.

EPI-CONVERGENT SMOOTHING WITH APPLICATIONS TO CONVEX COMPOSITE FUNCTIONS

JAMES V. BURKE AND TIM HOHEISEL

ABSTRACT. Smoothing methods have become part of the standard tool set for the study and solution of nondifferentiable and constrained optimization problems as well as a range of other variational and equilibrium problems. In this note we synthesize and extend recent results due to Beck and Teboulle on infimal convolution smoothing for convex functions with those of X. Chen on gradient consistency for nonconvex functions. We use epi-convergence techniques to define a notion of epi-smoothing that allows us to tap into the rich variational structure of the subdifferential calculus for nonsmooth, nonconvex, and nonfinite-valued functions. As an illustration of the versatility and range of epi-smoothing techniques, the results are applied to the general constrained optimization for which nonlinear programming is a special case.

1 Introduction

A standard approach to solving nonsmooth and constrained optimization problems is to solve a related sequence of unconstrained smooth approximations [7, 8, 9, 21, 29, 33, 37, 48, 53]. The approximations are constructed so that cluster points of the solutions or stationary points of the approximating smooth problems are solutions or stationary points for the limiting nonsmooth or constrained optimization problem. In the setting of convex programming, there is now great interest in these methods in the very large-scale setting (e.g., see [26, 44, 48, 49]), where first-order methods for convex nonsmooth optimization have been very successful. At the same time, there are many recent applications of smoothing methods to general nonlinear programming, equilibrium, and mathematical programs with equilibrium constraints, e.g., see [10, 18, 19, 20, 22, 23, 31, 34, 35, 36]. This paper is concerned with synthesizing and expanding the ideas presented in two important recent papers on smoothing. The first is by Beck and Teboulle [7] which develops a smoothing framework for nonsmooth convex functions based on *infimal convolution.* The second is by Chen [21] which, among other things, studies the notion of *gradient consistency* for smoothing sequences. Our goal is to extend the ideas presented in [7] for convex functions to the class of *convex composite* functions and provide conditions under which this extension preserves the gradient consistency. Our primary tool in this analysis is the notion of variational convergence called *epi-convergence* [4, 5, 53]. Epi-convergence is ideally suited to the study of the variational properties of parametrized families of functions allowing, for example, the development of a calculus of smoothing functions which is essential for the

---

2010 *Mathematics Subject Classification.* 49J52, 49J53, 90C26, 90C30, 90C46.

Key words and phrases. smoothing method, subdifferential calculus, epi-convergence, infimal convolution, Moreau envelope, convex composite function, Karush-Kuhn-Tucker conditions.

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

applications to the nonlinear inverse problems that we have in mind [1, 2, 3]. Epi-smoothing is a weaker notion of smoothing than those considered in [7, Definition 2.1] where complexity results are one of the key contributions [7, Theorem 3.1]. It is the complexity results that require stronger notions of smoothing. On the other hand, our goal is to establish limiting variational properties in nonconvex applications, in particular, gradient consistency (see [21, Theorem 1] and [15, Theorem 4.5]).

We begin in Section 2 by introducing the notions of epigraphical and set-valued convergence upon which our analysis rests. We also introduce the tools from subdifferential calculus [53] that we use to establish gradient consistency. In Section 3, we define *epi-smoothing functions* and develop a calculus for these smoothing functions that includes basic arithmetic operations as well as composition. In Section 4, we give conditions under which the Beck and Teboulle [7] approach to smoothing via infimal convolution also gives rise to epi-smoothing functions that satisfy gradient consistency. These results are then applied to *Moreau envelopes* (e.g., see [53]) and *extended piecewise linear-quadratic functions.* In Section 5, we introduce convex composite functions an give conditions under which the epi-smoothing results of Section 4 can be extended to this class of functions. In Section 6, we conclude by applying the smoothing results for convex composite functions to general nonlinear programming problems.

Notation: Most of the notation used is standard. An element $x \in \mathbb{R}^n$ is understood as a column vector, and $\overline{\mathbb{R}} := [-\infty, +\infty]$ is the *extended real-line.* 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. The *null space* of A is the set

$$\operatorname{nul} A := \{x \in \mathbb{R}^n \mid Ax = 0\}.$$

By $I_{n\times n}$ we mean the $n \times n$ identity matrix and by $\mathsf{ones}(n, m)$ the $n \times m$ matrix each of whose entries is the number 1.

Unless otherwise stated, $\|\cdot\|$ denotes the *Euclidean norm* on $\mathbb{R}^n$ and $\|\cdot\|_1$ denotes the 1-*norm.* If $C \subset \mathbb{R}^n$ is nonempty and closed, the *Euclidean distance function* for C is given by

$$\operatorname{dist}(y \mid C) := \inf_{z\in C} \|y - z\|. \tag{1}$$

When C is convex it is easily established that the distance function is a convex function, and the optimization (1) has a unique solution $\Pi_C(y)$ which is called the *projection of* y *onto* C.

For a sequence $\{x^k\} \subset \mathbb{R}^n$ and a (nonempty) set $X \subset \mathbb{R}^n$ we abbreviate the fact that x^k converges to $\bar{x} \in \mathbb{R}^n$ and $x^k \in X$ for all $k \in \mathbb{N}$ by

$$x^k \to_X \bar{x}.$$

Moreover, for a function $f : \mathbb{R}^n \to \mathbb{R}$, define

$$x^k \to_f \bar{x} \quad :\Longleftrightarrow \quad x^k \to \bar{x} \quad \text{and} \quad f(x^k) \to f(\bar{x}).$$

This type of convergence coincides with ordinary convergence when f is continuous.

For 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.,

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

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 $\mathbb{R}^m$. The *graph* of S is the set

$$\operatorname{gph} S := \{(x, y) \mid y \in S(x)\},$$

which is equivalent to the classical notion when S is single-valued.

2 Preliminaries

In this section we review certain concepts from variational and nonsmooth analysis employed in the subsequent analysis. The notation is primarily based on [53].
For an extended real-valued function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ its *epigraph* is given by

$$\operatorname{epi} f := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \le \alpha\},$$

and its *domain* is the set

$$\operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\}.$$

The notion of the epigraph allows for very handy definitions of a number of properties for extended real-valued functions (see [41, 52, 53]).

Definition 2.1 (Closed, proper, convex functions)**.** *A function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is called* lower semicontinuous (lsc) *(or* closed*) if* $\operatorname{epi} f$ *is a closed set.* f *is called* convex *if* $\operatorname{epi} f$ *is a convex set. A convex function* f *is said to be* proper *if there exists* $x \in \operatorname{dom} f$ *such that* $f(x) \in \mathbb{R}$.

Note that these definitions coincide with the usual concepts for ordinary real-valued functions. Moreover, it holds that a convex function is always (locally Lipschitz) continuous on the (relative) interior of its domain [52, Theorem 10.4].
Furthermore, we point out that, in what follows, for an lsc, convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, we always exclude the case $f \equiv +\infty$, which means that we deal with proper functions.
An important function in this context is the *(convex) indicator function* of a set $C \subset \mathbb{R}^n$ given by $\delta(\cdot \mid C) : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with

$$\delta(x \mid C) = \begin{cases} 0 & \text{if} \quad x \in C, \\ +\infty & \text{if} \quad x \notin C. \end{cases}$$

The indicator function $\delta(\cdot \mid C)$ is convex if and only if C is convex, and $\delta(\cdot \mid C)$ is lsc if and only if C is closed.

A crucial role in our upcoming analysis is played by the concept of *epi-convergence*, which is now formally defined.

Definition 2.2 (Epi-convergence)**.** *We say that a sequence* $\{f_k\}$ *of functions* $f_k : \mathbb{R}^n \to \overline{\mathbb{R}}$ epi-converges *to* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *if*

$$\operatorname*{Lim}_{k\to\infty} \operatorname{epi} f_k = \operatorname{epi} f,$$

where a Painlevé-Kuratowski *notion of set-convergence as given by* [53, Definition 4.1] *is employed.*
In this case we write

$$\mathrm{e-}\lim f_k = f \quad \text{or} \quad f_k \overset{e}{\to} f.$$

Epi-convergence for sequences of convex functions goes back to Wijsman [58, 59], where it is called *infimal convergence*. The term epi-convergence arguably is due to Wets [57].
A handy characterization of epi-convergence is given by

$$f_k \overset{e}{\to} f \quad \iff \quad \forall \bar{x} \in \mathbb{R}^n \begin{cases} \forall \{x^k\} \to \bar{x}: & \liminf f_k(x^k) \geq f(\bar{x}), \\ \exists \{x^k\} \to \bar{x}: & \limsup f_k(x^k) \leq f(\bar{x}), \end{cases} \tag{2}$$

see [53, Proposition 7.2], which we invoke in several places. For extensive surveys of epi-convergence we refer the reader to [4] or [53, Chapter 7].

We make use of the *regular* and *limiting subdifferentials* to describe the variational behavior of nonsmooth functions. In constructing the limiting subdifferential, we employ the *outer limit* for a set-valued mapping, which we now define along with the *inner limit*:
For $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and $X \subset \mathbb{R}^n$ the *outer limit* of S at $\bar{x}$ *relative to* X is given by

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

and the *inner limit* of S at $\bar{x}$ *relative to* X is defined by

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

We say that S is *outer semicontinuous (osc)* at $\bar{x}$ *relative to* X if

$$\operatorname*{Lim\,sup}_{x \to_X \bar{x}} S(x) \subset S(\bar{x}).$$

In case that outer and inner limit coincide, we write

$$\operatorname*{Lim}_{x \to_X \bar{x}} S(x) := \operatorname*{Lim\,sup}_{x \to_X \bar{x}} S(x),$$

and say that S is *contiuous* at $\bar{x}$ *relative to* X.

Definition 2.3 (Regular and limiting subdifferential)**.** *Let* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *and* $\bar{x} \in \operatorname{dom} f$.

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

$$\hat{\partial} f(\bar{x}) := \left\{ v \mid f(x) \geq f(\bar{x}) + v^T(x - \bar{x}) + o(\|x - \bar{x}\|) \right\}.$$

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

$$\partial f(\bar{x}) := \operatorname*{Lim\,sup}_{x \to_f \bar{x}} \hat{\partial} f(x).$$

There are other ways to obtain the limiting subdifferential than the one described above, which goes back to Mordukhovich, e.g., cf. [45]. See [17] or [43] for a construction of the limiting subdifferential via *Dini-derivatives*.
It is a well-known fact, see [53, Proposition 8.12], that if $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex, both the limiting and the regular subdifferential coincide with the subdifferential of convex analysis, i.e.,

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

The above subdifferentials are closely tied to *normal cones*, in fact the *regular* and the *limiting normal cone*, see [53, Definition 6.3], of a closed set $C \subset \mathbb{R}^n$ at $\bar{x} \in C$ can be expressed as

$$\hat{N}(\bar{x} \mid C) = \hat{\partial}\delta(\bar{x} \mid C) \quad \text{and} \quad N(\bar{x} \mid C) = \partial\delta(\bar{x} \mid C),$$

see [53, Exercise 8.14].
An important concept in the context of subdifferentiation is *(subdifferential) regularity*. We say that $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is *(subdifferentially) regular* at $\bar{x} \in \operatorname{dom} f$ if

$$N((\bar{x}, f(\bar{x})) \mid \operatorname{epi} f) = \hat{N}((\bar{x}, f(\bar{x})) \mid \operatorname{epi} f).$$

Note that this regularity notion coincides with the one used in [24], see the discussion on page 61 in [24] in combination with [53, Corollary 6.29].

3 Epi-Smoothing Functions

In this section we lay out the general framework for the smoothing functions studied in this paper. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be lsc. We say $s_f : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$ is an *epi-smoothing function* for f if the following two conditions are satisfied:

(i) $s_f(\cdot, \mu_k)$ epi-converges to f for all $\{\mu_k\} \downarrow 0$, written

$$\mathrm{e-}\lim_{\mu \downarrow 0} s_f(\cdot, \mu) = f, \tag{3}$$

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

Note that (3) is always fulfilled, see [53, Theorem 7.11], under the following condition

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

which is called *continuous convergence* in [53]. As we will see in Section 4, however, continuous convergence can be an excessively strong assumption, especially when dealing with non-finite valued functions.
The following result provides an elementary calculus for epi-smoothing functions.

Proposition 3.1. *Let $g, h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be lsc and let s_g and s_h be epi-smoothing functions for g and h, respectively.*

a) *If s_g converges continuously to g, then $s_f := s_g + s_h$ is an epi-smoothing function for $f := g + h$.*
b) *If g is continuously differentiable, then $s_f := g + s_h$ is an epi-smoothing function $f := g + h$.*
c) *If $\lambda > 0$, then λs_g is an epi-smoothing function for λg.*
d) *If $A \in \mathbb{R}^{m \times n}$ has rank m and $b \in \mathbb{R}^m$, then $s_g(\cdot, \cdot) := s_g(A(\cdot) + b, \cdot)$ is an epi-smoothing function for $f := g(A(\cdot) + b)$.*

Proof. Item a) follows from [53, Theorem 7.46], while b) follows from a) and the fact that g is a continuously convergent epi-smoothing function for itself. Item c) is provided by [53, Exercise 7.8 d)]. Item d) is an immediate consequence of Theorem 3.2 and the discussion up front. □

To obtain a more powerful chain rule than the one given in item d) above, we need to invoke more refined tools from variational analysis. One such tool is *metric regularity* (e.g., see [17, 47, 53]), originally defined for set-valued mappings. For a single-valued mapping $F:\mathbb{R}^n\to\mathbb{R}^m$ we say that F is *metrically regular* at $\bar{x}\in\mathbb{R}^n$ if there exists $\gamma>0$ and neighborhoods W of $\bar{x}$ and V of $F(\bar{x})$ such that

$$\operatorname{dist}(x,F^{-1}(y))\le\gamma\|F(x)-y\|\quad\forall x\in W,y\in V.$$

We say that F is metrically regular, if it is metrically regular at every $\bar{x}\in\mathbb{R}^n$. In particular, F is metrically regular if it is a *locally Lipschitz homeomorphism* (e.g., see [53, Corollary 9.55]). Mordukhovich has shown that metric regularity can be fully characterized via the *coderivative criterion*, e.g., see [47, 53]. In the case of a single-valued, continuously differentiable map $F:\mathbb{R}^n\to\mathbb{R}^m$ the coderivative criterion reduces to the condition that $\operatorname{rank}F'(\bar{x})=m$, that is,

$$F \text{ is metrically regular at } \bar{x}\quad\Longleftrightarrow\quad\operatorname{rank}F'(\bar{x})=m.$$

Theorem 3.2. *Let $g:\mathbb{R}^m\to\mathbb{R}\cup\{+\infty\}$ and let s_g be an epi-smoothing function for g. Furthermore, let $F:\mathbb{R}^n\to\mathbb{R}^m$ be continuously differentiable and metrically regular. Then $s_f:=s_g(F(\cdot),\cdot)$ is an epi-smoothing function for $f:=g\circ F$.*

Proof. The smoothness properties are obvious from the assumptions. Next, let $\{\mu_k\}\downarrow$ be given and put $g_k:=s_g(\cdot,\mu_k)$ and $f_k:=g_k\circ F$. We need to show that $f_k\overset{e}{\to}f$. For this purpose, we invoke the characterization of epi-convergence as provided by (2). To this end, let $\bar{x}\in\mathbb{R}^n$ and $\{x^k\}\to\bar{x}$ be given. Then it follows from the fact that $g_k\overset{e}{\to}g$ and (2) that

$$\liminf_k f_k(x^k)=\liminf_k g_k(F(x^k))\ge g(F(\bar{x}))=f(\bar{x}).\tag{5}$$

Moreover, as $g_k\overset{e}{\to}g$, (2) yields a sequence $\{y^k\}\to\bar{y}:=F(\bar{x})$ such that

$$\limsup_k g_k(y^k)\le g(\bar{y}).$$

Since F is metrically regular at $\bar{x}$, we obtain a sequence $\{x^k\}\to\bar{x}$ such that $F(x^k)=y^k$ for all $k\in\mathbb{N}$. This, in turn, gives

$$\limsup_k f_k(x^k)=\limsup_k g_k(y^k)\ge\bar{y}=f(\bar{x}).$$

This, together with (5) proves (2) for f_k with respect to f, and this concludes the proof. □

Although epi-convergence is arguably a mild condition, it still provides desirable convergence behavior for minimization in the following sense:

Theorem 3.3. [53, Theorem 7.33] *Suppose the sequence $\{f_k\}$ is* eventually level-bounded *(see* [53, p. 266]*), and $f_k\overset{e}{\to}f$ with f_k and f lsc and proper. Then*

$$\inf f_k\to\inf f\quad(finite).$$

Now, suppose a numerical algorithm produces sequences $\{x^k\}\to\bar{x}$ and $\{\mu_k\}\downarrow 0$ such that

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

A natural question to ask in this context is whether $\bar{x}$ is a critical point of f in the sense that $0 \in \partial f(\bar{x})$. A sufficient condition is, clearly, provided by

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

The next result shows that the converse inclusion is always valid if $s_f(\cdot, \mu) \xrightarrow{e} f$.

Lemma 3.4. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be lsc and s_f an epi-smoothing function for f. Then for $\bar{x} \in \operatorname{dom} f$ we have*

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

Proof. Let $v \in \partial f(\bar{x})$ be given. Since by assumption $\text{e}-\lim_{\mu \downarrow 0} s_f(\cdot, \mu) = f$ we may invoke [53, 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

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

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 assertion follows. $\square$

A major contribution of this paper is the construction of smoothing functions having the property that

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

at any point $\bar{x} \in \operatorname{dom} f$. This condition implies the notion of gradient consistency defined in [21, Equation (4)] which is obtained by taking the convex hull on both sides of this equation. However, since all of the functions we consider are subdifferentially regular, Lemma 3.4 implies that (6) is equivalent to gradient consistency.

4 Epi-Smoothing via Infimal Convolution

In this section we show that the class of smoothing functions for nonsmooth, convex and lsc functions introduced in [7] fits into the framework layed out in Section 3. As a by-product, we show that *Moreau envelopes* fulfill the requirements of our smoothing setup.

The approach taken in [7] is based on *infimal convolution* [6, 41, 42, 52, 53]. Given two (extended real-valued) functions $f_1, f_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$ the *inf-convolution* (or *epi-sum*, see Lemma 4.2 b) in this context) is the function $f_1 \# f_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$ defined by

$$(f_1 \# f_2)(x) := \inf_{u \in \mathbb{R}^n} \{f_1(u) + f_2(x - u)\}.$$

In what follows we assume that

(A) $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is proper, lsc, and convex, and
(B) $\omega : \mathbb{R}^n \to \mathbb{R}$ is convex and continuously differentiable with Lipschitz gradient.

Moreover, for $\mu > 0$, define the function $\omega_\mu : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ by

$$\omega_\mu(y) := \mu \omega\left(\frac{y}{\mu}\right).$$

Obviously, ω_μ is also convex and continuously differentiable with Lipschitz gradient. In [7], the authors consider the (convex) function

$$(g\#\omega_\mu)(x) = \inf_{u\in\mathbb{R}^n} \left\{ g(u) + \mu\omega\Big(\frac{x-u}{\mu}\Big) \right\} \quad (\mu > 0)$$

as a smoothing function for g. We now investigate conditions on ω for which the inf-convolution $g\#\omega_\mu$ serves as an epi-smoothing function in the sense of Section 3. In this context, the notion of *coercivity* plays a key role where it arises as a natural assumption on the function ω. Several different notions of coercivity occur in the literature. We now define those useful to our study.

Definition 4.1 (Coercive functions)**.** *Let $f:\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be lsc and convex.*

a) *f is called* 0-coercive *if*

$$\lim_{\|x\|\to\infty} f(x) = +\infty.$$

b) *f is called* 1-coercive *if*

$$\lim_{\|x\|\to\infty} \frac{f(x)}{\|x\|} = +\infty.$$

The first result establishes important properties of the function $g\#\omega_\mu$.

Lemma 4.2. *If ω is 1-coercive (or 0-coercive and g bounded from below) the following holds:*

a) *$g\#\omega_\mu$ is finite-valued, i.e., $g\#\omega_\mu : \mathbb{R}^n \to \mathbb{R}$, and for all $x\in\mathbb{R}^n$ we have*

$$(g\#\omega_\mu)(x) = \min_{u\in\mathbb{R}^n} \left\{ g(u) + \mu\omega\Big(\frac{x-u}{\mu}\Big) \right\}$$

i.e.,

$$\operatorname*{argmin}_{u\in\mathbb{R}^n} \left\{ g(u) + \mu\omega\Big(\frac{x-u}{\mu}\Big) \right\} \neq \emptyset.$$

b) *We have*

$$\operatorname{epi} g\#\omega_\mu = \operatorname{epi} g + \operatorname{epi} \omega_\mu.$$

c) *$g\#\omega_\mu$ is continuously differentiable with*

$$\nabla(g\#\omega_\mu)(x) = \nabla\omega\Big(\frac{x-u_\mu(x)}{\mu}\Big) = \nabla\omega_\mu(x-u_\mu(x)) \quad \forall x\in\mathbb{R}^n,$$

where $u_\mu(x) \in \operatorname{argmin}_{u\in\mathbb{R}^n} \left\{ g(u) + \mu\omega\Big(\frac{x-u}{\mu}\Big) \right\}$.

Proof. The assertion that

$$(g\#\omega_\mu)(x) < +\infty \quad \forall x\in\mathbb{R}^n$$

is due to the fact that ω is finite-valued and $g \not\equiv +\infty$. Moreover, ω_μ obviously inherits the respective coercivity properties from ω. Hence, the remainder of a) follows immediately from [6, Proposition 12.14].
In turn, b) follows from a) and [6, Proposition 12.8 (ii)].
Item c) is an immediate consequence of a) together with [7, Theorem 4.2 (c)]. □

The following auxiliary result, which is key for establishing epigraphical limit behavior of $g\#\omega_\mu$, states that the epigraphical limit of ω_μ for $\mu \downarrow 0$ is $\delta(\cdot \mid \{0\})$ if and only if ω is 1-coercive.

Lemma 4.3. *ω is 1-coercive if and only if*

$$\operatorname{e-}\lim_{\mu\downarrow 0}\omega_\mu = \delta(\cdot \mid \{0\}).$$

Proof. First, let ω be 1-coercive:
We start by showing that $\operatorname{Lim\,sup}_{\mu\downarrow 0}\operatorname{epi}\omega_\mu \subset \{0\}\times\mathbb{R}_+ = \operatorname{epi}\delta(\cdot \mid \{0\})$.
To this end, let $(\bar z,\bar\alpha)\in \operatorname{Lim\,sup}_{\mu\downarrow 0}\operatorname{epi}\omega_\mu$. Then there exist sequences $\{z^k\}\to\bar z$, $\{\alpha_k\}\to\bar\alpha$ and $\{\mu_k\}\downarrow 0$ such that

$$\mu_k\omega\Big(\frac{z^k}{\mu_k}\Big) \le \alpha_k \quad \forall k\in\mathbb{N}. \tag{7}$$

This can be written as

$$\omega\Big(\frac{z^k}{\mu_k}\Big) \le \frac{\alpha_k}{\mu_k} \quad \forall k\in\mathbb{N}.$$

It is immediately clear from this representation, that $\bar\alpha\ge 0$, since otherwise the right-hand side would tend to $-\infty$, while the left-hand side remains either convergent on a subsequence (if $\{\frac{z^k}{\mu_k}\}$ is bounded) or tends to $+\infty$ (if $\{\frac{z^k}{\mu_k}\}$ is unbounded).
Now, suppose that $\bar z\ne 0$. Then $\{\frac{z^k}{\mu_k}\}$ is unbounded and (7) can be rewritten as

$$\frac{\omega\Big(\frac{z^k}{\mu_k}\Big)}{\|\frac{z^k}{\mu^k}\|} \le \frac{\alpha_k}{\|z^k\|} \quad \forall k\in\mathbb{N}.$$

By the 1-coercivity of ω the left-hand side tends to $+\infty$, while the right-hand side is bounded, which is a contradiction. Hence, we have proven that $\bar z = 0$ and $\bar\alpha\ge 0$, which shows that, in fact, $\operatorname{Lim\,sup}_{\mu\downarrow 0}\operatorname{epi}\omega_\mu\subset\{0\}\times\mathbb{R}_+$.
We now show that $\operatorname{Lim\,inf}_{\mu\downarrow 0}\operatorname{epi}\omega_\mu \supseteq \{0\}\times\mathbb{R}_+$. For these purposes, let $\bar\alpha\ge 0$ and $\{\mu_k\}\downarrow 0$ be given. Then choose $z^k := 0$ and $\alpha_k := \bar\alpha + \mu_k\omega(0) \ge \omega_{\mu_k}(z^k)$. Then $(z^k,\alpha_k)\in\operatorname{epi}\omega_{\mu_k}$ for all $k\in\mathbb{N}$ and $(z^k,\alpha_k)\to(0,\bar\alpha)$. This shows that $\operatorname{Lim\,inf}_{\mu\downarrow 0}\operatorname{epi}\omega_\mu\supseteq\{0\}\times\mathbb{R}_+$.
Putting together all the pieces of information, we see that

$$\operatorname*{Lim}_{\mu\downarrow 0}\operatorname{epi}\omega_\mu = \operatorname{epi}\delta(\cdot \mid \{0\}),$$

i.e.,

$$\operatorname{e-}\lim_{\mu\downarrow 0}\omega_\mu = \delta(\cdot \mid \{0\}).$$

Now, suppose that ω is not 1-coercive. Then there exists an unbounded sequence $\{x^k\}$ such that either

$$\frac{\omega(x^k)}{\|x^k\|}\to -\infty$$

or $\Big\{\frac{\omega(x^k)}{\|x^k\|}\Big\}$ is bounded. Put $\mu_k := \frac{1}{\|x^k\|}\to 0$. Then

$$\omega_{\mu_k}\Big(\frac{x^k}{\|x^k\|}\Big) = \frac{\omega(x^k)}{\|x^k\|},$$

and we have

$$\Big(\frac{x^k}{\|x^k\|},\omega_{\mu_k}\Big(\frac{x^k}{\|x^k\|}\Big)\Big)\in\operatorname{epi}\omega_{\mu_k} \quad \forall k\in\mathbb{N}. \tag{8}$$

If $\frac{\omega(x^k)}{\|x^k\|} \to -\infty$, we infer that ω_{u_k} does not converge epigraphically at all (in particular not to $\delta(\cdot \mid \{0\})$) from (2), since we have $\liminf_{k\to\infty} \omega_{\mu_k}\left(\frac{x^k}{\|x^k\|}\right) \to -\infty$. In case that $\left\{\frac{\omega(x^k)}{\|x^k\|}\right\}$ is bounded, we may assume w.l.g. that

$$\frac{\omega(x^k)}{\|x^k\|} \to \bar{\omega}$$

for some $\bar{\omega} \in \mathbb{R}$. Then we infer from (8) that

$$(\bar{x}, \bar{\omega}) \in \operatorname*{Lim\,sup}_{k\to\infty} \operatorname{epi} \omega_{\mu_k}$$

with $\bar{x} \neq 0$ being an accumulation point of $\left\{\frac{x^k}{\|x^k\|}\right\}$. But $(\bar{x}, \bar{\omega}) \notin \operatorname{epi} \delta(\cdot \mid \{0\})$, which concludes the proof. □

The following lemma establishes monotonicity properties for the family of functions $g\#\omega_\mu$, which come into play in Section 5.

Lemma 4.4. *If $\omega(0) \leq 0$, then for all $x \in \mathbb{R}^n$ the function $\mu \mapsto (g\#\omega_\mu)(x)$ is nondecreasing on $\mathbb{R}_{++}$ and bounded by $g(x)$ from above.*

Proof. Let $y \in \mathbb{R}^n$. Then for $\mu_1 > \mu_2 > 0$ we have

$$\begin{aligned} \omega\Big(\frac{y}{\mu_1}\Big) &= \omega\Big(\frac{\mu_2}{\mu_1}\frac{y}{\mu_2} + \Big(1 - \frac{\mu_2}{\mu_1}\Big)0\Big) \\ &\leq \frac{\mu_2}{\mu_1}\omega\Big(\frac{y}{\mu_2}\Big) + \Big(1 - \frac{\mu_2}{\mu_1}\Big)\omega(0) \\ &\leq \frac{\mu_2}{\mu_1}\omega\Big(\frac{y}{\mu_2}\Big). \end{aligned}$$

Multiplying by μ_1 yields

$$\omega_{\mu_1}(y) \leq \omega_{\mu_2}(y) \quad \forall y \in \mathbb{R}^n,$$

and hence for $x \in \mathbb{R}^n$ arbitrarily given, we have

$$g(u) + \omega_{\mu_1}(x - u) \leq g(u) + \omega_{\mu_2}(x - u) \quad \forall u \in \mathbb{R}^n.$$

Taking the infimum over all $u \in \mathbb{R}^n$ gives

$$(g\#\omega_{\mu_1})(x) \leq (g\#\omega_{\mu_2})(x),$$

which concludes the proof due the choice of μ_1 and μ_2. □

The following result establishes the desired epi-convergence properties of the inf-convolutions. Note that, to our knowledge, we cannot deduce it from known results such as [53, Proposition 7.56] or [5, Theorem 4.2], since our assumptions do not meet the requirements for the application of these results. In particular, we do not assume g to be bounded from below.

Proposition 4.5. *If ω is 1-coercive, then*

$$\mathrm{e}\!-\!\lim_{\mu\downarrow 0} g\#\omega_\mu = g.$$

Proof. The fact that $\operatorname{Lim\,inf}_{\mu\downarrow 0} \operatorname{epi} g\#\omega_\mu \supseteq \operatorname{epi} g$ follows immediately from [53, Theorem 4.29 a)] when applied to the respective epigraphs.
Therefore, it is enough to show that $\operatorname{Lim\,sup}_{\mu\downarrow 0} \operatorname{epi} g\#\omega_\mu \subset \operatorname{epi} g$.
To this end, pick $(\bar{x}, \bar{\alpha}) \in \operatorname{Lim\,sup}_{\mu\downarrow 0} \operatorname{epi} g\#\omega_\mu$ arbitrarily. Then there exist sequences $\{\mu_k\} \downarrow 0, \{x^k\} \to \bar{x}$ and $\alpha_k \to \bar{\alpha}$ such that

$$(g\#\omega_{\mu_k})(x^k) \leq \alpha_k \quad \forall k \in \mathbb{N}. \tag{9}$$

With

$$u^k \in \operatorname*{argmin}_{u\in\mathbb{R}^n} \Big\{ g(u) + \mu_k \omega\Big(\frac{x^k - u}{\mu_k}\Big)\Big\},$$

(9) can be written as

$$g(u^k) + \mu_k \omega\Big(\frac{x^k - u^k}{\mu_k}\Big) \leq \alpha_k \quad \forall k \in \mathbb{N}. \tag{10}$$

Using the fact, cf. [6, Theorem 9.19], that the convex, lsc function g is minorized by an affine function, say $x \mapsto b^T x + \beta$, this leads to

$$b^T u^k + \beta + \mu_k \omega\Big(\frac{x^k - u^k}{\mu_k}\Big) \leq \alpha_k \quad \forall k \in \mathbb{N}.$$

If we assume that $\{u^k\}$ does not convergence to $\bar{x}$, we can rewrite this (for k sufficiently large) as

$$\frac{\omega\Big(\frac{x^k - u^k}{\mu_k}\Big)}{\|\frac{x^k - u^k}{\mu_k}\|} \leq \frac{\alpha_k - b^T u^k - \beta}{\|x^k - u^k\|}.$$

Whether $\{u^k\}$ is unbounded or not, we obtain a contradiction, since the left-hand side tends to $+\infty$, as ω is 1-coercive, while the right-hand side remains bounded.

Hence, $\{u^k\} \to \bar{x}$. We now claim that $g(u^k) \not\to +\infty$, and hence, in particular, $\bar{x} \in \operatorname{dom} g$. If this were not the case, we invoke [6, Theorem 9.19] again to get an affine minorant of ω, say $x \mapsto c^T x + \gamma$, and infer from (10) that

$$g(u^k) + c^T(u^k - x^k) + \mu_k \gamma \leq \alpha_k \quad \forall k \in \mathbb{N}.$$

This, however, leads to a contradiction if $g(u^k) \to +\infty$ since $c^T(u^k - x^k) + \mu_k\gamma \to 0$ and $\alpha_k \to \bar{\alpha} < +\infty$. Thus, we have shown that $\{g(u^k)\}$ is bounded from above. Since g is lsc and $u^k \to \bar{x}$, we also know that $\liminf_{k\to\infty} g(u^k) \geq g(\bar{x})$. Hence, we may as well assume that $g(u^k) \to \hat{g} \geq g(\bar{x})$ and, in particular, we have $\bar{x} \in \operatorname{dom} g$. We now infer from (10) that

$$(x^k - u^k, \alpha_k - g(u^k)) \in \operatorname{epi} \omega_{\mu_k} \quad \forall k \in \mathbb{N}.$$

Since $x^k - u^k \to 0$ and $\alpha_k - g(u^k) \to \alpha - \hat{g}$, Lemma 4.3 implies

$$(0, \bar{\alpha} - \hat{g}) \in \operatorname*{Lim\,sup}_{\mu\downarrow 0} \operatorname{epi} \omega_\mu \subset \operatorname{epi} \delta(\cdot \mid \{0\}).$$

This immediately gives

$$g(\bar{x}) \leq \hat{g} \leq \bar{\alpha},$$

i.e., $(\bar{x}, \bar{\alpha}) \in \operatorname{epi} g$, which concludes the proof. □

We are now in a position to state the main result of this section.

Theorem 4.6. *If ω is 1-coercive then the function $s_g : (x,\mu) \mapsto (g\#\omega_\mu)(x)$ is an epi-smoothing function for g with*

$$\operatorname{gph} \nabla_x s_g(\cdot,\mu) \underset{\mu\downarrow 0}{\to} \operatorname{gph} \partial g,$$

and hence, in particular,

$$\operatorname*{Lim\,sup}_{\mu\downarrow 0,\, x\to\bar{x}} \nabla_x s_g(x,\mu) = \partial g(\bar{x}) \quad \forall \bar{x} \in \operatorname{dom} g.$$

Proof. Due to Propostion 4.5, we have $\text{e}-\lim_{\mu\downarrow 0} s_g(\cdot,\mu) = \text{e}-\lim_{\mu\downarrow 0} g\#\omega_\mu = g$. The smoothness properties of $\nabla_x s_g(\cdot,\mu) = \nabla g\#\omega_\mu$ follow from Lemma 4.2. The remaining assertion is an immediate consequence of *Attouch's Theorem*, see [53, Theorem 12.35]. This concludes the proof. □

Moreau Envelopes The most prominent choice for ω is given by

$$\omega := \frac{1}{2}\|\cdot\|^2.$$

The resulting inf-convolution of ω_μ with an lsc function $g : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is called the *Moreau envelope* or *Moreau-Yosida regularization* of g and is denoted by $e_\mu g$, i.e.,

$$e_\mu g(x) = \inf_w \Big\{ g(w) + \frac{1}{2\mu}\|w-x\|^2 \Big\}.$$

The set-valued map $P_\mu g : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ given by

$$P_\mu g(x) := \operatorname*{argmin}_w \Big\{ g(w) + \frac{1}{2\mu}\|w-x\|^2 \Big\}$$

is called the *proximal mapping* for g.
The following properties of Moreau envelopes and proximal mappings of convex functions are well known, see [52, 53] or [41].

Proposition 4.7. *Let $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be lsc and convex and $\mu > 0$. Then the following holds:*

a) *$P_\mu f$ is single-valued and Lipschitz continuous.*
b) *$e_\mu f$ is convex and smooth with Lipschitz gradient $\nabla e_\mu f$ given by*

$$\nabla e_\mu f(x) = \frac{1}{\mu}[x - P_\mu f(x)].$$

c) $\operatorname{argmin} f = \operatorname{argmin} e_\mu f$.

In view of item c) it is possible to recover the minimzers of a (possibly nonsmooth) convex function by those of its Moreau envelope. Hence, it is not even necessary to drive the smoothing parameter to zero.
Since the function $x \mapsto \frac{1}{2}\|x\|^2$ is 1-coercive, the following result can be formulated as a corollary of Theorem 4.6.

Corollary 4.8. *Let $g : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be lsc and convex. Then $s_g : (x,\mu) \mapsto e_\mu g(x)$ is an epi-smoothing function for g with*

$$\operatorname*{Lim\,sup}_{\mu\downarrow 0,\, x\to\bar{x}} \nabla_x s_g(x,\mu) = \partial g(\bar{x}) \quad \forall \bar{x} \in \operatorname{dom} g.$$

When g is lsc and convex, the fact that $e_\mu g$ epi-converges to g as $\mu \downarrow 0$ is well known (cf. the discussion in [53] after Proposition 7.4).

Extended Piecewise Linear-Quadratic Functions (EPLQ) [53] EPLQ functions play a key role in a wide variety of applications, e.g., *signal denoising* [25, 26], *model selection* [55], *compressed sensing* [27, 28, 38], *robust statistics* [40], *Kalman filtering* [1, 2, 32], and *support vector classifiers* [30, 51, 54]. Examples include arbitrary gauge functionals [53] (e.g., norms), the *Huber penalty* [7, 40], the *hinge loss function* [30, 51, 54], and the *Vapnik penalty* [39, 56]. For an overview of these functions and their statistical properties see [3, 53]. In this section, we show that the Moreau envelope mapping $g \mapsto e_\mu g$ maps the class of EPLQ functions to itself in a very natural way.

Definition 4.9. *The convex function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be extended piecewise linear-quadratic if for some positive integer m there exists a nonempty closed convex set $U \subset \mathbb{R}^m$ (typically polyhedral), an injective matrix $R \in \mathbb{R}^{n\times m}$, a symmetric and positive semi-definite matrix $B \in \mathbb{R}^{m\times m}$, and a vector $b \in \mathbb{R}^m$ such that*

$$g(x) := \theta_{(U,B,R,b)}(x) := \sup_{u\in U} \langle u,\, Rx - b\rangle - \frac{1}{2}u^T Bu. \tag{11}$$

If $m = n$, $R = I$, and $b = 0$, then g is said to be piecewise linear-quadratic *(PLQ).*

Example 4.10 (Examples of EPLQ functions)**.**

(1) *Norms: Let $\|\cdot\|_*$ be a norm with closed unit ball $\mathbb{B}_*$. Then $\|\cdot\|_* = \theta_{(\mathbb{B}_*^\circ,0,I,0)}$, where $\mathbb{B}_*^\circ := \{v \mid \langle v,\, u\rangle \le 1\ \forall\, u \in \mathbb{B}_*\}$.*
(2) *The Huber penalty: Let $\kappa > 0$. Then $\theta_{([-\kappa,\kappa]^n,I,I,0)}$ is the Huber penalty with threshold κ.*
(3) *The Vapnik penalty: Let $\epsilon > 0$ and define $U = [0,1]^{2n}$, $R = [I_{n\times n},\, -I_{n\times n}]^T$, and $b = \epsilon\,\mathsf{ones}(2n,1)$, then $\theta_{(U,0,T,b)}$ is the Vapnik penalty with threshold ϵ.*

Proposition 4.11. *Let $\theta_{(U,B,R,b)}$ be an extended piecewise linear-quadratic function. If B is positive definite or U is bounded, then*

$$e_\mu \theta_{(U,B,R,b)} = \theta_{(U,\hat{B},R,b)},$$

where $\hat{B} = B + \mu RR^T$. Moreover, for each $x \in \mathbb{R}^n$ there exists a saddle-point $(\bar{u},\bar{v}) \in U \times \mathbb{R}^n$ for the closed proper concave-convex saddle-function [52, Section 33]

$$K(u,v) := \langle Rv - b,\, u\rangle - \frac{1}{2}u^T Bu + \frac{1}{2\mu}\|x - v\|^2 - \delta(u \mid U)$$

satisfying $e_\mu g(x) = K(\bar{u},\bar{v})$.

Proof. Regardless of the choice of x, K is coercive in v for each $u \in U$, and if B is positive definite or U is bounded, then $-K$ is coercive in u for each $v \in \mathbb{R}^n$. Hence, by [52, Theorem 37.6], for every $x \in \mathbb{R}^n$, K has a saddle-point $(\bar{u},\bar{v}) \in U \times \mathbb{R}^n$ satisfying

$$\begin{aligned} e_\mu g(x) &= \inf_{v\in\mathbb{R}^n}\sup_{u\in U} K(u,v) \\ &= K(\bar{u},\bar{v}) \\ &= \sup_{u\in U}\inf_{v\in\mathbb{R}^n} K(u,v). \end{aligned}$$

To complete the proof observe that the problem

$$\inf_{v\in\mathbb{R}^n} K(u,v) = -\left[\langle b,\, u\rangle + \frac{1}{2}u^T Bu\right] + \inf_{v\in\mathbb{R}^n}\left[\left\langle v,\, R^T u\right\rangle + \frac{1}{2\mu}\|x - v\|^2\right]$$

has a unique solution at $v(x,u) = x - \mu R^T u$. Plugging this solution into K gives $e_\mu g(x) = \sup_{u\in U} K(u, v(x,u)) = \theta_{(U,\hat{B},R,b)}(x)$. □

Example 4.12 (Lasso-Problem)**.** *Given $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ with $m << n$, consider the nonsmooth optimization problem*

$$\min_x f(x) := \frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1, \tag{12}$$

where $\lambda > 0$. This problem is known in the literature as the Lasso-Problem*, see* [28, 55].

The objective function f is the sum of two convex functions, one is smooth and the other is a nonsmooth PLQ function. By Proposition 3.1, an epi-smoothing function for f can be obtained by computing the Moreau envelope for the 1-norm. This envelope is obtained from the proximal mapping which in this case is commonly referred to in the literature as soft thresholding [25, 26]*. An easy computation shows that*

$$P_\mu\|\cdot\|_1(x) = \begin{cases} x_i + \mu & \text{if} \quad x_i < -\mu, \\ x_i - \mu & \text{if} \quad x_i > \mu, \\ 0 & if \quad |x_i| \le \mu. \end{cases}$$

5 Convex Composite Functions

An important and powerful class of nonsmooth, nonconvex functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is given by

$$f(x) := g(H(x)) \quad \forall x \in \mathbb{R}^n, \tag{13}$$

where $g : \mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ is lsc and convex and $H : \mathbb{R}^n \to \mathbb{R}^m$ (twice) continuously differentiable. These functions go by the name *convex composite*, see, e.g., [11, 12] or [16], and are closely related to *amenable* functions, see [53, Definition 10.32]. Suppose one has an epi-smoothing function s_g of g, then it is a natural question to ask whether $s_f(\cdot,\cdot) := s_g(H(\cdot),\cdot)$ is an epi-smoothing function of f. That is, do the smoothing properties of s_g (with respect to g) carry over to smoothing properties of s_f (with respect to f)? In particular, does the epi-convergence of $s_g(\cdot,\mu)$ to g imply the epi-convergence of $s_f(\cdot,\mu)$ to f ? To clarify this connection, we start with an easy observation for which we give a self-contained proof (an alternative proof can be obtained by applying [53, Formula 4(8)] to the respective epigraphs and the function $F(x,\alpha) := (H(x),\alpha)$ satisfying $\operatorname{epi} f = F^{-1}(\operatorname{epi} g)$).

Lemma 5.1. *Let s_g be an epi-smoothing function for g, and define $s_f(\cdot,\cdot) := s_g(H(\cdot),\cdot)$. Then*

$$\operatorname*{Lim\,sup}_{\mu\downarrow 0} \operatorname{epi} s_f(\cdot,\mu) \subset \operatorname{epi} f.$$

Proof. Let $(\bar{x},\bar{\alpha}) \in \operatorname{Lim\,sup}_{\mu\downarrow 0} \operatorname{epi} s_f(\cdot,\mu)$. Then there exist sequences $\{x^k\} \to \bar{x}, \{\alpha_k\} \to \bar{\alpha}$ and $\{\mu_k\} \downarrow 0$ such that

$$s_g(H(x^k),\mu_k) \le \alpha_k \quad \forall k \in \mathbb{N},$$

i.e.,

$$(H(x^k),\alpha_k) \in \operatorname{epi} s_g(\cdot,\mu_k) \quad \forall k \in \mathbb{N}.$$

Since $(H(x^k),\alpha^k) \to (H(\bar{x}),\bar{\alpha})$ we get from the epi-convergence of $s_g(\cdot,\mu)$ to g that

$$(H(\bar{x}),\bar{\alpha}) \in \operatorname{epi} g,$$

which immediately yields

$$(\bar{x}, \bar{\alpha}) \in \operatorname{epi} f.$$

This proves the result. □

We point out that in the previous result, as well as in the following two results, only continuity of H and no smoothness assumption is needed.

Proposition 5.2. *Let s_g be an epi-smoothing function for g such that $s_g(y, \mu)$ is nondecreasing as $\mu \downarrow 0$ for all $y \in \mathbb{R}^m$. Then for $s_f(\cdot, \cdot) := s_g(H(\cdot), \cdot)$ we have*

$$\mathrm{e}\!-\!\lim_{\mu \downarrow 0} s_f(\cdot, \mu) = f.$$

Proof. Due to Lemma 5.1, it suffices to show that

$$\operatorname*{Lim\,inf}_{\mu \downarrow 0} \operatorname{epi} s_f(\cdot, \mu) \supseteq \operatorname{epi} f.$$

To this end, let $(\bar{x}, \bar{\alpha}) \in \operatorname{epi} f$, i.e., $g(H(\bar{x})) \leq \bar{\alpha}$. Now, let $\{\mu_k\} \downarrow 0$ be given. In view of the monotonicity assumption we get $s_g(H(\bar{x}), \mu_k) \leq \bar{\alpha}$ and hence

$$(\bar{x}, \bar{\alpha}) \in \operatorname{epi} s_f(\cdot, \mu_k) \quad \forall k \in \mathbb{N}.$$

With the choice $x^k := \bar{x}$ and $\alpha_k := \bar{\alpha}$ it follows immediately that

$$(\bar{x}, \bar{\alpha}) \in \operatorname*{Lim\,inf}_{\mu \downarrow 0} \operatorname{epi} s_f(\cdot, \mu),$$

which concludes the proof. □

Corollary 5.3. *If, in the setting of Section 4, ω is 1-coercive with $\omega(0) \leq 0$, then for $s_g(\cdot, \mu) := g \# \omega_\mu$ we have*

$$\mathrm{e}\!-\!\lim_{\mu \downarrow 0} s_g(H(\cdot), \mu) = g \circ H.$$

Proof. The assertion follows immediately from Lemma 4.4 and Proposition 5.2. □

In the following result we employ the limiting normal cone for a (nonempty) convex set $C \subset \mathbb{R}^n$ at $\bar{x} \in C$, which is given by, cf. [53, Theorem 6.9],

$$N(\bar{x} \mid C) = \{v \in \mathbb{R}^n \mid v^T(x - \bar{x}) \leq 0 \quad \forall x \in C\}.$$

In our setting, C is the domain of an lsc, convex function $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, which is closed and convex.

Lemma 5.4. *Let $\{g_k\}$ be a sequence of lsc, convex functions $g_k : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ converging epi-graphically to $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$. Furthermore, let $\{z^k\}$ be an unbounded sequence such that $z^k \in \partial g_k(y^k)$ for all $k \in \mathbb{N}$ for some $\{y^k\} \to \bar{y} \in \operatorname{dom} g$. Then every accumulation point of $\left\{\frac{z^k}{\|z^k\|}\right\}$ lies in $N(\bar{y} \mid \operatorname{dom} g)$.*

Proof. Let $\bar{z}$ be an accumulation point of $\left\{\frac{z^k}{\|z^k\|}\right\}$. W.l.g. we can assume that $\frac{z^k}{\|z^k\|} \to \bar{z}$. Moreover, let $y \in \operatorname{dom} g$ be given. Since $\mathrm{e}\!-\!\lim_{k \to \infty} g_k = g$, we may invoke (2) to obtain a sequence $\{\hat{y}^k\} \to y$ such that $\limsup_{k \to \infty} g_k(\hat{y}^k) \leq g(y)$. Since, by assumption, $z^k \in \partial g(y^k)$ for all $k \in \mathbb{N}$, we infer

$$g_k(\hat{y}^k) - g_k(y^k) \geq (z^k)^T(\hat{y}^k - y^k) \quad \forall k \in \mathbb{N}.$$

Dividing by $\|z^k\|$ yields

$$\frac{g_k(\hat{y}^k) - g_k(y^k)}{\|z^k\|} \geq \frac{(z^k)^T}{\|z^k\|}(\hat{y}^k - y^k) \to \bar{z}^T(y - \bar{y}).$$

To prove the assertion it suffices to see that the numerator of the left-hand side of the above inequality is bounded from above at least on a subsequence. This, however, is true due to the choice of $\{\hat{y}^k\}$ and (2). □

A standard assumption in the context of convex composite functions, cf. [16], is the *basic contstraint qualification* which is formally stated in the following definition.

Definition 5.5 (Basic constraint qualification)**.** *Let f be given as in* (13)*. Then f is said to satisfy the* basic constraint qualification (BCQ) *at a point $\bar{x} \in \operatorname{dom} f$ if*

$$N(H(\bar{x}) \mid \operatorname{dom} g) \cap \operatorname{nul} H'(\bar{x})^T = \{0\}.$$

Note that, in the setting of (13), BCQ always holds at a point $\bar{x} \in \operatorname{dom} f$ where $H'(\bar{x})^T$ has full column rank. Moreover, BCQ is always fulfilled when g is finite-valued, since then $\operatorname{dom} g = \mathbb{R}^m$ and thus, $N(H(\bar{x}) \mid \operatorname{dom} g) = \{0\}$ for all $\bar{x} \in \mathbb{R}^n$. The BCQ is important since it guarantees a rich subdifferential calculus for the composition $f = g \circ H$.

Lemma 5.6. [53, Theorem 10.6] *Let f be given as in* (13)*. If BCQ is satisfied at $\bar{x} \in \operatorname{dom} f$, then f is (subdifferentially) regular at $\bar{x}$ and we have*

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

Theorem 5.7. *Let s_g be an epi-smoothing function for g. If $s_f(\cdot,\cdot) := s_g(H(\cdot),\cdot)$ is an epi-smoothing function for $f := g \circ H$, then*

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

for all $\bar{x} \in \operatorname{dom} f$ at which the BCQ holds.

Proof. We need only show that $\operatorname{Lim\,sup}_{\mu \downarrow 0, x \to \bar{x}} \nabla_x s_f(H(x),\mu) \subset \partial f(\bar{x})$, since the Lim inf-inclusion is clear from Lemma 3.4.

To this end, let $v \in \operatorname{Lim\,sup}_{\mu \downarrow 0, x \to \bar{x}} \nabla_x s_f(H(x),\mu)$ be given. Then there exist sequences $\{x^k\} \to \bar{x}$ and $\{\mu_k\} \downarrow 0$ such that

$$H'(x^k)^T \nabla_x s_g(H(x^k),\mu_k) = \nabla_x s_f(x^k,\mu_k) \to v. \tag{14}$$

Put $z^k := s_g(H(x^k),\mu_k)\,(k \in \mathbb{N})$. If $\{z^k\}$ were unbounded, then w.l.g. $\{\frac{z^k}{\|z^k\|}\} \to \bar{z} \neq 0$, and we infer from (14) that

$$\bar{z} \in \operatorname{nul} H'(\bar{x})^T.$$

On the other hand, Lemma 5.4 tells us that $\bar{z} \in N(H(\bar{x}) \mid \operatorname{dom} g)$, thus,

$$0 \neq \bar{z} \in N(H(\bar{x}) \mid \operatorname{dom} g) \cap \operatorname{nul} H'(\bar{x})^T,$$

which contradicts BCQ. Hence, $\{z^k\}$ is bounded and converges at least on a subsequence, and due to Attouch's theorem [53, Theorem 12.35] the limit (accumulation point) lies in $\partial g(H(\bar{x}))$. Using this and the fact that H' is continuous, we get

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

where the equality is due to Lemma 5.6. This concludes the proof. □

Corollary 5.8. *Let s_g be an epi-smoothing function for g, and suppose ω is 1-coercive with $\omega(0) \leq 0$. Then $s_f(\cdot,\cdot) := s_g(H(\cdot),\cdot)$ is an epi-smoothing function for $f := g \circ H$ and*

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

for all $\bar{x} \in \operatorname{dom} f$ *at which the BCQ holds.*

Proof. The result follows immediately from Corollary 5.3 and Theorem 5.7. □

We point out that, unlike in the convex case in Theorem 4.6, where we obtain the gradient consistency condition directly via Attouch's theorem, we cannot derive it in this case from a generalized version of Attouch's theorem for convex composite functions as it is presented in [50, Theorem 2.1], since we do not meet the assumptions there.

6 Constrained Optimization

We now apply the results of the previous section the constrained optimization problem

$$\begin{array}{ll} \text{minimize} & \phi(x) \\ \text{subject to} & h(x) \in C, \end{array} \tag{15}$$

where $\phi : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ are smooth mappings and $C \subset \mathbb{R}^m$ is a nonempty closed convex set. This is an example of a convex composite optimization problem [11, 12, 16] where the composite function $f = g \circ H$ is given by

$$g(\gamma, y) := \gamma + \delta(y \mid C) \quad \text{and} \quad H(x) := \begin{bmatrix} \phi(x) \\ h(x) \end{bmatrix}.$$

In this case, g is the sum of a smooth convex function, $g_1(\gamma, y) := \gamma$, and a nonsmooth convex function $g_2(\gamma, y) := \delta(y \mid C)$. Hence, by Proposition 3.1, we can obtain an epi-smoothing function for g by only smoothing the g_2 term. A straightforward computation shows that

$$e_\mu g_2(y) = \frac{1}{2\mu} \operatorname{dist}^2(y \mid C).$$

Therefore, by Corollary 5.3,

$$s_f(x, \mu) = \phi(x) + \frac{1}{2\mu} \operatorname{dist}^2(h(x) \mid C) \tag{16}$$

is an epi-smoothing function for f. This is one of the classical smoothing functions for constrained optimization [33]. The BCQ becomes the condition

$$\operatorname{nul} h'(x)^T \cap N(h(x) \mid C) = \{0\}. \tag{17}$$

In the case where $C = \{0\}^s \times \mathbb{R}^{m-s}_-$, the function (16) is the classical least-squares smoothing function for nonlinear programming, and (17) reduces to the *Mangasarian-Fromovitz constraint qualification* (e.g., see [53, Example 6.40]).

Corollary 5.8 tells us that at every point $\bar{x}$ with $h(\bar{x}) \in C$ we have

$$\operatorname*{Lim\,sup}_{\mu \downarrow 0, x \to \bar{x}} \nabla_x s_f(x, \mu) = \nabla \phi(\bar{x}) + h'(\bar{x})^T N(h(\bar{x}) \mid C),$$

whenever condition (17) holds at $\bar{x}$, where, by Proposition 4.7,

$$\nabla_x s_f(x, \mu) = \nabla \phi(x) + h'(x)^T \left(\frac{h(x) - \Pi_C(h(x))}{\mu} \right).$$

The results of Section 5 allow us to make powerful statements about algorithms that use the epi-smoothing function (16) to solve the optimization problem (15). We begin by studying the case of cluster points that are feasible for (15).

Theorem 6.1. *Let s_f be as in* (16) *with ϕ, h, and C satisfying the hypotheses specified in* (15)*. Let $\{x^k\} \subset \mathbb{R}^n$ and $\{\mu_k\} \downarrow 0$ satisfy $\left\|\nabla_x s_f(x^k, \mu_k)\right\| \downarrow 0$. Then every feasible cluster point $\bar{x}$ of $\{x^k\}$ at which* (17) *is satisfied, is a* Karush-Kuhn-Tucker point *for* (15)*, i.e.,*

$$0 \in \partial f(\bar{x}) = \nabla\phi(\bar{x}) + h'(\bar{x})^T N\left(h(\bar{x}) \mid C\right).$$

Proof. Lemma 5.6 implies that $\partial f(\bar{x}) = \nabla\phi(\bar{x}) + h'(\bar{x})^T N\left(h(\bar{x}) \mid C\right)$. Hence, by Corollary 5.8, $\bar{x}$ is a KKT point for (15). □

Theorem 6.1 tells us that the feasible cluster points of sequences of approximate stationary points of s_f are KKT points, but, from and algorithmic perspective, this does not give us a mechanism for testing proximity to optimality via standard optimality conditions. That is, it does not show how to approximate the multiplier vector. This is addressed by the following corollary.

Corollary 6.2. *Let s_f, ϕ, h, C, $\{x^k\}$, and $\{\mu_k\}$ be as in Theorem 6.1, and let $\bar{x}$ be a cluster point of $\{x^k\}$ at which $h(\bar{x}) \in C$ and* (17) *is satisfied. If $J \subset \mathbb{N}$ is a subsequence for which $x^k \to_J \bar{x}$, then the associated subsequence $\{y^k\}_J$, where*

$$y^k := \frac{h(x^k) - \Pi_C(h(x^k))}{\mu_k} \quad \forall k \in \mathbb{N},$$

remains bounded and every cluster point $\bar{y}$ is such that $(\bar{x}, \bar{y})$ is a Karush-Kuhn-Tucker pair *for* (15)*, i.e.,*

$$0 = \nabla\phi(\bar{x}) + h'(\bar{x})^T\bar{y} \quad \textit{with} \quad \bar{y} \in N(h(\bar{x}) \mid C).$$

Proof. Let $J \subset \mathbb{N}$ and $\bar{x}$ be as in the statement of the corollary. Theorem 6.1 tells us that $\bar{x}$ is a KKT point for (15), i.e., $0 \in \partial f(\bar{x}) = \nabla\phi(\bar{x}) + h'(\bar{x})^T N\left(h(\bar{x}) \mid C\right)$. We first show that the subsequence $\{y^k\}_J$ given above is necessarily bounded.

Suppose, to the contrary, that the sequence is not bounded. Then there is a further subsequence $\hat{J} \subset J$ such that $\left\|y^k\right\| \uparrow_{\hat{J}} +\infty$. With no loss in generality we may assume that there is a unit vector $\tilde{y}$ such that $y^k / \left\|y^k\right\| \to_{\hat{J}} \tilde{y}$. Since $y^k \in N\left(\Pi_C(h(x^k)) \mid C\right)$ for all k, the outer semicontinuity of the normal cone operator $z \mapsto N\left(z \mid C\right)$ relative to C, cf. [53, Proposition 6.6], implies that $\tilde{y} \in N\left(h(\bar{x}) \mid C\right)$. Dividing $\left\|\nabla_x s_f(x^k, \mu_k)\right\|$ by $\left\|y^k\right\|$ and taking the limit over $\hat{J}$ gives $h'(\bar{x})^T\tilde{y} = 0$. But this contradicts the BCQ (17) since $\tilde{y}$ is a unit vector. Therefore, the sequence $\{y^k\}_J$ is bounded.

Let $\bar{y}$ be any cluster point of the sequence $\{y^k\}_J$ (at least one such cluster point must exist since this sequence is bounded). As above, $\bar{y} \in N\left(h(\bar{x}) \mid C\right)$, and by the hypotheses, $0 = \nabla\phi(\bar{x}) + h'(\bar{x})^T\bar{y}$. Hence, $\bar{x}$ is a KKT point for (15) and $\bar{y}$ is an associated KKT multiplier. □

We now address the case of infeasible cluster points, i.e., cluster points $\bar{x}$ for which $h(\bar{x}) \notin C$. To understand this case, we must first review the subdifferential properties of the distance function $\mathrm{dist}(\cdot \mid C)$ and the associated convex composite function

$$\psi(x) := \mathrm{dist}(h(x) \mid C).$$

First, recall from [14, Proposition 3.1] that

$$\partial\mathrm{dist}(y \mid C) = \begin{cases} N\left(y \mid C\right) \cap \mathbb{B} & \text{if} \quad y \in C, \\ N\left(y \mid C + \mathrm{dist}(y \mid C)\mathbb{B}\right) \cap \mathrm{bdry}(\mathbb{B}) & \text{if} \quad y \notin C, \end{cases} \tag{18}$$

where bdry($\mathbb{B}$) is the boundary of the unit ball, and, by [53, Example 8.53], we also have

$$\partial \text{dist}(y \mid C) = N\left(y \mid C + \text{dist}(y \mid C)\mathbb{B}\right) \cap \text{bdry}(\mathbb{B}) = \left\{ \frac{y - \Pi_C(y)}{\text{dist}(y \mid C)} \right\} \quad \forall\, y \notin C. \tag{19}$$

In addition, from [12, Equation 2.4], ψ is subdifferentially regular on $\mathbb{R}^n$ with

$$\partial \psi(x) = h'(x)^T \partial \text{dist}(h(x) \mid C). \tag{20}$$

These formulas yield the following result.

Theorem 6.3. *Let s_f, ϕ, h, C, $\{x^k\}$, and $\{\mu_k\}$ be as in Theorem 6.1, and let $\bar{x}$ be a cluster point of $\{x^k\}$ at which $h(\bar{x}) \notin C$. Then $0 \in \partial\psi(\bar{x})$.*

Proof. Let $J \subset \mathbb{N}$ be such that $x^k \to_J \bar{x}$. Since $\left\|\nabla_x s_f(x^k, \mu_k)\right\| \downarrow 0$, we have $\mu_k \left\|\nabla_x s_f(x^k, \mu_k)\right\| \downarrow 0$, and consequently

$$h'(x^k)^T (h(x^k) - \Pi_C(h(x^k))) \to 0.$$

Hence, by the continuity of Π_C and (19), $0 \in \partial\psi(\bar{x})$. ☐

Theorem 6.3 shows that any algorithm that drives $\nabla_x s_f(x^k, \mu_k)$ to zero as $\mu_k \downarrow 0$ performs admirably even when the problem (15) is itself infeasible. That is, in the absence of feasibility, it naturally tries to locate a *nonfeasible stationary point* for (15) as defined in [13]. It may happen that the original problem is feasible while all cluster points are nonfeasible stationary points. This can be rectified by placing a further restriction on how the iterates $\{x^k\}$ are generated.

Proposition 6.4. *Let C, ϕ, h, and s_f be as in* (15) *and* (16)*, and let $\mu_k \downarrow 0$. Suppose that there is a known feasible point $\tilde{x}$ for* (15)*. If $\{x^k\}$ is a sequence for which $s_f(x^k, \mu_k) \le s_f(\tilde{x}, \mu_k) = \phi(\tilde{x})$ for all $k = 1, 2, \ldots$, then every cluster point of $\{x^k\}$ must be feasible for* (15)*.*

Proof. Let $\bar{x}$ be a cluster point of $\{x^k\}$ and let $J \subset \mathbb{N}$ be such that $x^k \to_J \bar{x}$. If $\bar{x}$ is not feasible, then $\frac{1}{2\mu_k}\text{dist}^2(h(x^k) \mid C) \to_J +\infty$. But $s_f(x^k, \mu_k) = \phi(x^k) + \frac{1}{2\mu_k}\text{dist}^2(h(x^k) \mid C) \le \phi(\tilde{x})$ giving the contradiction $\phi(x^k) \to_J -\infty$. ☐

In fact, without further hypotheses, feasibility might not be attained in the limit even in the prototypical example of convex composite optimization, the Gauss-Newton method for solving nonlinear systems of equations. It is often the case that the additional hypotheses employed are related to the BCQ (17). One way to understand the role of nonfeasible stationary points and their effect on computation is through constraint qualifications that apply to nonfeasible points. These constraint qualifications extend (17) to points on the whole space. Among the many possible extensions one might consider, we use one from the geometry of the subdifferential in (18). We say that the *extended constraint qualification* (ECQ) for (15) is satisfied if

$$\text{nul}\, h'(x)^T \cap N\left(h(x) \mid C + \text{dist}(h(x) \mid C)\mathbb{B}\right) = \{0\}. \tag{21}$$

Note that this condition is well defined on all of $\mathbb{R}^n$ and reduces to (17) when $h(x) \in C$. When $h(x) \notin C$, it is easily seen that $0 \in \partial\psi(x)$ if and only if (21) is not satisfied. Hence, if one assumes that ECQ is satisfied at all iterates, then nonfeasible cluster points cannot exist. For example, if $C = \{0\}$, then a standard global constraint qualification is to assume that $h'(x)$ is everywhere surjective, i.e.,

$\operatorname{nul} h'(x)^T = \{0\}$ for all x. This implies (21) which simply says that $h'(x)^T h(x) \neq 0$ whenever $h(x) \neq 0$ and $h'(x)$ is surjective whenever $h(x) = 0$.

7 Final Remarks

In this paper we have synthesized the infimal convolution smoothing ideas proposed by Beck and Teboulle in [7] with the notion of gradient consistency defined by Chen in [21]. To achieve this we make use of epi-convergence techniques that are well suited to the study of the variational properties of parametrized families of functions. Using epi-convergence, we defined the notion of epi-smoothing for which we established a rudimentary calculus. Epi-smoothing is a weakening of the kinds of smoothing studied in [7] where the focus is on convex optimization and the derivation of complexity results which necessitate stronger forms of smoothing. We then applied the epi-smoothing ideas to study the epi-smoothing properties of convex composite functions, a very broad and important class of nonconvex functions. In particular, we showed that general constrained optimization falls within this class. Using the epi-smoothing calculus, we easily derived the convergence properties of a classical smoothing approach to constrained optimization establishing the convergence properties even in the case when the underlying optimization problem is not feasible. This application demonstrates the power of these ideas as well as their ease of use.

8 Acknowledgements

The authors thank Prof. Christian Kanzow for helpful comments on earlier drafts of the paper.

References

[1] A. Aravkin, B. Bell, J.V. Burke, and G.Pillonetto: *An ℓ_1-Laplace robust Kalman smoother.* IEEE Transactions on Automatic Control, 2011.

[2] A. Aravkin, B.M. Bell, J.V. Burke, and G. Pillonetto: *Learning using state space kernel machines.* In "Proceedings of the IFAC World Congress 2011", Milan, Italy, 2011.

[3] A. Aravkin, J.V. Burke, and G. Pillonetto: *A statistical and computational theory for robust/sparse estimation and Kalman smoothing using log-concave piecewise linear-quadratic densities.* Submitted to the Annals of Statistics, 2011.

[4] H. Attouch: *Variational Convergence of Functions and Operators.* Applicable Mathematics Series, Pitman Advanced Publishing Program, Boston, London, Melbourne, 1984.

[5] H. Attouch and R.J.-B. Wets: *Epigraphical analysis.* Ann. Inst. H. Poincaré Anal. Non Linéare 6, 1989, pp. 73–100.

[6] H.H. Bauschke and P.L. Combettes: *Convex analysis and Monotone Operator Theory in Hilbert Spaces.* CMS Books in Mathematics, Springer-Verlag, New York, 2011.

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

[8] A. Ben-Tal and M. Teboulle: *A smoothing technique for nondifferentiable optimization problems.* In Optimization, Lecture Notes in Math. 1405, S. Dolecki, ed., Springer-Verlag, New York, 1989, pp. 1–11.

[9] D.P. Bertsekas: *Nondifferentiable optimization via approximation.* Mathematical Programming Studies, 1975, pp. 1-25.

[10] W. Bian and X. Chen: *Smoothing neural network for constrained non-Lipschitz optimizationwith applications.* IEEE Transactions on Neural Networks and Learning Systems 23, 2012, pp.399–411.

[11] J.V. Burke: *Descent methods for composite nondifferentiable optimization problems.* Mathematical Programming 33, 1985, pp. 260–279.

[12] J.V. Burke: *Second order necessary and sufficient conditions for convex composite NDO.* Mathematical Programming 38, 1987, pp. 287–302.
[13] J.V. Burke: *A sequential quadratic programming method for potentially infeasible mathematical programs.* Journal of Mathematical Analysis and Applications 139, 1987, pp. 319–351.
[14] J.V. Burke: *An exact penalization viewpoint of constrained optimization.* SIAM J. Control and Optim. 29, pp. 968–998.
[15] J.V. Burke, T. Hoheisel, and C. Kanzow *Gradient consistency for integral-convolution smoothing functions.* Preprint, 2012.
[16] J.V. Burke and R.A. Poliquin: *Optimality conditions for non-finite valued convex composite functions.* Mathematical Programming 57, 1992, pp. 103–120.
[17] J.M. Borwein and A.S. Lewis: *Convex Analysis and Nonlinear Optimization. Theory and Examples.* CMS Books in Mathematics, Springer-Verlag, New York, 2000.
[18] B. Chen and X. Chen: *A global and local superlinear continuation-smoothing method for P0 and R0 NCP or monotone NCP.* SIAM Journal on Optimization 9, 1999, pp. 624–645.
[19] C. Chen and O.L. Mangasarian: *A class of smoothing functions for nonlinear and mixed complementarity problems.* Mathematical Programming 71, 1995, pp. 51–70.
[20] X. Chen: *Smoothing methods for complementarity problems and their applications: a survey.* Journal of the Operations Research Society of Japan 43, 2000, pp. 32–46.
[21] X. Chen: *Smoothing methods for nonsmooth, nonconvex minimization.* Mathematical Programming 134, 2012, pp. 71–99.
[22] X. Chen and M. Fukushima: *A smoothing method for a mathematical program with P-matrix linear complementarity constraints.* Computational Optimization and Applications 27, 2004, pp. 223–246.
[23] X. Chen, Z. Nashed, L. Qi: *Smoothing methods and semismooth methods for nondifferentiable operator equations.* SIAM Journal on Numerical Analysis 38, 2000, pp.1200–1216.
[24] F.H. Clarke: *Optimization and Nonsmooth Analysis.* John Wiley & Sons, New York, 1983.
[25] P.L. Combettes and V.R. Wajs: *Signal recovery by proximal forward-backward splitting.* Multiscale Modeling and Simulation 4(4), 2005, pp. 1168–1200.
[26] D. Donoho: *De-noising by soft thresholding.* IEEE Transactions on Information Theory 41, 1995, pp. 613–627.
[27] D. Donoho: *Compressed sensing.* IEEE Transactions on Information Theory 52(4), 2006, pp. 1289–1306.
[28] B. Efron, T. Hastie, L. Johnstone, and R. Tibshirani: *Least angle regression.* Annals of Statistics 32, 2004, pp. 407–499.
[29] Y.M. Ermoliev, V.I. Norkin, and R.J-B. Wets: *The minimization of semi-continuous functions: mollifier subgradients.* SIAM Journal on Control and Optimization 33, 1995, pp. 149-167.
[30] T. Evgeniou, M. Pontil, and T. Poggio: *Regularization networks and support vector machines.* Advances in Computational Mathematics 13, 2000, pp. 1–150.
[31] F. Facchinei, H. Jiang, and L. Qi: *A smoothing method for mathematical programs with equilibrium constraints.* Mathematical Programming 85, 1999, pp. 107–134.
[32] S. Farahmand, G.B. Giannakis, and D. Angelosante: *Doubly robust smoothing of dynamical processes via outlier sparsity constraints.* IEEE Transactions on Signal Processing 59, 2011, pp. 4529–4543.
[33] A.V. Fiacco and G.P. McCormick: *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* SIAM Classics in Applied Mathematics, 1987.
[34] M. Fukushima, Z.-Q. Luo and P. Tseng: *Smoothing functions for second-order-cone complementarity problems.* SIAM Journal on Optimization 12, 2002, pp. 436–460.
[35] M. Fukushima and J.-S. Pang: *Convergence of a smoothing continuation method for mathematical programs with complementarity constraints.* In: Thera, M., Tichatschke, R. (eds.) Lecture Notes in Economics and Mathematical Systems, vol. 477, 1999, pp. 99–110. Springer, Berlin.
[36] S.A. Gabriel and J.J. More: *Smoothing of mixed complementarity problems.* In: Ferris, M.C., Pang, J.S. (eds.) Complementarity and Variational Problems: State of the Art, 1997, pp. 105–116. SIAM, Philadelphia.
[37] H. Halkin: *Mathematical Programming without differentiability.* In Calculus of Variations and Control Theory, D. Russell ed., pp. 279–287, Academic Press, New York, 1976.

[38] T.J. Hastie and R.J. Tibshirani. *Generalized additive models.* In "Monographs on Statistics and Applied Probability", Volume 43, Chapman and Hall, London, UK, 1990.
[39] T.J. Hastie, R.J. Tibshirani, and J. Friedman: *The Elements of Statistical Learning. Data Mining, Inference and Prediction.* Springer, Canada, 2001.
[40] P.J. Huber: *Robust Statistics.* John Wiley & Sons, New York, 1981.
[41] J.-B. Hiriart-Urrruty and C. Lemaréchal: *Convex Analysis and Minimization Algorithms I.* Second Corrected Printing, Springer, Berlin, Heidelberg, 1996.
[42] J.-B. Hiriart-Urrruty and C. Lemaréchal: *Fundamentals of Convex Analysis.* Grundlehren Text Editions, Springer, Berlin, Heidelberg, 2001.
[43] A.D. Ioffe: *Approximate subdifferentials and applications I: The finite dimensional case.* Transactions of the American Mathematical Society 281, 1984, pp. 389–416.
[44] Z. Liu, R.D.C. Monteiro, and A. Nemirovski *Large-scale semidefinite programming via saddle point mirror-prox algorithm.* Mathematical Programming 109, 2007, pp. 211-237.
[45] 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.
[46] B.S. Mordukhovich: *Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems.* Soviet Mathematical Doklady, 22, pp. 526–530.
[47] B. S. Mordukhovich: *Variational Analysis and Generalized Differentiation I. Basic Theory.* A Series of Comprehensive Studies in Mathematics, Vol. 330, Springer, Berlin, Heidelberg, 2006.
[48] Y. Nesterov: *Smooth minimization of non-smooth functions.* Mathematical Programming 103, 2005, pp. 127-152.
[49] Y. Nesterov: *Smoothing technique and its applications in semidefinite optimization.* Mathematical Programming 110, 2007, pp. 245–259.
[50] R.A. Poliquin: *An extension of Attouch's theorem and its application to second-order epi-differentiation of convexly composite functions.* Transactions of the American Mathematical Society 332, 1992, pp. 861–874.
[51] M. Pontil and A. Verri: *Properties of support vector machines.* Neural Computation 10, 1998, pp. 955–974.
[52] R.T. Rockafellar: *Convex Analysis.* Princeton University Press, Princeton, New Jersey, 1970.
[53] R.T. Rockafellar and R.J.-B. Wets: *Variational Analysis.* A Series of Comprehensive Studies in Mathematics, Vol. 317, Springer, Berlin, Heidelberg, 1998.
[54] B. Schölkopf, A.J. Smola, R.C. Williamson, and P.L. Bartlett: *New support vector algorithms.* Neural Computation 12, 2000, pp. 1207–1245.
[55] R. Tibshirani: *Regression shrinkage and selection via the LASSO.* Journal of the Royal Statistical Society, Series B 58(1), 1996, pp. 267–288.
[56] V. Vapnik: *Statistical Learning Theory.* John Wiley & Sons, New York, 1998.
[57] R.J-B. Wets: *Convergence of convex functions, variational inequalities and convex optimization problems.* In "Variational Inequalities and Complementarity Problems", edited by R. Cottle, F. Giannessi and J.-L. Lions, Wiley, New York, 1980.
[58] R.A. Wijsman: *Convergence of sequences of convex sets, cones and functions.* Bulletin of the American Mathematical Society 70, 1964, pp. 186–188.
[59] R.A. Wijsman: *Convergence of sequences of convex sets, cones and functions II.* Transactions of the American Mathematical Society 123, 1966, pp. 32–45.

University of Washington, Department of Mathematics, Box 354350 Seattle, Washington 98195-4350
E-mail address: burke@math.washington.edu

University of Würzburg, Institute of Mathematics, Campus Hubland Nord, Emil-Fischer-Strasse 30, 97074 Würzburg, Germany
E-mail address: hoheisel@mathematik.uni-wuerzburg.de