Algorithmic Operations Research Vol.2 (2007) 75–82

IDENTIFYING ACTIVE MANIFOLDS

W.L. Hare [a]

[a] Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

A.S. Lewis [b]

[b]School of ORIE, Cornell University, Ithaca, NY 14853, USA.

Abstract

Determining the "active manifold" for a minimization problem is a large step towards solving the problem. Many researchers have studied under what conditions certain algorithms identify active manifolds in a finite number of iterations. In this work we outline a unifying framework encompassing many earlier results on identification via the Subgradient (Gradient) Projection Method, Newton-like Methods, and the Proximal Point Algorithm. This framework, prox-regular partial smoothness, has the advantage of not requiring convexity for its conclusions, and therefore extends many of these earlier results.

Key words: Nonsmooth Optimization, Nonconvex Optimization, Active Constraint Identification, Prox-regular, Partly Smooth

1. Introduction

Our theme in this work is the idea of an "active manifold". To motivate our terminology, consider what is perhaps the canonical example of a nonsmooth function in optimization:

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

where each function $f_i : \mathbf{R}^m \to \mathbf{R}$ is twice continuously differentiable. Consider a local minimizer $\bar{x} \in \mathbf{R}^m$ satisfying the classical second-order sufficient conditions: the set of "active" gradients $\{\nabla f_i(\bar{x}) : i \in I\}$ (where $I = \{i : f_i(\bar{x}) = f(\bar{x})\}$) is linearly independent and contains zero in the relative interior of its convex hull, and f restricted to the "active manifold"

$$\mathcal{M} \;=\; \{x \in \mathbf{R}^m : f_i(x) = f_j(x) \text{ for all } i,j \in I\}$$

grows quadratically around $\bar{x}$. If we somehow knew the active manifold (or equivalently the index set I), the problem of locally minimizing the nonsmooth function f reduces to solving a set of $m+|I|+1$ smooth equations in $m+|I|+1$ variables $x \in \mathbf{R}^m,\ \lambda \in \mathbf{R}^I,\ \mu \in \mathbf{R}$:

$$\sum_{i\in I} \lambda_i \nabla f_i(x) = 0$$
$$\sum_{i\in I} \lambda_i = 1$$
$$f_i(x) = \mu \quad (i \in I).$$

Some traditional methods for constrained optimization aim precisely to estimate the active set I. Our aim in this work is to study how a variety of fundamental algorithms "identify" a suitably generalized notion of "active manifold" (the generalization we use is formally defined in Definition 2..2).

Viewed in this light, the study of active manifolds dates back to at least 1976, when it was noted that if a minimization problem of the form $\min_{x\in S} f(x)$ had particularly favorable structure, certain algorithms would terminate after a finite number of iterations, [1] [2]. More precisely, an active manifold consisting of a single point could be identified in a finite number of iterations. We say an algorithm *identifies* an active manifold if all iterates of the algorithm must lie on the active manifold after a finite number of iterations.

Early examples of active manifold identification imposed restrictive conditions (S being a box, for exam-

Email: W.L. Hare [whare@ irmacs.sfu.ca], A.S. Lewis [aslewis@orie.cornell.edu].

[1] Work of W.L. Hare is supported in part by CNPq (Brazil) grant 150234/2004-0

[2] Work of A.S. Lewis is supported in part by National Science Foundation (USA) Grant DMS-0504032.

ple). By the early 1990's it was found that, even in the absence of finite termination, for favorably structured problems the active manifold could be identified by various algorithms [3–12]. In these later works the "favorable structure" requirements for the problem were greatly relaxed and more general notions of active manifold were considered.

In this paper we consider a particular notion of active manifold, and study the identification results that follow. We do not aim to develop new algorithms that identify active manifolds, (although we address this to a small degree), instead our goal is to show how previous results on active manifold identification can be understood in a unifying framework: *prox-regular partial smoothness*. In many cases this extends previously known results by removing the convexity assumptions on the constraint sets and objective functions.

The idea of prox-regularity was first introduced by Poliquin and Rockafellar in [13]; we use an equivalent definition [14, Thm 1.3]. Simply put, a set is prox-regular at a point if the projection mapping is single-valued near that point. Both convex sets and sets defined by a finite number of smooth constraints are prox-regular [15, Ex 13.30] and [15, Ex 13.33] (with [15, Def 10.29]), so prox-regularity includes most (if not all) of the constraint sets S examined in previous approaches to active manifold identification. As such, prox-regularity is an elegant way to unify properties of convex or smoothly constrained sets.

Partial smoothness was introduced in [16] to study stability properties of active manifolds. The full definition appears in Section 2. of this paper: for now we observe, loosely speaking, that a set is partly smooth at a point along an active manifold if the normal cone at points on the manifold behaves continuously and a certain regularity condition holds.

A subsequent paper [17] showed that when prox-regularity and partial smoothness are combined, the projection mapping is not only single-valued (as ensured by prox-regularity), but is even smooth and can be used in a certain sense to identify the active manifold of the partly smooth set or function [17, Thm 3.3, Thm 4.1 & Thm 5.3]. In this paper we apply this result to various optimization algorithms and show how, as a consequence, many of the previous results on finite constraint identification can be recaptured. Specifically we show how results on *Subgradient (Gradient) Projection Methods* [1,3,4,10,11] *Newton-like Methods* [5,11], and the *Proximal Point Algorithm* [2,8,12] can all be understood in this single framework.

It should be noted that, although our framework does remove the assumptions of convexity from the finite identification process, our goal is not so much a broader framework as a more unified theory. With the exceptions of [5] and [11], previous results on active manifold identification focused on a single algorithm. Neither [5] nor [11] consider the proximal point method or the subgradient projection method.

Computational practice is not our primary concern here: most of the algorithms we analyze are conceptual rather than implementable. When outlining iterative methods, we do not discuss step size choices and stopping criteria, instead focusing on the core idea of the algorithm. In particular, we sidestep the key question of the convergence of particular algorithms, by simply assuming convergence. Nonetheless, the analysis sheds interesting light on both conceptual algorithms and practical variants, and our assumptions are no stronger than the earlier frameworks we seek to unify.

1.1. *Notation*

We follow the notation of [15] and refer there for many basic results.

We denote the distance of a point $x \in \mathbf{R}^m$ to a set $S \subset \mathbf{R}^m$ and the projection of the point onto the set by

$$\operatorname{dist}(x, S) := \inf\{|x - s| : s \in S\}$$

and

$$P_S(x) := \arg\min\{|x - s| : s \in S\}.$$

Here, $|\cdot|$ denotes the Euclidean norm.

We also 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* and the *(limiting) normal cone*, to a set S at a point $\bar{x} \in S$, defined by

$$\hat{N}_S(\bar{x}) := \hat{\partial}\delta_S(\bar{x}) \quad \text{and} \quad N_S(\bar{x}) := \partial\delta_S(\bar{x}),$$

(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, $\hat{\partial} f(\bar{x}) = \partial f(\bar{x})$.

We refer to a point $\bar{x}$ as a *critical point* for a regular function f if $0 \in \partial f(\bar{x})$. If $0 \in \operatorname{rint} \partial f(\bar{x})$ (where rint denotes the relative interior) we call $\bar{x}$ a *nondegenerate critical point*, and if $0 \in \operatorname{int}\partial f(\bar{x})$ we call $\bar{x}$ a *strict critical point*. When we consider minimizing a $\mathcal{C}^1$ function f over a constraint set S, the condition of being a *critical point*, *nondegenerate critical point*, or *strict critical point* become, respectively, $-\nabla f(\bar{x}) \in N_S(\bar{x})$, $-\nabla f(\bar{x}) \in \operatorname{rint} N_S(\bar{x})$ and $-\nabla f(\bar{x}) \in \operatorname{int} N_S(\bar{x})$.

2. Building Blocks and Tools

The primary goal of this paper is to develop a framework that encompasses many of the past results on active manifold identification. The framework we develop is based on two ideas: prox-regularity and partial smoothness. We define these concepts next. We begin with prox-regularity.

Definition 2..1 (Prox-regularity) *A closed set $S \subseteq \mathbf{R}^n$ is* prox-regular *at a point $\bar{x} \in S$ if the projection mapping P_S is single valued near $\bar{x}$.*

A lower semi-continuous function $f : \mathbf{R}^m \to \bar{\mathbf{R}}$ is prox-regular *at a point $\bar{x}$ with $f(\bar{x})$ finite if its epigraph is prox-regular at $(\bar{x}, f(\bar{x}))$.*

Like prox-regularity, we define partial smoothness in terms of sets and then form the definition for functions via epigraphs.

Definition 2..2 (Partly Smooth) *A set $S \subset \mathbf{R}^m$ is* partly smooth *at a point $\bar{x} \in S$ relative to a set $\mathcal{M} \subseteq S$ if $\mathcal{M}$ is a smooth ($\mathcal{C}^2$) manifold about $\bar{x}$ and the following properties hold:*

(i) *$S \cap \mathcal{M}$ is a neighbourhood of $\bar{x}$ in $\mathcal{M}$;*
(ii) *S is regular at all points in $\mathcal{M}$ near $\bar{x}$;*
(iii) *$N_{\mathcal{M}}(\bar{x}) \subseteq N_S(\bar{x}) - N_S(\bar{x})$; and*
(iv) *the normal cone map $N_S(\cdot)$ is continuous at $\bar{x}$ relative to $\mathcal{M}$.*

We then refer to $\mathcal{M}$ as the active manifold *(of partial smoothness).*

If a function $f : \mathbf{R}^m \to \bar{\mathbf{R}}$ is finite at $\bar{x}$, we call it partly smooth *at $\bar{x}$ relative to a set $\mathcal{M}$ if $\mathcal{M}$ is a smooth manifold about $\bar{x}$ and* epi f *is partly smooth at $(\bar{x}, f(\bar{x}))$ relative to $\widehat{\mathcal{M}} := \{(x, f(x)) : x \in \mathcal{M}\}$.*

Note that the definition of partly smooth functions implicitly forces f to be smooth on $\mathcal{M}$, as otherwise $\widehat{\mathcal{M}}$ is not a manifold. The original definition of partial smoothness based directly on the function can be found in [16], while the equivalence of this definition to the original can be found in [17, Thm 5.1].

As mentioned, all convex sets are prox-regular. The next example shows that with a standard constraint qualification, sets defined by a finite number of smooth constraints are also partly smooth.

Example 2..1 (Finitely Constrained Sets) Consider the set

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

where $g_i \in \mathcal{C}^2$.

For any point $\bar{x} \in S$ define $A_S(\bar{x}) := \{i : g_i(\bar{x}) = 0\}$. If the *active gradients* of S at $\bar{x}$, $\{\nabla g_i(\bar{x}) : i \in A(\bar{x})\}$, form a linearly independent set, then S is prox-regular at $\bar{x}$ and partly smooth there relative to the active manifold

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

([13, Cor 2.12] and [16, 6.3]). □

A second example of prox-regular partial smoothness is generated by examining strict critical points.

Example 2..2 (Strict Critical Points) If the set $S \subseteq \mathbf{R}^n$ is regular at the point $\bar{x} \in S$ and the normal cone $N_S(\bar{x})$ has interior, then S is partly smooth at $\bar{x}$ relative to the manifold $\{\bar{x}\}$.

Indeed, as $\{\bar{x}\}$ is a singleton conditions (i) and (iv) hold true. Condition (ii) is given, while condition (iii) follows from $N_{\mathcal{M}}(\bar{x}) = \mathbf{R}^n$ and $N_S(\bar{x})$ having interior. □

Examples 2..1 and 2..2 show that the class of prox-regular partly smooth sets encompasses a large collection of commonly studied constraint sets. Both of these examples are easily transferable to functions.

Our use of prox-regular partial smoothness hinges on the next theorem, a small extension of [17, Thm 4.1 & 5.3].

Theorem 1 (Active Manifold Identification) *Consider a set S that is partly smooth 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*

$$x_k \to \bar{x} \text{ and } d_k \to \bar{n} \tag{1}$$

then

$$\mathrm{dist}(d_k, N_S(x_k)) \to 0 \;\; \text{iff} \;\; x_k \in \mathcal{M} \;\; \text{for all large } k. \tag{2}$$

Similarly, suppose the function f *is partly smooth at the point* $\bar{x}$ *relative to the manifold* $\mathcal{M}$*, and prox-regular there, with* $0 \in \mathrm{rint}\,\partial f(\bar{x})$*. If* $y_k \to \bar{x}$ *and* $f(y_k) \to f(\bar{x})$*, then*

$$\mathrm{dist}(0, \partial f(y_k)) \to 0 \text{ if and only if } y_k \in \mathcal{M} \text{ for all large } k.$$

Proof: $(\Rightarrow)$ See Theorem 4.1 of [17].

$(\Leftarrow)$ Note if $x_k \in \mathcal{M}$ for all k large, then condition (iv) of partial smoothness implies $N_S(x_k) \to N_S(\bar{x})$. Applying regularity (condition (ii) of partial smoothness) and [15, Cor 4.7] we see $\mathrm{dist}(d_k, N_S(x_k)) = \mathrm{dist}(0, N_S(x_k) - d_k) \to \mathrm{dist}(0, N_S(\bar{x}) - \bar{n}) = 0$.

The case of functions is found in [17, Thm 5.3]. □

3. Algorithms

3.1. *Subgradient (Gradient) Projection Methods*

As outlined in [1, p. 174], the gradient projection algorithm first appeared in the mid 1960's through the works of Goldstein (1964) and Levitin and Polyak (1965). The idea is to use gradient information to determine a descent direction, and then apply projections to maintain feasibility. Extending the idea to nondifferentiable functions is loosely a matter of replacing gradient vectors with subgradient vectors. This leads to the algorithm defined by

$$x_{k+1} \in P_S(x_k - s_k w_k), \tag{3}$$

where $s_k > 0$ is a step size and $w_k \in \partial f(x_k)$. When the function f is $\mathcal{C}^1$ the term gradient projection is used, since the subdifferential consists of the single gradient vector.

In [1] Bertsekas shows that, with several assumptions on the directional derivatives and Hessian mapping of a function, the active manifold of a minimization problem over a "box" can be identified in a finite number of iterations [1, Prop 3]. In this case the active manifold is the face of the box on which the solution lies.

In 1987 Dunn replaced the assumptions of Bertsekas on the directional derivative and Hessian mapping with the restriction that the minimum be a "uniformly isolated zero" and a nondegenerate critical point [3, Thm 2.1]. His work shows that, under these conditions, the active manifold for a linearly constrained minimization problem could be finitely identified [3, Thm 2.1]. Calamai and Moré expanded Dunn's results into an if and only if statement on active manifold identification for linear constraint sets [4, Thm 4.1]. Like Dunn, the work of Calamai and Moré assumes nondegeneracy.

In 1993 Wright moved beyond linear constraints and considered what he called *identifiable surfaces*. Identifiable surfaces are defined to be manifolds contained in the convex constraint set with respect to which the normal cone is continuous [11, Def 2]. (Although Wright never uses the term "manifold", his definition clearly makes use of one.) Wright showed the projected gradient method would identify such surfaces in a finite number of iterations when the algorithm converged to a nondegenerate critical point.

In the case of nondifferentiable functions, active manifold identification for the *subgradient projection method* has been studied by Flåm [10]. In particular, [10, Thm 3.1 & 4.1] shows that a nondegeneracy assumption leads to finite identification of the active manifold for constraint sets defined by a finite number of smooth constraints via the subgradient projection method. Like Theorem 2 below, these results require the assumption that the subgradients used to generate the iterates of the subgradient projection algorithm converge.

Theorem 2 below encompasses the results of Bertsekas, Dunn, Calamai and Moré, Flåm and Wright, and extends them to a nonconvex setting. Like them, we assume that a collection of iterates generated via the subgradient projection method converge to nondegenerate critical point and that the step sizes used are bounded below. For gradient projection no further assumptions are required, while in the case of subgradient projection we assume (like Flåm) that the subgradient vectors also converge. Under these conditions if the constraint set is prox-regular and partly smooth then finite identification of the active manifold occurs. The sets used by Bertsekas, Dunn, Calamai and Moré, and Flåm are easily confirmed to be both prox-regular and partly smooth while the active manifolds they consider are precisely the active manifold of partial smoothness. The equivalence of identifiable surfaces to convex partly smooth sets is shown in [16, Thm 6.3].

Theorem 2 *For a function* f *and a constraint set* S*, suppose the Subgradient Projection method is used to* *create iterates $\{x_k\}$ which converge to $\bar{x}$. Further suppose that either $f \in \mathcal{C}^1$, or that these iterates are generated at each iteration by using subgradient vectors $w_k \in \partial f(x_k)$ that converge to $\bar{w}$. If the step size s_k satisfies $\liminf_k s_k > 0$, then*

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

In this case, if S is prox-regular at $\bar{x}$ and partly smooth there relative to a manifold $\mathcal{M}$ and $-\bar{w} \in \operatorname{rint} N_S(\bar{x})$, then $x_k \in \mathcal{M}$ for all large k.

Proof: First note that $f \in \mathcal{C}^1$ can be considered a specific example of the converging subgradient case (set $w_k = \nabla f(x_k) \to \nabla f(\bar{x}) = \bar{w}$). As such we only consider the case when the iterates are created by using subgradient vectors $w_k \in \partial f(x_k)$ which converge to $\bar{w}$.

Recall, if $y \in P_S(x)$ then $x - y \in N_S(y)$ [15, Ex 6.16 & Prop 6.5]. Applying this to the iteration point x_{k+1} immediately yields

$$x_k - s_k w_k - x_{k+1} \in N_S(x_{k+1}),$$

so (as N_S is a cone)

$$\frac{1}{s_k}(x_k - x_{k+1}) - w_k \in N_S(x_{k+1}).$$

Thus we have,

$$\begin{aligned}\operatorname{dist}(-w_{k+1}, N_S(x_{k+1})) &\leq |-w_{k+1} - (\tfrac{1}{s_k}(x_k - \\ &\qquad x_{k+1}) - w_k)| \\ &\leq \tfrac{1}{s_k}|x_{k+1} - x_k| + \\ &\qquad |w_k - w_{k+1}|.\end{aligned}$$

As x_k and w_k converge, and s_k is bounded below, the right hand side converges to 0. Thus $\operatorname{dist}(-w_k, N_S(x_k)) \to 0$ as desired.

Theorem 1, with $d_k = -w_k$ and $\bar{n} = -\bar{w}$, completes the proof. □

3.2. *Newton-like Methods*

We now turn our attention to *Newton-like* methods. This classical method can be outlined as follows: find

$$\tilde{x}_k \in \arg\min_x \{\langle \nabla f(x_k), x - x_k\rangle + \frac{1}{2}\langle x - x_k, H_k(x - x_k)\rangle : x \in S\},$$

then set $x_{k+1} = x_k + s_k(\tilde{x}_k - x_k)$, for some step size s_k (H_k is selected to somehow approximate the Hessian of f while simultaneously ensuring convergence).

In 1988 Burke and Moré examined Newton-like methods and the idea of "open facets", a generalization of polyhedral faces to any surface of a set that locally appears flat, proving that such methods identify the open facet on which a nondegenerate critical point lies in a finite number of iterations [5, Thm 4.1]. In 1993 Wright extended the work of Burke and Moré from open facets to identifiable surfaces.

Theorem 3 below encompasses both of these works, as all open facets are identifiable surfaces, and identifiable surfaces are equivalent to partial smoothness in the convex case [16, Thm 6.3]. It further extends the above results by replacing the condition of convexity with prox-regularity. We assume that the sequence of iterates created by the Newton-like method converges to a nondegenerate critical point, that the step sizes are eventually equal to 1, and the matrices H_k are bounded in norm. All of these conditions are also assumed in [5] and [11].

In [9], Al-Khayyal and Kyparisis showed that if the optimal point of a convex constrained optimization problem was a strict critical point then convergent algorithms could be modified to ensure finite convergence to the solution. Their technique involved creating a Newton-like point for every iteration, and proving that, if the original algorithm converges, then these points must converge finitely [9, Thm 3.1]. Although Example 2..2 strongly suggests that this result can be encompassed in Theorem 3 below, the convergence of the Newton-like points relies heavily on the local strict convexity of the constraint set and cannot be easily assured without it. A solution to this problem can be found in [21].

Theorem 3 *Consider the problem of minimizing a $\mathcal{C}^1$ function f over a constraint set S. Suppose a Newton-like method is used to generate a sequence of iterates x_k that converge to $\bar{x} \in S$. If eventually the step size s_k is always 1, and the matrices H_k are uniformly bounded, then*

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

In this case, if S is prox-regular at $\bar{x}$ and partly smooth there relative to a manifold $\mathcal{M}$ and $-\nabla f(\bar{x}) \in \operatorname{rint} N_S(\bar{x})$ then $x_k \in \mathcal{M}$ for all large k.

Proof: Without loss of generality we assume the step size is 1 for all k.

Define the function

$$q_k(x) := \nabla f(x_k)(x - x_k) + \frac{1}{2}\langle x - x_k, H_k(x - x_k)\rangle.$$

Note that the algorithm begins by finding $\tilde{x}_k$ a minimizer of q_k over S. Thus $-\nabla q_k(\tilde{x}_k) \in N_S(\tilde{x}_k)$ [15, Thm 8.15]. This yields

$$\begin{aligned} \mathrm{dist}(\,-\nabla f(\tilde{x}_k), N_S(\tilde{x}_k)) &\le |\nabla f(\tilde{x}_k) - \nabla q_k(\tilde{x}_k)| \\ &= |\nabla f(\tilde{x}_k) - \nabla f(x_k) - H_k(\tilde{x}_k - x_k)| \\ &\le |\nabla f(\tilde{x}_k) - \nabla f(x_k)| + ||H_k||\, |\tilde{x}_k - x_k|. \end{aligned}$$

Since the step size is always one, $x_{k+1} = \tilde{x}_k$, thus $\mathrm{dist}(-\nabla f(x_k), N_S(x_k))$

$$\le |\nabla f(x_{k+1}) - \nabla f(x_k)| + ||H_k||\, |x_{k+1} - x_k|. \quad (4)$$

As H_k is bounded, $f \in \mathcal{C}^1$, and x_k converges we must have the right hand side converge to zero. Thus $\mathrm{dist}(-\nabla f(x_k), N_S(x_k)) \to 0$.

Applying Theorem 1 with $d_k = -\nabla f(x_k)$ and $\bar{n} = -\nabla f(\bar{x})$ completes the proof. □

3.3. *Proximal Point Methods*

To study the proximal point algorithm we define the *proximal* (or *Moreau*) *envelope* and the *proximal point mapping*:

$$\begin{aligned} e_R(x) &:= \min_y\{f(y) + \tfrac{R}{2}|y-x|^2\}, \\ P_R(x) &:= \mathrm{argmin}_y\{f(y) + \tfrac{R}{2}|y-x|^2\}. \end{aligned}$$

A function is called *prox-bounded* if there exists some point and scalar for which the proximal envelope is finite.

The proximal point method was first introduced by Martinet in [18, Sec 4]. The method selects its iterates by solving

$$x_{k+1} \in P_R(x_k).$$

In 1976 Rockafellar showed some of the first finite convergence results for the proximal point algorithm [2]. Specifically, strict critical points of a convex function can be identified via a finite number of iterations of the proximal point algorithm. He further showed that in the case when the function is polyhedral, the proximal point algorithm identifies the active face of the polyhedron of the minimization problem regardless of the behaviour of the subdifferential [2, Prop 8].

Ferris furthered this work by considering convex functions which grew sharply in directions away from the set of minima (see [8, Def 1]). For such functions the proximal point method was shown to converge in a finite number of iterations [8, Thm 6]. Ferris's ideas are captured in the sharpness conditions of partial smoothness.

More recently, work by Mifflin and Sagastizábal on "fast tracks" has shown that the proximal point method identifies fast tracks for a convex function in a finite number of iterations [12]. In [19] it is shown that convex functions containing a fast track are always prox-regular and partly smooth, with the active manifold of partial smoothness being equivalent to the fast track.

Theorem 4 below unifies the identification aspects of the works of Rockafellar, Ferris, and Mifflin and Sagastizábal, and extends them to a nonconvex setting. Unlike Rockafellar's work, Theorem 4 makes the assumption that the proximal point algorithm converges. In the nonconvex case this is necessary, as without this assumption the iterates may converge to an alternate critical point. The works of Ferris and Mifflin and Sagastizábal, show that when x_k is sufficiently close to the minimal value, the next iterate identifies the active manifold. This is equivalent to the convergence assumption we use. Like Theorem 4, Rockafellar, Ferris, and Mifflin and Sagastizábal all use nondegenerate critical points; furthermore, all the functions they consider are convex (therefore prox-regular) and partly smooth.

Theorem 4 *Suppose the function f is prox-bounded, and prox-regular at the point $\bar{x}$. Suppose the proximal point algorithm is used to generate a sequence of iterates x_k that converge to $\bar{x}$. If $R > 0$ is sufficiently large then*

$$f(x_k) \to f(\bar{x}) \;\; \textit{and} \;\; \mathrm{dist}(0, \partial f(x_k)) \to 0. \quad (5)$$

If f is furthermore partly smooth at $\bar{x}$ relative to a manifold $\mathcal{M}$, and $0 \in \mathrm{rint}\,\partial f(\bar{x})$, then for any point z sufficiently close to $\bar{x}$ one finds $P_R(z) \in \mathcal{M}$. Thus $x_k \in \mathcal{M}$ for all large k.

It is worth noting that in the case of a convex function the phrase 'R sufficiently large' reduces to '$R > 0$'.

Before proving Theorem 4 we require a lemma relating proximal points and prox-regularity.

Lemma 5 [20, Thm 2.3] *Suppose the function f is prox-bounded and prox-regular at $\bar{x}$ and that $0 \in \partial f(\bar{x})$. Then for R sufficiently large*

(i) *the proximal envelope e_R is $\mathcal{C}^1$ near $\bar{x}$ with $e_R(\bar{x}) = f(\bar{x})$ and,*
(ii) *the proximal point mapping P_R is single valued and Lipschitz continuous near $\bar{x}$, and satisfies $P_R(\bar{x}) = \{\bar{x}\}$.*

As in Theorem 4, in the case of a convex function 'R sufficiently large' reduces to '$R > 0$'. We now proceed with the proof of Theorem 4.

Proof of Theorem 4: We begin by selecting R large enough that Lemma 5 may be applied.

Consider any sequence of points z_k converging to $\bar{x}$. (Since Lemma 5 applies we assume the related proximal points are unique, and set $y_k = P_R(z_k)$.)

By Lemma 5 (ii) we know $y_k \to \bar{x}$, thus $\frac{R}{2}|y_k - z_k|^2 \to 0$. Combining $e_R(z_k) = f(y_k) + \frac{R}{2}|y_k - z_k|^2$ with Lemma 5 (i) then shows that $f(y_k) \to f(\bar{x})$.

Next notice, as $y_k \in \operatorname{argmin}\{f(y) + \frac{R}{2}|y - z_k|^2\}$, we must have $0 \in \partial(f(\cdot) + \frac{R}{2}|\cdot - z_k|^2)(y_k)$ for each k. This simplifies to

$$0 \in \partial f(y_k) + R(y_k - z_k).$$

As $R(y_k - z_k)$ converges to 0 we have $\operatorname{dist}(0, \partial f(y_k)) \to 0$.

Equation (5) is the special case when $z_k = x_k$. The rest of the result follows from Theorem 1. □

4. Necessity of Nondegeneracy

Section 3. unifies many of the previous results on active manifold identification under the framework of prox-regular partial smoothness. To do this it repeatedly makes use of Theorem 1, and therefore nondegenerate critical points. In this section we provide three simple examples showing the necessity of nondegeneracy in identifying active manifolds.

Example 4..1 (Nondegeneracy and GradientProjection)
Consider the problem,

$$\min\{y : |(x, y)| \leq 1, x \geq 0\}.$$

The constraint set $S := \{|(x, y)| \leq 1, x \geq 0\}$ is convex and partly smooth at the point $(\bar{x}, \bar{y}) := (0, -1)$ relative to the active manifold $\mathcal{M} := \{(0, -1)\}$. Although this point is the unique minimizer for $f(x, y) = y$ over S, it is not a nondegenerate critical point.

Suppose we approached the problem via the gradient projection method, and that iterate (x_k, y_k) is located on the set $\widehat{S} := \{(x, y) : |(x, y)| = 1, y < 0, x > 0\} \subseteq S$. Then the next iteration yields,

$$(x_{k+1}, y_{k+1}) = P_S((x_k, y_k) - s_k(0, 1))$$

$$= \frac{(x_k, y_k - s_k)}{|(x_k, y_k - s_k)|} \in \widehat{S}.$$

Although this converges in limit, it will never identify the active manifold of the problem, $\mathcal{M}$, as x_{k+1} is never equal to 0. □

Example 4..2 (Nondegeneracy and Newton) Consider the problem

$$\min\{x^2 + x^3 : x \geq 0\}.$$

The constraint set $S := \{x \geq 0\}$ is convex and partly smooth at the point $\bar{x} := 0$ relative to the active manifold $\mathcal{M} := \{0\}$. Although this point is the unique minimizer for $f(x) = x$ over S, it is not a nondegenerate critical point.

Suppose we approached the problem via Newton's method, using the exact Hessian and a constant step size of 1. Then, given any current iterate x_k, we obtain

$$\begin{aligned} x_{k+1} = \tilde{x}_k &= \operatorname{argmin}\{\langle \nabla f(x_k), x - x_k\rangle + \\ &\quad \tfrac{1}{2}\langle x - x_k, \nabla^2 f(x_k)(x - x_k)\rangle : x \geq 0\} \\ &= \tfrac{-2x_k - 3x_k^2}{2 + 6x_k} + x_k = \tfrac{3x_k^2}{2 + 6x_k}. \end{aligned}$$

Thus x_k converges to $\bar{x}$, but never identifies the active manifold of the problem, $\mathcal{M}$. □

Example 4..3 (Nondegeneracy and Proximal Points)
Consider the problem

$$\min f(x) \quad \text{where} \quad f(x) := \begin{cases} -x & x \leq 0 \\ x^3 & x \geq 0 \end{cases}$$

Then f is convex and partly smooth at the point $\bar{x} = 0$ relative to the manifold $\mathcal{M} = \{0\}$. Although this point is the unique minimizer of f, it is not a nondegenerate critical point.

Suppose we approached the problem via the proximal point algorithm. Then given an iterate $x_k \in (0, 1)$ the next iterate $x_{k+1} = P_R f(x_k)$ is equal to

$$\begin{aligned} x_{k+1} &= \operatorname{argmin}_{y \geq 0}\{y^3 + \tfrac{R}{2}(y - x_k)^2\} \\ &= \tfrac{1}{6}(\sqrt{R^2 + 12Rx_k} - R). \end{aligned}$$

As R and x_k are both strictly positive, x_{k+1} is strictly positive. Therefore, regardless of how close x_k is to the point $\bar{x}$, the next iterate never identifies the active manifold of the problem $\mathcal{M}$. □

5. Concluding Remarks

In this paper we developed a unifying framework encompassing many earlier results on identification via the Subgradient (Gradient) Projection Method, Newton-like Methods, and the Proximal Point Algorithm. This framework, prox-regular partial smoothness, has the advantage that it does not demand convexity of the constraint set, nor of the objective function, and therefore extends the results of [1–5,7,8,10–12] to a nonconvex setting. Finally, we provide three key examples which demonstrate the need for nondegenerate critical points for the finite identification of active manifolds.

References

[1] Bertsekas, D. P., *On the Goldstein-Levitin-Polyak gradient projection method,* IEEE Transactions Automatic Control, Vol. AC-21, No. 2, pp. 174–184, 1976.

[2] Rockafellar, R. T., *Monotone operators and the proximal point algorithm,* SIAM Journal on Control and Optimization, Vol. 14, No. 5, pp. 877–898, 1976.

[3] Dunn, J. C., *On the convergence of projected gradient processes to singular critical points,* Journal of Optimization Theory and Applications, Vol. 55, No. 2, pp. 203–216, 1987.

[4] Calamai, P. H., and Moré, J. J., *Projected gradient methods for linearly constrained problems,* Mathematical Programming, Vol. 39, No. 1, pp. 93–116, 1987.

[5] Burke, J. V., and Moré, J. J., *On the identification of active constraints,* SIAM Journal on Numerical Analysis, Vol. 25, No. 5, pp. 1197–1211, 1988.

[6] Lemaire, B., *The proximal algorithm,* New Methods in Optimization and Their Industrial Uses (Pau/Paris, 1987), Vol. 87 of Internat. Schriftenreihe Numer. Math., pp. 73–87, Birkhäuser, Basel, 1989.

[7] Burke, J., *On the identification of active constraints. II. The nonconvex case,* SIAM Journal on Numerical Analysis, Vol. 27, No. 4, pp. 1081–1103, 1990.

[8] Ferris, M. C., *Finite termination of the proximal point algorithm,* Mathematical Programming, Vol. 5, No. 3 (Ser. A), pp. 359–366, 1991.

Received 19 March, 2007; revised 24 May 2007; accepted 29 May 2007

[9] Al-khayyal, F., and kyparisis, J., *Finite convergence of algorithms for nonlinear programs and variational inequalities,* Journal of Optimization Theory and Applications, Vol. 70, No. 2, pp. 319–332, 1991.

[10] Flåm, S. D., *On finite convergence and constraint identification of subgradient projection methods,* Mathematical Programming, Vol. 57, No. 3 (Ser. A), pp. 427–437, 1992.

[11] Wright, S. J., *Identifiable surfaces in constrained optimization,* SIAM Journal on Control and Optimization, Vol. 31, No. 4, pp. 1063–1079, 1993.

[12] Mifflin, R., and sagastizábal *Proximal points are on the fast track,* Journal of Convex Analysis, Vol. 9, No. 2, pp. 563–579, 2002. Special issue on optimization (Montpellier, 2000).

[13] Poliquin, R. A., and Rockafellar, R. T., *Prox-regular functions in variational analysis,* Transactions of the American Mathematical Society, Vol. 348, No. 5, pp. 1805–1838, 1996.

[14] Poliquin, R. A., Rockafellar, R. T., and THIBAULT, L., *Local differentiability of distance functions,* Transactions of the American Mathematical Society, Vol. 352, No. 11, pp. 5231–5249, 2000.

[15] Rockafellar, R. T., and Wets, R. J. B., *Variational Analysis,* Springer-Verlag, Berlin, Germany, 1998.

[16] Lewis, A. S., *Active sets, nonsmoothness, and sensitivity,* SIAM Journal on Optimization, Vol. 13, No. 3, pp. 702–725, 2003.

[17] Hare, W. L., and Lewis, A. S., *Identifying active constraints via partial smoothness and prox-regularity,* Journal of Convex Analysis, Vol. 11, pp. 251-266, 2004.

[18] Martinet, B., *Régularisation d'inéquations variationnelles par approximations successives,* Rev. Française Informat. Recherche Opérationnelle, Vol. 4 (Ser. R-3), pp. 154–158, 1970.

[19] Hare, W. L., *Functions and sets of smooth substructure: relationships and examples,* Journal of Computational Optimization and Applications, Vol 33, No. 2/3, to appear, 2006.

[20] Poliquin, R. A., and Rockafellar, R. T., *Generalized hessian properties of regularized nonsmooth functions,* SIAM Journal on Optimization, Vol. 6, No. 4, pp. 1121–1137, 1996.

[21] Hare, W. L., *A proximal method for identifying active manifolds,* submitted to *Computational Optimization and Applications.*