x\mathcal{M}$. Choose a $\mathcal{C}^1$ curve $\gamma:(-\epsilon,\epsilon)\to\mathcal{M}$ satisfying $\gamma(0)=x$ and $\dot{\gamma}(0)=\xi$. The definition of a Fréchet subgradient gives

$$f(\gamma(t))-f(x)\geq\langle g,\gamma(t)-x\rangle+o(\|\gamma(t)-x\|).$$

Since $f|_{\mathcal{M}}$ is $\mathcal{C}^1$ around x, dividing by $t>0$ and letting $t\downarrow 0$ yields

$$\mathrm{D}(f|_{\mathcal{M}})(x)[\xi]\geq\langle g,\xi\rangle,$$

while dividing both sides by $t<0$ and letting $t\uparrow 0$ yields the reverse inequality. Therefore,

$$\mathrm{D}(f|_{\mathcal{M}})(x)[\xi]=\langle g,\xi\rangle=\langle P_{T_x\mathcal{M}}(g),\xi\rangle.$$

By the definition of the Riemannian gradient, it follows that

$$\nabla_{\mathcal{M}}f(x)=P_{T_x\mathcal{M}}(g).$$

Hence, any two elements of $\hat{\partial}f(x)$ have the same projection onto $T_x\mathcal{M}$, proving (2.1).

B.3. Proof of Lemma 2.5. Let

$$T\coloneqq T_x\mathcal{M}, \qquad N\coloneqq N_x\mathcal{M}, \qquad g\coloneqq\nabla_{\mathcal{M}}f(x)\in T.$$

By Lemma 2.1, after shrinking the neighborhood of x if necessary, there is a $\mathcal{C}^1$ map

$$h:T\cap B_\rho(0)\to N$$

such that $h(0)=0$, $\nabla h(0)=0$, and

$$\Phi(u)\coloneqq x+u+h(u), \qquad u\in T\cap B_\rho(0),$$

parametrizes $\mathcal{M}$ near x. Since $h(u)=o(\|u\|)$, we have

$$\|\Phi(u)-x\|=\|u+h(u)\|=\|u\|+o(\|u\|). \tag{B.1}$$

Define

$$F:T\cap B_\rho(0)\to\mathbb{R}, \qquad F(u)\coloneqq f(\Phi(u)).$$

Since $f|_{\mathcal{M}}$ and Φ are $\mathcal{C}^1$, the function F is $\mathcal{C}^1$ near 0 in the vector space T. Moreover,

$$\mathrm{D}\Phi(0)[u]=u+\nabla h(0)u=u, \qquad u\in T.$$

By the chain rule and the definition of the Riemannian gradient,

$$\mathrm{D}F(0)[u]=\mathrm{D}(f|_{\mathcal{M}})(x)\big[\mathrm{D}\Phi(0)[u]\big]=\mathrm{D}(f|_{\mathcal{M}})(x)[u]=\langle g,u\rangle, \qquad u\in T.$$

Therefore the first-order Taylor expansion of F at 0 is

$$F(u)=F(0)+\langle g,u\rangle+o(\|u\|). \tag{B.2}$$

We first prove the upper bound. If x is a local minimizer of $f|_{\mathcal{M}}$, then

$$|\nabla(f|_{\mathcal{M}})|(x)=0\leq\|g\|.$$

Otherwise, let $y_r\in\mathcal{M}$ satisfy $y_r\to x$ and $y_r\neq x$. Writing $y_r=\Phi(u_r)$ with $u_r\to 0$ and using (B.1)–(B.2), we obtain

$$\frac{f(x)-f(y_r)}{\|x-y_r\|}=\frac{-\langle g,u_r\rangle+o(\|u_r\|)}{\|u_r\|+o(\|u_r\|)}\leq\frac{\|g\|+o(1)}{1+o(1)}.$$

Taking the limsup over all such sequences yields

$$|\nabla(f|_{\mathcal{M}})|(x)\leq\|g\|.$$

For the reverse inequality, if $g=0$, the upper bound and nonnegativity of the slope already imply equality. Assume $g\neq 0$ and set $v=-g/\|g\|\in T$. For $t>0$ small, let $y_t=\Phi(tv)\in\mathcal{M}$. Using (B.1)–(B.2) again, we get

$$\frac{f(x)-f(y_t)}{\|x-y_t\|}=\frac{t\|g\|+o(t)}{t+o(t)}\to\|g\| \qquad (t\downarrow 0).$$

Hence $|\nabla(f|_{\mathcal{M}})|(x)\geq\|g\|$. The two inequalities prove the claim.

Appendix C. Proofs of Results in Section 4.

C.1. Proof of Lemma 4.7. Fix $x \in \mathcal{M}$ sufficiently close to $\bar{x}$. By Proposition 3.2, par $\hat{\partial} f(x) = N_x\mathcal{M}$. Thus, for any $g, g' \in \hat{\partial} f(x)$, $g - g' \in N_x\mathcal{M}$, so that

$$P_{T_x\mathcal{M}}(g - g') = 0.$$

Hence the tangential projection is independent of the choice of subgradient. Let $p = P_{\operatorname{aff} \hat{\partial} f(x)}(0)$. The characterization of the Euclidean projection onto an affine space gives $p \perp N_x\mathcal{M}$, while $g - p \in N_x\mathcal{M}$ for every $g \in \hat{\partial} f(x)$. Consequently,

$$P_{T_x\mathcal{M}}(g) = p = P_{\operatorname{aff} \hat{\partial} f(x)}(0).$$

C.2. Proof of Lemma 4.8. By (2.2), it suffices to show that $\nabla_{\mathcal{M}} f(x) \in \operatorname{ri} \hat{\partial} f(x)$ for all $x \in \mathcal{M}$ near $\bar{x}$. Prox-regularity at $\bar{x}$ for 0 gives $0 \in \partial_P f(\bar{x}) \subseteq \hat{\partial} f(\bar{x})$, so (2.2) implies $\nabla_{\mathcal{M}} f(\bar{x}) = 0$. Since $f|_{\mathcal{M}}$ and $\mathcal{M}$ are $\mathcal{C}^1$-smooth, the Riemannian gradient is continuous, and hence

$$\nabla_{\mathcal{M}} f(x) \to 0 \qquad \text{as } x \to \bar{x} \text{ in } \mathcal{M}. \tag{C.1}$$

For $x \in \mathcal{M}$ near $\bar{x}$, the set $\hat{\partial} f(x)$ is nonempty, closed, and convex. By Proposition 3.2 and Lemma 4.7 and (2.2),

$$\operatorname{aff} \hat{\partial} f(x) = \nabla_{\mathcal{M}} f(x) + N_x\mathcal{M}.$$

If $N_{\bar{x}}\mathcal{M} = \{0\}$, this identity shows that $\hat{\partial} f(x) = \{\nabla_{\mathcal{M}} f(x)\}$ for all nearby $x \in \mathcal{M}$, and the conclusion follows. Otherwise, suppose for a contradiction that there are $x_k \to \bar{x}$ in $\mathcal{M}$ such that $\nabla_{\mathcal{M}} f(x_k) \notin \operatorname{ri} \hat{\partial} f(x_k)$. Applying convex separation in the normal space $N_{x_k}\mathcal{M}$ to the translated set $\hat{\partial} f(x_k) - \nabla_{\mathcal{M}} f(x_k)$, we obtain unit vectors $w_k \in N_{x_k}\mathcal{M}$ with

$$\langle w_k, v - \nabla_{\mathcal{M}} f(x_k) \rangle \le 0 \qquad \forall v \in \hat{\partial} f(x_k). \tag{C.2}$$

Passing to a subsequence and using continuity of the normal spaces, assume that $w_k \to w \in N_{\bar{x}}\mathcal{M}$ with $\|w\| = 1$.

Sharpness and nondegeneracy at $\bar{x}$ give $\delta w \in \hat{\partial} f(\bar{x})$ for some $\delta > 0$. Choose δ small enough that δw lies in the neighborhoods of 0 where both inner semicontinuity and prox-regularity in Definition 3.1 apply. Inner semicontinuity provides $v_k \in \partial f(x_k)$ with $v_k \to \delta w$. Since $f(x_k) \to f(\bar{x})$, prox-regularity at $\bar{x}$ for 0 implies $v_k \in \partial_P f(x_k) \subseteq \hat{\partial} f(x_k)$ for all sufficiently large k. Applying (C.2) to v_k and using (C.1), we obtain

$$0 \ge \lim_{k\to\infty} \langle w_k, v_k - \nabla_{\mathcal{M}} f(x_k) \rangle = \delta \|w\|^2 = \delta > 0,$$

a contradiction.

C.3. Proof of Lemma 4.9. By Lemmas 4.7 and 4.8 and (2.2),

$$\nabla_{\mathcal{M}} f(x) = P_{\operatorname{aff} \hat{\partial} f(x)}(0) \in \hat{\partial} f(x).$$

Thus the projection of 0 onto the affine hull already belongs to $\hat{\partial} f(x)$, and is also its unique minimum-norm element. The two asserted identities follow.