

On Almost Smooth Functions and Piecewise Smooth Functions¹

Liqun Qi

Department of Mathematics
City University of Hong Kong
Kowloon Tong, Kowloon, Hong Kong
E-mail: maqilq@cityu.edu.hk

and

Paul Tseng

Department of Mathematics
The University of Washington
Seattle, WA 98195-4350, USA
E-mail: tseng@math.washington.edu

February 10, 2006 (revised June 6, 2006)

Abstract. Piecewise smooth (PS) functions are perhaps best-known examples of semismooth functions, which play key roles in the solution of nonsmooth equations and nonsmooth optimization. Recently, there have emerged other examples of semismooth functions, including the p -norm function ($1 < p < \infty$) defined on \mathbf{R}^n with $n \geq 2$, NCP functions, smoothing/penalty functions, and integral functions. These semismooth functions share the special property that their smooth point sets are locally connected around their nonsmooth points. By extending a result of Rockafellar, we show that the smooth point set of a PS function cannot have such a property. This shows that the above functions, though semismooth, are not PS. We call such functions *almost smooth* (AS). We show that the B-subdifferential of an AS function at a point has either one or infinitely many elements, which contrasts with PS functions whose B-subdifferential at a point has only a finite number of elements. We derive other useful properties of AS functions and sufficient conditions for a function to be AS. These results are then applied to various smoothing/penalty functions and integral functions.

Key Words. Nonsmooth function, piecewise smooth function, semismooth function, subdifferential, Newton method.

AMS Subject Classification. 26A27, 26B05, 26B25, 26B35, 41A29, 49J52, 52A41, 65D15, 90C25

¹This is a revision of a 2002 paper titled “An Analysis of Piecewise Smooth Functions and Almost Smooth Functions.” This work is supported by the Hong Kong Research Grant Council for the first author and by National Science Foundation Grants CCR-9731273, DMS-0511283 for the second author.

1 Introduction

In the past two decades, nonsmooth functions have emerged to play important roles in optimization [10, 15, 21, 42]. These functions, particularly those defined on finite-dimensional spaces, tend to be close to being smooth (i.e., continuously differentiable) in the sense that they are continuous everywhere and smooth almost everywhere. A well-known class of functions of this type are the **piecewise smooth (PS)** functions, which are locally representable as a selection from a finite collection of smooth functions; see Definition 1. These functions and their applications have been well studied [1, 21, 22, 23, 28, 39, 43]. Examples include the 1-norm function and the ∞ -norm function and, more generally, piecewise linear functions composed with smooth functions. The PS functions have the nice properties that they are **locally Lipschitzian** (i.e., **strictly continuous** [10, 42]), and **semismooth** [25]. These properties ensure that these functions have generalized Jacobians [10] and B-subdifferentials [31]. Moreover, nonsmooth equations involving these functions can be efficiently solved by nonsmooth Newton methods [36]. Recently, there emerged many interesting nonsmooth functions that have similarly nice properties. These include the p -norm function on \mathbf{R}^n with $1 < p < \infty$ and $n \geq 2$, the Fischer-Burmeister function [16] and other nonlinear complementarity (NCP) functions [15, 20, 24, 32, 45], smoothing/penalty functions [5, 6, 7, 8, 15, 35], and integral functions involving splines and projection onto the nonnegative reals [12, 13, 18, 34]. Are these functions PS also? If yes, then they can be treated within a well studied unifying framework. If not, then can they be treated systematically within some new framework?

By using an observation of Pang and Ralph [28] that the B-subdifferential of a PS function at a point has only finitely many distinct elements, Dontchev, Qi, and Qi [13] showed that the aforementioned integral functions are not PS due to their B-subdifferentials at the origin having infinitely many elements. This proof is specialized and does not appear to extend to other functions. This motivated a conjecture by the first author that a function defined on \mathbf{R}^n with $n \geq 2$ is not PS if it is smooth everywhere except at one point. This conjecture was recently proved by Rockafellar [41], showing that the nonsmooth point set of a PS function defined on \mathbf{R}^n ($n \geq 2$) cannot be a singleton (or, more generally, isolated points); see Theorem 1(c). Motivated by the above results, we make further studies in this paper of the nonsmooth point sets of PS functions, and use our results to formulate and study a new class of ‘nice’ nonsmooth functions, which we call “almost smooth” functions, that are not PS but encompasses the aforementioned functions like the p -norm function, smoothing/penalty functions, and integral functions.

Our first contribution is a generalization of the result of Rockafellar [41]. We show that the smooth point set of a PS function f defined on \mathbf{R}^n ($n \geq 2$) is locally disconnected around each point where f is not **strictly differentiable** [42]; see Theorem 2(c). We also give a characterization of strict differentiability for PS functions. A corollary of this result is that the smooth point sets of f cannot be locally connected around all the nonsmooth points; see Corollary 1. Intuitively, the nonsmooth point set of a PS function f defined on \mathbf{R}^n partitions \mathbf{R}^n into multiple connected components, with f being smooth on the interior of each component. For example, when $n = 2$, the nonsmooth point set can be the union of

lines and curves. All such lines and curves should either exclude their endpoints or extend to “infinity” without endpoints.

Our second contribution is a systematic treatment of nonsmooth functions such as the aforementioned p -norm function ($1 < p < \infty$, $n \geq 2$), NCP functions, smoothing/penalty functions, and integral functions, within a common framework. In particular, these functions are not only locally Lipschitzian and semismooth, they share the additional property that their smooth point sets are locally connected around all the nonsmooth points. Thus these functions are *not* PS. We call a function **weakly almost smooth** if it is locally Lipschitzian and has this additional property; see Definition 3. We call it **almost smooth** (respectively, **strongly almost smooth**) if, in addition, it is semismooth (respectively, strongly semismooth); see Definition 4. In what follows, we will often abbreviate “almost smooth” as “AS”. We study the subdifferential properties of AS functions. In particular, we show that the B-subdifferential of a weakly AS function at a point contains either a single element (in this case it is strictly differentiable at that point) or infinitely many elements; see Theorem 3. This property further distinguishes AS functions from PS functions. We define the **principal part** of the B-subdifferential of a locally Lipschitzian function, a notion also used by Klatte and Kummer [21, Eq. (6.30)] in analyzing nonsmooth Newton methods. For PS functions, the B-subdifferential coincides with its principal part. We show that this is also true for a weakly AS function that is smooth everywhere except at isolated points; see Theorem 4. We also show that if f is smooth everywhere except at a point and f is positively homogeneous about that point, then f is AS. If in addition ∇f is locally Lipschitzian everywhere except at that point, then f is strongly AS; see Theorem 5. This provides an easy check for positively homogeneous functions to be AS.

In Section 4, we make further studies of AS functions, with more examples and applications. In Subsections 4.1 and 4.2, the AS functions include the p -norm function defined on \mathbf{R}^n ($1 < p < \infty$, $n \geq 2$), certain NCP functions, and integral function associated with convex best interpolation. In Subsection 4.3, we derive general conditions for a class of smoothing/penalty functions for NCP and constrained optimization to be AS. We then apply them to various examples, including the Chen-Mangasarian class of smoothing functions and the exponential penalty function, to show they are AS. In Subsection 4.4, we study conditions for a certain integral function to be strongly AS. While particular instances of AS functions have been studied in disparate contexts [13, 16, 30, 32, 35], this is the first systematic study of such functions.

Throughout this paper, for any $x \in \mathbf{R}^n$, we use $\|x\|$ to denote the Euclidean norm of x . For any $\bar{x} \in \mathbf{R}^n$ and $\varepsilon > 0$, we denote the open Euclidean ball $\mathbf{B}_\varepsilon(\bar{x}) = \{x \in \mathbf{R}^n : \|x - \bar{x}\| < \varepsilon\}$. We use “:=” to mean “define”. \mathbf{R}_+ and \mathbf{R}_- denote, respectively, the nonnegative reals and the nonpositive reals. A subset $S \subseteq \mathbf{R}^n$ is **connected** if, for any two points y and z in S , there exists a continuous function $x : [0, 1] \rightarrow \mathbf{R}^n$ satisfying $x(0) = y$, $x(1) = z$, and $x(t) \in S$ for all $t \in [0, 1]$. For a closed set $S \subseteq \mathbf{R}^n$ and $x \in \mathbf{R}^n$, we denote $\text{dist}(x, S) := \min_{s \in S} \|x - s\|$. For any $a \in \mathbf{R}$, $a_+ = \max\{0, a\}$.

Throughout, **differentiability** will always mean Fréchet differentiability, which for locally Lipschitzian functions is equivalent to Gâteaux differentiability [10]. For a locally

Lipschitzian function f defined on an open set $O \subseteq \mathbf{R}^n$, a property stronger than differentiability is **strict differentiability** [10, 42] (i.e., ∇f is continuous when approached from the set of differentiable points), while a property stronger than strict differentiability is **smoothness** (i.e., ∇f exists and is continuous). We denote

$$\begin{aligned} F_f &:= \{x \in O : f \text{ is differentiable at } x\}, \\ S_f &:= \{x \in O : f \text{ is strictly differentiable at } x\}, \\ X_f &:= \{x \in O : f \text{ is smooth at } x\}. \end{aligned}$$

Then $F_f \supseteq S_f \supseteq X_f$.

Let f be a real-valued locally Lipschitzian function defined on a nonempty open set $O \subseteq \mathbf{R}^n$. The **B-subdifferential** of f at a point $x \in O$ [31] is defined as

$$\partial_B f(x) := \left\{ \lim_{\substack{x^k \rightarrow x \\ x^k \in F_f}} \nabla f(x^k) \right\}.$$

It is well known that $\partial_B f(x)$ is nonempty and compact for all $x \in O$ [31]. Moreover, f is strictly differentiable at x if and only if $\partial_B f(x) = \{\nabla f(x)\}$. f is **semismooth** at $x \in O$ if in addition f is directionally differentiable at x and

$$f(x+h) - f(x) - g^T h = o(\|h\|) \quad \forall g \in \partial_B f(x+h).$$

f is **strongly semismooth** if $o(\|h\|)$ can be replaced by $O(\|h\|^2)$.

2 Piecewise Smooth Functions

We begin with a definition of a PS function.

Definition 1 *A real-valued function f defined on a nonempty open set $O \subseteq \mathbf{R}^n$ is **piecewise smooth (PS)** on O if it is continuous on O and there exists a finite collection of smooth functions $f_i : O \rightarrow \mathbf{R}$, $i = 1, \dots, m$, such that*

$$f(x) \in \{f_i(x) : i \in \{1, \dots, m\}\} \quad \forall x \in O.$$

*Such a collection (not necessarily unique) is called a **representation** for f on O .*

A PS function is locally Lipschitzian and semismooth. We next give the definition of a minimal local representation for a PS function, a term used by Rockafellar [41]. Scholtes used the term “essentially active selection” [43, Section 4.1].

Definition 2 *Let f be a real-valued PS function defined on a nonempty open set $O \subseteq \mathbf{R}^n$, and let f_1, \dots, f_m be a representation for f on O . A collection $\{f_i\}_{i \in I}$ with $I \subseteq \{1, \dots, m\}$ forms a **local representation** for f at a $\bar{x} \in O$ if there exists $\varepsilon > 0$ such that*

$$f(x) \in \{f_i(x) : i \in I\} \quad \forall x \in O \cap \mathbf{B}_\varepsilon(\bar{x}). \quad (1)$$

*A local representation for f at \bar{x} is **minimal** if no proper subcollection forms a local representation for f at \bar{x} .*

The results of Rockafellar can be stated in full as follows.

Theorem 1 (Rockafellar 2003) *Let f be a real-valued PS function defined on a nonempty open set $O \subseteq \mathbf{R}^n$. For any $\bar{x} \in O$ and any minimal local representation $\{f_i\}_{i \in I}$ for f at \bar{x} , the following results hold.*

- (a) *For every $i \in I$, there is an open set $O_i \subseteq O$ such that $\bar{x} \in \text{cl}O_i$ and $f \equiv f_i$ on O_i .*
- (b) *If f is differentiable at \bar{x} , then there exists $i \in I$ such that $f(\bar{x}) = f_i(\bar{x})$ and $\nabla f(\bar{x}) = \nabla f_i(\bar{x})$.*
- (c) *If $n \geq 2$ and f is smooth on $O \setminus \{\bar{x}\}$, then f is smooth at \bar{x} and $\nabla f(\bar{x}) = \nabla f_i(\bar{x})$ for all $i \in I$.*

Part (c) of Theorem 1 is the main result in [41], whose proof uses parts (a) and (b). In fact, parts (a) and (b) were shown earlier by Scholtes [43, Propositions 4.1.1 and 4.1.3]. A related result of Scholtes [43, Proposition 4.1.5] shows that a PS function defined on a nonempty open set O is smooth on some open dense subset of O .

We extend Theorem 1(c) below by relaxing the assumption that f is smooth everywhere except possibly at some point to f is smooth on a dense subset which is locally connected around some point. Our proof uses Theorem 1(a),(b), as well an argument involving connected graph. An alternative proof that may be viewed as a direct extension of Rockafellar's approach is given in Section 5.

Theorem 2 *Let f be a real-valued PS function defined on a nonempty open set $O \subseteq \mathbf{R}^n$ with representation f_1, \dots, f_m . For any $\bar{x} \in O$ and any minimal local representation $\{f_i\}_{i \in I}$ for f at \bar{x} , the following results hold.*

- (a) *Let $D_1 \subseteq O$ be any open set containing \bar{x} such that*

$$f(x) \in \{f_i(x)\}_{i \in I} \quad \forall x \in D_1.$$

For each nonempty $J \subseteq I$, there exists a (possibly empty) open set $O_J \subseteq D_1$ such that

$$f(x) = f_i(x) \quad \forall i \in J, \quad f(x) \neq f_i(x) \quad \forall i \in I \setminus J$$

and $\cup_{J \subseteq I} \text{cl}O_J = \text{cl}D_1$.

- (b) *If f is differentiable at \bar{x} , then f is strictly differentiable at \bar{x} if and only if $\nabla f(\bar{x}) = \nabla f_i(\bar{x})$ for all $i \in I$.*
- (c) *Suppose $n \geq 2$ and f is smooth on a subset $X \subseteq O$ with $\text{cl}X \supseteq O$ and $\bar{x} \in O \setminus X$. Suppose that there exists an $\bar{\varepsilon} > 0$ such that*

$$X \cap \mathbf{B}_\varepsilon(\bar{x}) \text{ is connected} \quad \forall \varepsilon \in (0, \bar{\varepsilon}). \quad (2)$$

Then f is strictly differentiable at \bar{x} , and $\nabla f(\bar{x}) = \nabla f_i(\bar{x})$ for all $i \in I$.

Proof. (a) For $k = 1, 2, \dots, \text{card}I$, denote $\mathcal{J}_k := \{J \subseteq I : \text{card}J = k\}$ and, for each $J \in \mathcal{J}_k$, define $C_J := \{x \in D_k : f(x) = f_i(x) \forall i \in J\}$, $O_J := D_k \setminus (\cup_{J' \in \mathcal{J}_k \setminus \{J\}} C_{J'})$, and define inductively

$$D_{k+1} := D_k \setminus (\cup_{J \in \mathcal{J}_k} \text{cl}O_J).$$

It can be seen by induction that, for all $k = 1, 2, \dots, \text{card}I$, D_k and O_J are open (since each C_J is closed relative to D_k) and $O_J \subset C_J$ for all $J \in \mathcal{J}_k$. Moreover, we have $f(x) \neq f_i(x)$ for all $i \notin J$ and all $x \in O_J$. Also, it can be seen that $\cup_{J \subseteq I} \text{cl}O_J = \text{cl}D_1$.

(b) Suppose that $\bar{x} \in F_f$ and $\nabla f(\bar{x}) = \nabla f_i(\bar{x})$ for all $i \in I$. Since $\{f_i\}_{i \in I}$ is a minimal local representation for f at \bar{x} , there exists $\varepsilon > 0$ satisfying (1). Thus, for any sequence $\{x^k\} \subseteq F_f$ converging to \bar{x} , we have $x^k \in \mathbf{B}_\varepsilon(\bar{x})$ for all k sufficiently large, in which case Theorem 1(b) implies $\nabla f(x^k) = \nabla f_{i^k}(x^k)$ for some $i^k \in I$. Since ∇f_i is continuous at \bar{x} and $\nabla f_i(\bar{x}) = \nabla f(\bar{x})$ for all $i \in I$, this shows that $\nabla f(x^k) \rightarrow \nabla f(\bar{x})$. Thus f is strictly differentiable at \bar{x} .

Suppose f is strictly differentiable at \bar{x} . By Theorem 1(a), for each $i \in I$, there exists an open set $O_i \subseteq O$ such that $\bar{x} \in \text{cl}O_i$ and $f \equiv f_i$ on O_i . Then there exists a sequence $\{x^k\} \subseteq O_i$ converging to \bar{x} and $\nabla f \equiv \nabla f_i$ on $O_i \subseteq F_f$. Since ∇f_i is continuous, this implies $\nabla f(x^k) \rightarrow \nabla f_i(\bar{x})$. Since f is strictly differentiable at \bar{x} , we must also have $\nabla f(x^k) \rightarrow \nabla f(\bar{x})$. Thus $\nabla f_i(\bar{x}) = \nabla f(\bar{x})$ for all $i \in I$.

(c) Let D_1 and $C_J, O_J, J \subseteq I$, be defined as in (a). Let $\mathcal{V} := \{J \subseteq I : \bar{x} \in \text{cl}O_J\}$. For $\bar{\varepsilon} > 0$ sufficiently small, we have $\mathbf{B}_{\bar{\varepsilon}}(\bar{x}) \subseteq D_1$ and

$$\bigcup_{J \in \mathcal{V}} \text{cl}U_J = \text{cl}\mathbf{B}_{\bar{\varepsilon}}(\bar{x}), \quad (3)$$

where we define $U_J := O_J \cap \mathbf{B}_{\bar{\varepsilon}}(\bar{x})$. For each $J \in \mathcal{V}$, U_J is a nonempty open subset of C_J . By taking $\bar{\varepsilon}$ smaller if necessary, we can assume that, for every $(J, J') \in \mathcal{A} := \{(J, J') : J, J' \in \mathcal{V}, J \neq J', (\text{cl}U_J \cap \text{cl}U_{J'}) \cap X \neq \emptyset\}$, we have

$$((\text{cl}U_J \cap \text{cl}U_{J'}) \cap \mathbf{B}_{\bar{\varepsilon}}(\bar{x})) \cap X \neq \emptyset \quad \forall \varepsilon \in (0, \bar{\varepsilon}]. \quad (4)$$

Since (2) holds, by taking $\bar{\varepsilon}$ even smaller if necessary, we can further assume that $\mathbf{B}_{\bar{\varepsilon}}(\bar{x}) \cap X$ is connected for every $\varepsilon \in (0, \bar{\varepsilon}]$. Notice that this set is also nonempty since X is dense in O .

For each $J \in \mathcal{V}$, U_J is a nonempty open subset of C_J . Since X is dense in O , then $U_J \cap X$ is dense in U_J so $U_J \cap X$ is nonempty. For each $i \in J$, since f_i is smooth and equal to f on U_J , then f is smooth on U_J and $\nabla f(x) = \nabla f_i(x)$ for all $x \in U_J$. Since f is smooth on X and f_i is smooth on O so that ∇f and ∇f_i are continuous on $\text{cl}U_J \cap X$, this implies

$$\nabla f(x) = \nabla f_i(x) \quad \forall x \in \text{cl}U_J \cap X, \quad \forall i \in J. \quad (5)$$

For each $(J, J') \in \mathcal{A}$, we have from (4) that there exists a sequence $\{x^k\}_{k=1}^\infty \subseteq (\text{cl}U_J \cap X) \cap (\text{cl}U_{J'} \cap X)$ converging to \bar{x} . Then, (5) implies $\nabla f_i(x^k) = \nabla f_j(x^k)$ for all $i, j \in J \cup J'$ and all k . Since $\nabla f_i, \nabla f_j$ are continuous, we have in the limit as $k \rightarrow \infty$ that $\nabla f_i(\bar{x}) =$

$\nabla f_j(\bar{x})$ for all $i, j \in J \cup J'$. Since $\mathbf{B}_\varepsilon(\bar{x}) \cap X$ is nonempty and connected, then (3) implies $\bigcup_{J \in \mathcal{V}} (\text{cl}U_J \cap X)$ is nonempty and connected, so that the (undirected) graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ is connected. (Notice that $(J, J') \in \mathcal{A}$ if and only if $J \neq J'$ and $(\text{cl}U_J \cap X) \cap (\text{cl}U_{J'} \cap X) \neq \emptyset$.) Hence $\nabla f_i(\bar{x}) = \nabla f_j(\bar{x})$ for all $i, j \in \bigcup_{J \in \mathcal{V}} J$. For each $i \in I$, we have $\{i\} \in \mathcal{V}$ (else f_i would be superfluous in the minimal local representation), so $\bigcup_{J \in \mathcal{V}} J = I$.

Thus f is differentiable at \bar{x} and $\nabla f(\bar{x}) = \nabla f_i(\bar{x})$ for all $i \in I$. By (b), f is strictly differentiable at \bar{x} . \square

It can be seen from its proof that Theorem 2(a) holds for any nonempty open set $D_1 \subseteq O$ and any representation $\{f_i\}_{i \in I}$ for f on D_1 . Taking $D_1 = O$, it shows that, for a real-valued PS function f defined on a nonempty open set $O \subseteq \mathbf{R}^n$, X_f contains an open subset that is dense in O . This gives an alternative proof of [43, Proposition 4.1.5].

Theorem 1(b) shows that if a PS function f is differentiable at \bar{x} , and $\{f_i\}_{i \in I}$ is a minimal local representation for f at \bar{x} , then $\nabla f(\bar{x}) = \nabla f_i(\bar{x})$ for some $i \in I$. Theorem 2(b) shows that if in addition $\nabla f(\bar{x}) = \nabla f_i(\bar{x})$ for all $i \in I$, then f is strictly differentiable at \bar{x} . This additional condition cannot be dropped, as the following example shows. Geometrically, f can fail to be strictly differentiable at \bar{x} when there exists some $i \in I$ and open set O_i such that $\bar{x} \in \text{cl}O_i$, $f \equiv f_i$ on O_i and $\nabla f_i(\bar{x})$ is a normal vector to $\text{cl}O_i$ at \bar{x} .

Example 1 Consider the PS function

$$f(x_1, x_2) = \begin{cases} x_2 & \text{if } |x_2| \leq (x_1)^2, \\ (x_1)^2 & \text{if } x_2 > (x_1)^2, \\ -(x_1)^2 & \text{if } x_2 < -(x_1)^2, \end{cases}$$

defined on $O = \mathbf{R}^2$. It can be seen that f is differentiable at the origin with $\nabla f(0, 0) = 0$, but is not strictly differentiable there. In particular, for $x^k = (\frac{1}{k}, 0)$, $k = 1, 2, \dots$, we have $x^k \rightarrow (0, 0)$ while $\nabla f(x^k) = (0, 1) \not\rightarrow \nabla f(0, 0)$. Notice that $(0, 1)$ is a normal vector to $\{(x_1, x_2) : |x_2| < (x_1)^2\}$ at $(0, 0)$. Here, $F_f = \mathbf{R}^2 \setminus \{(x_1, x_2) : |x_2| = (x_1)^2, x_1 \neq 0\}$ and $S_f = X_f = F_f \setminus \{(0, 0)\}$.

Theorem 2(c) shows that, for a real-valued PS function defined on open set $O \subseteq \mathbf{R}^n$, each point of $O \setminus S_f$ lies in some open ball whose intersection with X_f is disconnected. For example, when $n = 2$, $O \setminus S_f$ cannot include any ‘‘relative boundary point’’ of $O \setminus X_f$. By using Theorem 2(b), we have the following generalization of Theorem 1(c).

Corollary 1 Let f be a real-valued PS function defined on a nonempty open set $O \subseteq \mathbf{R}^n$ ($n \geq 2$). Suppose f is smooth on a subset $X \subseteq O$ with $\text{cl}X \supseteq O$. Suppose also that, for every $\bar{x} \in O \setminus X$, there exists an $\bar{\varepsilon} > 0$ such that (2) holds. Then f is smooth on O .

Proof. By Theorem 2(c), f is strictly differentiable on O . Then $S_f = F_f = O$. Since ∇f is continuous at each point in S_f when approached from F_f , this implies ∇f is continuous on O , so f is smooth on O . \square

Lastly, Theorem 2(c) is false if we drop the assumption of $\text{cl}X \supseteq O$ or the existence of $\bar{\varepsilon} > 0$ such that (2) holds. This is illustrated with the example below.

Example 2 Consider the PS function

$$f(x_1, x_2) = \max\{0, x_1\}$$

defined on $O = \mathbf{R}^2$. For $X = \{(x_1, 0) : x_1 > 0\}$ and $\bar{x} = (0, 0)$, we have that (2) holds with $\bar{\varepsilon} = 1$ but f is not differentiable at \bar{x} . For $X = \{(x_1, x_2) : x_1 \neq 0\}$ and $\bar{x} = (0, 0)$, we see that $\text{cl}X \supseteq O$ but f is not differentiable at \bar{x} .

3 Almost Smooth Functions

Theorems 1(c), 2(c) and Corollary 1 describe topological properties of the nonsmooth point set of a PS function. In particular, Corollary 1 shows that a function *cannot* be PS if its smooth point set is locally connected around all its nonsmooth points. As we mentioned in Section 1, nonsmooth functions that possess this property, in addition to being locally Lipschitzian and semismooth, are surprisingly abundant. These include the p -norm function defined on \mathbf{R}^n ($1 < p < \infty$, $n \geq 2$), NCP functions, smoothing/penalty functions, integral functions. For some of these functions, the nonsmooth point set comprise isolated points (e.g., the sum of p -norm functions each composed with affine mappings). For others, the nonsmooth point set has dimension less than $n - 1$ but greater than 0 (e.g., the exponential penalty function in Example 5, whose nonsmooth point set is the union of $\frac{1}{2}(n - 1)(n - 2)$ subspaces of dimension $n - 2$; and the integral function in Subsection 4.4, whose nonsmooth point set is a manifold of dimension $n - 2$). In this and next section, we study this class of nonsmooth functions, which we call almost smooth (AS) functions.

Definition 3 Let f be a real-valued locally Lipschitzian function defined on a nonempty open set $O \subseteq \mathbf{R}^n$. Then f is **weakly almost smooth** if, for every $\bar{x} \in O \setminus X_f$, there is an $\bar{\varepsilon} > 0$ such that

$$X_f \cap \mathbf{B}_\varepsilon(\bar{x}) \text{ is nonempty and connected } \quad \forall \varepsilon \in (0, \bar{\varepsilon}). \quad (6)$$

f is **basic weakly almost smooth** if $n \geq 2$ and $O \setminus X_f$ comprises isolated points.²

The above definition of a weakly AS function f , motivated by (2), implies that X_f is dense in O , and is locally connected around each point not in X_f . Note that when $n = 1$, a weakly AS function is a smooth function. When $n \geq 2$, a weakly AS function is either smooth or else, by Corollary 1, it is not PS. Also, a basic weakly AS function f is weakly AS.

According to Pang and Ralph [28] (also see [43, Proposition A.4.1]), the B-subdifferential of a PS function at a point contains only a finite number of elements. We show below that, for a weakly AS function f , the elements of $\partial_B f(x)$ are *not isolated* if f is not strictly differentiable at x . This implies that the B-subdifferential of a weakly AS function f at a point contains either a single element (i.e., f is strictly differentiable at that point) or infinitely many elements. Thus, the subdifferential structure of a weakly AS function is very different from that of a PS function.

²In other words, for every $\bar{x} \in O \setminus X_f$, there is an open ball \mathbf{B} centered at \bar{x} such that $X_f \cap \mathbf{B} = \mathbf{B} \setminus \{\bar{x}\}$.

Theorem 3 *Let f be a real-valued weakly AS function defined on a nonempty open set $O \subseteq \mathbf{R}^n$ with $n \geq 2$. Let $\bar{x} \in O \setminus S_f$. Then, $\partial_B f(\bar{x})$ has infinitely many elements and none of its elements is isolated in it.*

Proof. Since f is locally Lipschitz continuous, $\partial_B f(\bar{x})$ is nonempty. Since $\bar{x} \notin S_f$, $\partial_B f(\bar{x})$ has at least two elements. Suppose that $\partial_B f(\bar{x})$ has an isolated element g . We show below that this yields a contradiction.

Let $G := \partial_B f(\bar{x}) \setminus \{g\}$. Since $\partial_B f(\bar{x})$ has at least two elements, $G \neq \emptyset$. Since $\partial_B f(\bar{x})$ is closed and g is isolated in $\partial_B f(\bar{x})$, G is closed. Then the scalar quantity

$$\delta := \text{dist}(g, G) \tag{7}$$

is positive. Let

$$X_\varepsilon := X_f \cap \mathbf{B}_\varepsilon(\bar{x}).$$

Since f is smooth on X_f , the definition of $\partial_B f(\bar{x})$ implies

$$\limsup_{\varepsilon \downarrow 0} \nabla f(X_\varepsilon) \subseteq \partial_B f(\bar{x}).$$

This convergence is uniform in ε in the sense that

$$\sup_{x \in X_\varepsilon} \text{dist}(\nabla f(x), \partial_B f(\bar{x})) \rightarrow 0 \quad \text{as } \varepsilon \downarrow 0.$$

Thus, there exists $\bar{\varepsilon} > 0$ such that for any $\varepsilon \in (0, \bar{\varepsilon})$, we have $\mathbf{B}_\varepsilon(\bar{x}) \subseteq X_f$ and

$$\sup_{x \in X_\varepsilon} \text{dist}(\nabla f(x), \partial_B f(\bar{x})) \leq \delta/3 \quad \forall \varepsilon \in (0, \bar{\varepsilon}). \tag{8}$$

Let $D_g := \{x \in X_\varepsilon : \|\nabla f(x) - g\| \leq \delta/3\}$ and $D_G := \{x \in X_\varepsilon : \text{dist}(\nabla f(x), G) \leq \delta/3\}$. Moreover, (7) implies that D_g and D_G are disjoint, while (8) and $G = \partial_B f(\bar{x}) \setminus \{g\}$ imply that

$$X_\varepsilon = D_g \cup D_G.$$

But X_ε is connected. Thus, one of D_g and D_G should be empty. This contradicts the assumption that $g \in \partial_B f(\bar{x})$ and the fact that $G \neq \emptyset$. \square

For any real-valued locally Lipschitzian function f defined on a nonempty open set $O \subseteq \mathbf{R}^n$ with X_f dense in O , we define the **principal part** of the B-subdifferential as

$$\partial_P f(x) := \left\{ \lim_{\substack{x^k \rightarrow x \\ x^k \in X_f}} \nabla f(x^k) \right\}.$$

Since X_f is dense in O and $X_f \subseteq F_f$, it readily follows that $\partial_P f(x)$ is nonempty and compact for all $x \in O$, and $\partial_P f(x) \subseteq \partial_B f(x)$. In fact, $\partial_P f$ had been used by Klatter and Kummer [21, Eq. (6.30)] in the context of Newton map for a **pseudo-smooth** function $f : \mathbf{R}^n \rightarrow \mathbf{R}$, i.e., f is locally Lipschitzian and $O \setminus X_f$ is open and dense in \mathbf{R}^n . Thus $\partial_P f$ is of practical interest.

If f is PS, then it is known [28, Lemma 2], [43, Proposition A.4.1] that $\partial_B f(x) = \{\nabla f_i(x)\}_{i \in I}$ for any $x \in O$, where $\{f_i\}_{i \in I}$ is any minimal local representation for f at x . (This can also be proved using Theorem 1(a),(b).) Moreover, Theorem 1(a) shows that $\partial_P f(x) \supseteq \{\nabla f_i(x)\}_{i \in I}$, so in fact

$$\partial_P f(x) = \partial_B f(x) \quad \forall x \in O. \quad (9)$$

Is this property of PS functions shared by weakly AS functions? The following theorem shows that the answer is yes for a basic weakly AS function. For a general weakly AS, this question remains open.

Theorem 4 *Let f be a real-valued basic weakly AS function defined on a nonempty open set $O \subseteq \mathbf{R}^n$. Then (9) holds.*

Proof. Since $O \setminus X_f$ comprises isolated points, X_f is dense in O . It suffices to show that (9) holds for all $x \in O \setminus X_f$. Consider any $\bar{x} \in O \setminus X_f$. Since f is basic weakly AS, there exists an open ball \mathbf{B} centered at \bar{x} such that

$$X_f \cap \mathbf{B} = \mathbf{B} \setminus \{\bar{x}\}.$$

If $\bar{x} \notin F_f$, then $F_f \cap \mathbf{B} = X_f \cap \mathbf{B}$ so that (9) holds for $x = \bar{x}$. Otherwise $\bar{x} \in F_f$, so that $F_f \cap \mathbf{B} = (X_f \cap \mathbf{B}) \cup \{\bar{x}\}$. We prove below that $\nabla f(\bar{x}) \in \partial_P f(\bar{x})$ and hence, by $F_f \cap \mathbf{B} = (X_f \cap \mathbf{B}) \cup \{\bar{x}\}$, $\partial_P f(\bar{x}) = \partial_B f(\bar{x})$.

Without loss of generality, assume $\bar{x} = 0$ and $\nabla f(\bar{x}) = 0$. We argue by contradiction. Suppose $0 \notin \partial_P f(\bar{x})$. Then there exist $\bar{\epsilon} > 0$ and $\rho > 0$ such that $\mathbf{B}_{\bar{\epsilon}}(0) \subseteq \mathbf{B}$ and

$$\|\nabla f(x)\| \geq \rho \quad \forall x \in \mathbf{B}_{\bar{\epsilon}}(0) \setminus \{0\}. \quad (10)$$

Since f is locally Lipschitzian at \bar{x} , then Theorem 9.13 in [42] implies that, by taking $\bar{\epsilon}$ sufficiently small, we can assume that ∇f is uniformly bounded on $\mathbf{B}_{\bar{\epsilon}}(0)$, i.e., there exists $M \geq 0$ such that

$$\|\nabla f(x)\| \leq M \quad \forall x \in \mathbf{B}_{\bar{\epsilon}}(0). \quad (11)$$

Consider any $\epsilon \in (0, \bar{\epsilon}/3)$. Denote $R_\epsilon := \{x \in \mathbf{R}^n : \epsilon \leq \|x\| \leq 3\epsilon\}$. Then $R_\epsilon \subset \mathbf{B}_{\bar{\epsilon}}(0) \setminus \{0\} \subset \mathbf{B} \setminus \{0\}$. Let x_ϵ be any point satisfying $\|x_\epsilon\| = 2\epsilon$. Since f is smooth on $\mathbf{B} \setminus \{0\}$, then ∇f is defined and continuous on R_ϵ . Since R_ϵ is compact, ∇f is uniformly continuous on R_ϵ . Thus, by choosing an integer $k \geq 1$ sufficiently large (k depends on ϵ), we have

$$\|\nabla f(x) - \nabla f(y)\| \leq \epsilon \quad \text{whenever } x, y \in R_\epsilon, \|x - y\| \leq \delta, \quad (12)$$

where $\delta := \frac{\epsilon}{k}$. Define

$$x^0 := x_\epsilon \quad \text{and} \quad x^{j+1} := x^j + \frac{\delta}{M} \nabla f(x^j), \quad j = 0, 1, \dots, k-1, \quad \text{and} \quad y_\epsilon := x^k.$$

We have that $\|x^{j+1} - x^j\| = \frac{\epsilon}{Mk} \|\nabla f(x^j)\|$. Since the distance from x_ϵ to the boundary of R_ϵ is ϵ and $\|\nabla f(x^j)\| \leq M$ whenever $x^j \in R_\epsilon$, an induction argument shows that

$$\|x^{j+1} - x^j\| \leq \delta \quad \text{and} \quad [x^j, x^{j+1}] \subseteq R_\epsilon, \quad j = 0, 1, \dots, k-1. \quad (13)$$

(Recall that $[x, y]$ denotes the closed line segment joining x and y .) Then, for each j , we can apply the mean value theorem to conclude that there exists $z^j \in [x^j, x^{j+1}]$ such that

$$\begin{aligned} f(x^{j+1}) - f(x^j) &= \nabla f(z^j)^T (x^{j+1} - x^j) \\ &= \frac{\delta}{M} \nabla f(z^j)^T \nabla f(x^j) \\ &= \frac{\delta}{M} (\nabla f(z^j) - \nabla f(x^j))^T \nabla f(x^j) + \frac{\delta}{M} \|\nabla f(x^j)\|^2 \\ &\geq -\frac{\delta}{M} \|\nabla f(z^j) - \nabla f(x^j)\| \|\nabla f(x^j)\| + \frac{\delta}{M} \|\nabla f(x^j)\|^2 \\ &\geq -\delta\epsilon + \frac{\delta}{M} \rho^2, \end{aligned}$$

where the last inequality uses (10)–(13). Summing this inequality over $j = 0, 1, \dots, k-1$ yields

$$f(y_\epsilon) - f(x_\epsilon) \geq -k\delta\epsilon + k\frac{\delta}{M}\rho^2 = -\epsilon^2 + \epsilon\frac{\rho^2}{M} = o(\epsilon) + \epsilon\frac{\rho^2}{M}.$$

However, this contradicts our hypothesis that f is differentiable at 0 with $\nabla f(0) = 0$ and the fact that $x_\epsilon, y_\epsilon \in R_\epsilon$, so that

$$f(x_\epsilon) - f(0) = o(\|x_\epsilon\|) = o(\epsilon), \quad f(y_\epsilon) - f(0) = o(\|y_\epsilon\|) = o(\epsilon),$$

implying $f(y_\epsilon) - f(x_\epsilon) = o(\epsilon)$. □

In general, a locally Lipschitzian function f with dense smooth point set need not satisfy (9), as the following example shows.

Example 3 *Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be the function whose graph is obtained by taking the line through $(-1, 1)$ with slope -2 , reflecting it up (with slope 2) whenever it hits the x -axis and down (with slope -2) whenever it hits the parabola $y = x^2$. This defines $f(x)$ for all $x < 0$. We define $f(x)$ symmetrically (about the y -axis) for all $x > 0$, and define $f(0) = 0$. Thus, the graph of f is bounded between the x -axis and the parabola $y = x^2$ and comprises infinitely many line segments with slopes of either 2 or -2 .*

It can be calculated that these line segments have endpoints $\pm a_1, \pm a_2, \dots$, where

$$a_1 = \frac{1}{2}, \quad a_{2k} = \varphi_{\text{even}}(a_{2k-1}), \quad a_{2k+1} = \varphi_{\text{odd}}(a_{2k}), \quad k = 1, 2, \dots,$$

and we define $\varphi_{\text{even}}(x) := \sqrt{2x+1} - 1$, $\varphi_{\text{odd}}(x) := x - x^2/2$.

It is readily seen that f is locally Lipschitzian on \mathbf{R} and is differentiable at 0, with $\nabla f(0) = 0$, but f is not continuously differentiable at 0. Thus, $F_f = \mathbf{R} \setminus \{\pm a_1, \pm a_2, \dots\}$ and

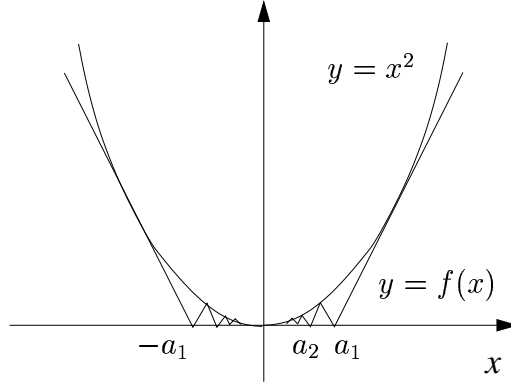


Figure 1: Graph of function f for Example 3.

$X_f = F_f \setminus \{0\}$. Then $\partial_P f(0) = \{-2, 2\}$ while $\partial_B f(0) = \{-2, 2, 0\}$. Since a PS function f defined on an open set O must satisfy (9), this proves that f is not PS. Since $n = 1$ and f is not smooth, f is also not weakly AS. It can be further checked that f is not semismooth at 0, though f is pseudo-smooth in the sense of [21].

Definition 4 A real-valued function f defined on a nonempty open set $O \subseteq \mathbf{R}^n$ is **(basic) almost smooth** if f is (basic) weakly almost smooth and semismooth. If in addition f is strongly semismooth, then f is **(basic) strongly almost smooth**.

In general, a weakly AS function need not be semismooth. For example, let $f : \mathbf{R}^n \rightarrow \mathbf{R}$, $n \geq 2$, be defined by

$$f(x) = \begin{cases} \|x\|^2 \sin\left(\frac{1}{\|x\|}\right), & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

This function is locally Lipschitzian, differentiable everywhere. It is smooth everywhere except at the origin, so it is basic weakly AS but not PS. In particular, $\partial_B f(0)$ is the unit ball, which contains infinitely many elements. It is readily seen that

$$f(h) - f(0) - f'(h; h) = \|h\| \cos\left(\frac{1}{\|h\|}\right) + O(\|h\|^2) \quad \forall h \neq 0,$$

so f is not semismooth at 0.

Since sum and product of locally Lipschitzian semismooth functions are also locally Lipschitzian semismooth, the following properties of AS functions readily follow:

P1 If f_1, f_2 are real-valued AS functions defined on an open set $O \subseteq \mathbf{R}^n$ and $O \setminus X_{f_1}, O \setminus X_{f_2}$ are manifolds of dimension less than $n - 1$,³ then $f_1 + f_2$ and $f_1 f_2$ are both AS functions defined on O .

³We say $Y \subset O$ is a manifold of dimension $k \leq n$ if $Y = \{x \in O : F(x) = 0\}$ for some smooth mapping $F : O \rightarrow \mathbf{R}^k$ whose Jacobian has rank $n - k$ on O .

P2 If f is a real-valued AS function defined on an open set $O \subseteq \mathbf{R}^n$ and F is a smooth 1-to-1 mapping from an open set $O' \subseteq \mathbf{R}^n$ to O , then $f \circ F$ is an AS function defined on O' .

P3 If f is a real-valued basic AS function defined on an open set $O \subseteq \mathbf{R}^n$ and F is a smooth mapping from an open set $O' \subseteq \mathbf{R}^m$ to O with the Jacobian of F having rank greater than 1 on O' , then $f \circ F$ is an AS function defined on O' .

Klatte and Kummer [21, page 128] had proposed a **locally PC¹** function as a generalization of PS functions and the Euclidean norm function. In particular, $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is locally PC¹ if (i) f is pseudo-smooth; (ii) f is selected from a finite collection of continuous functions f_1, \dots, f_m with f_i smooth on some open set $O_i \subseteq \mathbf{R}^n$ and ∇f_i uniformly continuous on each bounded subset of O_i ; and (iii) for each $\bar{x} \in \mathbf{R}^n$, there is an open ball \mathbf{B} centered at \bar{x} such that $x \in X_f \cap \mathbf{B}$ implies $f(\bar{x}) = f_i(\bar{x})$, $f(x) = f_i(x)$, $\nabla f(x) = \nabla f_i(x)$, and $]\bar{x}, x[\subseteq O_i$ for some $i \in \{1, \dots, m\}$. (Here $]\bar{x}, x[= \{t\bar{x} + (1-t)x : 0 < t < 1\}$.) It is shown in [21, Theorem 6.18] that locally PC¹ functions have properties similar to semismoothness but with $\partial_P f$ in place of $\partial_B f$. Moreover, locally PC¹ functions include PS functions, the composition of the Euclidean norm with a linear mapping, and the composition of a smooth mapping with two PC¹ functions [21, Lemma 6.17]. In fact, conditions (ii) and (iii) hold (with $m = 1$) whenever ∇f is uniformly continuous on each bounded subset of X_f and $O \setminus X_f$ is polyhedral. If $O \setminus X_f$ is not polyhedral, then (iii) may fail to hold. For example, the function

$$f(x_1, x_2, x_3) = \sqrt{((x_1)^2 + (x_2)^2 - x_3)^2 + (x_2)^2}$$

is pseudo-smooth with $X_f = \mathbf{R}^3 \setminus \{(x_1, 0, x_3) : (x_1)^2 = x_3\}$. Let \bar{x} be the origin. For any $0 < \varepsilon \leq 1$, we have $x = (\varepsilon/2, 0, (\varepsilon/2)^3) \in \mathbf{B}_\varepsilon(\bar{x})$ but $]\bar{x}, x[\not\subseteq X_f$. Thus, f does not appear to be locally PC¹. On the other hand, f is semismooth and X_f has dimension 1, so f is AS. This also follows from property P3 and the observation that f is the composition of the Euclidean norm with a smooth mapping whose Jacobian has rank greater than 1.

As a referee noted, apart from $\partial_P f$ and $\partial_B f$, there is another subdifferential, first studied by Mordukhovich [26] and now playing a central role in variational analysis [27, 42]. For a real-valued locally Lipschitzian function f defined on a nonempty open set $O \subseteq \mathbf{R}^n$, this subdifferential is defined as

$$\partial f(x) := \left\{ \lim_{\substack{x^k \rightarrow x \\ x^k \in R_f, v^k \in \hat{\partial} f(x^k)}} v^k \right\},$$

where $R_f := \{x \in O : \hat{\partial} f(x) \neq \emptyset\}$ and $\hat{\partial} f(x) := \{v \in \mathbf{R}^n : f(y) \geq f(x) + v^T(y - x) + o(\|y - x\|)\}$. Since $F_f \subseteq R_f$, it follows that $\partial_B f(x) \subseteq \partial f(x)$ for all $x \in O$. If f is basic weakly AS, then $\partial f(x) = \partial_B f(x) \cup \hat{\partial} f(x)$. As examples, for $f(x) = \|x\|$, $\hat{\partial} f(0)$ is the closed unit ball; while for $f(x) = -\|x\|$, $\hat{\partial} f(0) = \emptyset$. For f in Example 3, $\hat{\partial} f(0) = \{0\}$ and $\partial f(0) = [-2, 2]$. In general, what can we say about ∂f when f is weakly AS or (basic) AS? This is an interesting question for future investigation.

By Theorem 4, a basic weakly AS function shares with PS functions the subdifferential property (9). Example 1 shows that a differentiable PS function need not be smooth. What

about a basic AS function? Since such a function is smooth everywhere except at isolated points and is semismooth there, might differentiability imply smoothness? The following example shows that this is false. The motivation for the example comes from the observation that if f is differentiable and semismooth at the origin 0 and (without loss of generality) $\nabla f(0) = 0$, then for any sequence x^k converging to 0 , we have $\nabla f(x^k)^T x^k / \|x^k\| \rightarrow 0$. Thus, if $\nabla f(x^k) \not\rightarrow 0$, then $\nabla f(x^k)$ must be perpendicular to $x^k / \|x^k\|$ asymptotically. The example constructs a function f with such a property.

Example 4 We define f below. Let $\varphi : \mathbf{R}^2 \rightarrow \mathbf{R}$ be any smooth function satisfying (i) $\varphi(x) = 0$ whenever $\|x\| \geq 1$, and (ii) $\nabla_1 \varphi(\hat{x}) \neq 0$ for some \hat{x} satisfying $\|\hat{x}\| < 1$. (Here $\nabla_1 \varphi$ denotes the partial derivative with respect to x_1 .) For $k = 2, 3, \dots$, we scale and translate φ to obtain the function

$$\varphi_k(x) := \frac{1}{k} \varphi \left(k \left(\frac{1}{x_1} - k^2, \frac{1}{x_2} - k \right) \right).$$

Notice that φ_k is smooth and vanishes outside of

$$X_k := \left\{ (x_1, x_2) : \left\| \left(\frac{1}{x_1} - k^2, \frac{1}{x_2} - k \right) \right\| < \frac{1}{k} \right\}.$$

Moreover, X_k , $k = 2, 3, \dots$, are disjoint. In particular, observe that $x \in X_k$ implies $|\frac{1}{x_1} - k^2| < \frac{1}{k}$ and $|\frac{1}{x_2} - k| < \frac{1}{k}$ or, equivalently, $1/(k^2 + \frac{1}{k}) < x_1 < 1/(k^2 - \frac{1}{k})$ and $1/(k + \frac{1}{k}) < x_2 < 1/(k - \frac{1}{k})$. Now, define

$$f(x) := \begin{cases} (x_1)^2 \varphi_k(x) & \text{if } x \in X_k \text{ for some } k \in \{2, 3, \dots\}; \\ 0 & \text{else.} \end{cases}$$

Roughly speaking, the graph of f comprises a sequence of “bumps” along the parabola $x_1 = (x_2)^2$, with each bump of radius approximately x_2 and height $O((x_1)^2 x_2)$.

We have that $X_k \subset \mathbf{R}_{++}^2$ for all $k = 2, 3, \dots$, so it follows from the chain rule that f is smooth at every $x \in \text{int} X_k$ for each k . Since φ_k is smooth and has value 0 on the boundary of X_k , it is also readily seen that f is smooth on the boundary of X_k . Also, clearly f is smooth at every x outside of $(\cup_{k=2}^{\infty} X_k) \cup \{(0, 0)\}$. Thus, f is smooth on $\mathbf{R}^2 \setminus \{(0, 0)\}$.

We claim that f is differentiable at $(0, 0)$ with $\nabla f(0, 0) = (0, 0)$. Since $f(0, 0) = 0$, it suffices to show that $f(x) = o(\|x\|)$ for $x \neq (0, 0)$. For $x \notin (\cup_{k=2}^{\infty} X_k) \cup \{(0, 0)\}$, this is clearly true. For $x \in X_k$ for some $k \in \{2, 3, \dots\}$, we have that x_1 is in the order of $1/k^2$ and x_2 is in the order of $1/k$, so that $x_1 = O((x_2)^2)$. Since φ_k is bounded, this yields $(x_1)^2 \varphi_k(x) = o(x_1 + x_2)$ and hence $f(x) = o(\|x\|)$.

We claim that f is semismooth at $(0, 0)$. Since $f(0, 0) = 0$ and $f(x) = o(\|x\|)$, it suffices to show that $\nabla f(x)^T x = o(\|x\|)$ for $x \neq (0, 0)$. We have that $\nabla f(x) = (0, 0)$ if $x \notin \cup_{k=2}^{\infty} X_k$, so this holds immediately. If $x \in X_k$ for some $k \in \{2, 3, \dots\}$, then we have from the chain rule and the form of φ_k that

$$\begin{aligned} \nabla_1 f(x) &= 2x_1 \varphi_k(x) - \nabla_1 \varphi \left(k \left(\frac{1}{x_1} - k^2, \frac{1}{x_2} - k \right) \right), \\ \nabla_2 f(x) &= -\frac{(x_1)^2}{(x_2)^2} \nabla_2 \varphi \left(k \left(\frac{1}{x_1} - k^2, \frac{1}{x_2} - k \right) \right). \end{aligned}$$

Since x_2 is in the order of $1/k$ and $\varphi_k(x) = O(\frac{1}{k})$, then $\varphi_k(x) = O(x_2)$. Thus $2x_1\varphi_k(x) = O(x_1x_2) = o(\|x\|)$. Since $x_1 = O((x_2)^2)$, then $(x_1)^2/(x_2)^2 = O(x_1)$ and $x_1 = O(\|x\|^2)$. Then, the boundedness of $\nabla_1\varphi$ and $\nabla_2\varphi$ yield that $\nabla_1f(x) \cdot x_1 = O(x_1) = O(\|x\|^2)$ and $\nabla_2f(x) \cdot x_2 = O(x_1x_2) = O(\|x\|^2)$. Thus $\nabla f(x)^T x = O(\|x\|^2)$.

The above argument also shows that ∇f is bounded. Thus, by the mean value theorem or Theorem 9.13 in [42], f is locally Lipschitzian everywhere.

Finally, for each $k \in \{2, 3, \dots\}$, define $x^k = (x_1^k, x_2^k)$ by

$$x_1^k = \frac{1}{k^2 + \hat{x}_1/k}, \quad x_2^k = \frac{1}{k + \hat{x}_2/k}.$$

Then $x^k \in X_k$ and we have from the above formula for ∇_1f that

$$\nabla_1f(x^k) = 2x_1^k\varphi_k(x^k) - \nabla_1\varphi(\hat{x}).$$

Thus $\nabla_1f(x^k) \rightarrow \nabla_1\varphi(\hat{x}) \neq 0$ while $x^k \rightarrow (0, 0)$ as $k \rightarrow \infty$. Since $\nabla_1f(0, 0) = 0$, this shows that f is not smooth at $(0, 0)$.

A basic AS function need not be pseudo-smooth in the sense of [21]. For an example, let $\varphi : \mathbf{R}^2 \rightarrow \mathbf{R}$ be a function that (i) vanishes outside the unit Euclidean ball and (ii) is smooth everywhere except at the origin and is differentiable and semismooth there. Such a φ can be constructed by modifying the function f in Example 4 which satisfies (ii). Define $f : \mathbf{R}^2 \rightarrow \mathbf{R}$ by

$$f(x) = \sum_{k=1}^{\infty} \frac{1}{k2^k} \varphi \left(2^{k+2}(x - (2^{-k}, 0)) \right).$$

It can be verified that $\nabla f(x) = \frac{4}{k} \nabla \varphi \left(2^{k+2}(x - (2^{-k}, 0)) \right)$ if $\|x - (2^{-k}, 0)\| \leq 1/2^{k+2}$ for some k and otherwise $\nabla f(x) = 0$. Moreover, f is locally Lipschitzian and semismooth everywhere and ∇f is continuous at the origin. Thus $\mathbf{R}^2 \setminus X_f = \{(2^{-k}, 0)\}_{k=1,2,\dots}$ comprises isolated points, so f is basic AS. However, X_f is not open, so f is not pseudo-smooth.

The next theorem shows that if a function f defined on \mathbf{R}^n ($n \geq 2$) is smooth everywhere except at one point and is positively homogeneous about that point, then f is basic AS. If in addition ∇f is locally Lipschitzian everywhere outside that point, then f is basic strongly AS. The p -norm function on \mathbf{R}^n with $1 < p < \infty$ and the Fischer-Burmeister function are examples of such functions; see Subsection 4.1 for further discussions.

Theorem 5 *Let f be a real-valued function defined on \mathbf{R}^n . Suppose that f is smooth everywhere except at some $\bar{x} \in \mathbf{R}^n$. Also, suppose that f is positively homogeneous about \bar{x} , i.e., for any $h \in \mathbf{R}^n$ and $t \in \mathbf{R}_+$,*

$$f(\bar{x} + th) = tf(\bar{x} + h).$$

Then f is Lipschitz continuous on \mathbf{R}^n and is strongly semismooth at \bar{x} . If furthermore ∇f is locally Lipschitzian on $\mathbf{R}^n \setminus \{\bar{x}\}$, then f is a strongly semismooth function.

Proof. First, we prove that f is Lipschitz continuous on \mathbf{R}^n . Clearly, $f(\bar{x}) = 0$. Since f and ∇f are continuous on the compact unit sphere $S^0 := \{x \in \mathbf{R}^n : \|x - \bar{x}\| = 1\}$, then there exists scalar $L > 0$ such that $|f(x)| \leq L$ and $\|\nabla f(x)\| \leq L$ for all $x \in S^0$. Using the fact that the arlength of the geodesic on S^0 between any two points $x, y \in S^0$ is at most $\frac{\pi}{2}\|x - y\|$, it is readily shown by parameterizing the geodesic that $|f(x) - f(y)| \leq \frac{\pi}{2}L\|x - y\|$. Consider any $h, k \in \mathbf{R}^n$. If $h \neq 0$ and $k \neq 0$, we have

$$\begin{aligned} f(\bar{x} + h) - f(\bar{x} + k) &= f\left(\bar{x} + \|h\|\frac{h}{\|h\|}\right) - f\left(\bar{x} + \|k\|\frac{k}{\|k\|}\right) \\ &= \|h\|f\left(\bar{x} + \frac{h}{\|h\|}\right) - \|k\|f\left(\bar{x} + \frac{k}{\|k\|}\right) \\ &= (\|h\| - \|k\|)f\left(\bar{x} + \frac{h}{\|h\|}\right) + \|k\|\left(f\left(\bar{x} + \frac{h}{\|h\|}\right) - f\left(\bar{x} + \frac{k}{\|k\|}\right)\right), \end{aligned}$$

and hence

$$\begin{aligned} |f(\bar{x} + h) - f(\bar{x} + k)| &\leq \left|\|h\| - \|k\|\right|L + \|k\|\frac{\pi}{2}L\left\|\frac{h}{\|h\|} - \frac{k}{\|k\|}\right\| \\ &\leq \left|\|h\| - \|k\|\right|L + \|k\|\frac{\pi}{2}L\left(\left\|\frac{h}{\|h\|} - \frac{h}{\|k\|}\right\| + \left\|\frac{h}{\|k\|} - \frac{k}{\|k\|}\right\|\right) \\ &= \left|\|h\| - \|k\|\right|(L + \frac{\pi}{2}L) + \frac{\pi}{2}L\|h - k\| \\ &\leq \|h - k\|(L + \frac{\pi}{2}L) + \frac{\pi}{2}L\|h - k\|. \end{aligned}$$

If $h \neq 0$ and $k = 0$, then

$$|f(\bar{x} + h) - f(\bar{x} + k)| = \|h\|\left|f\left(\bar{x} + \frac{h}{\|h\|}\right)\right| \leq \|h\|L = \|h - k\|L.$$

The case of $h = 0$ and $k \neq 0$ can be treated similarly.

For any $h \in \mathbf{R}^n$ with $h \neq 0$, f is smooth at $\bar{x} + h$ and

$$\begin{aligned} f(\bar{x} + h) - f(\bar{x}) - \nabla f(\bar{x} + h)^T h &= f(\bar{x} + h) - f'(\bar{x} + h; h) \\ &= f(\bar{x} + h) - \lim_{t \downarrow 0} \frac{f(\bar{x} + h + th) - f(\bar{x} + h)}{t} \\ &= 0. \end{aligned}$$

This shows that f is strongly semismooth at \bar{x} . Since f is smooth everywhere except at \bar{x} , then f is semismooth everywhere. If ∇f is locally Lipschitzian on $\mathbf{R}^n \setminus \{\bar{x}\}$, then f is strongly semismooth everywhere, i.e., it is a strongly semismooth function. \square

4 Further Results, Examples, and Applications

In this section, we make further studies of AS functions, with more examples and applications.

4.1 p -Norm Functions and NCP Functions

The p -norm function $f : \mathbf{R}^n \rightarrow \mathbf{R}$, defined by

$$f(x) := \|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

for $1 < p < \infty$, is smooth everywhere except at the origin. In particular, $X_f = F_f = \mathbf{R}^n \setminus \{0\}$. It is positively homogeneous about the origin. Hence, when $n \geq 2$, by Theorem 5, it is a basic AS function. If $p \geq 2$, then it is basic strongly AS. More generally, the function

$$f(x) := \sum_{i=1}^N \|F_i(x)\|_{p_i}$$

where $1 < p_i < \infty$, $n_i \geq 2$, $F_i : \mathbf{R}^m \rightarrow \mathbf{R}^{n_i}$ is a smooth mapping whose Jacobian has rank greater than 1 on \mathbf{R}^m , is an AS function defined on \mathbf{R}^m . These p -norm functions arise frequently in nonlinear optimization [38, 47].

An **NCP function** is a function $f : \mathbf{R}^2 \rightarrow \mathbf{R}$ with the property that $f(a, b) = 0$ if and only if $a \geq 0$, $b \geq 0$, $ab = 0$. A well-known NCP function closely related to 2-norm is the **Fischer-Burmeister** function [16], defined by

$$f(a, b) := \|(a, b)\|_2 - a - b \quad \forall (a, b) \in \mathbf{R}^2.$$

This function is basic strongly AS. It is used extensively in the solution of nonlinear complementarity and variational inequality problems [11, 14, 15, 16, 19, 33, 48]. Qi [32] showed that the Fischer-Burmeister function and its several variants, such as the Tseng-Luo NCP function [24, 45] and the Kanzow-Kleinmichel NCP function [20], are smooth away from the origin and are strongly semismooth at the origin.⁴ The Fischer-Burmeister function and these variants are irrational. Qi [32] proposed a class of piecewise rational NCP functions having the same strongly semismooth property. All these functions are basic strongly AS.

4.2 Convex best Interpolation

By using the first-order optimality condition and duality, a certain convex best interpolation problem may be reformulated equivalently as a system of *nonsmooth equations*

$$F(\lambda) = d, \tag{14}$$

where the i th component of F has the form

$$F_i(\lambda) := \int_a^b \left(\sum_{l=1}^N \lambda_l B_l(t) \right)_+ B_i(t) dt.$$

Here B_i is the normalized B-spline of order two associated with the problem data.

⁴This can also be deduced from [21, Theorem 6.18(ii)] and Theorem 4.

Irvine, Marin and Smith [18] proposed in 1986 a Newton-type method for solving the equation (14). They observed fast convergence in numerical experiments and raised the question of theoretically estimating the rate of convergence. Dontchev, Qi and Qi [12] answered this question by proving that F is semismooth. Dontchev, Qi and Qi [13] further proved that F is strongly semismooth, based on which quadratic convergence of the above method was proved. Specifically, they proved that

$$\begin{aligned} F_1(\lambda) &= \Phi_1(\lambda_1) + \Psi_1(\lambda_1, \lambda_2), \\ F_i(\lambda) &= \Gamma_i(\lambda_{i-1}, \lambda_i) + \Psi_i(\lambda_i, \lambda_{i+1}), \quad i = 2, \dots, N-1, \\ F_N(\lambda) &= \Gamma_N(\lambda_{N-1}, \lambda_N) + \Phi_2(\lambda_N), \end{aligned}$$

where Φ_1 and Φ_2 are piecewise linear, and each Γ_i and Ψ_i are strongly semismooth, and smooth with a Lipschitzian gradient away from the origin. Moreover, they showed that each Γ_i and Ψ_i are not PS, and neither is F . This also follows from Theorem 1(c) or Corollary 1. Using the terminology in this paper, we see that Γ_i and Ψ_i are basic strongly AS functions.

4.3 Smoothing/Penalty Functions

A popular technique to deal with nonsmooth functions is to approximate them by smooth functions [9, 35]. Suppose a function $g : \mathbf{R}^n \rightarrow \mathbf{R}$ is nonsmooth. We choose a locally Lipschitzian function $f : \mathbf{R}^{n+1} \rightarrow \mathbf{R}$ such that

$$f(x, 0) = g(x) \quad \forall x \in \mathbf{R}^n$$

and f is smooth at all (x, t) with $t \neq 0$. Then, instead of g , we work with f at $t \neq 0$ and drive $t \rightarrow 0$. Such a function f is called a **smoothing function** of g .

Since a smoothing function f is locally Lipschitzian and is smooth at all (x, t) with $t \neq 0$, it is readily seen that f satisfies the definition of a weakly AS function *except* for possibly the local connectedness of X_f at each $(\bar{x}, 0) \in \mathbf{R}^{n+1} \setminus X_f$, i.e., property (6). The local connectedness of X_f needs to be checked separately since in general we only know that X_f contains $\{(x, t) \in \mathbf{R}^{n+1} : t \neq 0\}$, which is not locally connected around its boundary points. We will also be interested in smoothing functions that are (strongly) AS, which means checking (strong) semismoothness of f . We do this below for some popular classes of smoothing/penalty functions.

Let $\psi : \mathbf{R}^n \rightarrow \mathbf{R}$ be a smooth convex function. Consider the function $f : \mathbf{R}^{n+1} \rightarrow \mathbf{R}$ defined by

$$f(x, t) := \begin{cases} |t|\psi(x/|t|) & \text{if } t \neq 0, \\ \psi_\infty(x) & \text{if } t = 0, \end{cases} \quad (15)$$

where ψ_∞ denotes the **recession function** of ψ , i.e., $\psi_\infty(x) := \lim_{t \downarrow 0} t\psi(x/t)$ [40, Corollary 8.5.2]. It is known that ψ_∞ is proper closed convex and positively homogeneous about the origin. Two well-known examples of ψ are

$$\psi(x) = \sum_{i=1}^n \sqrt{1 + x_i^2}, \quad (16)$$

$$\psi(x) = \ln \left(\sum_{i=1}^n e^{x_i} \right), \quad (17)$$

with $\psi_\infty(x) = \|x\|_1$ and $\psi_\infty(x) = \max\{x_1, \dots, x_n\}$, respectively. The exponential penalty function ([3, 4, 29, 30], [40, page 68]) corresponds to $f(x, t)$ with ψ given by (17). The **Chen-Mangasarian** class of smoothing functions for $g(x) = x_+ = \max\{0, x\}$ corresponds to $f(x, t)$, with $\psi : \mathbf{R} \mapsto \mathbf{R}$ being a smooth convex function satisfying $\lim_{x \rightarrow -\infty} \psi(x) = \lim_{x \rightarrow \infty} \psi(x) - x = 0$ [7, 8, 46]. This class includes the popular **Chen-Harker-Kanzow-Smale (CHKS) function** [5, 17, 44], which is a special case of $f(x, t)$ with

$$\psi(x) = \frac{x + \sqrt{x^2 + 1}}{2}.$$

It also includes the “neural network” function [7, 8], which is a special case of $f(x, t)$ with

$$\psi(x) = \ln(e^x + 1).$$

Here $\psi_\infty(x) = \max\{0, x\}$.

The relationship between ψ_∞ and the parameterized smoothing/penalty function $t\psi(\cdot/t)$ has been much studied, particular in the context of penalty functions for constrained optimization; see [3, 8, 46] and references therein. Moreover, it is known that f is a convex function on $\mathbf{R}^n \times \mathbf{R}_+$ [40, Theorem 8.2]. It readily follows that f is a convex function on $\mathbf{R}^n \times \mathbf{R}_-$ also. However, f need not be a convex function on \mathbf{R}^{n+1} . The theorem below shows that f is convex on \mathbf{R}^{n+1} provided ψ_∞ is majorized by ψ .

Theorem 6 *Let $\psi : \mathbf{R}^n \rightarrow \mathbf{R}$ be a convex function satisfying $\psi_\infty(x) \leq \psi(x)$ for all $x \in \mathbf{R}^n$. Then f defined by (15) is a convex function on \mathbf{R}^{n+1} .*

Proof. Fix any $(x, t), (y, s) \in \mathbf{R}^{n+1}$. We will prove that

$$f((1 - \lambda)(x, t) + \lambda(y, s)) \leq (1 - \lambda)f(x, t) + \lambda f(y, s) \quad \forall 0 \leq \lambda \leq 1.$$

If $t \geq 0, s \geq 0$, this follows from the convexity of f on $\mathbf{R}^n \times \mathbf{R}_+$. If $t \leq 0, s \leq 0$, this follows from the convexity of f on $\mathbf{R}^n \times \mathbf{R}_-$. Suppose $t > 0$ and $s < 0$. (The case of $t < 0$ and $s > 0$ can be treated similarly.) Let $\lambda_0 := t/(t - s)$, which satisfies $(1 - \lambda_0)t + \lambda_0 s = 0$. Direct calculation shows that $(1 - \lambda_0)t = \lambda_0(-s) = \beta$, where $\beta := -st/(t - s)$. Thus,

$$\begin{aligned} f((1 - \lambda_0)x + \lambda_0 y, 0) &= \psi_\infty((1 - \lambda_0)x + \lambda_0 y) \\ &= \psi_\infty\left(\beta \frac{x}{t} + \beta \frac{y}{(-s)}\right) \\ &= 2\beta \psi_\infty\left(\frac{1}{2}\left(\frac{x}{t} + \frac{y}{(-s)}\right)\right) \\ &\leq \beta \left(\psi_\infty\left(\frac{x}{t}\right) + \psi_\infty\left(\frac{y}{(-s)}\right)\right) \end{aligned}$$

$$\begin{aligned}
&\leq \beta \left(\psi \left(\frac{x}{t} \right) + \psi \left(\frac{y}{(-s)} \right) \right) \\
&= (1 - \lambda_0)t\psi \left(\frac{x}{t} \right) + \lambda_0(-s)\psi \left(\frac{y}{(-s)} \right) \\
&= (1 - \lambda_0)f(x, t) + \lambda_0f(y, s), \tag{18}
\end{aligned}$$

where the third equality is due to ψ_∞ being positively homogeneous about the origin. For $0 < \lambda < \lambda_0$, we have $(1 - \lambda)t + \lambda s > 0$ and hence the convexity of f on $\mathbf{R}^n \times \mathbf{R}_+$ yields

$$\begin{aligned}
f((1 - \lambda)(x, t) + \lambda(y, s)) &= f \left(\left(1 - \frac{\lambda}{\lambda_0} \right) (x, t) + \frac{\lambda}{\lambda_0} ((1 - \lambda_0)x + \lambda_0 y, 0) \right) \\
&\leq \left(1 - \frac{\lambda}{\lambda_0} \right) f(x, t) + \frac{\lambda}{\lambda_0} f((1 - \lambda_0)x + \lambda_0 y, 0) \\
&\leq \left(1 - \frac{\lambda}{\lambda_0} \right) f(x, t) + \frac{\lambda}{\lambda_0} ((1 - \lambda_0)f(x, t) + \lambda_0 f(y, s)) \\
&= (1 - \lambda)f(x, t) + \lambda f(y, s),
\end{aligned}$$

where the second inequality uses (18). Similarly, for $\lambda_0 < \lambda < 1$, we have $(1 - \lambda)t + \lambda s < 0$ and hence the convexity of f on $\mathbf{R}^n \times \mathbf{R}_-$ yields

$$\begin{aligned}
f((1 - \lambda)(x, t) + \lambda(y, s)) &= f \left(\left(1 - \frac{1 - \lambda}{1 - \lambda_0} \right) (y, s) + \frac{1 - \lambda}{1 - \lambda_0} ((1 - \lambda_0)x + \lambda_0 y, 0) \right) \\
&\leq \left(1 - \frac{1 - \lambda}{1 - \lambda_0} \right) f(y, s) + \frac{1 - \lambda}{1 - \lambda_0} f((1 - \lambda_0)x + \lambda_0 y, 0) \\
&\leq \left(1 - \frac{1 - \lambda}{1 - \lambda_0} \right) f(y, s) + \frac{1 - \lambda}{1 - \lambda_0} ((1 - \lambda_0)f(x, t) + \lambda_0 f(y, s)) \\
&= (1 - \lambda)f(x, t) + \lambda f(y, s),
\end{aligned}$$

where the second inequality uses (18). □

Since a real-valued convex function is locally Lipschitzian and directionally differentiable everywhere [40], it follows from Theorem 6 that f is locally Lipschitzian and directionally differentiable everywhere provided ψ_∞ is real-valued and majorized by ψ . In particular, this holds for the examples (16), (17), as well as the CM class of smoothing functions. Notice that ψ being positive-valued is not sufficient for ψ_∞ to be majorized by ψ . An example is $\psi(x) = \sqrt{1 + x^2} - c$ with $0 < c < 1$.

Suppose f is locally Lipschitzian and directionally differentiable everywhere. Since f is smooth everywhere except possibly on $\mathbf{R}^n \times \{0\}$, then f is semismooth provided that, for any $\bar{x} \in \mathbf{R}^n$,

$$f(x, t) - f(\bar{x}, 0) - f'((x, t); (x - \bar{x}, t)) = o(\|x - \bar{x}\| + |t|) \tag{19}$$

for all (x, t) in some neighborhood of $(\bar{x}, 0)$. Similarly, f is strongly semismooth provided that

$$f(x, t) - f(\bar{x}, 0) - f'((x, t); (x - \bar{x}, t)) = O(\|x - \bar{x}\|^2 + |t|^2). \tag{20}$$

If $t = 0$, then $f(x, t) = \psi_\infty(x)$ and $f'((x, t); (x - \bar{x}, t)) = \psi'_\infty(x; (x - \bar{x}))$. Thus, provided that ψ_∞ is semismooth (respectively, strongly semismooth), (19) (respectively, (20)) would hold for all $(x, 0)$ in some neighborhood of $(\bar{x}, 0)$. For any (x, t) with $t \neq 0$, the following lemma gives a simplification of the left-hand side of (19).

Lemma 1 *Let $\psi : \mathbf{R}^n \rightarrow \mathbf{R}$ be a smooth convex function. Let f be defined by (15). For any $\bar{x} \in \mathbf{R}^n$ and any $(x, t) \in \mathbf{R}^{n+1}$ with $t \neq 0$,*

$$0 \geq f(x, t) - f(\bar{x}, 0) - \nabla f(x, t)^T(x - \bar{x}, t) = \nabla\psi(x/|t|)^T\bar{x} - \psi_\infty(\bar{x}). \quad (21)$$

Proof. For any (x, t) with $t > 0$, we have from (15) that $\nabla f(x, t) = (\nabla\psi(x/t), \psi(x/t) - \nabla\psi(x/t)^T x/t)$. This together with the convexity of f on $\mathbf{R}^n \times \mathbf{R}_+$ yields

$$\begin{aligned} 0 &\geq f(x, t) - f(\bar{x}, 0) - \nabla f(x, t)^T(x - \bar{x}, t) \\ &= t\psi(x/t) - \psi_\infty(\bar{x}) - \left(\nabla\psi(x/t)^T(x - \bar{x}) + (\psi(x/t) - \nabla\psi(x/t)^T x/t)t \right) \\ &= \nabla\psi(x/t)^T\bar{x} - \psi_\infty(\bar{x}). \end{aligned}$$

For $t < 0$, we have $\nabla f(x, t) = (\nabla\psi(-x/t), -\psi(-x/t) - \nabla\psi(-x/t)^T x/t)$. This together with the convexity of f on $\mathbf{R}^n \times \mathbf{R}_-$ yields

$$\begin{aligned} 0 &\geq f(x, t) - f(\bar{x}, 0) - \nabla f(x, t)^T(x - \bar{x}, t) \\ &= -t\psi(-x/t) - \psi_\infty(\bar{x}) - \left(\nabla\psi(-x/t)^T(x - \bar{x}) + (-\psi(-x/t) - \nabla\psi(-x/t)^T x/t)t \right) \\ &= \nabla\psi(-x/t)^T\bar{x} - \psi_\infty(\bar{x}). \end{aligned}$$

(The argument for the case of $t < 0$ can alternatively be deduced from the case of $t > 0$ by using $f(x, t) = f(x, -t)$ for all (x, t) .) \square

Example 5 *Suppose ψ is given by (17). Then $\nabla\psi(x) = \frac{1}{e^{x_1} + \dots + e^{x_n}}(e^{x_1}, \dots, e^{x_n})$. Without loss of generality, assume that $\bar{x}_1 = \max\{\bar{x}_1, \dots, \bar{x}_n\} = \psi_\infty(\bar{x})$ and let $\mathcal{I} := \{i \in \{1, \dots, n\} : \bar{x}_i \neq \bar{x}_1\}$ and $\Delta := \bar{x}_1 - \max_{i \in \mathcal{I}} \bar{x}_i$. Then, whenever $\|x - \bar{x}\|_\infty \leq \Delta/4$, we have $x_i - x_1 \leq -\Delta/2$ for all $i \in \mathcal{I}$ so that*

$$\begin{aligned} \nabla\psi(x/|t|)^T\bar{x} - \psi_\infty(\bar{x}) &= \frac{\sum_{i=1}^n e^{x_i/|t|}\bar{x}_i}{\sum_{i=1}^n e^{x_i/|t|}} - \bar{x}_1 \\ &= \frac{\bar{x}_1 + \sum_{i \neq 1} e^{(x_i - x_1)/|t|}\bar{x}_i}{1 + \sum_{i \neq 1} e^{(x_i - x_1)/|t|}} - \bar{x}_1 \\ &= \frac{\sum_{i \in \mathcal{I}} e^{(x_i - x_1)/|t|}(\bar{x}_i - \bar{x}_1)}{1 + \sum_{i \neq 1} e^{(x_i - x_1)/|t|}} \\ &\geq e^{-0.5\Delta/|t|} \sum_{i \in \mathcal{I}} (\bar{x}_i - \bar{x}_1). \end{aligned}$$

This together with (21) shows that f defined by (15) satisfies

$$\left| f(x, t) - f(\bar{x}, 0) - \nabla f(x, t)^T (x - \bar{x}, t) \right| \leq e^{-0.5\Delta/|t|} \sum_{i \in \mathcal{I}} (\bar{x}_1 - \bar{x}_i) \quad \forall t > 0.$$

Notice that $\lim_{t \rightarrow 0} e^{-0.5\Delta/|t|}/|t|^k = 0$ for any $k > 0$ and, in particular, for $k = 2$. Since ψ_∞ is piecewise linear and hence strongly semismooth, then (20) holds, i.e., f is strongly semismooth at $(\bar{x}, 0)$. In fact, f is semismooth of order k at $(\bar{x}, 0)$ for any $k > 0$.

Since f is twice continuously differentiable at any (\bar{x}, \bar{t}) with $\bar{t} \neq 0$, this shows that f is strongly semismooth on \mathbf{R}^{n+1} [30].

Next, we show that

$$X_f = \{(x, t) : x \in \mathbf{R}^n, t \neq 0\} \cup \{(x, 0) : \nu(x) = 1\},$$

where $\nu(x) := \text{Card}\{i \in \{1, \dots, n\} : \psi_\infty(x) = x_i\}$. To see this, note that f is smooth at every (x, t) with $t \neq 0$ and f is not differentiable at every $(x, 0)$ with $\nu(x) \geq 2$. Thus, it suffices to verify that f is smooth at every $(x, 0)$ with $\nu(x) = 1$. Fix any $\bar{x} \in \mathbf{R}^n$ with $\nu(\bar{x}) = 1$. Without loss of generality we can assume that $\bar{x}_1 > \bar{x}_i$ for $i = 2, \dots, n$. Then, it is straightforward to verify that f is differentiable at $(\bar{x}, 0)$ with $\nabla f(\bar{x}, 0) = (1, 0, \dots, 0)$. For all (x, t) with $t > 0$, we have that

$$\begin{aligned} \nabla f(x, t) &= (\nabla \psi(x/t), \psi(x/t) - \nabla \psi(x/t)^T x/t) \\ &= \left(\frac{(e^{x_1/t}, \dots, e^{x_n/t})}{\sum_{i=1}^n e^{x_i/t}}, \psi(x/t) - \frac{(e^{x_1/t} x_1/t + \dots + e^{x_n/t} x_n/t)}{\sum_{i=1}^n e^{x_i/t}} \right) \\ &= \left(\frac{(1, \dots, e^{(x_n - x_1)/t})}{1 + \sum_{i \neq 1} e^{(x_i - x_1)/t}}, \ln \left(1 + \sum_{i \neq 1} e^{(x_i - x_1)/t} \right) + \frac{x_1}{t} - \frac{x_1/t + \dots + e^{(x_n - x_1)/t} x_n/t}{1 + \sum_{i \neq 1} e^{(x_i - x_1)/t}} \right). \end{aligned}$$

As $(x, t) \rightarrow (\bar{x}, 0)$, we have $x_i - x_1 \rightarrow \bar{x}_i - \bar{x}_1 < 0$ and hence $e^{(x_i - x_1)/t}/t \rightarrow 0$ for all $i \neq 1$. Then the above formula yields that $\nabla f(x, t) \rightarrow (1, 0, \dots, 0)$. For all (x, t) with $t < 0$, the same conclusion is reached. For all $(x, 0)$ with x sufficiently near \bar{x} , we have that $\nu(x) = 1$ and hence f is differentiable at $(x, 0)$ with $\nabla f(x) = (1, 0, \dots, 0)$. This verifies that f is smooth at every $(x, 0)$ with $\nu(x) = 1$.

Finally, we verify that X_f is locally connected, which would establish that f is strongly AS. Fix any $\bar{x} \in \mathbf{R}^n$ with $\nu(\bar{x}) \geq 2$ and any $\varepsilon > 0$. $X_f \cap \mathbf{B}_\varepsilon((\bar{x}, 0))$ is nonempty since it contains $(\bar{x}, \varepsilon/2)$. Consider any two points (x^1, t^1) and (x^2, t^2) in $X_f \cap \mathbf{B}_\varepsilon((\bar{x}, 0))$. If t^1 and t^2 have the same positive or negative sign, then the line segment joining them lies in $X_f \cap \mathbf{B}_\varepsilon((\bar{x}, 0))$. The same is true if exactly one of t^1 and t^2 is nonzero. If both t^1 and t^2 are zero, then the line segments from (x^1, t^1) to $(\bar{x}, \varepsilon/2)$ and then from $(\bar{x}, \varepsilon/2)$ to (x^2, t^2) would lie in $X_f \cap \mathbf{B}_\varepsilon((\bar{x}, 0))$. If $t^1 > 0$ and $t^2 < 0$, then the line segments from (x^1, t^1) to $(x^3, 0)$ and then from $(x^3, 0)$ to (x^2, t^2) would lie in $X_f \cap \mathbf{B}_\varepsilon((\bar{x}, 0))$, where x^3 is a small perturbation of \bar{x} so that $\nu(x^3) = 1$ and $(x^3, 0) \in \mathbf{B}_\varepsilon((\bar{x}, 0))$. Thus, $X_f \cap \mathbf{B}_\varepsilon((\bar{x}, 0))$ is nonempty and connected.

Notice that f is not basic AS when $n \geq 2$ since $\mathbf{R}^{n+1} \setminus X_f = \{(x, 0) \in \mathbf{R}^{n+1} : x_i = x_j \text{ for some } 1 \leq i < j \leq n\}$, being the union of $\frac{1}{2}n(n-1)$ subspaces of dimension $n-1$, is not comprised of isolated points.

Suppose $\psi : \mathbf{R} \rightarrow \mathbf{R}$ is smooth convex and ψ_∞ is real-valued. Then it is known that ψ' is a nondecreasing function and $\psi_\infty(x) = \max\{\alpha x, \beta x\}$, where $\alpha := \lim_{x \rightarrow -\infty} \psi'(x) > -\infty$ and $\beta := \lim_{x \rightarrow \infty} \psi'(x) < \infty$. We note that f is semismooth at $(\bar{x}, 0)$ for any $\bar{x} \in \mathbf{R}$ if and only if

$$\lim_{x \rightarrow -\infty} (\psi'(x) - \alpha)x = 0, \quad \lim_{x \rightarrow \infty} (\beta - \psi'(x))x = 0. \quad (22)$$

This is due to ψ_∞ being piecewise linear (so that (20) holds for all $(x, 0)$ near $(\bar{x}, 0)$), (21), and the observation that

$$\psi'(x/|t|)\bar{x} - \psi_\infty(\bar{x}) = \begin{cases} (\psi'(x/|t|) - \alpha)\bar{x} & \text{if } \bar{x} < 0, \\ 0 & \text{if } \bar{x} = 0, \\ (\psi'(x/|t|) - \beta)\bar{x} & \text{if } \bar{x} > 0. \end{cases}$$

Similarly, for any $1 < k \leq 2$, we have that f defined by (15) is semismooth of order k at $(\bar{x}, 0)$ for any $\bar{x} \in \mathbf{R}$ if and only if

$$\limsup_{x \rightarrow -\infty} (\psi'(x) - \alpha)|x|^k < \infty, \quad \limsup_{x \rightarrow \infty} (\beta - \psi'(x))x^k < \infty.$$

Suppose that, in addition to (22), we have $\alpha \neq \beta$ and

$$\lim_{x \rightarrow -\infty} \psi(x) - \alpha x = 0, \quad \lim_{x \rightarrow \infty} \psi(x) - \beta x = 0. \quad (23)$$

Then we claim that

$$X_f = \mathbf{R}^2 \setminus \{(0, 0)\}.$$

To see this, note that f is smooth at every (x, t) with $t \neq 0$ and, due to $\alpha \neq \beta$, f is not differentiable at $(0, 0)$. Thus, it suffices to verify that f is smooth at every $(x, 0)$ with $x \neq 0$. Fix any $\bar{x} \in \mathbf{R}$ with $\bar{x} > 0$. It is straightforward to verify using (23) that f is differentiable at $(\bar{x}, 0)$ with $\nabla f(\bar{x}, 0) = (\beta, 0)$. For all (x, t) with $t > 0$, we have

$$\nabla f(x, t) = (\psi'(x/t), \psi(x/t) - \psi'(x/t)x/t).$$

As $(x, t) \rightarrow (\bar{x}, 0)$, we have $x/t \rightarrow \infty$ and hence $\psi'(x/t) \rightarrow \beta$. Also, (22) and (23) imply that $\psi(x/t) - \psi'(x/t)x/t \rightarrow 0$. Thus, $\nabla f(x, t) \rightarrow (\beta, 0)$. For all (x, t) with $t < 0$, the same conclusion is reached. For all $(x, 0)$ with x sufficiently near \bar{x} , we have that $x > 0$ and hence f is differentiable at $(x, 0)$ with $\nabla f(x) = (\beta, 0)$. This verifies that f is smooth at every $(x, 0)$ with $x > 0$. The case of $x < 0$ can be treated similarly.

We illustrate the above results with three examples below.

Example 6 Suppose $\psi(x) = \sqrt{1+x^2}$ with $x \in \mathbf{R}$. Then, $\psi_\infty(x) = |x|$ and $\alpha = -1$ and $\beta = 1$. Moreover,

$$(\psi'(x) - \alpha)x^2 = \left(\frac{x}{\sqrt{1+x^2}} + 1 \right) x^2 = \frac{x^2}{\sqrt{1+x^2}(\sqrt{1+x^2} - x)} \rightarrow \frac{1}{2} \quad \text{as } x \rightarrow -\infty.$$

$$(\beta - \psi'(x))x^2 = \left(1 - \frac{x}{\sqrt{1+x^2}} \right) x^2 = \frac{x^2}{\sqrt{1+x^2}(\sqrt{1+x^2} + x)} \rightarrow \frac{1}{2} \quad \text{as } x \rightarrow \infty.$$

Thus, by the above fact, f defined by (15) is strongly semismooth at $(\bar{x}, 0)$. Since f is twice continuously differentiable at any (\bar{x}, \bar{t}) with $\bar{t} \neq 0$, then f is strongly semismooth on \mathbf{R}^{n+1} . Also, ψ_∞ is majorized by ψ and (23) holds, so f is locally Lipschitzian, directionally differentiable, and smooth everywhere except at $(0, 0)$. Thus f is basic strongly AS. In particular, this shows that the CHKS function is basic strongly AS. Since f is positively homogeneous about $(0, 0)$, Theorem 5 shows that f is in fact Lipschitz continuous.

Example 7 Suppose $\psi : \mathbf{R} \rightarrow \mathbf{R}$ is given by

$$\psi(x) = \begin{cases} 2x + \frac{e}{\ln x} & \text{if } x \geq e, \\ e^{x-e} + 3e - 1 & \text{if } x \leq e. \end{cases}$$

This is a smooth convex function and, in particular, its derivative

$$\psi'(x) = \begin{cases} 2 - \frac{e}{(\ln x)^2 x} & \text{if } x \geq e, \\ e^{x-e} & \text{if } x \leq e, \end{cases}$$

is an increasing function. Also, $\psi_\infty(x) = 2 \max\{0, x\}$ and $\alpha = 0, \beta = 2$. We have

$$\begin{aligned} (\psi'(x) - \alpha)x &= e^{x-e}x \rightarrow 0 \quad \text{as } x \rightarrow -\infty, \\ (\beta - \psi'(x))x &= \frac{e}{(\ln x)^2} \rightarrow 0 \quad \text{as } x \rightarrow \infty, \end{aligned}$$

so (22) holds. Thus, by the above result, f defined by (15) is semismooth at $(\bar{x}, 0)$ for any $\bar{x} \in \mathbf{R}$. Hence f is semismooth on \mathbf{R}^2 . However, it is easily seen that $(\beta - \psi'(x))x^k \rightarrow \infty$ for any $k > 0$, so f is not semismooth of order k at $(\bar{x}, 0)$. Also, ψ_∞ is majorized by ψ and (23) holds, so f is locally Lipschitzian, directionally differentiable, smooth everywhere except at $(0, 0)$. Thus f is basic AS. Since f is positively homogeneous about $(0, 0)$, Theorem 5 shows that f is in fact Lipschitz continuous.

Example 8 Suppose $\psi : \mathbf{R} \rightarrow \mathbf{R}$ is given by

$$\psi(x) = \begin{cases} 2x - e \ln x & \text{if } x \geq e, \\ e^{x-e} + e - 1 & \text{if } x \leq e. \end{cases}$$

This is a smooth convex function and, in particular, its derivative

$$\psi'(x) = \begin{cases} 2 - \frac{e}{x} & \text{if } x \geq e, \\ e^{x-e} & \text{if } x \leq e, \end{cases}$$

is an increasing function. Also, $\psi_\infty(x) = 2 \max\{0, x\}$ and $\alpha = 0, \beta = 2$. We have

$$(\beta - \psi'(x))x = e \not\rightarrow 0 \quad \text{as } x \rightarrow \infty.$$

Thus f is not semismooth at $(\bar{x}, 0)$ when $\bar{x} > 0$.

As Example 8 shows, ψ being smooth convex and positive-valued is not sufficient for f given by (15) to be semismooth.

4.4 Integral Functions

The result of Dontchev, Qi and Qi [13] cited in Subsection 4.1 was proved by showing that an integral function involving B-splines is strongly semismooth. This result was then extended by Qi and Yin [37] to a more general class of functions; also see [34, Section 5]. In particular, they showed that if g is a continuous function on $[a, b]$ ($-\infty < a < b < \infty$) and u, v are two real-valued strongly semismooth functions on \mathbf{R}^n , then the integral function $f : \mathbf{R}^n \rightarrow \mathbf{R}$ defined by

$$f(x) := \int_a^b (tu(x) + v(x))_+ g(t) dt \quad (24)$$

is strongly semismooth on \mathbf{R}^n . This result was used in [37] to prove quadratic convergence of a Newton-type method proposed by Andersson, Elfving, Iliev and Vlaxhkova [2] for interpolation of convex scattered data. Notice that, although the integrand in (24) is strongly semismooth, this alone is not sufficient for the integral function to be strongly semismooth, as is shown by a counterexample in [37].

Since the integrand in (24) is PS for each t , is the function f defined by (24) a PS function also? In the case where u and v are linear, i.e., $u(x) = \bar{u}^T x$, $v(x) = \bar{v}^T x$ for some $\bar{u}, \bar{v} \in \mathbf{R}^n$, Qi and Yin proved the following result [37].

- (a) If \bar{u} and \bar{v} are linearly dependent, then f is piecewise linear and hence PS.
- (b) If \bar{u} and \bar{v} are linearly independent, then f is smooth and strongly semismooth on $\mathbf{R}^n \setminus \{x \in \mathbf{R}^n : \bar{u}^T x = \bar{v}^T x = 0\}$. If in addition $g \not\equiv 0$ on $[a, b]$, then f is strongly semismooth on \mathbf{R}^n and $X_f = F_f = \mathbf{R}^n \setminus \{x \in \mathbf{R}^n : \bar{u}^T x = \bar{v}^T x = 0\}$.

It readily follows from Definition 4 that if \bar{u}, \bar{v} are linearly independent, then f is basic strongly AS for $n = 2$ and f is strongly AS but not basic strongly AS for $n > 2$ (since $X_f = F_f$ is a subspace of dimension $n - 2$). Then, by Corollary 1, f is not PS whenever \bar{u} and \bar{v} are linearly independent.

What about nonlinear u, v ? We study this general case below. Suppose $u, v : \mathbf{R}^n \rightarrow \mathbf{R}$ are smooth and g is continuous on $[a, b]$ ($-\infty < a < b < \infty$). It can be shown that f defined by (24) is locally Lipschitzian and semismooth on \mathbf{R}^n ; see [12]. Moreover, by using the formula (24), it is straightforward to show that f is smooth at every $x \in \mathbf{R}^n$ such that $u(x), v(x)$ are not both zero. In particular,

$$\nabla f(x) = \begin{cases} 0 & \text{if } u(x) = 0, v(x) < 0, \\ \int_a^b (t\nabla u(x) + \nabla v(x))g(t)dt & \text{if } u(x) = 0, v(x) > 0, \\ \int_{\max\{a, r(x)\}}^{\max\{b, r(x)\}} (t\nabla u(x) + \nabla v(x))g(t)dt & \text{if } u(x) > 0, \\ \int_{\min\{a, r(x)\}}^{\min\{b, r(x)\}} (t\nabla u(x) + \nabla v(x))g(t)dt & \text{if } u(x) < 0, \end{cases}$$

where $r(x) := -v(x)/u(x)$. Thus $\mathbf{R}^n \setminus X_f \subseteq Y := \{x \in \mathbf{R}^n : u(x) = v(x) = 0\}$, so, in order to verify that f is AS, it suffices to show that for every $\bar{x} \in Y$ there is a $\bar{\varepsilon} > 0$

satisfying (6). Fix any $\bar{x} \in Y$. Suppose that $\nabla u(x), \nabla v(x)$ are linearly independent for all $x \in \mathbf{R}^n$ ($n \geq 2$). It follows from the implicit function theorem that there exist indices $1 \leq i < j \leq n$, open balls $\mathbf{B}_i \subset \mathbf{R}, \mathbf{B}_j \subset \mathbf{R}, \mathbf{B}_R \subset \mathbf{R}^{n-2}$ centered at $\bar{x}_i, \bar{x}_j, \bar{x}_R$,⁵ respectively, and differentiable functions $\varphi_i, \varphi_j : \mathbf{B}_R \rightarrow \mathbf{R}$ such that $x_i = \varphi_i(x_R), x_j = \varphi_j(x_R)$ if and only if $x \in Y, (x_i, x_j, x_R) \in \mathbf{B}_i \times \mathbf{B}_j \times \mathbf{B}_R$. Suppose that φ_i changes concavity finitely many times on each line segment lying in some open ball $\tilde{\mathbf{B}}_R \subseteq \mathbf{B}_R$ centered at \bar{x}_R and similarly for φ_j .⁶ (Roughly speaking, the manifold Y is not too “wiggly”.) Then we have that the open set

$$X := \mathbf{R}^n \setminus Y = \{(x_i, x_j, x_R) \in \mathbf{B}_i \times \mathbf{B}_j \times \tilde{\mathbf{B}}_R : x_i \neq \varphi_i(x_R) \text{ or } x_j \neq \varphi_j(x_R)\}$$

is nonempty and connected. This is because, for any two points $(\tilde{x}_i, \tilde{x}_j, \tilde{x}_R)$ and $(\hat{x}_i, \hat{x}_j, \hat{x}_R)$ in X , the line segment joining them intersects Y at (x_i, x_j, x_R) if and only if the line segment joining $(\tilde{x}_i, \tilde{x}_R)$ and (\hat{x}_i, \hat{x}_R) intersects the graph of φ_i at (x_i, x_R) and the line segment joining $(\tilde{x}_j, \tilde{x}_R)$ and (\hat{x}_j, \hat{x}_R) intersects the graph of φ_j at (x_j, x_R) . By our assumption on φ_i , the number of such intersection points on its graph is finite and similarly for φ_j . Thus, perturbing either \tilde{x}_i or \hat{x}_i changes the x_R component of each intersection point on the graph of φ_i while leaving the intersection points on the graph of φ_j unchanged. Since φ_i is continuous, this means that, for sufficiently small perturbation, the x_R component of each intersection point on the graph of φ_i differs from the x_R component of each intersection point on the graph of φ_j . Thus the perturbed line segment does not intersect Y and lies entirely in X . This shows that there is a $\varepsilon > 0$ satisfying (6). Hence f is AS under the above assumption on the (implicit) functions φ_i and φ_j defined around each $\bar{x} \in Y$. If in addition u and v are strongly semismooth, then the result of Qi and Yin [37] implies that f is strongly AS. By Corollary 1, f is either smooth or not PS.

Example 9 Suppose $n \geq 3$ and $u(x) = x_1 + p(x_3, \dots, x_n), v(x) = x_2 + q(x_3, \dots, x_n)$, where $p, q : \mathbf{R}^{n-2} \rightarrow \mathbf{R}$ are polynomial functions. Then u, v are infinitely differentiable and $\nabla u(x), \nabla v(x)$ are linearly independent for all $x \in \mathbf{R}^n$. Moreover, we can take $i = 1, j = 2$ and $\varphi_i \equiv -p, \varphi_j \equiv -q$. Then φ_i, φ_j are polynomial functions and hence they change concavity finitely many times on any line segment in \mathbf{R}^{n-2} . Thus, by the above discussion, if g is continuous on $[a, b]$, then f defined by (24) is strongly AS.

It is an open question whether we can relax the above assumption of finitely many concavity changes on the (implicit) functions φ_i, φ_j . In particular, complication arises if the graphs of these functions can have infinitely many intersection points with a line segment.

5 Appendix

In this appendix, we give an alternative proof of Theorem 2(c) that is a more direct extension of Rockafellar’s proof.

⁵For any $x \in \mathbf{R}^n$, we denote by x_R the reduced vector in \mathbf{R}^{n-2} obtained by dropping x_i, x_j from x .

⁶More precisely, for any $x_R, y_R \in \tilde{\mathbf{B}}_R$, the derivative of the function $t \mapsto \varphi_i((1-t)x_R + ty_R)$ changes sign a finite number of times over $0 \leq t \leq 1$, and the same holds for φ_j in place of φ_i .

Let $\bar{\varepsilon} > 0$ be small enough such that (1) holds with $\varepsilon = \bar{\varepsilon}$ and $\text{cl}X_{\bar{\varepsilon}} \subset O$, where we define

$$X_{\varepsilon} := X \cap \mathbf{B}_{\varepsilon}(\bar{x}) \quad \forall \varepsilon \in (0, \bar{\varepsilon}).$$

Notice that X_{ε} is nonempty for every $\varepsilon > 0$ since X is dense in O . For each $i \in I$, let

$$D_i := \{x \in O : f(x) = f_i(x), \nabla f(x) = \nabla f_i(x)\}.$$

On X_{ε} , the mapping ∇f is continuous, and it agrees with ∇f_i on $D_i \cap X_{\varepsilon}$. For each $x \in X_{\varepsilon}$, $\{f_i\}_{i \in I}$ forms a local representation for f at x and hence, by Theorem 1(b), $x \in D_i$ for some $i \in I$. This implies that the sets D_i , $i \in I$, cover X_{ε} for all $\varepsilon \in (0, \bar{\varepsilon})$. By Theorem 1(a), for each $i \in I$ there is an open set O_i such that $\bar{x} \in \text{cl}O_i$ and $f \equiv f_i$ on O_i . The latter implies f is smooth on O_i and $\nabla f \equiv \nabla f_i$ on O_i . Hence $D_i \supseteq O_i \cap \mathbf{B}_{\varepsilon}(\bar{x}) \neq \emptyset$ for all $\varepsilon \in (0, \bar{\varepsilon})$, so that $\bar{x} \in \text{cl}D_i$. Since D_i , $i \in I$, cover X_{ε} for all $\varepsilon \in (0, \bar{\varepsilon})$ and ∇f_i is continuous on each D_i , whose closure contains \bar{x} , we have

$$\limsup_{\varepsilon \downarrow 0} \nabla f(X_{\varepsilon}) \subseteq \{\nabla f_i(\bar{x}) : i \in I\}.$$

On the other hand, ∇f is continuous on X_{ε} and X_{ε} is nonempty and connected by (2), so the image $\nabla f(X_{\varepsilon})$ is nonempty and connected. Hence the set $\nabla f(X_{\varepsilon})$, which is uniformly bounded as $\varepsilon \downarrow 0$, must converge to some particular element of $\{\nabla f_i(\bar{x}) : i \in I\}$, say $\nabla f_{\bar{i}}(\bar{x})$ for some $\bar{i} \in I$. Thus, ∇f has a continuous extension from X to $X \cup \{\bar{x}\}$.

For each $i \in I$, since $O_i \cap X_{\varepsilon}$ is dense in $O_i \cap \mathbf{B}_{\varepsilon}(\bar{x})$ whose closure contains \bar{x} , then $\bar{x} \in \text{cl}(O_i \cap X_{\varepsilon})$. Since $\nabla f \equiv \nabla f_i$ on O_i , the continuity of ∇f_i on O_i implies

$$\nabla f_i(\bar{x}) = \limsup_{\varepsilon \downarrow 0} \nabla f(O_i \cap X_{\varepsilon}) \subseteq \limsup_{\varepsilon \downarrow 0} \nabla f(X_{\varepsilon}) = \nabla f_{\bar{i}}(\bar{x}). \quad (25)$$

Then, for any $x \in O \cap \mathbf{B}_{\varepsilon}(\bar{x})$, we have $f(x) = f_i(x)$ for some $i \in I$ and $f(\bar{x}) = f_i(\bar{x})$, $\nabla f_i(\bar{x}) = \nabla f_i(\bar{x})$. This, together with smoothness of f_i , implies

$$f(x) - f(\bar{x}) - \nabla f_{\bar{i}}(\bar{x})^T(x - \bar{x}) = f_i(x) - f_i(\bar{x}) - \nabla f_i(\bar{x})^T(x - \bar{x}) = o_i(h),$$

where $\lim_{\|h\| \rightarrow 0} o_i(h)/\|h\| = 0$. Since I is a finite set, this shows that f is differentiable at \bar{x} with $\nabla f(\bar{x}) = \nabla f_{\bar{i}}(\bar{x})$. The last claim of the theorem follows from (25).

Thus f is differentiable at \bar{x} and $\nabla f(\bar{x}) = \nabla f_i(\bar{x})$ for all $i \in I$. By (b), f is strictly differentiable at \bar{x} .

References

- [1] A. AGRACHEV, D. PALLASCHKE, AND S. SCHOLTES, *On Morse theory for piecewise smooth functions*, J. Dynam. Control Sys. 3 (1997), pp. 449-469.
- [2] L. E. ANDERSSON, T. ELFVING, G. ILIEV AND K. VLACHKOVA, *Interpolation of convex scattered data in \mathbf{R}^3 based upon an edged convex minimum norm network*, J. Approx. Theory 80 (1995), pp. 299-320.

- [3] A. AUSLENDER, *Penalty and barrier methods: a unified framework*, SIAM J. Optim. 10 (1999), pp. 211–230.
- [4] D. P. BERTSEKAS, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982. (Reprinted by Athena Scientific, Belmont, MA, 1996.)
- [5] B. CHEN AND P. T. HARKER, *A non-interior-point continuation method for linear complementarity problems*, SIAM J. Matrix Anal. Appl. 14 (1993), pp. 1168–1190.
- [6] B. CHEN AND N. XIU, *A global linear and local quadratic non-interior continuation method for nonlinear complementarity problems based on Chen-Mangasarian smoothing function*, SIAM J. Optim. 9 (1999), pp. 605–623.
- [7] C. CHEN AND O. L. MANGASARIAN, *Smoothing methods for convex inequalities and linear complementarity problems*, Math. Program. 71 (1995), pp. 51–69.
- [8] C. CHEN AND O. L. MANGASARIAN, *A class of smoothing functions for nonlinear and mixed complementarity problems*, Comput. Optim. Appl. 5 (1996), pp. 97–138.
- [9] X. CHEN, L. QI, AND D. SUN, *Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities*, Math. Comput. 67 (1998), pp. 519–540.
- [10] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, NY, 1983. (Reprinted by SIAM, Philadelphia, PA, 1990.)
- [11] T. DE LUCA, F. FACCHINEI AND C. KANZOW, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Program. 75 (1996), pp. 407–439.
- [12] A. L. DONTCHEV, H. D. QI AND L. QI, *Convergence of Newton’s method for convex best interpolation*, Numer. Math. 87 (2001), pp. 435–456.
- [13] A. L. DONTCHEV, H. D. QI AND L. QI, *Quadratic convergence of Newton’s method for convex interpolation and smoothing*, Constr. Approx. 19 (2003), pp. 123–143.
- [14] F. FACCHINEI AND C. KANZOW, *A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems*, Math. Program. 76 (1997), pp. 493–512.
- [15] F. FACCHINEI AND J.-S. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, New York, NY, 2003.
- [16] A. FISCHER, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Math. Program. 76 (1997), pp. 513–532.

- [17] K. HOTTA AND A. YOSHISE, *Global convergence of a class of non-interior-point algorithms using Chen-Harker-Kanzow functions for nonlinear complementarity problems*, Math. Program. 86 (1999), pp 105–133.
- [18] L. D. IRVINE, S. P. MARIN AND P. W. SMITH, *Constrained interpolation and smoothing*, Constr. Approx. 2 (1986), pp. 129–151.
- [19] H. JIANG, L. QI, X. CHEN AND D. SUN, *Semismoothness and superlinear convergence in nonsmooth optimization and nonsmooth equations*, Nonlinear optimization and applications (Erice, 1995), 197–212, G. Di Pillo and F. Giannessi eds. Plenum Publishing Corporation, New York, NY, 1996.
- [20] C. KANZOW AND H. KLEINMICHEL, *A new class of semismooth Newton-type methods for nonlinear complementarity problems*, Comput. Optim. Appl. 11 (1998), pp. 227–251.
- [21] D. KLATTE AND B. KUMMER, *Nonsmooth Equations in Optimization*, Kluwer, Dordrecht, The Netherlands, 2002.
- [22] L. KUNTZ AND S. SCHOLTES, *Structural analysis of nonsmooth mappings, inverse functions, and metric projections*, J. Math. Anal. Appl. 188 (1994), pp. 346–386.
- [23] L. KUNTZ AND S. SCHOLTES, *Qualitative aspects of the local approximation of a piecewise differentiable function*, Nonlin. Anal. 25 (1995), pp. 197–215.
- [24] Z. Q. LUO AND P. TSENG, *A new class of merit functions for the nonlinear complementarity problem*, Complementarity and Variational Problems: State of the Art, 204–225, M. C. Ferris and J. S. Pang, eds. SIAM, Philadelphia, PA, 1997.
- [25] R. MIFFLIN, *Semismoothness and semiconvex functions in constrained optimization*, SIAM J. Control Optim. 15 (1977), pp. 959–972.
- [26] B. S. MORDUKHOVICH, *Maximum principle in the problem of time optimal response with nonsmooth constraints*, J. Appl. Math. Mech. 40 (1976), pp. 960–969.
- [27] B. S. MORDUKHOVICH, *Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications*, Springer-Verlag, Berlin, 2006.
- [28] J.-S. PANG AND D. RALPH, *Piecewise smoothness, local invertibility, and parametric analysis of normal maps*, Math. Oper. Res., 21 (1996), pp. 401–426.
- [29] J. M. PENG AND Z. LIN, *A non-interior continuation method for generalized linear complementarity problems*, Math. Program. 86 (1999), pp. 533–563.
- [30] H. D. QI AND L. LIAO, *A smoothing Newton method for extended vertical linear complementarity problems*, SIAM J. Matrix Anal. Appl., 21 (1999), pp. 45–66.
- [31] L. QI, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res. 18 (1993), pp. 227–244.

- [32] L. QI, *Regular pseudo-smooth NCP and BVIP functions and globally and quadratically convergent generalized Newton methods for complementarity and variational inequality problems*, Math. Oper. Res. 24 (1999), pp. 440–471.
- [33] L. QI AND H. JIANG, *Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations*, Math. Oper. Res. 22 (1997), pp. 301–325.
- [34] L. QI, A. SHAPIRO, AND C. LING, *Differentiability and semismoothness properties of integral functions and their applications*, Math. Program. 102 (2005), pp. 223–248.
- [35] L. QI, D. SUN AND G. ZHOU, *A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities*, Math. Program. 87 (2000), pp. 1–35.
- [36] L. QI AND J. SUN, *A nonsmooth version of Newton's method*, Math. Program. 58 (1993), pp. 353–367.
- [37] L. QI AND H. YIN, *A strongly semismooth integral function and its application*, Comput. Optim. Appl. 25 (2003), pp. 223–246.
- [38] L. QI AND G. ZHOU, *A smoothing Newton method for minimizing a sum of Euclidean norms*, SIAM J. Optim. 11 (2000), pp. 389–410.
- [39] D. RALPH AND S. SCHOLTES, *Sensitivity analysis of composite piecewise smooth equations*, Math. Program. 76 (1997), pp. 593–612.
- [40] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- [41] R. T. ROCKAFELLAR, *A property of piecewise smooth functions*, Comput. Optim. Appl. 25 (2003), pp. 247–250.
- [42] R. T. ROCKAFELLAR AND R. J. B. WETS, *Variational Analysis*, Springer, New York, NY, 1998.
- [43] S. SCHOLTES, *Introduction to piecewise differentiable equations*, Habilitation Thesis, Preprint No. 54/1994, Institut für Statistik und Mathematische Wirtschaftstheorie, University of Karlsruhe, Karlsruhe, Germany, 1994.
- [44] S. SMALE, *Algorithms for solving equations*, Proceedings of the International Congress of Mathematicians, 172-195, Berkeley, California, CA, 1986.
- [45] P. TSENG, *Global behaviour of a class of merit functions for the nonlinear complementarity problem*, J. Optim. Theory Appl. 89 (1996), pp. 17-37.

- [46] P. TSENG, *Analysis of a non-interior continuation method based on Chen-Mangasarian smoothing functions for complementarity problems*, Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, 381–404, M. Fukushima and L. Qi, eds. Kluwer Academic Publisher, Norwell, USA, 1999.
- [47] G. A. WATSON, *On the Gauss-Newton method for l_1 orthogonal distance regression*, IMA J. Numer. Anal. 22 (2002), pp. 345–357.
- [48] N. YAMASHITA AND M. FUKUSHIMA, *Modified Newton methods for solving semismooth reformulations of monotone complementarity problems*, Math. Program. 76 (1997), pp. 469–491.