GRADIENT CONSISTENCY FOR INTEGRAL-CONVOLUTION SMOOTHING FUNCTIONS[1]

James V. Burke[2], Tim Hoheisel[3], and Christian Kanzow[3]

Preprint 309 September 2012

[2] University of Washington
Department of Mathematics
Box 354350
Seattle, Washington 98195-4350
e-mail: burke@math.washington.edu

[3] University of Würzburg
Institute of Mathematics
Campus Hubland Nord
Emil-Fischer-Str. 30
97074 Würzburg, Germany
e-mail: hoheisel@mathematik.uni-wuerzburg.de
kanzow@mathematik.uni-wuerzburg.de

September 4, 2012

[1]This research was partially supported by the DFG (Deutsche Forschungsgemeinschaft) under grant HO 4739/1-1.

Abstract. Chen and Mangasarian (1995) developed smoothing approximations to the plus function built on integral-convolution with density functions. X. Chen (2012) has recently picked up this idea constructing a large class of smoothing functions for nonsmooth minimization through composition with smooth mappings. In this paper, we generalize this idea by substituting the plus function for an arbitrary finite max-function. Calculus rules such as inner and outer composition with smooth mappings are provided, showing that the new class of smoothing functions satisfies, under reasonable assumptions, *gradient consistency*, a fundamental concept coined by Chen (2012). In particular, this guarantees the desired limiting behavior of critical points of the smooth approximations.

Key Words: Smoothing method, Subdifferential calculus, integral-convolution, piecewise-affine function

Mathematics Subject Classification (2010): 49J52,90C26, 90C30

1 Introduction

An unconstrained optimization problem takes the form

$$\min_{x \in \mathbb{R}^n} f(x), \tag{1}$$

for a function $f : \mathbb{R}^n \to \mathbb{R}$. In case that f is continuously differentiable, see, e.g., [1] or [21], powerful numerical solution methods have been introduced and successfully employed. Things are more subtle when the traditional smoothness assumption is not satisfied. An important class of potentially nonsmooth functions are the *convex* ones, see [13] for an overview on solution methods for convex optimization. A generalization of convexity on the one hand and continuous differentiability on the other, is *local Lipschitz continuity*, cf., e.g., [11] for solution procedures for locally Lipschitz problems.

A well-recognized technique for the numerical solution of (1) in the nonsmooth case is to replace f by a smooth approximation, and solve a sequence of smooth problems, while driving the approximation closer and closer to the original function, with the intention to approximate minimizers (critical points) of f by those of the smooth approximations. This technique of replacing a nonsmooth problem by a sequence of smooth problems is, in general, known as *smoothing* and it has been employed extensively for several different kinds of problems, see, e.g., [2, 5, 3, 6, 9, 12, 20], or the recent survey [6] which includes an extensive list of references. Certain smoothing methods are also closely related to the class of interior-point methods, cf. [16].

In this paper, we are concerned with a class of smoothing functions for *finite max-functions*, see Section 4 for a formal definition, which are special *piecewise affine mappings.* These approximations are shown to be well-behaved under both outer and inner composition with smooth functions such that a satisfactory calculus can be built up, and hence a class of smoothing approximations for a broad class of nonsmooth, locally Lipschitz functions is obtained.

Following [6] and [9], respectively, the smoothing functions for the finite max-functions are constructed via integral-convolution with special *density functions*.

A major aspect of the analysis consists in showing that the smoothing functions considered satisfy *gradient consistency*, see Section 3. Gradient consistency, as defined in [6], is a fundamental tool for establishing limiting stationarity properties of smoothing methods for optimization. In particular, it guarantees that (accumulation points of) sequences of first-order critical points of the smooth approximations yield critical points of the original function f.

This paper can be viewed as an extension to parts of the recent paper [6] by Chen, in which the author builds up an analysis of smoothing approximations build on the *plus function*, see Section 2. The plus function is a special case of a finite max-function. Moreover, we fill a void which arises from an insufficient proof of [6, Theorem 1 (i)], see Section 5 and [7]. The latter result is key for establishing gradient consistency for composite smoothing functions, and hence of fundamental importance. Not withstanding the insuffiency of the current proof of [6, Theorem 1 (i)], we conjecture that the assertion is valid, yet not achievable via a chain rule approach without the assumptions discussed in the sequel.

The organization of the paper is as follows: In Section 2 we review some concepts from nonsmooth analysis which are employed in the sequel. In Section 3 we lay out a general framework for the smoothing functions. Section 4 establishes the class of smoothing functions for finite-max functions and provides calculus rules for compositions with smooth mappings. We close with some final remarks in Section 5. In particular, we compare [6, Theorem 1 (i)] with our main theorem.

Most of the notation used is standard. An element $x \in \mathbb{R}^n$ is understood as a column vector. The symbol $\mathbb{R}^n_+$ denotes the set of all vectors whose components are nonnegative. The space of all real $m \times n$-matrices is denoted by $\mathbb{R}^{m\times n}$, and for $A \in \mathbb{R}^{m\times n}$, A^T is its transpose, and $\operatorname{rank} A$ denotes its rank. An $n \times n$ diagonal matrix D with the vector x on its diagonal is denoted by

$$D = \operatorname{diag}(x) = \operatorname{diag}(x_i).$$

The *Euclidean norm* on $\mathbb{R}^n$ is denoted by $\|\cdot\|$, i.e.,

$$\|x\| = \sqrt{x^T x} \quad \forall x \in \mathbb{R}^n.$$

The closed *Euclidean ball* centered around $\bar{x} \in \mathbb{R}^n$ with radius $r \geq 0$ is denoted by $B_r(\bar{x})$, i.e.,

$$B_r(\bar{x}) := \{x \in \mathbb{R}^n \mid \|x - \bar{x}\| \leq r\}.$$

For a set $S \subset \mathbb{R}^n$ its *convex hull* is denoted by $\operatorname{conv} S$. Given a real-valued function $f : \mathbb{R}^n \to \mathbb{R}$ differentiable at $\bar{x}$, the *gradient* is given by $\nabla f(\bar{x})$ which is understood as a column vector. For a function $F : \mathbb{R}^n \to \mathbb{R}^m$ differentiable at $\bar{x}$, the *Jacobian* of F at $\bar{x}$ is denoted by $F'(\bar{x})$, i.e.,

$$F'(\bar{x}) = \begin{pmatrix} \nabla F_1(\bar{x})^T \\ \vdots \\ \nabla F_m(\bar{x})^T \end{pmatrix} \in \mathbb{R}^{m\times n},$$

whereas $\nabla F(\bar{x})$ is the transposed Jacobian. In order to distinguish between single- and set-valued maps, we write $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ to indicate that S maps vectors from $\mathbb{R}^n$ to subsets of $\mathbb{R}^m$. Finally, the symbol $x^k \to_X \bar{x}$ indicates that $\{x^k\}$ is a sequence converging to the limit point $\bar{x}$ such that all iterates x^k belong to a set $X \subset \mathbb{R}^n$.

2 Preliminaries

In this section we review certain concepts from variational and nonsmooth analysis, which will be used in the subsequent analysis. At this, the notation is mainly based on [23].

A major role is played by different kinds of subdifferentials as a tool for dealing with nonsmoothness of the functions considered. To this end, we commence by introducing the so-called *regular* and *limiting subdifferential*. In the definition of the limiting subdifferential, we employ the *outer limit* for a set-valued mapping, which we now define: For $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and $X \subset \mathbb{R}^n$, we define the *outer limit*

$$\operatorname*{Lim\,sup}_{x \to_X \bar{x}} S(x) := \left\{ v \mid \exists \{x^k\} \to_X \bar{x}, \exists \{v^k\} \to v : v^k \in S(x^k) \quad \forall k \in \mathbb{N} \right\}.$$

Definition 2.1 (Regular and limiting subdifferential)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous and $\bar{x} \in \mathbb{R}^n$.*

a) The regular subdifferential *of f at $\bar{x}$ is the set given by*

$$\hat{\partial} f(\bar{x}) := \big\{ v \mid \liminf_{x \to \bar{x}} \frac{f(x) - f(\bar{x}) - v^T(x - \bar{x})}{\|x - \bar{x}\|} \geq 0 \big\}.$$

b) The limiting subdifferential *of f at $\bar{x}$ is the set given by*

$$\begin{aligned} \partial f(\bar{x}) &:= \operatorname*{Lim\,sup}_{x \to \bar{x}} \hat{\partial} f(x) \\ &= \big\{ v \mid \exists \{x^k\} \to \bar{x}, \exists \{v^k\} \to v : v^k \in \hat{f}(x^k) \;\; \forall k \in \mathbb{N} \big\}. \end{aligned}$$

Note that there are different ways of obtaining the limiting subdifferential than the one described above, which basically goes back to Mordukhovich, cf. [17]. In this context, see [15] (or [4]) for a construction of the limiting subdifferential via *Dini-derivatives.*

A very important class of potentially nonsmooth, nonconvex functions are the locally Lipschitz ones. We call a function $F : \mathbb{R}^n \to \mathbb{R}^m$ *locally Lipschitz* at $\bar{x} \in \mathbb{R}^n$ if there exist $\varepsilon > 0$ and $L \geq 0$ such that

$$\|F(x) - F(y)\| \leq L\|x - y\|$$

for all $x, y \in B_\varepsilon(\bar{x})$. For such an F, we may define the so-called *generalized Jacobian*, which goes back to Clarke, see [10, Definition 2.6.1], in a way that is opened up by *Rademacher's Theorem*, see [4, Theorem 9.1.2] or [23, Theorem 9.60], which yields that the complement of the set

$$D_F := \{x \mid F \,\text{is differentiable at}\, x\}$$

has *Lebesgue measure* 0, and hence the set

$$\bar{\nabla} F(\bar{x}) := \Big\{ V \mid \exists \{x^k\} \subset D_F : x^k \to \bar{x} \,\wedge\, F'(x^k) \to V \Big\}$$

is well-defined, even compact, cf. [23, Theorem 9.62] or [10, p. 63]. Note that the set $\bar{\nabla} F(\bar{x})$ itself is usually called the *B-subdifferential* of F at $\bar{x}$, see, e.g., [22], though we will not use this terminology here.

Definition 2.2 (Generalized Jacobian)**.** *For a locally Lipschitz function $F : \mathbb{R}^n \to \mathbb{R}^m$ we define the* generalized Jacobian *of F at $\bar{x}$ by*

$$\bar{\partial} F(\bar{x}) := \operatorname{conv} \bar{\nabla} F(\bar{x}).$$

For the special case $m = 1$ we actually recover the *Clarke subdifferential* with the above definition, see [10] for an extensive treatment. However, in this case, to be consistent with the generalized Jacobian, the elements from the Clarke subdifferential are row vectors, but we prefer to think of them as column vectors, so everytime a generalized Jacobian is involved, we have to transpose accordingly. The Clarke subdifferential of a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ can alternatively be obtained via the limiting subdifferential. In fact, see [23, Theorem 9.61], we have

$$\bar{\partial} f(\bar{x}) = \operatorname{conv} \partial f(\bar{x}) \quad \forall \bar{x} \in \mathbb{R}^n \tag{2}$$

in this case.

An important concept in the context of subdifferentiation is *(subdifferential) regularity*, which we define only for the locally Lipschitz case. For the general case, cf. [23, Definition 7.25].

Definition 2.3 (Subdifferential regularity)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz. Then we say that f is* (subdifferentially) regular *at $\bar{x} \in \mathbb{R}^n$ if*

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

If the above equality holds for all $\bar{x} \in \mathbb{R}^n$, we simply say that f is (subdifferentially) regular.

Note that this notion of regularity actually coincides with the one given by [10, Definition 2.3.4], cf. [10, Theorem 2.4.9 (ii)] in connection with [23, Definition 7.25].

A prominent class of (potentially nonsmooth) locally Lipschitz, regular functions are the *convex* ones. Here we refer to [4] or [14] for the fact that (finite-valued) convex functions are locally Lipschitz, and to [10, Proposition 2.3.6 b)] or [23, Example 7.27] to see that they are indeed regular.

It is a well-known fact, see [23, Proposition 8.12], that in case $f : \mathbb{R}^n \to \mathbb{R}$ is convex, all subdifferentials from above coincide with the subdifferential of convex analysis, i.e.,

$$\bar{\partial} f(\bar{x}) = \partial f(\bar{x}) = \hat{\partial} f(\bar{x}) = \left\{ v \mid f(x) \geq f(\bar{x}) + v^T(x - \bar{x}) \quad \forall x \in \mathbb{R}^n \right\}$$

for all $\bar{x} \in \mathbb{R}^n$.

As an illustrating example for the concepts introduced above, we consider the function $(\cdot)_+ : \mathbb{R} \to \mathbb{R}$ given by

$$(t)_+ := \max\{t, 0\},$$

called the *plus function*, which appears, as a motivational special case, along the lines in Section 4.

Example 2.4 (Subdifferentiation of the plus function)**.**

a) Let $f_1 : \mathbb{R} \to \mathbb{R}$ be given by $f_1(t) := (t)_+$. Then the convexity of f_1 implies that

$$\hat{\partial} f_1(t) = \partial f_1(t) = \bar{\partial} f_1(t) = \begin{cases} 0 & \text{if} \quad t < 0, \\ [0,1] & \text{if} \quad t = 0, \\ 1 & \text{if} \quad t > 0, \end{cases} \tag{3}$$

whereas we have

$$\bar{\nabla} f_1(t) = \begin{cases} 0 & \text{if} \quad t < 0, \\ \{0,1\} & \text{if} \quad t = 0, \\ 1 & \text{if} \quad t > 0. \end{cases}$$

b) Let $f_2 : \mathbb{R} \to \mathbb{R}$ be given by $f_2(t) := -(t)_+$. Considering the only interesting point $t = 0$, an elementary calculation shows that

$$\hat{\partial} f_2(0) = \emptyset, \quad \partial f_2(0) = \bar{\nabla} f_2(0) = \{0, -1\} \quad \text{and} \quad \bar{\partial} f(0) = [-1, 0].$$

We close this section by introducing the *coderivative*, which goes back to Mordukhovich, see [18], which is a derivative concept for set-valued maps. Here we are only interested in a special case where $F : \mathbb{R}^n \to \mathbb{R}^m$ is single-valued and locally Lipschitz. Then the coderivative of F at $\bar{x}$ can be defined as the set-valued map $D^*F(\bar{x}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ given by

$$D^*F(\bar{x})(u) = \partial(u^T F)(\bar{x}),$$

where $u^T F : \mathbb{R}^n \to \mathbb{R}$, $(u^T F)(x) := \sum_{i=1}^m u_i F_i(x)$. This is, in fact, the so-called *scalarization formula*, see [19, Theorem 1.90] or [23, Proposition 9.24 (b)]. Furthermore, we have

$$\bar{\partial} F(\bar{x})^T u = \operatorname{conv} D^*F(\bar{x})(u) \quad \forall u \in \mathbb{R}^m,$$

see [23, Theorem 9.62]. If F is even continuously differentiable, we have

$$D^*F(\bar{x})(u) = \{F'(\bar{x})^T u\},$$

see [19, Theorem 1.38] or [23, Example 8.34].

3 The general smoothing setup

Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous. Then we call $s_f : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$ a *smoothing function* for f if the following assumptions are fulfilled:

- $s_f(\cdot, \mu)$ converges *continuously* to f in the sense of [23, Definition 5.41], i.e.,

$$\lim_{\mu \downarrow 0, x \to \bar{x}} s_f(x, \mu) = f(\bar{x}) \quad \forall \bar{x} \in \mathbb{R}^n, \tag{4}$$

- $s_f(\cdot, \mu)$ is continuously differentiable for all $\mu > 0$.

For algorithmic purposes, provided that one has sequences $\{x^k\} \to \bar{x}$ and $\{\mu_k\} \downarrow 0$ such that

$$\lim_{k \to \infty} \nabla s_f(x^k, \mu_k) \to 0,$$

the following question is of crucial importance:

Is $\bar{x}$ a critical point *of f in the sense that $0 \in \partial f(\bar{x})$ (or $0 \in \bar{\partial} f(\bar{x})$) ?*

The answer is positive if

$$\operatorname*{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu) \subset \partial f(\bar{x}),$$

where, for the sake of completeness, we recall that, according to the general definition of the outer limit, we have

$$\operatorname*{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu) = \big\{ v \mid \exists\{(x^k,\mu_k)\} \to (\bar{x},0) : \nabla_x s_f(x^k,\mu_k) \to v \big\}.$$

The next result shows that the converse inclusion is always valid if s_f is a smoothing function for f.

Lemma 3.1. *Let $f:\mathbb{R}^n \to \mathbb{R}$ be continuous and let s_f be a smoothing function for f. Then for $\bar{x} \in \mathbb{R}^n$ we have*

$$\partial f(\bar{x}) \subset \operatorname*{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu).$$

If further f is locally Lipschitz at $\bar{x}$, then

$$\bar{\partial} f(\bar{x}) \subset \operatorname{conv}\big\{ \operatorname*{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu) \big\}.$$

Proof. Let $v \in \partial f(\bar{x})$ be given. Since, by assumption, $s_f(\cdot,\mu)$ converges continuously to f, it converges, in particular, *epigraphically*, cf. [23, Theorem 7.11], and hence we may invoke [23, Corollary 8.47] in order to obtain sequences $\{\mu_k\} \downarrow 0, \{x^k\} \to \bar{x}$ and $\{v^k\}$ with $v^k \in \partial_x s_f(x^k,\mu_k)$ such that $v^k \to v$. Now, since $s_f(\cdot,\mu_k)$ is continuously differentiable by assumption, we have

$$v^k = \nabla_x s_f(x^k,\mu_k),$$

which identifies v as an element of $\operatorname{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu)$ and thus, the first inclusion follows. The second inclusion is an immediate consequence of the first one and the fact that $\operatorname{conv} \partial f(\bar{x}) = \bar{\partial} f(\bar{x})$ in the presence of local Lipschitz continuity, see (2). □

A trivial consequence is the following corollary.

Corollary 3.2. *Let $f:\mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz at $\bar{x}$ and let s_f be a smoothing function for f. Then we have the inclusions*

$$\partial f(\bar{x}) \subset \left\{ \begin{array}{c} \operatorname*{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu) \\ \bar{\partial} f(\bar{x}) \end{array} \right\} \subset \operatorname{conv}\big\{ \operatorname*{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu) \big\}.$$

It is clear that, in the locally Lipschitz setting, the condition

$$\operatorname*{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu) = \partial f(\bar{x}) \tag{5}$$

implies

$$\operatorname{conv}\big\{ \operatorname*{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu) \big\} = \bar{\partial} f(\bar{x}), \tag{6}$$

a condition which is called *gradient consistency* in [6]. In particular, both conditions coincide, when $\partial f(\bar{x}) = \bar{\partial} f(\bar{x})$ (which is the case if f is locally Lipschitz and subdifferentially regular).

The following example shows that, for locally Lipschitz f, condition (6) is in fact weaker than (5).

Example 3.3. *Let $f : \mathbb{R}^2 \to \mathbb{R}, f(a,b) := \min\{a,b\}$. Then $s_f((a,b),\mu) := \frac{1}{2}(a+b-\sqrt{(a-b)^2+4\mu})$ is a smoothing function for f, sometimes called the CHKS-function due to its origin from [8, 16, 25]. It holds that for all $a \in \mathbb{R}$ we have*

$$\partial f(a,a) = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \subsetneq \operatorname{conv} \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} = \bar{\partial} f(a,a),$$

but

$$\operatorname*{Lim\,sup}_{(x,y)\to(a,a),\mu\downarrow 0} \nabla s_f((x,y),\mu) = \operatorname{conv} \big\{ \operatorname*{Lim\,sup}_{(x,y)\to(a,a),\mu\downarrow 0} \nabla s_f((x,y),\mu)(a,a) \big\} = \bar{\partial} f(a,a).$$

The following result is the main motivation for the analysis in Section 4. In this context, for $f : \mathbb{R}^n \to \mathbb{R}$ locally Lipschitz, we call $\bar{x} \in \mathbb{R}^n$ *Clarke-stationary*, *C-stationary* for short, if $0 \in \bar{\partial} f(\bar{x})$.

Theorem 3.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and let s_f be a smoothing function for f. Furthermore let $\{x^k\} \subset \mathbb{R}^n$ and $\{\mu_k\} \downarrow 0$ such that*

$$\|\nabla_x s_f(x^k, \mu_k)\| \leq c\mu_k \quad \forall k \in \mathbb{N}, \tag{7}$$

for some $c > 0$. Then every accumulation point $\bar{x}$ of $\{x^k\}$ such that the gradient consistency condition (6) holds at $\bar{x}$ is a C-stationary point of f.

Proof. Let $\bar{x}$ be a limit point of a subsequence $\{x^k\}_K$ such that gradient consistency holds at $\bar{x}$. Since $\{\mu_k\}_K \downarrow 0$ on the same subsequence, we can deduce from (7) that

$$0 \in \operatorname*{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu).$$

This implies

$$0 \in \operatorname{conv} \big\{ \operatorname*{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu) \big\} = \bar{\partial} f(\bar{x})$$

by the gradient consistency assumption. □

We point out that, in particular, the *smoothing gradient method* proposed in [6] fits into the framework of Theorem 3.4, cf. [6, Theorem 2].

We close this section with the remark that Theorem 3.4 can be refined in the following sense: Suppose that, in the situation of the latter theorem, the stronger condition (5) holds at $\bar{x}$. Then it follows from the previous proof that $0 \in \partial f(\bar{x})$, which is (without regularity) a tighter property than C-stationarity (typically called *M-stationarity* in the corresponding literature). However, we are not aware of a class of (non-regular) functions, for which (5) holds. Example 3.3 displays our impression that, in the non-regular case, the gradient consistency (6) is substantially weaker, hence much more likely to hold than (5) in the smoothing setup described above. This is also confirmed by the analysis in the upcoming section.

4 Smoothing via integral-convolution

In this section we generalize (and to a certain extent correct, see [7]) the results from [6, Section 3-4], in the sense that we do not entirely focus on the plus function.

For these purposes, let $p : \mathbb{R} \to \mathbb{R}$ be a *finite max-function*, i.e.,

$$p(t) = \max_{i=1,\dots,r} \{f_i(t)\}$$

where $f_i : \mathbb{R} \to \mathbb{R}$ is affine linear, i.e.,

$$f_i(t) = a_i t + b_i$$

with scalars $a_i, b_i \in \mathbb{R}$ for all $i = 1, \dots, r$ ($r \in \mathbb{N}$). Note that p is, in particular, *piecewise affine*, hence (globally) Lipschitz, see [24, Proposition 2.2.7] and convex, cf. [14, Proposition B 2.1.2]. Moreover, it can be seen (cf. Figure 1) that, after skipping all indices which do not contribute in the maximization, and after reordering the remaining indices, we can assume without loss of generality that

$$a_1 < a_2 < \dots < a_{r-1} < a_r, \tag{8}$$

and there exists a partition of the real line

$$-\infty = t_1 < t_2 < \dots < t_r < t_{r+1} = +\infty$$

such that

$$a_i t_{i+1} + b_i = a_{i+1} t_{i+1} + b_{i+1} \quad \forall i = 1, \dots, r-1, \tag{9}$$

and

$$p(t) = \begin{cases} a_1 t + b_1 & \text{if} \quad t \le t_2, \\ a_i t + b_i & \text{if} \quad t \in [t_i, t_{i+1}] \qquad (i \in \{2, \dots, r-1\}), \\ a_r t + b_t & \text{if} \quad t \ge t_r. \end{cases} \tag{10}$$

Let $\rho : \mathbb{R} \to \mathbb{R}$ be a piecewise continuous, symmetric *density function*, i.e.,

$$\rho(t) = \rho(-t) \quad (t \in \mathbb{R}) \quad \text{and} \quad \int_{\mathbb{R}} \rho(t)\, dt = 1, \tag{11}$$

such that

$$\rho \ge 0 \quad \text{and} \quad \int_{\mathbb{R}} |t| \rho(t)\, dt < +\infty. \tag{12}$$

We denote the *distribution function* that goes with the density ρ by F, i.e., $F : \mathbb{R} \to [0, 1]$ is given by

$$F(x) = \int_{-\infty}^{x} \rho(t)\, dt.$$

In particular, since ρ is piecewise continuous, F is continuous with

$$\lim_{x \to +\infty} F(x) = 1 \quad \text{and} \quad \lim_{x \to -\infty} F(x) = 0.$$

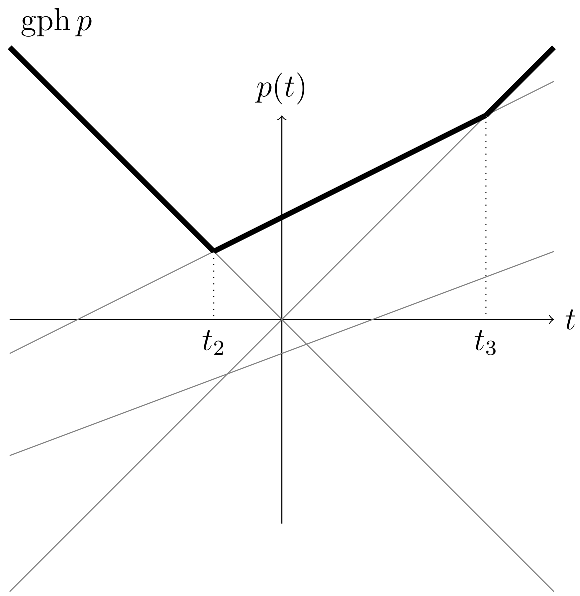

Figure 1: Illustration for the choices in (8)-(10)

Lemma 4.1. *Let $p : \mathbb{R} \to \mathbb{R}$ be a finite max-function. Furthermore, let $\rho : \mathbb{R} \to \mathbb{R}_+$ be a piecewise continuous function satisfying* (11) *and* (12)*. Then the convolution*

$$s_p(t, \mu) := \int_{\mathbb{R}} p(t - \mu s)\rho(s)\, ds$$

is a (well-defined) smoothing function for p with

$$\operatorname*{Lim\,sup}_{t\to\bar{t},\mu\downarrow 0} \frac{d}{dt} s_p(t, \mu) = \partial p(\bar{t}) \quad \forall \bar{t} \in \mathbb{R}.$$

Proof. The fact that $s_p(t, \mu)$ exists for all $\mu > 0$ is a consequence of the conditions imposed on ρ in (11) and (12) and the representation of p from (8)-(10), which we can assume without loss of generality.

Next we show that $\lim_{t\to\bar{t},\mu\downarrow 0} s_p(t, \mu) = p(\bar{t})$ for all $\bar{t} \in \mathbb{R}$. This is due to the fact that

$$\begin{aligned} |p(\bar{t}) - s_p(t, \mu)| &= \Big| \int_{\mathbb{R}} p(\bar{t})\rho(s)\, ds - \int_{\mathbb{R}} p(t - \mu s)\rho(s)\, ds \Big| \\ &\le \int_{\mathbb{R}} |p(\bar{t}) - p(t - \mu s)|\rho(s)\, ds \\ &\le L_p \int_{\mathbb{R}} |\bar{t} - t + \mu s|\rho(s)\, ds \\ &\le L_p|\bar{t} - t| + L_p\mu \int_{\mathbb{R}} |s|\rho(s)\, ds, \end{aligned}$$

where L_p is the (global) Lipschitz constant of p. Taking into account assumption (12), it follows that the last expression tends to 0 as $t \to \bar{t}$ and $\mu \downarrow 0$. This shows that $s_p(\cdot, \mu)$ converges continuously to p.

We will now compute the derivative of $s_p(\cdot, \mu)$ for some fixed $\mu > 0$. To this end, recall that, without loss of generality, we may assume that p admits a representation as in (10). Hence, we get

$$s_p(t, \mu) = \sum_{i=1}^{r} \int_{\frac{t-t_{i+1}}{\mu}}^{\frac{t-t_i}{\mu}} [a_i(t - \mu s) + b_i]\rho(s)\, ds,$$

where we have $\frac{t-t_1}{\mu} \equiv +\infty$ and $\frac{t-t_{r+1}}{\mu} \equiv -\infty$. Thus, we obtain

$$\frac{d}{dt} s_p(t, \mu) = \sum_{i=1}^{r} \frac{d}{dt} \Big[\int_{\frac{t-t_{i+1}}{\mu}}^{\frac{t-t_i}{\mu}} [a_i(t - \mu s) + b_i]\rho(s)\, ds \Big], \tag{13}$$

where the existence of the corresponding derivatives follows, e.g., from the Leibniz integral rule with variable limits. More precisely, this rule allows us to compute the derivatives explicitly. For the summands $i = 2, \dots, r-1$, we obtain

$$\begin{aligned} &\frac{d}{dt} \Big[\int_{\frac{t-t_{i+1}}{\mu}}^{\frac{t-t_i}{\mu}} [a_i(t - \mu s) + b_i]\rho(s)\, ds \Big] \\ &= \; a_i \int_{\frac{t-t_{i+1}}{\mu}}^{\frac{t-t_i}{\mu}} \rho(s)\, ds + (a_i t_i + b_i) \frac{\rho(\frac{t-t_i}{\mu})}{\mu} - (a_i t_{i+1} + b_i) \frac{\rho(\frac{t-t_{i+1}}{\mu})}{\mu}. \end{aligned}$$

For the summand $i = 1$ we compute

$$\frac{d}{dt} \Big[\int_{\frac{t-t_2}{\mu}}^{+\infty} [a_1(t - \mu s) + b_1]\rho(s)\, ds \Big] = a_1 \int_{\frac{t-t_2}{\mu}}^{+\infty} \rho(s)\, ds - (a_1 t_2 + b_1) \frac{\rho(\frac{t-t_2}{\mu})}{\mu},$$

and for $i = r$ we get

$$\frac{d}{dt} \Big[\int_{-\infty}^{\frac{t-t_r}{\mu}} [a_r(t - \mu s) + b_r]\rho(s)\, ds \Big] = a_r \int_{-\infty}^{\frac{t-t_r}{\mu}} \rho(s)\, ds + (a_r t_{r+1} + b_r) \frac{\rho(\frac{t-t_r}{\mu})}{\mu}.$$

Inserting these expressions in (13) and exploiting the fact that $a_i t_{i+1} + b_i = a_{i+1} t_{i+1} + b_{i+1}$ $(i = 1, \dots, r-1)$ (see (9)), we obtain

$$\begin{aligned} &\frac{d}{dt} s_p(t, \mu) \\ &= \; \sum_{i=1}^{r} a_i \int_{\frac{t-t_{i+1}}{\mu}}^{\frac{t-t_i}{\mu}} \rho(s)\, ds \\ &= \; a_1 \Big(1 - F\Big(\frac{t - t_2}{\mu}\Big)\Big) + \sum_{i=2}^{r-1} a_i \Big(F\Big(\frac{t - t_i}{\mu}\Big) - F\Big(\frac{t - t_{i+1}}{\mu}\Big)\Big) + a_r F\Big(\frac{t - t_r}{\mu}\Big) \end{aligned} \tag{14}$$

due to the telescoping structure of the resulting sum. In particular, since F is continuous, so is $\frac{d}{dt}s_p(\cdot,\mu)$ for all $\mu > 0$. Altogether, we have shown that s_p is a smoothing function for p.

In order to verify the remaining assertion, first note that, in view of (8)-(10), we have

$$\partial p(\bar t) = \begin{cases} \{a_1\} & \text{if} \quad \bar t < t_2, \\ \{a_i\} & \text{if} \quad \bar t \in (t_i, t_{i+1})\,(i = 2,\dots,r-1), \\ \{a_r\} & \text{if} \quad \bar t > t_r, \\ [a_i, a_{i+1}] & \text{if} \quad \bar t = t_{i+1}\,(i = 1,\dots,r-1). \end{cases} \tag{15}$$

Now, recall that Lemma 3.1 guarantees that $\operatorname{Lim\,sup}_{t\to\bar t,\mu\downarrow 0} \frac{d}{dt}s_p(t,\mu) \supset \partial p(\bar t)$, as s_p is a smoothing function for p. In order to see the converse inclusion, let $v \in \operatorname{Lim\,sup}_{t\to\bar t,\mu\downarrow 0} \frac{d}{dt}s_p(t,\mu)$ be given. Then there exist sequences $\{\mu_k\} \downarrow 0$ and $\{t_k\} \to \bar t$ such that

$$\frac{d}{dt}s_p(t_k,\mu_k) \to v.$$

Clearly, if $\bar t \in (t_j, t_{j+1})$ for some $j \in \{2,\dots,r-1\}$, we obtain

$$F\Big(\frac{t_k - t_i}{\mu_k}\Big) - F\Big(\frac{t_k - t_{i+1}}{\mu_k}\Big) \to 0 \quad \forall i \neq j,$$

$$F\Big(\frac{t_k - t_j}{\mu_k}\Big) - F\Big(\frac{t_k - t_{j+1}}{\mu_k}\Big) \to 1 - 0 = 1,$$

and

$$1 - F\Big(\frac{t_k - t_2}{\mu_k}\Big) \to 0, \qquad F\Big(\frac{t_k - t_r}{\mu_k}\Big) \to 0. \tag{16}$$

The representation (14) of the gradient of s_p therefore shows that $v = a_j$, hence we have $v = a_j \in \{a_j\} = \partial p(\bar t)$.

Furthermore, if $\bar t < t_2$, we infer that $v = a_1$ and, if $\bar t > t_r$ we get $v = a_r$, which yields $v \in \partial p(\bar t)$ also in these cases.

It remains to consider the cases where $\bar t = t_{j+1}$ for some $j \in \{1,\dots,r-1\}$. First consider the case where $j \in \{2,\dots,r-2\}$. Then

$$F\Big(\frac{t_k - t_i}{\mu_k}\Big) - F\Big(\frac{t_k - t_{i+1}}{\mu_k}\Big) \to 0 \quad \forall i \notin \{j, j+1\},$$

and since $F : \mathbb{R} \to [0,1]$, we get (at least on a subsequence)

$$F\Big(\frac{t_k - t_j}{\mu_k}\Big) - F\Big(\frac{t_k - t_{j+1}}{\mu_k}\Big) \to 1 - \lambda$$

and

$$F\Big(\frac{t_k - t_{j+1}}{\mu_k}\Big) - F\Big(\frac{t_k - t_{j+2}}{\mu_k}\Big) \to \lambda,$$

for some $\lambda \in [0,1]$. Using once more the limit conditions from (16) as well as the representation (14), we obtain $v = a_j(1-\lambda) + a_{j+1}\lambda \in [a_j, a_{j+1}] = \partial p(\bar t)$.

Finally, the arguments are similar if $\bar t = t_1$ or $\bar t = t_{r-1}$, hence we skip the details. $\square$

Note that the former result is still valid for a function p, which admits a piecewise-affine representation as given by (9) and (10), without demanding (8). The latter condition corresponds to convexity and hence, regularity of p, which will be needed already in the following result.

Corollary 4.2. *Let $p : \mathbb{R} \to \mathbb{R}$ be a finite max-function, $h : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable and let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) := p(h(x))$. Then, if s_p is defined as in Lemma 4.1, the function $s_f(\cdot,\cdot) := s_p(h(\cdot),\cdot)$ is a smoothing function for f with*

$$\operatorname*{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu) = \partial f(\bar{x}) \quad \forall \bar{x} \in \mathbb{R}.$$

In particular, the gradient consistency property (6) *holds.*

Proof. The fact that s_f is a smoothing function for f is due to the fact that s_p has this property with respect to p and h is continuously differentiable. Hence, given $\bar{x} \in \mathbb{R}^n$, the inclusion $\operatorname{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu) \supset \partial f(\bar{x})$ is clear from Lemma 3.1.

In order to establish the converse inclusion, note first that

$$\partial f(\bar{x}) = \nabla h(\bar{x}) \partial g(h(\bar{x})),$$

since p is convex, hence regular and h is smooth, cf. [23, Theorem 10.6]. Let $v \in \operatorname{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu)$. Then, there exist sequences $\{x^k\} \to \bar{x}$ and $\{\mu_k\} \downarrow 0$ such that

$$\nabla h(x^k) \frac{d}{dt} s_p(h(x^k),\mu_k) = \nabla_x s_f(x^k,\mu_k) \to v. \tag{17}$$

Since we have (see (14))

$$\begin{aligned} &\frac{d}{dt} s_p(h(x^k),\mu_k) \\ &= a_1\Big(1 - F\Big(\frac{h(x^k)-t_2}{\mu_k}\Big)\Big) + \sum_{i=2}^{r-1} a_i\Big(F\Big(\frac{h(x^k)-t_i}{\mu_k}\Big) - F\Big(\frac{h(x^k)-t_{i+1}}{\mu_k}\Big)\Big) \\ &\quad + a_r F\Big(\frac{h(x^k)-t_r}{\mu_k}\Big), \end{aligned}$$

and $F : \mathbb{R} \to [0,1]$, the sequence $\{\frac{d}{dt} s_p(h(x^k),\mu_k)\}$ is bounded. Hence Lemma 4.1 implies that $\{\frac{d}{dt} s_p(h(x^k),\mu_k)\}$ converges (at least on a subsequence) to some element $\tau \in \partial g(h(\bar{x}))$. It therefore follows from (17) that $v = \nabla h(\bar{x})\tau \in \nabla h(\bar{x})\partial g(h(\bar{x})) = \partial f(\bar{x})$. This concludes the proof. □

To prepare the main theorem of this section, we need the following preliminary result.

Lemma 4.3. *Let $H : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable as well as $G : \mathbb{R}^m \to \mathbb{R}^m$ given by*

$$G(y) := (\varphi_i(y_i))_{i=1}^m,$$

where $\varphi_i : \mathbb{R} \to \mathbb{R}\, (i = 1,\dots,m)$ is regular. Then for $F := G \circ H$ the following holds:

a) For all $\bar{x} \in \mathbb{R}^n$

$$\bar{\partial} F(\bar{x})^T d = H'(\bar{x})^T \bar{\partial} G(H(\bar{x}))d \quad \forall d \in \mathbb{R}^m_+.$$

b) If $\operatorname{rank} H'(\bar{x}) = m$*, it holds that*

$$\bar{\partial} F(\bar{x})^T d = H'(\bar{x})^T \bar{\partial} G(H(\bar{x}))d \quad \forall d \in \mathbb{R}^m.$$

Proof. a) For $d \in \mathbb{R}^m_+$ and $\bar{x} \in \mathbb{R}^n$ we have

$$\begin{aligned}
\bar{\partial} F(\bar{x})^T d &= \operatorname{conv} D^* F(\bar{x})(d) \\
&= \operatorname{conv} \partial(d^T F)(\bar{x}) \\
&= \bar{\partial}(d^T F)(\bar{x}) \\
&= \bar{\partial}\big[\sum_{i=1}^m d_i F_i\big](\bar{x}) \\
&= \sum_{i=1}^m d_i \bar{\partial} F_i(\bar{x}) \\
&= \sum_{i=1}^m d_i \partial\varphi_i(H_i(\bar{x}))\nabla H_i(\bar{x}) \\
&= H'(\bar{x})^T \bar{\partial} G(H(\bar{x}))d.
\end{aligned}$$

At this, the first equality follows from [23, Theorem 9.62], the second one is the scalarization formula, see Section 2. The third equality uses the fact that the Clarke subdifferential is the convex hull of the limiting subdifferential, cf. (2). The fourth equality is just the definition of the function $x \mapsto (d^T F)(x)$, and the fifth one is due to the fact that the functions F_i are regular by [10, Theorem 2.3.9 (iii)], and hence the sum rule from [10, Proposition 2.3.3] holds with equality, see [10, Corollary 3] (note that $d \geq 0$ is required here!). The sixth equality is once again the chain rule from [10, Theorem 2.3.9 (iii)], and the final line is just a short-hand form of the previous one.

b) If $H'(\bar{x})$ has rank m it follows from [19, Theorem 1.66] that for $d \in \mathbb{R}^m$ we have

$$D^* F(\bar{x})(d) = D^*(G \circ H)(\bar{x})(d) = H'(\bar{x})^T D^* G(H(\bar{x}))(d).$$

Taking the convex hull and using [23, Theorem 9.62] yields

$$\begin{aligned}
\bar{\partial} F(\bar{x})^T d &= \operatorname{conv} D^* F(\bar{x})(d) \\
&= \operatorname{conv} \big\{H'(\bar{x})^T D^* G(H(\bar{x}))(d)\big\} \\
&= H'(\bar{x})^T \operatorname{conv} \big\{D^* G(H(\bar{x}))(d)\big\} \\
&= H'(\bar{x})^T \bar{\partial} G(H(\bar{x}))d.
\end{aligned}$$

This completes the proof $\square$

The following example shows that if in the above theorem $d \in \mathbb{R}^m$ has negative components and $H'(\bar{x})$ is not onto, the desired assertion may fail. It is this chain rule which is erroneously applied in the proof of [6, Theorem 1 (i)] without further assumptions, which leads to the insufffiency there. However, we point out, again, that we believe the assertion of [6, Theorem 1 (i)] to be true anyway.

Example 4.4 (Failure in Lemma 4.3)**.** *Consider the function* $F := G \circ H$ *with* $H : \mathbb{R} \to \mathbb{R}^2, H(x) := \binom{x}{x}$ *and* $G : \mathbb{R}^2 \to \mathbb{R}^2, G(y) := \binom{(y_1)_+}{(y_2)_+}$*, i.e.,* $F(x) = \binom{(x)_+}{(x)_+}$*. It follows that*

$$\bar{\partial} F(0) = \Big\{ \begin{pmatrix} a \\ a \end{pmatrix} \mid a \in [0,1] \Big\},$$

hence, for $d := \binom{1}{-1}$ *we have*

$$\bar{\partial} F(0)^T d = \{0\} \neq [-1, 1] = H'(0)^T \bar{\partial} G(H(0)) d.$$

The following is the main result of this section.

Theorem 4.5. *Let* $H : \mathbb{R}^n \to \mathbb{R}^m$ *and* $g : \mathbb{R}^m \to \mathbb{R}$ *be continuously differentiable and define* $f : \mathbb{R}^n \to \mathbb{R}$ *by* $f(x) := g(G(H(x)))$*, where*

$$G(y) := [p_i(y_i)]_{i=1}^m$$

and $p_i : \mathbb{R} \to \mathbb{R}\,(i = 1, \dots, m)$ *is a finite max-function. Then* $s_f : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$ *given by* $s_f(x, \mu) := g([s_{p_i}(H_i(x), \mu)]_{i=1}^m)$*, where* s_{p_i} *is given as in Lemma 4.1, is a smoothing function for* f*. If furthermore* $\bar{x} \in \mathbb{R}^n$ *is such that*

$$\nabla g(G(H(\bar{x}))) \in \mathbb{R}^m_+ \quad \text{or} \quad \operatorname{rank} H'(\bar{x}) = m$$

holds, then

$$\operatorname*{Lim\,sup}_{x \to \bar{x}, \mu \downarrow 0} \nabla_x s_f(x, \mu) \subset \bar{\partial} f(\bar{x}).$$

and hence

$$\operatorname{conv} \big\{ \operatorname*{Lim\,sup}_{x \to \bar{x}, \mu \downarrow 0} \nabla_x s_f(x, \mu) \big\} = \bar{\partial} f(\bar{x}),$$

i.e., the gradient consistency property (6) *holds.*

Proof. First, note that s_f is a smoothing function for f since s_{p_i} has this property with respect to p_i for all $i = 1, \dots, m$ and g and H are continuously differentiable.

Moreover, we compute the Clarke subdifferential of f at $\bar{x}$ as

$$\begin{aligned} \bar{\partial} f(\bar{x}) &= \bar{\partial}(G \circ H)(\bar{x})^T \nabla g(G(H(\bar{x}))) \\ &= H'(\bar{x})^T \bar{\partial} G(H(\bar{x}))^T \nabla g(G(H(\bar{x}))) \\ &= H'(\bar{x})^T \operatorname{diag}(\partial p_i(H_i(\bar{x}))) \nabla g(G(H(\bar{x}))), \end{aligned}$$

where the first equality is due to [10, Theorem 2.6.6], the second one follows (with $d = \nabla g(G(H(\bar{x})))$) from Lemma 4.3, and the third one exploits the componentwise structure of G.

Now, we show that $\operatorname{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu) \subset \bar{\partial} f(\bar{x})$ for all $\bar{x} \in \mathbb{R}^n$. To this end, we first note that

$$\nabla_x s_f(x,\mu) = H'(x)^T \mathrm{diag}\left(\frac{d}{dt} s_{p_i}(H_i(x),\mu)\right) \nabla g([s_{p_i}(H_i(x),\mu)]_{i=1}^m),$$

by the ordinary chain rule. Now, let $v \in \operatorname{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu)$ be given. Then there exist sequences $\{x^k\} \to \bar{x}$ and $\{\mu_k\} \downarrow 0$ such that

$$H'(x^k)^T \mathrm{diag}\left(\frac{d}{dt} s_{p_i}(H_i(x^k),\mu_k)\right) \nabla g([s_{p_i}(H_i(x^k),\mu_k)]_{i=1}^m) \to v.$$

Since, due to (14), we have

$$\begin{aligned} &\frac{d}{dt} s_{p_i}(H_i(x^k),\mu_k) = \\ &= a_1\Big(1 - F\Big(\frac{H_i(x^k) - t_2}{\mu_k}\Big)\Big) + \sum_{i=2}^{r-1} a_i\Big(F\Big(\frac{H_i(x^k) - t_i}{\mu_k}\Big) - F\Big(\frac{H_i(x^k) - t_{i+1}}{\mu_k}\Big)\Big) \\ &\quad + a_r F\Big(\frac{H_i(x^k) - t_r}{\mu_k}\Big), \end{aligned}$$

and $F : \mathbb{R} \to [0,1]$, the sequence $\{\mathrm{diag}\,(\frac{d}{dt} s_{p_i}(H_i(x^k),\mu_k))\}$ is bounded, hence convergent on a subsequence, with a cluster point $D \in \mathrm{diag}\,(\partial p_i(H_i(\bar{x})))$, due to Lemma 4.1, and hence

$$v = H'(\bar{x})^T D \nabla g(G(H_i(\bar{x}))) \in \bar{\partial} f(\bar{x}),$$

which gives the asserted inclusion. The remaining statements now follow from Lemma 3.1. □

We close this section with drawing the reader's attention to the following result which is an immediate consequence of Example 7.19 and Theorem 9.67 in [23].

Theorem 4.6. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz. Let $\psi : \mathbb{R}^n \to \mathbb{R}_+$ be continuous with $\int_{\mathbb{R}^n} \psi(x)\,dx = 1$ and such that the sets $B(\mu) := \{x \mid \phi(\frac{x}{\mu}) > 0\}$ form a bounded sequence and converge to $\{0\}$ as $\mu \downarrow 0$. Then the function s_f given by*

$$s_f(x,\mu) := \int_{\mathbb{R}^n} f(x-z)\frac{1}{\mu}\psi\big(\frac{z}{\mu}\big)\,dz$$

is a smoothing function for f with

$$\mathrm{conv}\operatorname*{Lim\,sup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x s_f(x,\mu) = \bar{\partial} f(\bar{x}) \quad \forall \bar{x} \in \mathbb{R}^n,$$

and when f is regular the convex hull is superfluous.

Although this is a very powerful and very general result it does not cover our analysis, since we do not restrict ourselves to *mollifiers* ψ which are continuous or have compact support, in the sense that is suggested by the boundedness condition above.

5 Final remarks

We investigated smoothing functions, based on integral-convolution, for a class of finite max-functions, which generalizes the analysis in [6] carried out for the plus function. In the main result it was shown that, under reasonable assumptions, (inner and outer) compositions with smooth functions fully agree with the framework layed out for the finite-max functions, and hence a satisfactory calculus is available.

Some words on the relation of Section 4 to the analysis in [6]: In [6, Theorem 1 (i)] it is stated that the assertions of our Theorem 4.5, when applied to $p_i := (\cdot)_+$ $(i = 1, \dots, m)$, were valid without any further assumptions on H or g. However, in the proof of [6, Theorem 1 (i)], the chain rule representation (using our notation)

$$\bar{\partial} f(\bar{x}) = H'(\bar{x})^T \bar{\partial} G(H(\bar{x})) \nabla g(G(H(\bar{x})))$$

is invoked, which is shown to be false by Example 4.4. On the other hand we conjecture that the assertions of [6, Theorem 1 (i)] are still valid, but for the general case considered in Theorem 4.5, we believe that the rank condition on $H'(\bar{x})$ is essential.

Acknowledgements

We gratefully acknowledge the open discussion with Prof. X. Chen [7] on the gap in the proof of [6, Theorem 1 (i)].

References

[1] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty: *Nonlinear Programming. Theory and Algorithms.* John Wiley & Sons, 1993 (second edition).

[2] A. Beck and M. Teboulle: *Smoothing and first order methods: A unified framework.* SIAM Journal on Optimization 22 (2), 2012, pp. 557–580.

[3] D.P. Bertsekas:: *Nondifferentiable optimization via approximation.* Mathematical Programming Study, 1975, pp. 1–25.

[4] J.M. Borwein and A.S. Lewis: *Convex Analysis and Nonlinear Optimization. Theory and Examples.* CMS Books in Mathematics, Springer-Verlag, New York, 2000.

[5] James V. Burke and Tim Hoheisel: *Epi-convergent smoothing with applications to convex composite functions.* Optimization Online, 2012.

[6] X. Chen: *Smoothing methods for nonsmooth, nonconvex minimization.* Mathematical Programming 134, 2012, pp. 71–99.

[7] X. Chen: Private Communictation. By email, 2012.

[8] B. Chen and P.T. Harker: *A non-interior-point continuation method for linear complementarity problems.* SIAM Journal on Matrix Analysis and Applications 14, 1993, pp. 1168–1190.

[9] C. Chen and O.L. Mangasarian: *A class of smoothing functions for nonlinear and mixed complementarity problems.* Computational Optimization and Applications 5, 1996, pp. 97–113.

[10] F.H. Clarke: *Optimization and Nonsmooth Analysis.* John Wiley & Sons, New York 1983.

[11] F. Facchinei and J.-S. Pang: *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volumes I and II.* Springer, New York, 2003.

[12] F. Facchinei, H. Jiang, and L. Qi: *A smoothing method for mathematical programs with equilibrium constraints.* Mathematical Programming 85, 1999, pp. 107–134.

[13] J.-B. Hiriart-Urrruty and C. Lemaréchal: *Convex Analysis and Minimization Algorithms I+II.* Second Corrected Printing, Springer, Berlin, Heidelberg, 1996.

[14] J.-B. Hiriart-Urrruty and C. Lemaréchal: *Fundamentals of Convex Analysis.* Grundlehren Tex Editions, Springer, Berlin, Heidelberg, 2001.

[15] A.D. Ioffe: *Approximate subdifferentials and applications I: The finite dimensional case.* Transactions of the American Mathematical Society 281, 1984, pp. 389–416.

[16] C. Kanzow: *Some noninterior continuation methods for linear complementarity problems.* SIAM Journal on Matrix Analysis and Applications 17, 1996, pp. 851–868.

[17] B.S. Mordukhovich: *Maximum principle in the optimal time control problem with nonsmooth constraints.* Journal of Applied Mathematics and Mechanics 40, 1976, pp. 960–969.

[18] B.S. Mordukhovich: *Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems.* Soviet Mathematical Doklady, 22, pp. 526–530.

[19] B.S. Mordukhovich: *Variational Analysis and Generalized Differentiation I. Basic Theory.* A Series of Comprehensive Studies in Mathematics, Vol. 330, Springer, Berlin, Heidelberg, 2006.

[20] Y. Nesterov: *Smooth minimization of non-smooth functions.* Mathematical Programming 103, 2005, pp. 127–152.

[21] J. Nocedal and S.J. Wright: *Numerical Optimization.* Springer Series in Operations Research, Springer-Verlag, New York, 1999.

[22] L. Qi: *Convergence analysis of some algorithms for solving nonsmooth equations.* Mathematics of Operations Research 18, 1993, pp. 227–244.

[23] R.T. Rockafellar and R.J.-B. Wets: *Variational Analysis.* A Series of Comprehensive Studies in Mathematics, Vol. 317, Springer, Berlin, Heidelberg, 1998.

[24] S. Scholtes: *Introduction to piecewise smooth equations.* Habilitation Thesis, University of Karlsruhe, 1994.

[25] *S. Smale: Algorithms for solving equations.* In: Proceedings of the International Congress of Mathematicians, Americal Mathematical Society, Providence, 1987, pp. 172–195.