Chapter 1
Identifying Active Manifolds in Regularization Problems

W. L. Hare

Abstract In 2009, Tseng and Yun [7], showed that the regularization problem of minimizing $f(x)+\|x\|_1$, where f is a $\mathscr{C}^2$ function and $\|x\|_1$ is the $\mathscr{L}_1$ norm of x, can be approached by minimizing the sum of a quadratic approximation of f and the $\mathscr{L}_1$ norm. We consider a generalization of this problem, in which the $\mathscr{L}_1$ norm is replaced by a more general nonsmooth function that contains an underlying smooth substructure. In particular, we consider the problem

$$\min_x \{f(x)+P(x)\}, \tag{1.1}$$

where f is $\mathscr{C}^2$ and P is prox-regular and partly smooth with respect to an active manifold $\mathscr{M}$ (the $\mathscr{L}_1$ norm satisfies these conditions.) We reexamine Tseng and Yun's algorithm in terms of active set identification, showing that their method will correctly identify the active manifold in a finite number of iterations. That is, after are finite number of iterations, all future iterates x^k will satisfy $x^k \in \mathscr{M}$. Furthermore, we confirm a conjecture of Tseng that, regardless of what technique is used to solve the original problem, the subproblem $p^k = \operatorname{argmin}_p\{\langle \nabla f(x^k), p\rangle + \frac{r}{2}|x^k - p|^2 + P(p)\}$ will correctly identify the active manifold in a finite number of iterations.

Key Words: Nonconvex Optimization, Active Constraint Identification, Prox-regular, Partly Smooth
AMS 2010 Subject Classification : Primary : 49K40, 65K05. Secondary : 52A30, 52A41, 90C53.

W. L. Hare
Dept. of Math, Stats, & Physics, UBC O, Kelowna, BC, V1V 1V7, Canada e-mail: warren.hare@ubc.ca

1.1 Introduction

In this work we consider the problem of minimizing the sum of two functions,

$$\min_x \{f(x) + P(x)\}, \tag{1.2}$$

where f is $\mathscr{C}^2$ and P is nonsmooth, but contains some underlying smooth substructure. One common example arises in $\mathscr{L}_1$ regularization problems,

$$\min_x \{f(x) + c\|x\|_1\}, \tag{1.3}$$

where $\|x\|_1$ is the $\mathscr{L}_1$ norm of x. A brief survey of $\mathscr{L}_1$ regularization problems and some applications can be found in the introduction to [7].

Recent work by Tseng and Yun [7] has suggested that one practical method to approach such a problem is to solve a sequence of quadratic approximation problems,

$$x^{k+1} = \min_x \{\langle \nabla f(x^k), x - x^k\rangle + (x - x^k)' H^k (x - x^k) + P(p)\}, \tag{1.4}$$

where H^k is an approximation to the Hessian of f. When P is well structured, such as the case when $P = c\|x\|_1$, this problem is easily solved ; potentially having a closed form solution. Convergence theory and numerical testing can be found in [7].

In this work we consider Tseng and Yun's method in terms of active manifold identification. In particular, we consider the case when P is *partly smooth* with respect to some manifold $\mathscr{M}$ containing $\bar{x} \in \operatorname{argmin}_x\{f(x) + P(x)\}$. We show that, under some conditions, all but a finite number of iterates will lie on the active manifold. In terms of $\mathscr{L}_1$ regularization, this means that if $\bar{x} \in \operatorname{argmin}_x\{f(x) + c\|x\|_1\}$ then for k sufficiently large, $x_i^k = 0$ if and only if $\bar{x}_i = 0$.

Tseng and Yun's method separates the original problem (1.2) into two pieces, the smooth portion and the nonsmooth portion, that are treated distinctly different. Similar ideas often occur in constrained optimization, where objective functions and constrain sets are treated differently. Tseng and Yun's separation technique could be viewed as an analog to such techniques for constrained optimization by rephrasing the problem as

$$\min_x \{f(x) + r : 0 \geq P(x) - r\}. \tag{1.5}$$

Recently, Hare [1] showed that the active manifold of a constraint set could be identified by examining a proximal subproblem. Along similar lines, Tseng conjectured that the subproblem

$$p^k = \operatorname{argmin}_p \{\langle \nabla f(x^k), p\rangle + \frac{r}{2}|x^k - p|^2 + P(p)\} \tag{1.6}$$

will correctly identifying the active constraints of the problem (1.2) in a finite number of iterations, regardless of the method used to solve the problem (1.2). Theorem 1.3.3 provides an affirmative proof for this conjecture.

The remainder of this work is organized as follows. In Section 1.2 we provide the background required to understand this work. In particular, Section 1.2 includes the definitions of *prox-regular* and *partly smooth functions*. In Section 1.3 we examine the active manifold identification properties of iterates generated by equations (1.4) and (1.6). A brief conclusion, consisting primarily of an e-mail from Tseng which prompted this work, appears in Section 1.4.

1.2 Definitions and Notations

In order to keep this work brief, we provide only the definitions and background necessary to understand its two main results (Theorems 1.3.1 and 1.3.3). In general we follow the notation of [6], with one notable exception.

We shall define the *Moreau envelope*, and its corresponding (potentially empty) *proximal point mapping*, of a function f at a point x with respect to a parameter r as

$$e_r f(x) := \inf\{f(y) + \frac{r}{2}|y-x|^2\} \tag{1.7}$$

$$\text{prox}_r f(x) := \text{argmin}\,\{f(y) + \frac{r}{2}|y-x|^2\} \tag{1.8}$$

where $|\cdot|$ is the usual Euclidean norm. (In [6, Def 1.23] the parameter r is place in the denominator of the quadratic penalty term '$\frac{1}{2r}$'.) We say that a function is *prox-bounded* if there exists some $r > 0$ and point x such that $e_r f(x)$ is finite. If a function is prox-bounded, then (for r sufficiently large) $e_r f$ is finite-value everywhere [6, ex 1.24].

A useful lemma, from [3], regarding proximal points is reproduced next.

Lemma 1.2.1 (tilting proximal points) *Let f be a proper lsc prox-bounded function and a v be a vector. Then for r sufficiently large and any x*

$$\min\{f(y) - \langle v, y\rangle + \frac{r}{2}|y-x|^2\} + \langle x, v\rangle + \frac{1}{2r}|v|^2 = e_r f(x + \frac{1}{r}v), \tag{1.9}$$

$$\text{argmin}\,\{f(y) - \langle v, y\rangle + \frac{r}{2}|y-x|^2\} = \text{prox}_r f(x + \frac{1}{r}v). \tag{1.10}$$

Proof. Lemma 2.2 of [3] can be rephrased to this form.

Following the notation of [6] we define the *regular normal cone* to a set S at a point $\bar{x} \in S$ as

$$\hat{N}_S(\bar{x}) = \{v : \langle v, x - \bar{x}\rangle \leq o(|x - \bar{x}|)\}, \tag{1.11}$$

and the *limiting normal cone* as

$$N_S(\bar{x}) = \limsup_{x \to \bar{x}} \hat{N}_S(x). \tag{1.12}$$

A set is *regular* at $\bar{x}$ if these two cones coincide.

The corresponding to functions we define the *regular subdifferential* of a function f at a point $\bar{x}$ where f is finite as

$$\hat{\partial} f(\bar{x}) := \{v \in \mathbf{R}^m : f(x) \geq f(\bar{x}) + \langle v, x - \bar{x}\rangle + o(|x - \bar{x}|)\} \tag{1.13}$$

and the *subdifferential*,

$$\partial f(\bar{x}) := \limsup_{x \to \bar{x},\ f(x) \to f(\bar{x})} \hat{\partial} f(x). \tag{1.14}$$

A function is *regular* at $\bar{x}$ if its epi-graph is regular at $(\bar{x}, f(\bar{x}))$. If f and g are regular at x then

$$\partial f(x) = \hat{\partial} f(x) \ \text{ and } \ \partial(f(x) + g(x)) = \partial f(x) + \partial g(x) \tag{1.15}$$

(see [6, Cor 8.11] and [6, Cor 10.9] respectively).

Prox-regularity will provide us with a framework for working with nonconvex functions.

Definition 1.2.2 (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*

$$f(x') \geq f(x) + \langle v, x' - x\rangle - \frac{\rho}{2}|x' - x|^2 \tag{1.16}$$

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 $v \in \partial f(\bar{x})$.*

It is clear that all convex functions are prox-regular. Moreover, any function f such that $f + \frac{r}{2}|\cdot|^2$ is convex is prox-regular. Prox-regularity further includes the broad class of functions known as *strongly amenable* [5, Def 2.4 & Prop 2.5] and *lower-$\mathscr{C}^2$ functions* [5, Ex 2.7]. In particular, and function that is composed of the maximum of a finite number of smooth functions is prox-regular [5, Ex 2.9].

Our framework for active manifolds will be partly smooth functions.

Definition 1.2.3 (partly smooth) *A function f is* partly smooth *at a point $\bar{x}$ relative to a set $\mathscr{M}$ containing $\bar{x}$ if $\mathscr{M}$ is a $\mathscr{C}^p$ manifold about $\bar{x}$ and :*

i. **(smoothness)** *$f|_{\mathscr{M}}$ is a $\mathscr{C}^2$ function near $\bar{x}$;*
ii. **(regularity)** *f is regular at all points $x \in \mathscr{M}$ near $\bar{x}$, with $\partial f(x) \neq \emptyset$;*
iii. **(sharpness)** *the affine span of $\partial f(\bar{x})$ is is a translate of $N_{\mathscr{M}}(\bar{x})$;*
iv. **(sub-continuity)** *∂f restricted to $\mathscr{M}$ is continuous at $\bar{x}$.*

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

First developed in [4], the idea of partly smooth functions provides a unifying framework for optimization research into functions where the minimum lies upon a active manifold. Most notably, the idea of partly smooth functions captures functions that are composed of the maximum of a finite number of smooth functions, provided a standard constraint qualification holds.

Example 1.2.4 (finite max functions) *Let*

$$P(x) := \max\{g_i(x) : i = 1, 2, \ldots, n\} \tag{1.17}$$

where g_i are $\mathscr{C}^2$ functions around the point $\bar{x}$. Then P is prox-regular at $\bar{x}$ [5, Ex 2.9].

Define the active set *for P at a point x by*

$$A_P(x) := \{i : g_i(x) = g(x)\}. \tag{1.18}$$

If that the set of all active gradients of P, $\{\nabla g_i(\bar{x}) : i \in A_P(\bar{x})\}$, is linearly independent, then [4, Cor 4.8] shows that P is partly smooth at $\bar{x}$ relative to the manifold

$$\mathscr{M} := \{x : A_P(x) = A_P(\bar{x})\}. \tag{1.19}$$

Relevant to this work, it is easy to confirm that the $\mathscr{L}_1$ norm is partly smooth and prox-regular.

Example 1.2.5 ($\mathscr{L}_1$ norm) *The $\mathscr{L}_1$ norm is convex and therefore prox-regular at any point. Also, the $\mathscr{L}_1$ norm is partly smooth with respect to the manifold*

$$\mathscr{M} = \{x : A_1(x) = A_1(\bar{x})\}, \tag{1.20}$$

where A_1 is the active set of the $\mathscr{L}_1$ norm : $A_1(x) := \{i : |x_i| = 0\}$ [4, p. 714].

With regards to this work, the strength of partly smooth functions lies in the ability of algorithms to identify their active manifolds. That is, under some conditions, many algorithms have the property that after a finite number of iterations all future iterates will be contained in the active manifold. The next theorem, reproduced from [2, Thm 5.3], captures this idea mathematically.

Theorem 1.2.6 (identifying active manifold) *Let the function F be prox-regular at $\bar{x}$ and partly smooth there relative to the manifold $\mathscr{M}$ with $0 \in \operatorname{rint} \partial f(\bar{x})$. Suppose $x^k \to \bar{x}$ and $F(x^k) \to F(\bar{x})$. Then*

$$x^k \in \mathscr{M} \text{ for all large } k \tag{1.21}$$

if and only if

$$\operatorname{dist}(0, \partial f(x^k)) \to 0. \tag{1.22}$$

1.3 Active Manifold Identification

1.3.1 Algorithmic Manifold Identification

Recall, this work concerns itself with the optimization problem

$$\min_x \{f(x) + P(x)\} \tag{1.23}$$

where $f \in \mathscr{C}^2$ and P is prox-regular and partly smooth. We first consider the algorithm proposed in [7], and show that, provided the algorithm converges and the approximate Hessians are bounded, it identifies the active manifold of P in a finite number of iterations. For detailed analysis on when the algorithm converges see [7]. Note that Assumption 1 of [7] implies the approximate Hessians are bounded.

Theorem 1.3.1 (identification via Tseng & Yun's algorithm) *Let $f \in \mathscr{C}^2$ and P be regular. Suppose $\bar{x} \in \operatorname{argmin}\{f(x) + P(x)\}$ and P is prox-regular at $\bar{x}$ and partly smooth there with respect to the manifold $\mathscr{M}$. Suppose iterates x^k are generated by solving the subproblem*

$$x^{k+1} \in \operatorname{argmin}_x \{\langle \nabla f(x^k), x - x^k \rangle + (x - x^k)' H^k (x - x^k) + P(x)\}, \tag{1.24}$$

where H^k is a sequence of positive definite matrices with $\|H^k\|$ bounded. Suppose iterates x^k converge to $\bar{x}$.

If $-\nabla f(\bar{x}) \in \operatorname{rint} \partial P(\bar{x})$ then $x^k \in \mathscr{M}$ for all k sufficiently large.

Proof. Notice that, as $f \in \mathscr{C}^2$, $F(x) = f(x) + P(x)$ is prox-regular at $\bar{x}$ and partly smooth there with respect to $\mathscr{M}$ [4, Cor 4.6]. Since P is regular $\bar{x}$ we have that $\partial F(x) = \nabla f(x) + \partial P(x)$, and $0 \in \operatorname{rint} \partial F(\bar{x})$. In order to apply Theorem 1.2.6 to F we must show $f(x^k) + P(x^k) \to f(\bar{x}) + P(\bar{x})$ and $\operatorname{dist}(0, \partial(f(x^k) + P(x^k)) \to 0$ (notice $x^k \to \bar{x}$ by assumption).

To see that $f(x^k) + P(x^k) \to f(\bar{x}) + P(\bar{x})$, first notice that for all k

$$f(\bar{x}) + P(\bar{x}) = \min_x \{f(x) + P(x)\} \leq f(x^k) + P(x^k), \tag{1.25}$$

so

$$f(\bar{x}) + P(\bar{x}) \leq \liminf_{k \to \infty} f(x^k) + P(x^k). \tag{1.26}$$

Next, notice that, as $x^{k+1} \in \operatorname{argmin}_x \{\langle \nabla f(x^k), x - x^k \rangle + (x - x^k)' H^k (x - x^k) + P(x)\}$, we have

$$\begin{aligned} &\langle \nabla f(x^k), x^{k+1} - x^k \rangle + (x^{k+1} - x^k)' H^k (x^{k+1} - x^k) + P(x^{k+1}) \\ \leq\ &\langle \nabla f(x^k), \bar{x} - x^k \rangle + (\bar{x} - x^k)' H^k (\bar{x} - x^k) + P(\bar{x}). \end{aligned} \tag{1.27}$$

Applying a limit in k, while noting $x^k \to \bar{x}$ and $\|H^k\|$ bounded, yields

$$\limsup_{k \to \infty} P(x^{k+1}) \leq P(\bar{x}). \tag{1.28}$$

As $f \in \mathscr{C}^2$ this implies (with equation (1.26)) that

$$\limsup_{k\to\infty} f(x^k)+P(x^k) \le f(\bar{x})+P(\bar{x}) \le \liminf_{k\to\infty} f(x^k)+P(x^k). \tag{1.29}$$

Which proves $f(x^k)+P(x^k) \to f(\bar{x})+P(\bar{x})$.

To see $\mathrm{dist}(0,\partial(f(x^k)+P(x^k)) \to 0$, notice that $x^{k+1} \in \mathrm{argmin}_x\{\langle \nabla f(x^k), x-x^k\rangle + (x-x^k)'H^k(x-x^k)+P(x)\}$ implies

$$\begin{aligned} 0 &\in \partial(\langle \nabla f(x^k), x-x^k\rangle + (x-x^k)'H^k(x-x^k)+P(x))(x^{k+1}) \\ 0 &\in \nabla f(x^k) + H^k(x^{k+1}-x^k) + \partial P(x^{k+1}) \\ -H^k(x^{k+1}-x^k) &\in \nabla f(x^k) + \partial P(x^{k+1}) \end{aligned} \tag{1.30}$$

Therefore

$$\begin{aligned} \mathrm{dist}(0,\partial(f(x^k)+P(x^k))) &= \mathrm{dist}(0,\nabla f(x^k)+P(x^k)) \\ &\le |-H^k(x^{k+1}-x^k)| \\ &\le \|H^k\||x^{k+1}-x^k|. \end{aligned} \tag{1.31}$$

Since $\|H^k\|$ is bounded and $|x^{k+1}-x^k| \to 0$, we have $\mathrm{dist}(0,\partial f(x^k)) \to 0$.

The result now follow from Theorem 1.2.6.

1.3.2 Manifold Identification via a Proximal Subproblem

In the paper [1], it was shown that the active manifold of a constraint set could be identified by inducing a proximal style subproblem. Tseng conjectured that a similar technique might work for problem (1.23). In particular, it was proposed that the subproblem

$$p^k = \mathrm{argmin}_p\{\langle \nabla f(x^k), p\rangle + \frac{r}{2}|x^k-p|^2 + P(p)\} \tag{1.32}$$

would identify the active manifold of the function P in a finite number of iterations. We next confirm this conjecture, by proving that if x^k converges to a critical point of problem (1.23) and P is a prox-regular partly smooth function, then excluding a finite number of iterations all points p^k will lie on the active manifold of P.

The proof will hinge on the following lemma.

Lemma 1.3.2 *Let $f \in \mathscr{C}^2$ and P be a regular prox-bounded function. Suppose $\bar{x} \in$ argmin$\{f(x)+P(x)\}$ and P is prox-regular at the point $\bar{x} - \frac{1}{r}\nabla f(\bar{x})$. Suppose the sequence of points x^k converges to $\bar{x}$, and consider a sequence of points*

$$p^k \in \mathrm{argmin}_p\{\langle \nabla f(x^k), p\rangle + \frac{r}{2}|p-x^k|^2 + P(p)\}. \tag{1.33}$$

If $r > 0$ is sufficiently large then the points p^k satisfy

i) $p^k \to \bar{x}$

ii) $P(p^k) + \langle \nabla f(\bar{x}), p^k \rangle \to P(\bar{x}) + \langle \nabla f(\bar{x}), \bar{x} \rangle$

iii) $\text{dist}(0, \partial(P + \langle \nabla f(\bar{x}), \cdot \rangle)(p^k) \to 0$

Proof. To ease discussion, define

$$\tilde{P}(x) = P(x) + \langle \nabla f(\bar{x}), x \rangle. \tag{1.34}$$

Since $f \in \mathscr{C}^2$ and P is regular, $\partial \tilde{P}(\bar{x}) = \nabla f(\bar{x}) + \partial P(\bar{x}) = \partial(f+P)(\bar{x})$. In particular, $0 \in \partial \tilde{P}(\bar{x})$, as $\bar{x} \in \text{argmin}\{f(x) + P(x)\}$.

Part i) $p^k \to \bar{x}$

Applying Lemma 1.2.1 to Equation (1.33) we see that

$$p^k = \text{prox}_r P(x^k - \frac{1}{r} \nabla f(x^k)). \tag{1.35}$$

Since $x^k \to \bar{x}$ and $f \in \mathscr{C}^2$ we have $x^k - \frac{1}{r}\nabla f(x^k) \to \bar{x} - \frac{1}{r}\nabla f(\bar{x})$. Since P is prox-regular at $\bar{x} - \frac{1}{r}\nabla f(\bar{x})$, for r sufficiently large the proximal point mapping is single-valued Lipschitz continuous in some neighbourhood of $\bar{x} - \frac{1}{r}\nabla f(\bar{x})$ [3, Thm 2.4]. Therefore we have p^k converges to some $\bar{p}$ with

$$\bar{p} = \text{prox}_r P(\bar{x} - \frac{1}{r}\nabla f(\bar{x})). \tag{1.36}$$

By Lemma 1.2.1, we see that

$$\begin{aligned} \text{prox}_r P(\bar{x} - \tfrac{1}{r}\nabla f(\bar{x})) &= \text{argmin}\{\langle \nabla f(\bar{x}), p \rangle + \tfrac{r}{2}|p - \bar{x}|^2 + P(p)\} \\ &= \text{argmin}\{\tilde{P}(p) + \tfrac{r}{2}|p - \bar{x}|^2\} \\ &= \text{prox}_r \tilde{P}(\bar{x}). \end{aligned} \tag{1.37}$$

Since $0 \in \partial \tilde{P}(\bar{x})$ we have that $\bar{x} \in \text{prox}_r \tilde{P}(\bar{x}) = \bar{p}$. Thus $\bar{p} = \bar{x}$, so $p^k \to \bar{x}$.

Part ii) $\tilde{P}(p^k) \to \tilde{P}(\bar{x})$

Since P is prox-regular at $\bar{x} - \frac{1}{r}\nabla f(\bar{x})$, for r sufficiently large the Moreau envelope is $\mathscr{C}^{1+}$ [3, Thm 2.4]. In particular, this implies that

$$e_r P(x^k - \frac{1}{r}\nabla f(x^k)) \to e_r P(\bar{x} - \frac{1}{r}\nabla f(\bar{x})). \tag{1.38}$$

Applying Lemma 1.2.1 and noting that the minimum for these proximal envelopes is achieved at p^k and $\bar{x}$ respectively, we see that

$$P(p^k) - \langle \nabla f(x^k), p^k \rangle + \frac{r}{2}|p^k - x^k|^2 + \langle x^k, \nabla f(x^k) \rangle + \frac{1}{2r}|\nabla f(x^k)|^2 \tag{1.39}$$

converges to

$$P(\bar{x}) - \langle \nabla f(\bar{x}), \bar{x} \rangle + \frac{r}{2}|\bar{x} - \bar{x}|^2 + \langle \bar{x}, \nabla f(\bar{x}) \rangle + \frac{1}{2r}|\nabla f(\bar{x})|^2 \tag{1.40}$$

Since $p^k \to \bar{x}$, $x^k \to \bar{x}$, and $f \in \mathscr{C}^2$ this shows that

$$P(p^k) \to P(\bar{x}), \text{ and } \tilde{P}(p^k) \to \tilde{P}(\bar{x}) \tag{1.41}$$

Part iii) $\mathrm{dist}(0, \partial\tilde{P}(p^k)) \to 0$

Since $p^k \in \mathrm{argmin}_p\{\langle \nabla f(x^k), p\rangle + \frac{r}{2}|x^k - p|^2 + P(p)\}$ we have for each k

$$\begin{aligned} 0 &\in \nabla f(x^k) + r(x^k - p^k) + \partial P(p^k) \\ -r(x^k - p^k) + \nabla f(\bar{x}) - \nabla f(x^k) &\in \nabla f(\bar{x}) + \partial P(p^k) \\ -r(x^k - p^k) + \nabla f(\bar{x}) - \nabla f(x^k) &\in \partial\tilde{P}(p^k). \end{aligned} \tag{1.42}$$

Since $p^k \to \bar{x}$, $x^k \to \bar{x}$, and $f \in \mathscr{C}^2$ this yields

$$\mathrm{dist}(0, \partial\tilde{P}(p^k)) \le r|x^k - p^k| + |\nabla f(\bar{x}) - \nabla f(x^k)| \to 0. \tag{1.43}$$

The proof now following easily.

Theorem 1.3.3 (Identification of $\mathscr{M}$ via sub-problem (1.33)**)** *Let $f \in \mathscr{C}^2$ and P be regular and prox-bounded. Suppose $\bar{x} \in \mathrm{argmin}\{f(x) + P(x)\}$ and P is prox-regular at the point $\bar{x} - \frac{1}{r}\nabla f(\bar{x})$. Suppose f is partly smooth at $\bar{x}$ relative to a manifold $\mathscr{M}$. Suppose the sequence of points x^k converge to $\bar{x}$, and consider the sequence of points p^k generated by sub-problem* (1.33)*.*

If $r > 0$ is sufficiently large and $-\nabla f(\bar{x}) \in \mathrm{rint}\,\partial P(\bar{x})$, then $p^k \in \mathscr{M}$ for all k sufficiently large.

Proof. Using $\tilde{P} = P(x) + \langle \nabla f(\bar{x}), x\rangle$ as before, we note that $\tilde{P}$ is partly smooth at $\bar{x}$ with respect to the same manifold $\mathscr{M}$ by [4, Cor 4.6]. Furthermore, $0 \in \mathrm{rint}\,\partial\tilde{P}(\bar{x})$. Finally, $\tilde{P}$ is prox-regular at the point $\bar{x}$ by [6, Ex 13.35]. Lemma 1.3.2 and Theorem 1.2.6 now combine to complete the proof.

1.4 Conclusion

On May 26th, 2009, Paul Tseng sent the following e-mail :

```
Almost forgot..

Instead of projecting onto the feasible set, another
type of active set identification arises in
compressed sensing or, more generally,
    min  f(x) + P(x),
where P(x) = max_i g_i(x), say, and f is smooth.
One can consider an analog of projection, namely,
    min_p  < f'(x), p > + r*|p-x|^2/2  + P(p)
and do active identification accordingly.  When P is
separable, as in the case of l_1-norm, this
decomposes and often has closed form solution. This
```

```
approach was used in my paper with Sangwoon Yun on
CGD method in the context of l_1-regularized
optimization, so P(x)=|x|_1. It helped to accelerate
the method on ill-conditioned problems.  This
is much more efficient than rewriting the original
problem as a smooth constrained problem
    min  f(x) + z    s.t.   g_i(x) <= z, i=1,...,m,
and then projecting onto the feasible set.  (In
general, projecting on to the feasible set is
expensive, unless the set has simple structure
like a simplex or a box.)  Results on active set
identification for smooth constrained optimization
should be extendable to this setting.

Paul
```

In this work we show that the algorithm developed by Tseng and Yun in [7] also finitely identifies active manifolds, and furthermore provide an affirmative proof to Tseng's conjecture above.

Littérature

1. W. L. Hare. A proximal method for identifying active manifolds. *Comput. Optim. Appl.*, 43(2) :295–306, 2009.
2. W. L. Hare and A. S. Lewis. Identifying active constraints via partial smoothness and prox-regularity. *J. Convex Anal.*, 11(2) :251–266, 2004.
3. W. L. Hare and R. A. Poliquin. Prox-regularity and stability of the proximal mapping. *J. Convex Anal.*, 14(3) :589–606, 2007.
4. A. S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM J. Optim.*, 13(3) :702–725 (electronic) (2003), 2002.
5. R. A. Poliquin and R. T. Rockafellar. Prox-regular functions in variational analysis. *Trans. Amer. Math. Soc*, 348(5) :1805–1838, 1996.
6. R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.
7. P. Tseng and S. Yun. A coordinate gradient descent method for nonsmooth separable minimization. *Math. Program.*, 117(1-2, Ser. B) :387–423, 2009.