Journal of Convex Analysis
Volume 11 (2004), No. 2, 251–266

Identifying Active Constraints via Partial Smoothness and Prox-Regularity

W. L. Hare
Department of Mathematics, Simon Fraser University,
Burnaby, BC V5A 1S6, Canada
whare@cecm.sfu.ca

A. S. Lewis*
Department of Mathematics, Simon Fraser University,
Burnaby, BC V5A 1S6, Canada
aslewis@sfu.ca, http://www.cecm.sfu.ca/~aslewis

Received December 3, 2002
Revised manuscript received November 12, 2003

Active set algorithms, such as the projected gradient method in nonlinear optimization, are designed to "identify" the active constraints of the problem in a finite number of iterations. Using the notions of "partial smoothness" and "prox-regularity" we extend work of Burke, Moré and Wright on identifiable surfaces from the convex case to a general nonsmooth setting. We further show how this setting can be used in the study of sufficient conditions for local minimizers.

Keywords: Nonlinear program, nonsmooth optimization, variational analysis, partly smooth, prox-regular, identifiable surface, projected gradient

2000 Mathematics Subject Classification: Primary: 91C30, 49K40, Secondary: 65K10

1. Introduction

In 1988 Burke and Moré [1] showed that certain simple algorithms would, in a finite number of iterations, "identify" the face of a polyhedral feasible region on which the solutions to an optimization problem occur. Specifically, if a sequence of points on the relative boundary of the feasible region approaches a point in the relative interior of the face, then the sequence eventually lies on the face. This result does not specify an algorithm, but simply lists the basic properties needed. These criteria are met by several algorithms, including natural versions of the projected gradient and Newton methods.

In a follow-up paper, Burke [2] generalized some of these results to the nonconvex case. His approach showed that by taking a local convex polyhedral approximation of a nonconvex set, certain algorithms could identify the active face of this approximation (although not the underlying set) in a finite number of iterations.

In 1993 Wright [10] extended Burke and Moré's work to nonpolyhedral sets. With the idea of "$\mathcal{C}^p$-identifiable surfaces", he was able to provide a large class of sets, and thus algorithms, that maintained this finite identification property. However he again restricted himself to convex sets. One of our aims in this work is to extend the results of Burke and Moré and those of Wright to the nonconvex setting.

*Research supported by NSERC.

ISSN 0944-6532 / $ 2.50

More recent work by Lewis [5] has led to a nonconvex version of $\mathcal{C}^p$-identifiable sets. Working from the notion of a "partly smooth function", [5] shows that $\mathcal{C}^p$-identifiable sets have partly smooth indicator functions. Partial smoothness asks that a function be smooth along some "active" manifold, and grows "sharply" in directions away from the manifold (see Definition 2.3 below). In the case of sets this growth requirement becomes a condition on the normal cone to the set. It is this definition that allows us to extend the research of Burke and Moré and of Wright.

In order to ensure the projection mapping naturally associated with an identifiable surface is well defined (at least locally), we must introduce an additional assumption to partial smoothness. "Prox-regularity", a condition analyzed at some length in [6], and also studied under a different name in [9], provides us with the extra tool we require. Combining prox-regularity with partial smoothness, we arrive at a broad collection of sets and functions that we show to have many desirable properties in optimization.

In order to motivate the finite identification results we desire, we briefly discuss general properties of projection mappings. In 1973 Holmes [4] examined the smoothness of the projection mapping for convex sets. Fitzpatrick and Phelps [3] continued this work in 1988. They showed that convex sets with $\mathcal{C}^p$ boundaries ($p \geq 2$) have $\mathcal{C}^{p-1}$ associated projection mappings, and gave conditions for this to be an if and only if statement. In our present work, Theorem 3.3 shows that this implication also holds true locally for partly smooth, prox-regular sets, and in fact that projecting onto such a set is equivalent (locally) to projecting onto a manifold contained in the set.

In Section 4 we use this result to show how the finite identification results of Burke and Moré, as well as those of Wright, can be extended to our broader class of sets. As all convex sets are prox-regular and both papers worked with what amounts to partly smooth sets, our results subsume those of both papers. Remarkably, we show that if a partly smooth function is prox-regular, then the corresponding active manifold must be *unique*.

In Sections 5 and 6 we shift away from sets, and begin looking at partly smooth prox-regular *functions*. To do this we show a useful theorem comparing partial smoothness for functions and their epigraphs. Just as a function is convex if and only if its epigraph is convex, a function is prox-regular if and only if its epigraph is prox-regular [6]. The analogous result for partial smoothness is equally intuitive, and shown in Section 5. Using these results it is not hard to shift our results about sets to the functional case.

A brief examination of the conditions required for Sections 4 and 5 easily shows that for partly smooth, prox-regular minimization problems and any "strict" critical point on an associated "active" manifold, all nearby critical points also lie on the active manifold. In Section 6 we examine how growth along the active manifold can ensure local minimality. In fact, for partly smooth prox-regular functions with second-order growth along the active manifold, any strict critical point is actually a local second-order minimizer.

In Section 7 we end with an example showing the necessity of prox-regularity throughout this analysis.

2. Notation

Throughout this paper we consider a function

$$f : \mathbf{R}^m \to [-\infty, +\infty].$$

We follow the notation of [8] and refer there for many basic results. Most notably we make use of the *regular* (or *Fréchet*) *subdifferential* of a function f at a point $\bar{x} \in \mathbf{R}^m$ where f is finite,

$$\hat{\partial} f(\bar{x}) := \{v \in \mathbf{R}^m : f(x) \geq f(\bar{x}) + \langle v, x - \bar{x}\rangle + o(|x - \bar{x}|)\}$$

(the regular subdifferential being empty at any point where f is infinite), and the *subdifferential*,

$$\partial f(\bar{x}) := \limsup_{x \to \bar{x},\ f(x) \to f(\bar{x})} \hat{\partial} f(x)$$

(also known as the *limiting Fréchet subdifferential*). Correspondingly we have the *regular* (or *Fréchet*) *normal cone* to a set S at a point $\bar{x} \in S$, which we can define by $\hat{N}_S(\bar{x}) := \hat{\partial}\delta_S(\bar{x})$, and the *(limiting) normal cone*, $N_S(\bar{x}) := \partial\delta_S(\bar{x})$, (where δ_S is the indicator function of S). Both normal cones are defined to be empty for any $\bar{x} \notin S$. We say S is *(Clarke) regular* at $\bar{x} \in S$ if it is locally closed at $\bar{x}$ and these two normal cones agree. Furthermore, we say f is *regular* at $\bar{x}$ if its epigraph

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

is regular at $(\bar{x}, f(\bar{x}))$: in this case, $\hat{\partial} f(\bar{x}) = \partial f(\bar{x})$.

As the title of the paper suggests we make abundant use of the concept of both prox-regularity and partial smoothness. We therefore go into their definitions in more detail. We begin with prox-regularity (see [6]).

Definition 2.1 (prox-regularity). A function f is *prox-regular* at a point $\bar{x}$ for a subgradient $\bar{v} \in \partial f(\bar{x})$ if f is finite at $\bar{x}$, locally lower semi-continuous around $\bar{x}$, and there exists $\rho > 0$ such that

$$f(x') \geq f(x) + \langle v, x' - x\rangle - \frac{\rho}{2}|x' - x|^2$$

whenever x and x' are near $\bar{x}$ with $f(x)$ near $f(\bar{x})$ and $v \in \partial f(x)$ is near $\bar{v}$. Further, f is *prox-regular* at $\bar{x}$ if it is prox-regular at $\bar{x}$ for every $v \in \partial f(\bar{x})$. A set S is *prox-regular* at a point $\bar{x} \in S$ (for a normal vector $\bar{v} \in N_S(\bar{x})$) if the indicator function of S is prox-regular there (for the appropriate subgradients). Thus S is prox-regular at $\bar{x} \in S$ for a normal vector $\bar{v} \in N_S(\bar{x})$ exactly when S is locally closed at $\bar{x}$, and there exists $\rho \geq 0$ such that

$$\langle v, x' - x\rangle \leq \frac{\rho}{2}|x' - x|^2$$

for all $x', x \in S$ near $\bar{x}$, and $v \in N_S(x)$ near $\bar{v}$.

The concept of prox-regularity in some ways extends the desirable properties of convexity to a broader class of functions. Many of these properties were developed in [6], then further extended in [7]. Our strongest use of prox-regularity is in its control over the projection mapping. This control is outlined in the next lemma.

Lemma 2.2 (prox-normal neighbourhood). *Suppose the set $S \subseteq \mathbf{R}^m$ is closed. Then S is prox-regular at the point $\bar{x} \in S$ if and only if the projection mapping P_S is single valued near $\bar{x}$. In this case there exists an open neighbourhood V of $\bar{x}$ on which the following properties hold:*

(*i*) *$P_S(\cdot)$ is single valued and Lipschitz continuous on V.*
(*ii*) *$P_S(\cdot) = (I + N_S)^{-1}(\cdot)$ on V.*
(*iii*) *For any point $x \in V$, all normal vectors to S in $(V - x)$ are "proximal normals": that is, for x and v in V, $v - x \in N_S(x)$ implies $x = P_S(v)$.*

Proof. The equivalence statement is shown in [7, Thm 1.3]. Parts (*i*) and (*ii*) can be found in [8, Ex 13.38], while part (*iii*) is a simple consequence. □

The if and only if condition of Lemma 2.2 shows that prox-regularity is exactly the condition one requires to ensure the projection mapping is well defined. Property (*iii*) further allows us great control over the projection mapping. This control is applied repeatedly throughout this paper, so it is useful to give the neighbourhood V a name: we refer to it as a *prox-normal neighbourhood* (of S at $\bar{x}$).

We now move on to partial smoothness and strong critical points. The purpose of the next definition is to identify structures of a nonsmooth function that aid in sensitivity analysis (see [5]). The analysis in [5] concerns the case for degree of smoothness $p = 2$: the theory for general p is a routine generalization. We recall that a set $\mathcal{M}$ is a $\mathcal{C}^p$ *manifold* about a point $\bar{x} \in \mathcal{M}$ if (locally) $\mathcal{M}$ can be described as the solution set of a collection of $\mathcal{C}^p$ equations with linearly independent gradients.

Definition 2.3 (partly smooth). A function f is $\mathcal{C}^p$*-partly smooth* at a point $\bar{x}$ relative to a set $\mathcal{M}$ containing $\bar{x}$ if $\mathcal{M}$ is a $\mathcal{C}^p$ manifold about $\bar{x}$ and:

(*i*) **(smoothness)** $f|_{\mathcal{M}}$ is a $\mathcal{C}^p$ function near $\bar{x}$;
(*ii*) **(regularity)** f is regular at all points $x \in \mathcal{M}$ near $\bar{x}$, with $\partial f(x) \neq \emptyset$;
(*iii*) **(sharpness)** the affine span of $\partial f(\bar{x})$ is a translate of $N_{\mathcal{M}}(\bar{x})$;
(*iv*) **(sub-continuity)** ∂f restricted to $\mathcal{M}$ is continuous at $\bar{x}$.

Further, a set S is $\mathcal{C}^p$*-partly smooth* at a point $\bar{x} \in S$ relative to a manifold $\mathcal{M}$ if its indicator function maintains this property. For both cases we refer to $\mathcal{M}$ as the *active manifold*.

For a discussion of the terminology "sharpness" see [5].

The sensitivity analysis of partly smooth functions revolves around the following idea. We denote the relative interior of a convex set C by $\operatorname{rint} C$.

Definition 2.4 (strong critical point). Let f be a partly smooth function at the point $\bar{x}$ relative to the manifold $\mathcal{M}$. Then we call $\bar{x}$ a *strong critical point* of f relative to $\mathcal{M}$ if

$$0 \in \operatorname{rint} \partial f(\bar{x})$$

and there exists $\varepsilon > 0$ such that

$$f(x) \geq f(\bar{x}) + \varepsilon |x - \bar{x}|^2 \tag{1}$$

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

The following are two good examples to keep in mind. The first demonstrates the abundance of partly smooth prox-regular functions and sets, and provides an insight as to what these functions and sets may look like. The second example shows where strong critical points may arise in optimization problems.

Example 2.5 (finite max functions and sets). Consider the function

$$f(x) := \max\{f_i(x) : i = 1, 2, \ldots, n\}$$

and set

$$S := \{x : g_i(x) \leq 0,\ i = 1, 2, \ldots, n\},$$

where f_i and g_i are $\mathcal{C}^2$ functions around the point $\bar{x}$, and $g_i(\bar{x}) \leq 0$ for each index i. Then [6, Ex 2.9 and Thm 3.5] can be easily combined to show both f and S are prox-regular at $\bar{x}$.

The *active set* for f at a point x is defined by $A_f(x) := \{i : f_i(x) = f(x)\}$. If we further include the assumption that the set of all active gradients of f, $\{\nabla f_i(\bar{x}) : i \in A_f(\bar{x})\}$, is linearly independent, then [5, Cor 4.8] shows f is partly smooth at $\bar{x}$ relative to the manifold

$$\mathcal{M}_f := \{x : A_f(x) = A_f(\bar{x})\}.$$

Similarly the *active set* for S at x is defined by $A_S(x) := \{i : g_i(x) = 0\}$. Adding the assumption of linear independence on the active gradients of S, $\{\nabla g_i(\bar{x}) : i \in A(\bar{x})\}$, then forces S to be partly smooth at $\bar{x}$ relative to the manifold

$$\mathcal{M}_g := \{x : A_S(x) = A_S(\bar{x})\}$$

(apply [5, 6.3] and [10, 2.4]).

Example 2.6 (smooth minimax). Consider functions $f_i \in \mathcal{C}^2$, and a point $(x, t) = (\bar{x}, f(\bar{x}))$ that is a local minimizer for the smooth nonlinear program

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & f_i(x) - t \leq 0 \quad (i = 1, 2, \ldots, n). \end{array}$$

Suppose the active constraint gradients are linearly independent, and satisfy strict complementarity and second-order sufficient conditions (see [8, Ex 13.25] for a discussion). Then it is classical to verify that $\bar{x}$ is a strong critical point of f relative to the manifold $\mathcal{M}_f$ (as defined in the previous example).

3. The Smooth Projection Theorem

We begin our examination into prox-regularity and partial smoothness with the case of sets. Our goal in this section is two-fold. First we seek to extend work started by Holmes [4] (1973) then continued by Fitzpatrick and Phelps [3] (1982) on the smoothness of the projection mapping. Secondly we provide the key result which allows much of this paper to be possible.

The work of Holmes showed for a convex set with a $\mathcal{C}^p$ boundary ($p \geq 2$), the projection mapping is $\mathcal{C}^{p-1}$. Fitzpatrick and Phelps showed with certain assumptions the reverse implication is true, then provided an example showing these requirements could not be loosened. This appeared to complete the topic, since without convexity the projection mapping is not a globally single-valued function.

By using Lemma 2.2 (prox-normal neighbourhood) and choosing points sufficiently close to a prox-regular set we can bypass this problem, and thus reopen the question of smoothness of the projection mapping. Theorem 3.3 (smooth projections) shows for a prox-regular set with a $\mathcal{C}^p$ boundary ($p \geq 2$) the projection map is $\mathcal{C}^{p-1}$ near the set. This corollary can be seen by noting the boundary is exactly the manifold required in the theorem.

However, the important part of Theorem 3.3 lies in its statement that, locally, projecting onto a partly smooth, prox-regular set is equivalent to projecting onto the corresponding active manifold. Uses of this are shown in Sections 4 and 5.

To achieve our goal we require two results. The main theorem follows these.

Lemma 3.1. *If $\mathcal{M}$ is a $\mathcal{C}^2$ manifold about 0 and the normal vector $\bar{y} \in N_{\mathcal{M}}(0)$ is sufficiently small then*

$$x \in \mathcal{M} \quad \Rightarrow \quad ||x - \bar{y}||^2 \geq ||\bar{y}||^2 + \frac{1}{2}||x||^2.$$

Proof. The required inequality is equivalent to

$$||x - 2\bar{y}||^2 \geq ||2\bar{y}||^2.$$

As $\mathcal{M}$ is $\mathcal{C}^2$ it is prox-regular at 0. Selecting $\bar{y}$ small enough that $2\bar{y}$ is in the prox-normal neighbourhood of S at 0 we have the projection of $2\bar{y}$ onto $\mathcal{M}$, $P_{\mathcal{M}}(2\bar{y})$, is 0 (by Lemma 2.2 (prox-normal neighbourhood), part (iii)). Now notice,

$$||x - 2\bar{y}||^2 \geq \min\{||z - 2\bar{y}||^2 : z \in \mathcal{M}\} = ||P_{\mathcal{M}}(2\bar{y}) - 2\bar{y}||^2 = ||2\bar{y}||^2,$$

as required. □

We also need the following result, essentially from [5].

Theorem 3.2 (parametric minimization). *Suppose the function $\rho : \mathbf{R}^k \times \mathbf{R}^m \to [-\infty, \infty]$ is $\mathcal{C}^p$-partly smooth at the point $(\bar{y}, \bar{z})$ relative to the manifold $\mathbf{R}^k \times \mathcal{M}$. If $\bar{z}$ is a strong critical point of $\rho_{\bar{y}}(\cdot) = \rho(\bar{y}, \cdot)$ relative to $\mathcal{M}$ then there exists neighbourhoods, U of $\bar{z}$ and V of $\bar{y}$ and a function $\Phi \in \mathcal{C}^{p-1}$ such that for all parameters $y \in V$,*

(i) $\Phi(\bar{y}) = \bar{z}$,
(ii) $\rho_y|_{\mathcal{M} \cap U}$ *has a unique critical point at* $\Phi(y)$,
(iii) $\Phi(y)$ *is a strong critical point of* $\rho_y|_{\mathcal{M} \cap U}$.

Proof. The proof of [5, Thm 5.7] is easily adaptable to this form. □

We are now ready to state and prove the key result.

Theorem 3.3 (smooth projections). *Let the set S be prox-regular and $\mathcal{C}^p$-partly smooth ($p \geq 2$) at the point $\bar{x}$ relative to the manifold $\mathcal{M}$. Then for any normal vector*

$$\bar{n} \in \operatorname{rint} N_S(\bar{x})$$

sufficiently small, there exists a neighbourhood of $\bar{x} + \bar{n}$ on which the projection mappings satisfy

$$P_{\mathcal{M}} \equiv P_S \in \mathcal{C}^{p-1}. \tag{2}$$

Proof. To begin note that by shifting the set we may assume without loss of generality that $\bar{x} = 0$.

Now define the function ρ via

$$\begin{array}{rcl} \rho : \mathbf{R}^m \times \mathbf{R}^m & \to & \bar{\mathbf{R}} \\ (n, x) & \mapsto & \frac{1}{2}||x - n||^2 + \delta_S(x), \end{array}$$

where δ_S is the indicator function of S. Assume $\bar{n} \in \operatorname{rint} N_S(0)$ is sufficiently small for Lemmas 2.2 (prox-normal neighbourhood) and 3.1 to hold on an open neighbourhood of $\bar{n}$. We claim ρ satisfies the conditions of Theorem 3.2 (parametric minimization) at $(\bar{n}, 0)$.

By [5, Prop 4.5], ρ is partly smooth at $(n, 0)$ relative to $\mathbf{R}^m \times \mathcal{M}$ for all vectors $n \in \mathbf{R}^m$. Since $\partial \rho_{\bar{n}}(0) = -\bar{n} + N_S(0)$, our choice of $\bar{n}$ ensures $0 \in \operatorname{rint} \partial \rho_{\bar{n}}(0)$. Lastly to see $(\bar{n}, 0)$ is a strong critical point note equation (1) is equivalent to

$$||x - \bar{n}||^2 \geq ||\bar{n}||^2 + 2\varepsilon ||x||^2 \quad \forall\, x \in \mathcal{M} \text{ near } \bar{x}.$$

This holds for $\varepsilon = \frac{1}{4}$ by Lemma 3.1.

Thus there exists a function $\Phi \in \mathcal{C}^{p-1}$ such that, for n near $\bar{n}$, $\Phi(n)$ is a strong critical point of ρ_n relative to $\mathcal{M}$ near 0. Thus $\Phi(n) \in \mathcal{M}$, and

$$0 \in \operatorname{rint} \partial \rho_n(\Phi(n)) = \operatorname{rint} \{(\Phi(n) - n) + N_S(\Phi(n))\}.$$

Therefore $n - \Phi(n) \in \operatorname{rint} N_S(\Phi(n))$.

As $\Phi \in \mathcal{C}^{p-1}$ near $\bar{n}$, we do not leave the prox-normal neighbourhood guaranteed by Lemma 2.2. Thus $n - \Phi(n) \in \operatorname{rint} N_S(\Phi(n))$ implies $n \mapsto P_S(n) = \Phi(n) = P_{\mathcal{M}}(n) \in \mathcal{C}^{p-1}$, completing the proof. $\square$

4. Identification for Partly Smooth Sets

Having established a useful understanding of the projection mapping for partly smooth prox-regular sets, we examine how this allows algorithms to identify the active manifold.

In 1988 Burke and Moré [1] showed how polyhedral faces of convex sets could be identified finitely. Their research was extended by Wright [10] who gave a convex version of Theorem 3.3 ([5, Thm 6.3] relates the results). He used this to provide an algorithm that identifies the active constraints in convex problems in a finite number of iterations. Although our arguments proceed in the style of Burke and Moré, the results also subsume those of Wright. This is easily seen as all convex sets are prox-regular.

We begin this section with a technical theorem stating the basic conditions required to ensure the manifold is identified in a finite number of steps. Theorem 4.3 characterizes which algorithms identify the manifold of a partly smooth prox-regular set. To see the applications of this theorem to optimization, we provide two simple, but illustrative, corollaries.

We begin with the main technical result.

Theorem 4.1 (finite identification for sets). *Consider a set S that is $\mathcal{C}^p$-partly smooth ($p \geq 2$) at the point $\bar{x}$ relative to the manifold $\mathcal{M}$ and prox-regular at $\bar{x}$. If the normal vector $\bar{n}$ is in* $\operatorname{rint} N_S(\bar{x})$*, and the sequences $\{x_k\}$ and $\{d_k\}$ satisfy*

$$x_k \to \bar{x}, \ d_k \to \bar{n}, \ \textit{and} \ \operatorname{dist}(d_k, N_S(x_k)) \to 0, \tag{3}$$

then

$$x_k \in \mathcal{M} \quad \textit{for all large } k.$$

Proof. Select a sequence of normal vectors $n_k \in N_S(x_k)$ such that $|d_k - n_k| \to 0$. Noting $\bar{n} \in N_S(\bar{x})$ implies $\lambda\bar{n} \in N_S(\bar{x})$ for any $\lambda > 0$ we may select $\lambda > 0$ such that Lemma 2.2 (prox-normal neighbourhood) and Theorem 3.3 (smooth projections) both hold for $\lambda\bar{n}$. Hence there exists a neighbourhood V of $\bar{x}+\lambda\bar{n}$ on which the projection mappings satisfy $P_S \equiv P_{\mathcal{M}} \in \mathcal{C}^{p-1}$. As $d_k \to \bar{n}$ and $|d_k - n_k| \to 0$, we must have $\lambda n_k \to \lambda\bar{n}$. Therefore for large k, $\lambda n_k + x_k$ will be in V. As $\lambda n_k \in N_S(x_k)$, the prox-normal property (Lemma 2.2 part (*iii*)) implies $x_k = P_S(\lambda n_k + x_k) = P_{\mathcal{M}}(\lambda n_k + x_k)$. Thus $x_k \in \mathcal{M}$ for all large k. □

Aside from the applications to optimization, the above theorem has surprising consequences for partial smoothness. Specifically, prox-regularity is strong enough to ensure the active manifold for a partly smooth set is unique. Section 7 contains an example that suggests this is a minimal condition that can guarantee the uniqueness of the active manifold.

Notice, in particular, that since convex sets are prox-regular, the previous ideas of Burke and Moré [1] and of Wright [10] used unique manifolds.

Corollary 4.2 (uniqueness of manifolds). *Consider a set S that is prox-regular at the point $\bar{x}$ and $\mathcal{C}^p$-partly smooth $(p \geq 2)$ there relative to each of the two manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$. Then near $\bar{x}$ we have $\mathcal{M}_1 \equiv \mathcal{M}_2$.*

Proof. For the sake of eventual contradiction, suppose there exists a sequence $\{x_k\}$ converging to $\bar{x}$ such that $x_k \in \mathcal{M}_1 \setminus \mathcal{M}_2$ for all k.

Select any normal vector $\bar{n} \in \operatorname{rint} N_S(\bar{x})$. As S is partly smooth relative to $\mathcal{M}_1$, we know the normal cones $N_S(x_k)$ converge to $N_S(\bar{x})$ [5, Prop 2.11 (iv)]. Thus there exists a sequence of normal vectors $n_k \in N_S(x_k)$ converging to $\bar{n}$. Applying Theorem 4.1 (finite identification for sets) with $\mathcal{M} \equiv \mathcal{M}_2$ shows $x_k \in \mathcal{M}_2$ for all large k. □

By replacing the sequence $\{d_k\}$ in the Theorem 4.1 (finite identification for sets) with gradients from a $\mathcal{C}^1$ function we can see how it applies to constrained optimization problems.

Theorem 4.3 (identification with constraints). *Consider a $\mathcal{C}^1$ function f. Let the set S be $\mathcal{C}^p$-partly smooth $(p \geq 2)$ at the point $\bar{x}$ relative to the manifold $\mathcal{M}$, and prox-regular there. Suppose $x_k \to \bar{x}$ and $-\nabla f(\bar{x}) \in \operatorname{rint} N_S(\bar{x})$. Then $x_k \in \mathcal{M}$ for all large k if and only if* $\operatorname{dist}(-\nabla f(x_k), N_S(x_k)) \to 0$.

Proof. $(\Rightarrow)$ If $x_k \in \mathcal{M}$ for all k sufficiently large, then S being partly smooth implies S is regular at x_k and $N_S(x_k) \to N_S(\bar{x})$ ([5, Prop 2.11 (ii) and (iv)]). As $f \in \mathcal{C}^1$ we know $-\nabla f(x_k) \to -\nabla f(\bar{x})$. Combining these facts with [8, Cor 4.7] yields the result.

$(\Leftarrow)$ Set $d_k = -\nabla f(x_k)$ and apply Theorem 4.1, noting $f \in \mathcal{C}^1$. □

Turning to more concrete optimization models, we return to Example 2.5 (finite max functions and sets). Applying the above result to a set defined via $\mathcal{C}^2$ constraints easily yields the following example.

Example 4.4 (identification with inequalities). Consider minimizing a $\mathcal{C}^1$ function f over a set

$$S := \{x : g_i(x) \le 0,\ i = 1, 2, \ldots, m\}$$

for $\mathcal{C}^2$ functions g_i with $\{\nabla g_i : i \in A(\bar{x})\}$ linearly independent, (where $A(x)$ is the active set $\{i : g_i(x) = 0\}$). Suppose $x_k \to \bar{x}$, and $-\nabla f(\bar{x}) \in \operatorname{rint} N_S(\bar{x})$, or equivalently, "strict complementarity" holds in the first order conditions:

$$-\nabla f(\bar{x}) = \sum_{i \in A(\bar{x})} \lambda_i \nabla g_i(\bar{x}), \quad \text{where } \lambda_i > 0 \text{ for all } i \in A(\bar{x}).$$

Then according to Theorem 4.3,

$$\operatorname{dist}(-\nabla f(x_k), N_S(x_k)) \to 0$$

if and only if

$$A(x_k) = A(\bar{x}) \ \text{ for all large } k.$$

(This example is also not too hard to work through by a direct argument.)

Simplifying Theorem 4.3 to the case of the active manifold being a single point gives the following corollary.

Corollary 4.5 (locally sharp minimizer). *If the conditions of Theorem* 4.3 *hold, and* $\mathcal{M} = \{\bar{x}\}$, *or equivalently* $N_S(\bar{x})$ *has interior, then* $x_k = \bar{x}$ *for all large* k *if and only if*

$$\operatorname{dist}(-\nabla f(x_k), N_{\mathcal{M}}(x_k)) \to 0.$$

Before moving on to functions we briefly explain how these results can be rephrased into the language of Burke and Moré. Theorem 4.3 (identification with constraints) discusses algorithms where the negative gradient of the objective comes arbitrarily close to the normal cone to the feasible set. This condition should be immediately recognizable from algorithms attempting to solve the first order necessary condition for $\min\{f : x \in S\}$. Another method of measuring the first order necessary condition is to examine the "projected (negative) gradient",

$$\nabla_S f(\bar{x}) := P_{T_S(\bar{x})}(-\nabla f(\bar{x})) \tag{4}$$

where $T_S(\bar{x})$ is the usual tangent cone (see [8, Def 6.1] for details). Of course for this definition to make sense, we require this projection to be unique. This is achieved when S is regular at $\bar{x}$, as is the case with prox-regularity. To see the interest in the projected gradient, notice it measures the distance of $-\nabla f(\bar{x})$ from satisfying first order necessity conditions. This is captured nicely in the following easy result.

Proposition 4.6 (approximate first order conditions). *If the function* f *is* $\mathcal{C}^1$ *and the set* S *is regular at the point* $\bar{x} \in S$, *then*

$$|\nabla_S f(x)| = \operatorname{dist}(-\nabla f(x), N_S(x)),$$

so in particular

$$\nabla_S f(\bar{x}) = 0 \iff \bar{x} \text{ is a critical point for } \min_S\{f\}.$$

Proof. Immediate from the fact that $T_S(\bar{x})$ and $N_S(\bar{x})$ are polar closed convex cones (see [1] for example). $\square$

Theorem 4.3 (identification with constraints) could now easily be rewritten in terms of the projected gradient.

5. Identification for Functions

The previous section shows that many algorithms solving

$$\min\{f(x) : x \in S\}$$

(for a smooth function f and a partly smooth prox-regular set S) could be expected to identify the active manifold in a finite number of steps. As this problem is often rephrased as $\min\{f + \delta_S\}$, we are naturally lead to the question of what can be said about minimizing a partly smooth *function*. To study this question, we make use of the equivalence of $\min\{f(x)\}$ to $\min\{r : (x, r) \in \operatorname{epi} f\}$. Using this idea we rewrite our previous results in terms of functions and their subdifferentials. This yields an elegant result on identifying manifolds for partly smooth functions.

We require the ability to move from functions to their epigraphs without losing prox-regularity or partial smoothness. Although we only require one direction, both prox-regular and partly smooth functions have exact correspondences with the prox-regularity and partial smoothness of their epigraphs. The prox-regular case was shown by Poliquin and Rockafellar in [6]. The second is not difficult and can be achieved by direct (though tedious) computation. We include a more elegant approach here.

Theorem 5.1 (partly smooth epigraph). *A given function f is $\mathcal{C}^p$-partly smooth at a point $\bar{x}$ relative to a manifold $\mathcal{M}$ if and only if $f|_{\mathcal{M}}$ is $\mathcal{C}^p$ around $\bar{x}$ and the set* $\operatorname{epi} f$ *is partly smooth at $(\bar{x}, f(\bar{x}))$ relative to the manifold $\{(x, f(x)) : x \in \mathcal{M}\}$.*

During the proof we make use of one simple lemma. In order to make the proof more readable we separate it out here.

Lemma 5.2. *Let the function f be smooth at the point $\bar{x}$ relative to the manifold $\mathcal{M}$. Define $\widehat{\mathcal{M}} := \{(x, f(x)) : x \in \mathcal{M}\}$. Then $\widehat{\mathcal{M}}$ is a manifold about $(\bar{x}, f(\bar{x}))$ and*

$$N_{\widehat{\mathcal{M}}}(\bar{x}, f(\bar{x})) = \{(n + \lambda\nabla g(\bar{x}), -\lambda) : n \in N_{\mathcal{M}}(\bar{x}),\ \lambda \in \mathbf{R}\}, \tag{5}$$

where g is any smooth function agreeing with f on $\mathcal{M}$.

Proof. Suppose $\mathcal{M}$ is defined locally by the smooth equations $h_i(x) = 0$ (for $i \in I$), where the set of gradients

$$G = \{\nabla h_i(\bar{x}) : i \in I\}$$

is linearly independent. Then the normal space $N_{\mathcal{M}}(\bar{x})$ is just the linear span of G [8, Ex. 6.8].

Using these functions, $\widehat{\mathcal{M}}$ is locally defined as the set of points $(x, t) \in \mathbf{R}^n \times \mathbf{R}$ satisfying the smooth equations

$$h_i(x) = 0 \quad (i \in I), \quad \text{and} \quad g(x) - t = 0.$$

Since the set of gradients

$$\widehat{G} = \{(\nabla h_i(\bar{x}), 0) : i \in I\} \cup \{(\nabla g(\bar{x}), -1)\}$$

is linearly independent, $\widehat{\mathcal{M}}$ is a manifold about $(\bar{x}, f(\bar{x}))$. The normal space $N_{\widehat{\mathcal{M}}}(\bar{x}, f(\bar{x}))$ is the linear span of $\widehat{G}$. The result follows. $\square$

It is worth noting before we begin the proof of Theorem 5.1, that the second half of this proof can be adapted to a direct if and only if approach. However, as the details get tiresome, we provide a calculus approach to the first half. The first half is also the more important direction for us, as when moving from sets to functions we use an indicator function approach as opposed to an epigraphical method.

Proof of Theorem 5.1. **($\Rightarrow$)** Let the function f be $\mathcal{C}^p$-partly smooth at the point $\bar{x}$ relative to the manifold $\mathcal{M}$. Then by definition f is $\mathcal{C}^p$ on $\mathcal{M}$. To prove the epigraphical formula we will construct the epigraph via basic calculus rules, then use previous results of [5]. Though the rules were proved in [5] for the case $p = 2$, the extension to general p is routine.

Let $P : \mathbf{R}^{m+1} \to \mathbf{R}^m$ be the projection mapping from $\mathbf{R}^{m+1}$ onto the first m coordinates. Next define the function h via

$$\begin{aligned} h :\ & \mathbf{R}^{m+1} \to \mathbf{R} \\ & (x, r) \mapsto f(P(x, r)) - r. \end{aligned}$$

Thus $\operatorname{epi} f = \{(x, r) : h(x, r) \leq 0\}$. Applying the chain rule ([5, Thm 4.2]), sum rule ([5, Cor 4.7]), and level set rule ([5, Thm 4.10]) shows $\operatorname{epi} f$ is $\mathcal{C}^p$-partly smooth at $(\bar{x}, f(\bar{x}))$ relative to $\{(x, f(x)) : x \in \mathcal{M}\}$.

($\Leftarrow$) Let the function f be $\mathcal{C}^p$ on the manifold $\mathcal{M}$ and the set $\operatorname{epi} f$ be partly smooth at the point $(\bar{x}, f(\bar{x}))$ relative to the manifold

$$\widehat{\mathcal{M}} = \{(x, f(x)) : x \in \mathcal{M}\}.$$

We will show the various conditions for f to be partly smooth are satisfied.

(Smoothness) By our assumptions f is $\mathcal{C}^p$ on $\mathcal{M}$.

(Regularity) By definition, f is regular at $\bar{x}$ if and only if $\operatorname{epi} f$ is regular at $(\bar{x}, f(\bar{x}))$.

(Sharpness) Let g be any $\mathcal{C}^p$ function that agrees with f on $\mathcal{M}$. For the case of $\nabla g(\bar{x}) = 0$, Lemma 5.2 reduces to

$$N_{\widehat{\mathcal{M}}}(\bar{x}, f(\bar{x})) = \{(n, -\lambda) : n \in N_{\mathcal{M}}(\bar{x}), \lambda \in \mathbf{R}\}.$$

By definition of the normal cone, and epi f's partial smoothness relative to $\widehat{\mathcal{M}}$ we know

$$N_{\widehat{\mathcal{M}}}(\bar{x}, f(\bar{x})) = \operatorname{span} \partial \delta_{\operatorname{epi} f}(\bar{x}, f(\bar{x})) = \operatorname{span} N_{\operatorname{epi} f}(\bar{x}, f(\bar{x})).$$

Combining these with [8, Thm 8.9] yields $N_{\mathcal{M}}(\bar{x}) = \operatorname{span} \partial f(\bar{x})$.

The case of $\nabla g(\bar{x}) \neq 0$ can easily be reduced to the previous case by noting the addition of a linear functional does not alter the active manifold ([5, Cor 4.7]). Thus f is partly smooth at $\bar{x}$ relative to $\mathcal{M}$ if and only if $f - \langle \nabla g, \cdot \rangle$ shares this property.

(Sub-continuity) As f is regular (above), using the semicontinuity terminology of [8, p. 152] we have ∂f is osc; therefore we need to show ∂f is isc relative to $\mathcal{M}$. That is for any subgradient $\bar{w} \in \partial f(\bar{x})$ and any sequence in $\mathcal{M}$, $x_n \to \bar{x}$, we must show that there exists subgradients $w_n \in \partial f(x_n)$ with $w_n \to \bar{w}$.

Since epi f is partly smooth at $(\bar{x}, f(\bar{x}))$ relative to $\widehat{\mathcal{M}}$, we have the indicator function $\delta_{\mathrm{epi}\, f}$ is partly smooth at $(\bar{x}, f(\bar{x}))$ relative to $\widehat{\mathcal{M}}$. Specifically $\delta_{\mathrm{epi}\, f}$ satisfies the sub-continuity condition at $(\bar{x}, f(\bar{x}))$ relative to $\widehat{\mathcal{M}}$. Since $N_{\mathrm{epi}\, f}(\bar{x}, f(\bar{x})) = \partial \delta_{\mathrm{epi}\, f}(\bar{x}, f(\bar{x}))$ we have $N_{\mathrm{epi}\, f}$ is continuous at $(\bar{x}, f(\bar{x}))$ relative to $\widehat{\mathcal{M}}$.

By [8, Thm 8.9] we have

$$\partial f(x) = \{w : (w, -1) \in N_{\mathrm{epi}\, f}(x, f(x))\}, \tag{6}$$

so $(\bar{w}, -1) \in N_{\mathrm{epi}\, f}(\bar{x}, f(\bar{x}))$. As $N_{\mathrm{epi}\, f}$ is continuous at $(\bar{x}, f(\bar{x}))$, and f is continuous relative to $\mathcal{M}$ there exists $(w_n, v_n) \in N_{\mathrm{epi}\, f}(x_n, f(x_n))$ with $(w_n, v_n) \to (\bar{w}, -1)$. Since $N_{\mathrm{epi}\, f}$ is a cone we can replace (w_n, v_n) by

$$(w_n/|v_n|, -1) \in N_{\mathrm{epi}\, f}(x_n, f(x_n)).$$

As $|v_n| \to 1$, we have $w_n/|v_n| \to \bar{w}$. Lastly notice that $w_n/|v_n| \in \partial f(x_n)$ by (6), thus showing the isc property. □

We are now ready to extend Theorem 4.3 (identification with constraints).

Theorem 5.3 (identification for functions). *Let the function f be $\mathcal{C}^p$-partly smooth ($p \geq 2$) at the point $\bar{x}$ relative to the manifold $\mathcal{M}$, and prox-regular there, with $0 \in \mathrm{rint}\, \partial f(\bar{x})$. Suppose $x_k \to \bar{x}$ and $f(x_k) \to f(\bar{x})$. Then*

$$x_k \in \mathcal{M} \quad \textit{for all large } k$$

if and only if

$$\mathrm{dist}(0, \partial f(x_k)) \to 0.$$

Proof. Let

$$S := \mathrm{epi}\, f,$$

and

$$\begin{aligned} g : \mathbf{R}^m \times \mathbf{R} &\to \mathbf{R} \\ (x, r) &\mapsto r. \end{aligned}$$

The previous two lemmas will now allow us to apply Theorem 4.3 (identification with constraints).

Let $z_k := (x_k, f(x_k))$, and $\bar{z} := (\bar{x}, f(\bar{x}))$. Notice

$$-\nabla g(\bar{z}) = (0, -1) \in \mathrm{rint}\, N_S(\bar{z}),$$

by [8, Thm 8.9]. Since Theorem 5.1 (partly smooth epigraphs) shows S is partly smooth at $\bar{z}$ relative to $\widehat{\mathcal{M}} := \{(x, f(x)) : x \in \mathcal{M}\}$, and [6, Thm 3.5] gives us prox-regularity at $\bar{z}$, we may apply Theorem 4.3 (identification with constraints). Thus we have

$$\begin{aligned} & x_k \in \mathcal{M} \text{ for all large } k \\ \Leftrightarrow\ & z_k \in \widehat{\mathcal{M}} \text{ for all large } k \\ \Leftrightarrow\ & \operatorname{dist}(-\nabla g(z_k), N_S(x_k, f(x_k))) \to 0 \\ \Leftrightarrow\ & \operatorname{dist}((0,-1), N_S(x_k, f(x_k))) \to 0 \\ \Leftrightarrow\ & \operatorname{dist}(0, \partial f(x_k)) \to 0, \end{aligned}$$

where the last equivalence follows from [8, Thm 8.9]. □

It is worth noting that we could reconstruct Theorem 4.3 (identification with constraints) for a function f and a set S by applying the above result to $f + \delta_S$. However this yields a slightly weaker result, as f must be $\mathcal{C}^2$ rather than $\mathcal{C}^1$. Notice, finally, that Corollary 4.2 (uniqueness of manifolds) has an exact analogue for functions.

6. Analysis of Critical Points

The past two sections suggest a consequence of partial smoothness and prox-regularity for the critical points of a minimization problem. Specifically suppose $\bar{x}$ is a *strict* critical point for the constrained optimization problem $\min\{f(x) : x \in S\}$, (by which we mean $-\nabla f(\bar{x}) \in \operatorname{rint} N_S(\bar{x})$). Now consider a sequence $\{x_k\}$ of critical points approaching $\bar{x}$. This sequence clearly satisfies

$$\operatorname{dist}(-\nabla f(x_k), N_S(x_k)) \to 0.$$

So if f is $\mathcal{C}^p$-partly smooth ($p \geq 2$) and prox-regular then x_k must lie on the active manifold for all large k. Using the same argument for the function case yields the following corollary to the identification results, Theorems 4.3 and 5.3.

Corollary 6.1. *Let the set S be $\mathcal{C}^2$-partly smooth at the point $\bar{x}$ relative to the manifold $\mathcal{M}$ and prox-regular there. Consider a function $f \in \mathcal{C}^1$ with $-\nabla f(\bar{x}) \in \operatorname{rint} N_S(\bar{x})$. Then locally all critical points of the constrained problem $\min\{f(x) : x \in S\}$ lie on $\mathcal{M}$.*

Similarly, suppose the function g is $\mathcal{C}^2$-partly smooth at the point $\bar{x}$ relative to the manifold $\mathcal{M}$ and prox-regular there, with $0 \in \operatorname{rint} \partial g(\bar{x})$. If g is continuous at $\bar{x}$ then locally all critical points of the unconstrained problem $\min\{g(x)\}$ lie on $\mathcal{M}$.

This result tells us that in our search for local minimizers of prox-regular partly smooth functions we can focus solely on the active manifold. As the definition of strong critical points demands strong growth along this manifold, we immediately know that no neighbouring critical points exist. Surprisingly, strong critical points of such functions are (locally) quadratic minimizers. We show this, and its strict critical point counterpart, in the next theorem.

Theorem 6.2 (sufficient optimality conditions). *Consider a function f, $\mathcal{C}^2$-partly smooth at the point $\bar{x}$ relative to the manifold $\mathcal{M}$, and prox-regular at $\bar{x}$.*

(*i*) *If $\bar{x}$ is a strict local minimizer of the restricted function $f|_{\mathcal{M}}$, and satisfies $0 \in \operatorname{rint} \partial f(\bar{x})$, then $\bar{x}$ is in fact an unconstrained strict local minimizer of f.*

(ii) *If $\bar{x}$ is a strong critical point of f relative to $\mathcal{M}$, then f grows at least quadratically near $\bar{x}$.*

Proof. Considering the two cases, we see both imply the existence of some function g such that

$$f(x) > f(\bar{x}) + g(x - \bar{x}) \tag{7}$$

for all $x \in \mathcal{M}$ near $\bar{x}$, with $x \neq \bar{x}$. Furthermore, in both cases it suffices to show

$$f(z) > f(\bar{x}) + g(z - \bar{x})$$

for all points $z \in \mathbf{R}^m$ near but not equal to $\bar{x}$. Indeed, for case (i) $g \equiv 0$; while for case (ii) $g = \varepsilon|\cdot|^2$ for some $\varepsilon > 0$.

For the sake of eventual contradiction, suppose there exists a sequence $z_k \to \bar{x}$ with

$$f(z_k) \leq f(\bar{x}) + g(z_k - \bar{x}) \text{ for all } k. \tag{8}$$

Equation (7) shows $z_k \notin \mathcal{M}$, so the projection of z_k onto $\mathcal{M}$, $x_k := P_{\mathcal{M}}(z_k)$ must differ from z_k. Thus the normal vectors

$$n_k := \frac{z_k - x_k}{|z_k - x_k|} \in N_{\mathcal{M}}(x_k) \tag{9}$$

are well defined, with $|n_k| = 1$. Dropping to a subsequence as necessary, and noting $\mathcal{M}$ is a $\mathcal{C}^2$ manifold, we may further assume $n_k \to \bar{n}$ for some normal vector $\bar{n} \in N_{\mathcal{M}}(\bar{x})$, as $k \to \infty$.

As $0 \in \operatorname{rint} \partial f(\bar{x})$ and f is partly smooth we know $N_{\mathcal{M}}(\bar{x}) = \mathbf{R}_+ \partial f(\bar{x})$. Thus there exists some $\lambda > 0$ so $\lambda \bar{n} \in \partial f(\bar{x})$. Moreover, ∂f is continuous along $\mathcal{M}$, so there exists $w_k \in \partial f(x_k)$ with $w_k \to \lambda \bar{n}$.

Since f is prox-regular at $\bar{x}$ for $\lambda \bar{n}$, there exists a constant $R > 0$ so for all large k we have

$$f(z_k) \geq f(x_k) + \langle w_k, z_k - x_k \rangle - R|z_k - x_k|^2, \tag{10}$$

and from equation (7) we also have

$$f(x_k) - g(x_k - \bar{x}) \geq f(\bar{x}). \tag{11}$$

Combining these with equation (8) we find

$$f(x_k) - g(x_k - \bar{x}) + g(z_k - \bar{x}) \geq f(x_k) + \langle w_k, z_k - x_k \rangle - R|x_k - z_k|^2.$$

which simplifies to

$$\left(\frac{g(z_k - \bar{x}) - g(x_k - \bar{x})}{|z_k - x_k|} \right) \geq \langle w_k, n_k \rangle - R|x_k - z_k|.$$

In both cases (i) and (ii), we note g is differentiable at the origin with gradient zero. Thus taking the limit as $k \to \infty$ yields

$$0 \geq \lambda \langle \bar{n}, \bar{n} \rangle = \lambda,$$

a contradiction. Thus the existence of z_k is impossible and the proof is complete. $\square$

Note this proof would work for any $\mathcal{C}^1$ function g with a critical point at the origin.

7. Necessity of Prox-Regularity

In this section we give an example of how our results fail when prox-regularity is removed.

Example 7.1 (Necessity of Prox-Regularity). Define the function

$$f(x,y) := \begin{cases} x^2 - y & (y \le 0) \\ \sqrt{x^4 + 2x^2y - y^2} & (0 < y < 2x^2) \\ 3x^2 - y & (2x^2 \le y \le 4x^2) \\ y - 5x^2 & (4x^2 < y). \end{cases}$$

In [5, Sec 7] it is shown that f is regular, locally Lipschitz, continuous everywhere, and $\mathcal{C}^2$ except on the manifolds

$$\begin{aligned} \mathcal{M}_1 &:= \{(x,y) : y = 0\} \\ \mathcal{M}_2 &:= \{(x,y) : y = 4x^2\}. \end{aligned}$$

Moreover f is partly smooth at $(0,0)$ relative to both $\mathcal{M}_1$ and $\mathcal{M}_2$. Furthermore in [5, Sec 7] it was shown that f has a strong critical point at $(0,0)$ with respect to $\mathcal{M}_1$, but $(0,0)$ is not a local minimum of f. Thus the conclusion of Theorem 6.2 (sufficient optimality conditions) fails.

Furthermore despite the fact $0 \in \operatorname{rint} \partial f(0,0)$ the conclusions of Theorem 5.3 (identification for functions) do not hold. To see this consider the sequence of points $x_k = (1/k, 0) \to (0,0)$ in $\mathcal{M}_1$. The subdifferential of f is explicitly found in [5, Sec 7], but it suffices to note that for all real x,

$$\partial f(x,0) = [(2x,-1),(2x,1)]$$

(the line segment between the points $(2x,-1)$ and $(2x,1)$). Thus we see $\operatorname{dist}(0, \partial f(x_k)) \to 0$, but $x_k \notin \mathcal{M}_2$ for all k. (We could similarly swap the roles of $\mathcal{M}_1$ and $\mathcal{M}_2$.)

The problem of course is that f is not prox-regular at $(0,0)$. This can be seen directly by noting that f is prox-regular at $(0,0)$ if and only if $\partial f + RI$ is locally monotone at $(0,0)$ for some $R > 0$ ([8, Thm 13.36], and the fact that f is continuous). Now consider points $(x,0)$ and $(x,4x^2)$ for x near 0. Referring again to [5, Sec 7] we have $\partial f(x,4x^2) = [(6x,-1),(-10x,1)]$, so

$$\begin{aligned} (2x,1) &\in \partial f(x,0) \\ (6x,-1) &\in \partial f(x,4x^2). \end{aligned}$$

For $\partial f + RI$ to be monotone we would require

$$\langle (2x - 6x, 2) + R(0,-4x^2), (0,-4x^2) \rangle \ge 0$$

which reduces to $(2 - 4Rx^2)(-4x^2) \ge 0$. This is clearly false when $x \ne 0$ is sufficiently small.

References

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

[2] J. V. Burke: On the identification of active constraints II: The nonconvex case, SIAM J Numer. Anal. 25 (1990) 1081–1102.

[3] S. Fitzpatrick, R. R. Phelps: Differentiability of the metric projection in Hilbert space, Trans. Am. Math. Soc. 270 (1982) 483–501.

[4] R. B. Holmes: Smoothness of certain metric projections of Hilbert space, Trans. Amer. Math. Soc. 184 (1973) 87–100.

[5] A. S. Lewis: Active sets, nonsmoothness and sensitivity, SIAM J. Optim. 13 (2003) 702–725.

[6] R. A. Poliquin, R. T. Rockafellar: Prox-regular functions in variational analysis: Trans. Amer. Math. Soc. 384 (1996) 1805–1838.

[7] R. A. Poliquin, R. T. Rockafellar, L. Thibault: Local differentiablity of distance functions, Trans. Amer. Math. Soc. 352 (2000) 5231–5249.

[8] R. T. Rockafellar, R. J.-B. Wets: Variational Analysis, Springer, New York (1998).

[9] A. S. Shapiro: Existence and differentiability of metric projections in Hilbert spaces, SIAM J. Optim. 4 (1994) 130–141.

[10] S. J. Wright: Identifiable surfaces in constrained optimization, SIAM J. Control Optimization 31 (1993) 1063–1079.