Set-Valued Var. Anal (2017) 25:1–23
DOI 10.1007/s11228-016-0362-y

Epi-convergence Properties of Smoothing by Infimal Convolution

James V. Burke[1] · Tim Hoheisel[2]

Received: 23 June 2015 / Accepted: 19 January 2016 / Published online: 24 January 2016

Abstract This paper concerns smoothing by infimal convolution for two large classes of functions: convex, proper and lower semicontinous as well as for (the nonconvex class of) convex-composite functions. The smooth approximations are constructed so that they epi-converge (to the underlying nonsmooth function) and fulfill a desirable property with respect to graph convergence of the gradient mappings to the subdifferential of the original function under reasonable assumptions. The close connection between epi-convergence of the smoothing functions and coercivity properties of the smoothing kernel is established.

Keywords Infimal convolution · Coercivity · Epi-convergence · Smoothing · Convex-composite · Attouch's theorem

Mathematics Subject Classification (2010) 49J52 · 49J53 · 90C25 · 90C26 · 90C46

1 Introduction

The idea of smoothing extended real-valued (hence nonsmooth) convex functions by means of infimal convolution with sufficiently smooth kernels dates back to the early works of

This research was partially supported by the DFG (Deutsche Forschungsgemeinschaft) under grant HO 4739/2-1 and by the *Prodi-Chair* program of the University of Würzburg

✉ Tim Hoheisel
hoheisel@mathematik.uni-wuerzburg.de

James V. Burke
burke@math.washington.edu

1 Department of Mathematics, University of Washington, Box 354350, Seattle, WA 98195-4350, USA

2 Institute of Mathematics, Campus Hubland Nord, University of Würzburg, Emil-Fischer-Strasse 30, 97074 Würzburg, Germany

Moreau [20]. The *Moreau envelope* (or *Moreau-Yosida regularization*) uses the squared Euclidean norm as a smoothing kernel, and is a well-established tool in convex analysis and optimization. In this note we study the epi-convergence properties of smooth approximations obtained through infimal convolution.

When the concept of *epi-convergence* arose in the 1960s, it was quickly proven that, in particular, the Moreau envelope converges epi-graphically to the convex function that it smoothes as the regularization/smoothing parameter goes to zero, see [1, 23]. Epi-convergence was designed so that epi-limits inherit certain desirable variational properties. In general, it is different from pointwise convergence, and strictly weaker than uniform convergence (on compact subsets). In a sense, it is the weakest convergence concept for sequences of functions which assures the (set-)convergence of minimizers and optimal values. This property makes it very attractive from an optimization perspective. Excellent accounts on this topic can be found in, e.g., [1] in infinite and [23, Ch. 7] in finite dimensions.

In [6], smoothing kernels (with Lipschitz gradient) different from the square of the Euclidean norm are employed to construct infimal convolution smoothing functions. Building on the work in [6], a notion of *epi-smoothing*, compatible with epi-convergence, is introduced in [8] where conditions for establishing the *gradient consistency* [12] of an epi-smoothing function are given. One of the major contributions of [8] is the construction of epi-convergent, gradient consistent smoothing functions for the important and broad class of *non*convex functions called *convex-composite* functions. Their construction is based on epi-convergent smooth approximations for convex functions, established as a preliminary result. In the context of epi-smoothing functions constructed through infimal convolution, we have recently discovered that the convex results in [8] were partially established earlier (under slightly different assumptions) by Strömberg in his doctoral dissertation [24]. In particular, Strömberg shows that, for the purposes driven by epi-convergence, Lipschitz continuity of the gradient of the smoothing kernels is not required . This hypothesis is used in [6] as part of the definition of *smoothable functions* (see [6, Definition 2.1]) in order to establish certain complexity properties which is the focus of that study.

In this paper we unify, refine, and extend results on the epi-convergence properties of smoothing by infimal convolution, and clarify the origins of these results. In particular, the authors highlight the important early result by Strömberg [24]. As noted, we restrict ourselves to infimal convolution smoothing schemes for extended real-valued convex and convex-composite functions. This includes as a special case the Moreau-Yosida regularization. This is distinct from [8] which considers a more general epi-smoothing framework that includes infimal convolution as an important special case. In addition, we omit consideration of other smoothing techniques such as integral convolution (cf. see [9, 12, 13, 17, 18]), Lasry-Lions regularization (cf. [4, 19, 25]), or self-dual smoothing [15].

More concretely, the objective of this paper is the study of parametrized families of smoothing functions $\{f_\lambda : \mathbb{R}^n \to \mathbb{R} \mid \lambda > 0\}$ built through infimal convolution that epi-converge to an underlying nonsmooth mapping $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. A primary goal is that this family of smooth approximations possesses the *gradient consistency* property defined by Chen in [12] at a given point $\bar{x} \in \operatorname{dom} f$, i.e.,

$$\limsup_{x\to\bar{x},\ \lambda\downarrow 0} \nabla f_\lambda(x) = \partial f(\bar{x}) \ . \tag{1}$$

In the fully convex case, gradient consistency is an immediate consequence of Attouch's Theorem [1] when the family $\{f_\lambda\}$ epi-converges to f since, in this case, we have graph convergence of the gradient mapping, $\operatorname{gph} \nabla f_\lambda \to \operatorname{gph} \partial f$ which is much stronger.

In the convex-composite case, the graph convergence of the sequence of gradients has to be established independently. We follow the approach taken in [22] and [14], where an extension of Attouch's Theorem is provided for certain classes of convex-composites.

We would also like to mention the early contribution [7], which is also in the spirit of the contents of this paper, but with a different approach.

The paper is organized as follows: In Section 2, we provide the necessary tools and terminology from variational analysis. Section 3 contains the main results on epi-convergent, infimal convolution based regularizations for convex functions as well as a discussion of and comparison with existing results. In particular, the crucial role of *supercoercivity* for obtaining epi-convergent approximations is highlighted. Here, the major tool for proving epi-convergence of the infimal convolutions is provided by the duality correspondence for epi-convergent sequences of convex functions established by Wijsman, see [26, 27]. Due to the special role of supercoercivity, Section 4 provides an elementary calculus for supercoercivity as well as examples of supercoercive (smooth and convex) kernels. In Section 5, the results from Section 3 are applied to the class of convex-composite functions, also providing a graph convergence result using the techniques of Poliquin [22] in the spirit of Attouch's Theorem.

We close with some final remarks in Section 6.

Notation: The notation used is quite standard. The extended real numbers are given by $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$.

The open Euclidean ball with radius $r > 0$ around $\bar{x} \in \mathbb{R}^n$ is denoted by $B_\varepsilon(\bar{x})$. Moreover, we put $\mathbb{B} := \mathrm{cl}\, B_1(0)$.

The standard inner product on $\mathbb{R}^n$ is denoted by $\langle \cdot, \cdot \rangle$, i.e.

$$\langle x, y \rangle = x^T y \quad \forall x, y \in \mathbb{R}^n.$$

For a differentiable function $H : \mathbb{R}^n \to \mathbb{R}^m$ its derivative is written as H', hence the Jacobian of H at $x \in \mathbb{R}^n$ is given by $H'(x) \in R^{m\times n}$. For the case $m = 1$, the gradient of H at $x \in \mathbb{R}^n$ is a column vector labeled by $\nabla H(x) \in \mathbb{R}^n$, i.e. $\nabla H(x) = H'(x)^T$.

For a matrix $A \in \mathbb{R}^{m\times n}$, its kernel will be denoted by $\ker A$.

Moreover, for a sequence $\{x^k \in \mathbb{R}^n\}$ and a infinite subset $K \subset \mathbb{N}$, we write

$$x_k \to_K \bar{x},$$

to indicate that the subsequence $\{x_k\}_{k\in K}$ converges to $\bar{x} \in \mathbb{R}^n$.

2 Preliminaries

For a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ its *epigraph* is given by

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

and its *domain* is the set

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

The notion of the epigraph allows for very handy definitions of a number of properties for extended real-valued functions (see,e.g., [5, 23]). For example, a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said be *lower semicontinuous (lsc)* (or *closed*) if $\mathrm{epi}\, f$ is a closed set and it is called *convex* if $\mathrm{epi}\, f$ is a convex set. A convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be *proper* if $\mathrm{dom}\, f \ne \emptyset$.

The main technique of our study is the smoothing of convex functions by *infimal convolution*. For two functions $f, g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, their *infimal convolution* (or *epi-sum*) is the function $f\#g : \mathbb{R}^n \to \overline{\mathbb{R}}$ defined by

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

The operation is referred to as epi-addition since

$$\{(x,\alpha) \mid (f\#g)(x) < \alpha\} = \{(x,\alpha) \mid f(x) < \alpha\} + \{(x,\alpha) \mid g(x) < \alpha\},$$

with $\operatorname{epi}(f\#g) = \operatorname{epi} f + \operatorname{epi} g$ whenever the infimum defining $f\#g$ is attained when finite [23, Exercise 1.28]. In our study, we focus on the infimal convolution over the class of closed proper convex functions:

$$\Gamma := \left\{f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\} \mid f \text{ proper, lsc and convex}\right\}.$$

An important extended real-valued function is the *indicator function* of a set $C \subset \mathbb{R}^n$ given by $\delta_C : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with

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

The indicator function δ_C is convex if and only if C is convex, and lsc if and only if C is closed.

For a proper function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, its *conjugate* $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is given by

$$f^*(y) := \sup_{x\in\operatorname{dom} f} \left\{x^T y - f(x)\right\}.$$

Note that, if f is proper, then $f^*, (f^*)^* \in \Gamma$, and $f = (f^*)^*$ if and only if $f \in \Gamma$, cf. [23, Theorem 11.1].

We now introduce epi-convergence concepts for sequences of extended real-valued functions. Let $\{f_k : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}\}$ be a sequence of functions, then the *epigraphical lower limit* of $\{f_k\}$ is given by

$$\text{e-}\liminf_{k\to\infty} f_k : x \in \mathbb{R} \mapsto \min\left\{\alpha \in \mathbb{R}\cup\{+\infty\} \;\middle|\; \exists x^k \to x : \liminf_{k\to\infty} f_k(x^k) = \alpha\right\},$$

whereas the *epi-graphical upper limit* is the function

$$\text{e-}\limsup_{k\to\infty} f_k : x \in \mathbb{R} \mapsto \min\left\{\alpha \in \mathbb{R}\cup\{+\infty\} \;\middle|\; \exists x^k \to x : \limsup_{k\to\infty} f_k(x^k) = \alpha\right\}.$$

If these funtions coincide we call $\{f_k\}$ *epi-convergent* to (the *epi-limit*)

$$\text{e-}\lim_{k\to\infty} f_k := \text{e-}\liminf_{k\to\infty} f_k (= \text{e-}\limsup_{k\to\infty} f_k).$$

Hence, $\{f_k\}$ *epi-converges* to a function $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ if and only if

$$\forall x \in \mathbb{R} \begin{cases} \forall \{x^k\} \to x : \liminf_{k\to\infty} f_k(x^k) \geq f(x), \\ \exists \{x^k\} \to x : \limsup_{k\to\infty} f_k(x^k) \leq f(x), \end{cases} \tag{2}$$

and we write

$$\text{e-}\lim_{k\to\infty} f_k = f \quad \text{or} \quad f_k \xrightarrow{e} f.$$

Two further notions of functional convergence will also be of use to us. The sequence $\{f_k\}$ is said to converge *continuously* to f if

$$\forall x \in \mathbb{R} \text{ and } \{x^k\} \to x : \quad \lim_{k\to\infty} f_k(x^k) = f(x),$$

and we write

$$\text{c-}\lim_{k\to\infty} f_k = f \quad \text{or} \quad f_k \xrightarrow{c} f.$$

Furthermore, $\{f_k\}$ is said to converge *pointwise* to f if

$$\forall x \in \mathbb{R}: \quad \lim_{k\to\infty} f_k(x) = f(x),$$

and we write

$$\text{p-}\lim_{k\to\infty} f_k = f \quad \text{or} \quad f_k \xrightarrow{p} f.$$

We extend these notions of functional convergence to families of functions $\{f_\lambda\}_{\{\lambda\downarrow 0\}}$ by requiring the respective convergence properties to hold for all sequences $\{\lambda_k\} \downarrow 0$. For instance,

$$f_\lambda \xrightarrow{e} f \quad \Longleftrightarrow \quad \forall\{\lambda_k\}\downarrow 0: \; f_{\lambda_k} \xrightarrow{e} f,$$

and so forth.

The notion of set-convergence that we are going to invoke is *Painlevé-Kuratowski set convergence*: For a sequence of sets $\{C^k\}$ with $C_k \subset \mathbb{R}^n$ for all $k \in \mathbb{N}$, we define the *outer limit* as

$$\operatorname*{Lim\,sup}_{k\to\infty} C^k := \left\{x \mid \exists K \subset \mathbb{N}(\text{infinite}), \{x^k\} \to_K x : x^k \in C^k \quad \forall k \in K\right\}$$

and the *inner limit* as

$$\operatorname*{Lim\,inf}_{k\to\infty} C^k := \left\{x \mid \exists k_0 \in \mathbb{N}, \{x^k\} \to x : x^k \in C^k \quad \forall k \geq k_0\right\}.$$

By definition, it is always the case that $\operatorname{Lim\,inf}_{k\to\infty} C^k \subset \operatorname{Lim\,sup}_{k\to\infty} C^k$. We say that $\{C^k\}$ converges if the outer and inner limit are equal, i.e.:

$$\operatorname*{Lim}_{k\to\infty} C^k := \operatorname*{Lim\,sup}_{k\to\infty} C^k = \operatorname*{Lim\,inf}_{k\to\infty} C^k.$$

Note that, by means of this convergence concept, we have

$$f_k \xrightarrow{e} f \quad \Longleftrightarrow \quad \operatorname{epi} f_k \to \operatorname{epi} f,$$

e.g., see [23].

Building on the notion of set-convergence in the sense of Painlevé-Kuratowski, we define the *outer limit* and *inner limit* for a set-valued mapping.

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

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

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

$$\begin{aligned}\operatorname*{Lim\,inf}_{x\to_X \bar{x}} S(x) &:= \bigcap_{\{x^k\}\to\bar{x}} \operatorname*{Lim\,inf}_{k\to\infty} S(x^k)\\ &= \left\{v \mid \forall\{x^k\} \to_X \bar{x}, \; \exists\{v^k\} \to v, k_0 \in \mathbb{N} : v^k \in S(x^k) \quad \forall k \geq k_0\right\}.\end{aligned}$$

In case that outer and inner limit coincide, we write

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

Definition 2.1 (Subdifferentials and Clarke regularity) 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 *proximal subdifferential* of f at $\bar{x}$ is the set given by

$$\partial_p f(x) := \left\{v \;\middle|\; \exists \rho, \delta > 0 : f(x) \geq f(\bar{x}) + v^T(x-\bar{x}) - \frac{\rho}{2}\|x-\bar{x}\|^2 \quad \forall x \in B_\delta(\bar{x})\right\}.$$

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

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

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

$$\partial^\infty f(\bar{x}) := \left\{v \mid \exists \{x^k\} \to_f \bar{x}, \{\mu_k\} \downarrow 0, v^k \in \hat{\partial} f(x^k) : \mu_k v^k \to v\right\} \cup \{0\}.$$

Clearly, by definition, for an arbitrary function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\partial_p f(\bar{x}) \subset \hat{\partial} f(\bar{x}) \subset \partial f(\bar{x}) \quad \forall \bar{x} \in \operatorname{dom} f.$$

Note that for a proper, convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$,

$$\partial f(\bar{x}) := \left\{v \in \mathbb{R}^n \;\middle|\; f(x) \geq f(\bar{x}) + v^T(x-\bar{x}) \quad \forall x \in \mathbb{R}^n\right\} = \hat{\partial} f(\bar{x}) = \partial_p f(x),$$

for all $\bar{x} \in \operatorname{dom} f$, see [23, Proposition 8.12]. For a convex set $C \subset \mathbb{R}^n$, the *normal cone* to C at $\bar{x} \in C$ is given by

$$N_C(\bar{x}) := \partial \delta_C(\bar{x}) = \left\{v \in \mathbb{R}^n \;\middle|\; v^T(x-\bar{x}) \leq 0 \quad \forall x \in C\right\}.$$

In turn, for $f \in \Gamma$ we have

$$N_{\operatorname{dom} f}(\bar{x}) = \partial^\infty f(\bar{x}) \quad \forall \bar{x} \in \operatorname{dom} f, \tag{3}$$

Rockafellar and Wets [23, Proposition 8.12].

3 Epi-Multiplication, Infimal Convolution, and Supercoercivity

Given $\omega \in \Gamma$, define

$$\omega_\lambda : x \in \mathbb{R}^n \mapsto \lambda \omega\left(\frac{x}{\lambda}\right) \quad (\lambda > 0). \tag{4}$$

In the literature, the operation

$$\lambda \star \omega := \omega_\lambda \quad (\lambda > 0),$$

is often called *epi-multiplication*, cf., e.g., [23, p. 24], which is closely related to *perspective functions*, see, e.g., [16, p. 90].

Note that

$$\omega_\lambda^* = \lambda \omega^* \quad (\lambda > 0), \tag{5}$$

cf. [23, Equation 11(3)].

Lemma 3.1 (Exercise 2.24 [23]) *Let $f, \omega \in \Gamma$. Then the following holds:*

a) $f\#\omega$ *is convex.*
b) ω_λ *is convex for all* $\lambda > 0$.
c) $\omega_{\lambda+\mu} = \omega_\lambda \# \omega_\mu$ *for all* $\lambda, \mu > 0$. *In particular, it holds that*

$$f\#\omega_\lambda\#\omega_\lambda = f\#\omega_{\lambda+\mu} \quad \forall\lambda, \mu > 0.$$

The next result, on which large parts of our reasoning relies, is due to Wijsman [26, 27]. See also [23, Theorem 11.37] for the version stated here.

Theorem 3.2 (Wijsman) *Given $f_k, f \in \Gamma$ ($k \in \mathbb{N}$) one has*

$$f_k \xrightarrow{e} f \quad \Longleftrightarrow \quad f_k^* \xrightarrow{e} f.$$

More generally, if $e\text{-}\liminf_{k\to\infty} f_k > -\infty$ *and there exists a bounded set* $B \subset \mathbb{R}^n$ *such that* $\limsup_{k\to\infty} \inf_B f_k < \infty$*, one has*

$$\liminf_{k\to\infty} f_k \geq f \iff \limsup_{k\to\infty} f_k^* \leq f^*$$
$$\limsup_{k\to\infty} f_k \leq f \iff \liminf_{k\to\infty} f_k^* \geq f^*$$

The next two results build the basis for establishing the intimate connection between epi-convergence of the epi-multiplication $\lambda \star \omega$ ($\omega \in \Omega$) to $\delta_{\{0\}}$ as $\lambda \downarrow 0$.

Lemma 3.3 *Let $\omega \in \Gamma$ with $0 \in \operatorname{dom}\omega$. Then for all $\{\lambda_k\} \downarrow 0$ and all $\bar{x} \in \mathbb{R}^n$, we have*

$$\liminf_{k\to\infty} \lambda_k \omega^*(x^k) \geq 0 \quad \forall\{x^k\} \to \bar{x},$$

i.e.

$$e\text{-}\liminf_{\lambda\downarrow 0} \lambda\omega^* \geq \sigma_{\{0\}}(\equiv 0).$$

Proof Let $\{\lambda_k\} \downarrow 0$, $\bar{x} \in \mathbb{R}^n$ and $\{x^k\} \to \bar{x}$. Then, noticing that

$$\inf \omega^* = -\omega(0) \in \mathbb{R},$$

see [23, Theorem 11.8 (a)], we have

$$\liminf_{k\to\infty} \lambda_k \omega^*(x^k) \geq \liminf_{k\to\infty} \lambda_k(-\omega(0)) = 0.$$

□

Corollary 3.4 *Let $\omega \in \Gamma$ with $0 \in \operatorname{dom}\omega$. Then*

$$e\text{-}\limsup_{\lambda\downarrow 0} \omega_\lambda \leq \delta_{\{0\}},$$

i.e. for all $\{\lambda_k\} \downarrow 0$ and all $\bar{x} \in \mathbb{R}^n$, we have

$$\exists\{x^k\} \to \bar{x} : \quad \limsup_{k\to\infty} \omega_{\lambda_k}(x^k) \leq \delta_{\{0\}}(\bar{x}).$$

Proof Let $\{\lambda_k\} \downarrow 0$. By Lemma 3.3 (and (5)) we know that $\text{e-}\liminf_{k\to\infty}(\omega_{\lambda_k})^* \geq \delta_{\{0\}}^*$. Moreover, by assumption and [23, Theorem 11.8 (a)],

$$-\infty < -\omega(0) = \inf w^* \leq w^*(v) < \infty,$$

for any $v \in \operatorname{dom} w^*$. We infer that

$$(\omega_{\lambda_k})^*(v) = \lambda_k \omega^*(v) \to 0,$$

where $\{v\}$ is clearly a bounded set. Hence $\limsup_{k\to\infty}(\omega_{\lambda_k})^*(v) = 0$. Therefore, the assertion follows from Wijsman's Theorem 3.2. □

The key assumption on ω for the remainder is *supercoercivity*. A function $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called *coercive* (or *0-coercive, level-bounded*) if

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

In turn, ϕ is said to be *supercoercive* (or *1-coercive*) if

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

The following result summarizes a collection of characterizations of supercoercivity useful to our study.

Lemma 3.5 *Let $\omega \in \Gamma$. Then the following statements are equivalent:*

i) *ω is supercoercive.*
ii) $\operatorname{dom} \omega^* = \mathbb{R}^n$.
iii) $\omega_\lambda \overset{e}{\to} \delta_{\{0\}} \quad (\lambda \downarrow 0)$.
iv) $\lambda\omega^* \overset{e}{\to} \sigma_{\{0\}} \equiv 0 \quad (\lambda \downarrow 0)$.
v) $\lambda\omega^* \overset{c}{\to} \sigma_{\{0\}} \equiv 0 \quad (\lambda \downarrow 0)$.

Proof The equivalence i) ⇔ ii) follows from [23, Theorem 11.8 (d)], while iii) ⇔ iv) follows from Theorem 3.2 and (5).

We now show that ii) ⇔ iv): Let $\{\lambda_k\} \downarrow 0$, and let us first assume that $\operatorname{dom} \omega^* = \mathbb{R}^n$: For $\bar{x} \in \mathbb{R}^n$, by Lemma 3.3, we have

$$\liminf_{k\to\infty} \lambda_k \omega^*(x^k) \geq 0 \quad \forall \{x^k\} \to \bar{x}.$$

On the other hand, for $\{x^k := \bar{x}\}$, we have

$$\limsup_{k\to\infty} \lambda_k \omega^*(x^k) = 0,$$

as $\bar{x} \in \operatorname{dom} \omega^* = \mathbb{R}^n$. This shows that $\text{e-}\lim_{\lambda\downarrow 0} \lambda\omega^* = 0$.

Now assume that $\operatorname{dom} \omega^* \subsetneq \mathbb{R}^n$, take $\bar{x} \notin \operatorname{cl}(\operatorname{dom} \omega^*)$.Then for all $\{x^k\} \to \bar{x}$, we have $x^k \notin \operatorname{dom} \omega^*$ for all k sufficiently large. Hence,

$$\limsup_{k\to\infty} \lambda_k \omega^*(x^k) = +\infty > 0,$$

so that $\text{e-}\limsup_{\lambda\downarrow 0} \lambda\omega^* \neq 0$, in particular $\text{e-}\lim_{\lambda\downarrow 0} \lambda\omega^* \neq 0$, which proves the converse implication.

To complete the proof, it suffices to show that $\lambda\omega^* \overset{e}{\to} \sigma_{\{0\}} \equiv 0$ implies $\lambda\omega^* \overset{c}{\to} \sigma_{\{0\}} \equiv 0$ (since the reverse implication is clear). However, this implication follows immediately from [23, Theorem 7.14/7.17] since $\operatorname{dom} \omega^* = \mathbb{R}^n$. □

Using the duality correspondences for epi-convergence due to Wijsman (Theorem 3.2), the implication i) ⇒ v) ⇒ iii) was already proved by Strömberg (see the proof of [24, Theorem 5.10 (c)]), while the equivalence *i*) ⇔ iii) was provided independently in [8, Lemma 4.3] with a more elementary technique of proof using the geometric characterization of epi-convergence via set-convergence (in the Painlevé-Kuratowski sense) of the epigraphs.

The following result is key.

Proposition 3.6 *Let $f, \omega \in \Gamma$. Then, for all $\lambda > 0$, the following hold:*

a) *We have*

$$(f\#\omega_\lambda)^* = f^* + \lambda\omega^*.$$

b) *If ω is supercoercive, then*

$$(f^* + \lambda\omega^*)^* = f\#w_\lambda.$$

c) *If ω is supercoercive (or coercive and f bounded from below), then $f\#\omega_\lambda$ is exact, that is,*

$$\forall x \in \mathbb{R}^n \ \exists u_\lambda(x): \ (f\#\omega_\lambda)(x) = f(u_\lambda(x)) + \omega_\lambda(x - u_\lambda(x)) \in \mathbb{R} \cup \{+\infty\},$$

or equivalently,

$$\underset{u\in\mathbb{R}^n}{argmin}\left\{f(u) + \lambda\omega\Big(\frac{x-u}{\lambda}\Big)\right\} \neq \emptyset \quad \forall x \in \mathbb{R}^n.$$

In particular, if ω is also finite-valued, $f\#\omega_\lambda$ is finite-valued (and convex), hence locally Lipschitz.

d) *If ω is supercoercive, then for $\bar{x} \in \operatorname{dom} f$ and*

$$u_\lambda(\bar{x}) \in \underset{u\in\mathbb{R}^n}{argmin}\left\{f(u) + \lambda\omega\Big(\frac{x-u}{\lambda}\Big)\right\}(\neq \emptyset)$$

we have

$$u_\lambda(\bar{x}) \to \bar{x} \quad and \quad f(u_\lambda(\bar{x})) \to f(\bar{x}) \quad (\lambda \downarrow 0).$$

Proof a) Follows immediately from [23, Theorem 11.23 (a)] (without acutally using the convexity and lower semicontinuity of f and ω) and (5).

b) This follows also from [23, Theorem 11.23 (a)] and (5), while using the fact that $\operatorname{dom}\omega^* = \mathbb{R}^n$ as ω is supercoercive (see Lemma 3.5).

c) The exactness result(s) follows from [5, Proposition 12.14]. If, in addition, ω is finite-valued, then for $\bar{x} \in \operatorname{dom} f \neq \emptyset$, we have

$$-\infty < (f\#\omega_\lambda)(x) \leq f(\bar{x}) + \omega_\lambda(x - \bar{x}) < +\infty \quad \forall x \in \mathbb{R}^n, \lambda > 0,$$

where the first (proper) inequality is due to the exactness. Hence, $f\#\omega_\lambda$ is finite-valued. Since it is convex, see Lemma 3.1 a), it is locally Lipschitz, see, e.g., [23, Example 9.14].

d) Let $x \mapsto a^T x + \beta$ be an affine minorant of f (see, e.g., [5, Theorem 9.19]) and $\bar{x} \in \operatorname{dom} f$. Then, for all $\lambda > 0$, we have

$$\begin{aligned} f(\bar{x}) &\geq (f\#\omega_\lambda)(\bar{x}) \\ &= f(u_\lambda(\bar{x})) + \lambda\omega\left(\frac{\bar{x} - u_\lambda(\bar{x})}{\lambda}\right) \\ &\geq a^T u_\lambda(\bar{x}) + \gamma + \lambda\omega\left(\frac{\bar{x} - u_\lambda(\bar{x})}{\lambda}\right) \\ &\geq -\|a\| \cdot \|u_\lambda(\bar{x})\| + \gamma + \lambda\omega\left(\frac{\bar{x} - u_\lambda(\bar{x})}{\lambda}\right). \end{aligned}$$

Here, the first inequality uses Lemma 3.7, where the last one uses Cauchy-Schwarz.

Rearranging and dividing by $\|\bar{x} - u_\lambda(\bar{x})\|$ (while neglecting (sub-)sequences $\{u_{\lambda_k}(\bar{x})\}$ with $u_{\lambda_k}(\bar{x}) = \bar{x}$ for all $k \in \mathbb{N}$) gives

$$\frac{f(\bar{x})}{\|\bar{x} - u_\lambda(\bar{x})\|} + \|a\| \cdot \frac{\|u_\lambda(\bar{x})\|}{\|\bar{x} - u_\lambda(\bar{x})\|} - \frac{\gamma}{\|\bar{x} - u_\lambda(\bar{x})\|} \geq \frac{\omega\left(\frac{\bar{x} - u_\lambda(\bar{x})}{\lambda}\right)}{\frac{\|\bar{x} - u_\lambda(\bar{x})\|}{\lambda}}. \tag{6}$$

Now, assume that $\{u_{\lambda_k}(\bar{x})\}$ for some sequence $(\lambda_k \downarrow 0)$ were unbounded (i.e. $\{\bar{x} - u_{\lambda_k}(x)\}$ and hence, $\left\{\frac{\bar{x} - u_{\lambda_k}(\bar{x})}{\lambda_k}\right\}$ unbounded). As $f(\bar{x}) < +\infty$ and ω is supercoercive, (6) gives the contradiction

$$+\infty > \|a\| \limsup_{k \to \infty} \frac{\|u_{\lambda_k}(\bar{x})\|}{\|\bar{x} - u_{\lambda_k}(\bar{x})\|} \geq \limsup_{k \to \infty} \frac{\omega\left(\frac{\bar{x} - u_{\lambda_k}(\bar{x})}{\lambda_k}\right)}{\frac{\|\bar{x} - u_{\lambda_k}(\bar{x})\|}{\lambda_k}} = +\infty.$$

Hence, $\{u_\lambda(\bar{x})\}$ $(\lambda \downarrow 0)$ is bounded. Using this fact again in (6), while assuming that $u_\lambda(\bar{x}) \not\to \bar{x}$ gives another contradiction, due to the 1-coercivity of ω. Hence, $u_\lambda(\bar{x}) \to \bar{x}$ as $\lambda \downarrow 0$, which proves the first assertion.

In order to see the second statement, we first point out that

$$\limsup_{\lambda \downarrow 0} \lambda \omega\left(\frac{\bar{x} - u_\lambda(\bar{x})}{\lambda}\right) \geq 0, \tag{7}$$

as ω is supercoercive. Hence, it follows that

$$\begin{aligned} f(\bar{x}) &\geq \limsup_{\lambda \downarrow 0} \left\{ f(u_\lambda(\bar{x})) + \lambda \omega\left(\frac{\bar{x} - u_\lambda(\bar{x})}{\lambda}\right)\right\} \\ &\geq \limsup_{\lambda \downarrow 0} f(u_\lambda(\bar{x})) \\ &\geq \liminf_{\lambda \downarrow 0} f(u_\lambda(\bar{x})) \\ &\geq f(\bar{x}). \end{aligned}$$

Here, the first inequality is, again, due to Lemma 3.7, the second one uses (7), while the last one follows from the fact that f is lsc and $u_\lambda(\bar{x}) \to \bar{x}$. □

Next we recall that the net $\{(f \# \omega_\lambda)(x)\}_{(\lambda > 0)}$ for $x \in \mathbb{R}^n$ is monotonic.

Lemma 3.7 [8, Lemma 4.4] *Let $f, \omega \in \Gamma$ such that $\omega(0) \leq 0$. Then for all $x \in \mathbb{R}^n$ the net $\{(f \# \omega_\lambda)(x)\}_{(\lambda > 0)}$ is monotonically increasing as $\lambda \downarrow 0$ and bounded above by $f(x)$.*

The following theorem captures the basic properties of epi-convergence for infimal convolution with supercoercive kernels in the convex case.

Theorem 3.8 *Let $f, \omega \in \Gamma$ with $0 \in \operatorname{dom} \omega$ and ω_λ $(\lambda > 0)$ defined as in (4). Then the following hold:*

a) $e\text{-}\liminf_{\lambda \downarrow 0}(f^* + \lambda \omega^*) \geq f^*$.
b) *If ω is supercoercive, then*

$$e\text{-}\lim_{\lambda \downarrow 0} f^* + \lambda \omega^* = f^*, \quad e\text{-}\lim_{\lambda \downarrow 0} f \# \omega_\lambda = f.$$

and

$$\operatorname{gph} \partial(f \# \omega_\lambda) \to \operatorname{gph} \partial f.$$

c) *If ω is supercoercive and $\omega(0) \le 0$, we also have*

$$\lim_{\lambda\downarrow 0}(f\#\omega_\lambda)(x) \uparrow f(x) \quad \forall x \in \mathbb{R}^n.$$

In particular,

$$p\text{-}\lim_{\lambda\downarrow 0} f\#\omega_\lambda = f.$$

Proof a) We know from Lemma 3.3 that $\text{e-}\liminf_{\lambda\downarrow 0} \lambda\omega^* \ge \sigma_{\{0\}} \equiv 0$. Moreover, as f^* is lsc, we also have $\text{e-}\lim_{\lambda\downarrow 0} f^* = f^*$. Hence, invoking [23, Theorem 7.46] (or, [10, Theorem 3.2]), we get

$$\text{e-}\liminf_{\lambda\downarrow 0}(f^* + \lambda\omega^*) \ge f^*.$$

b) If ω is coercive, we know, from Lemma 3.5, that $\text{c-}\lim_{\lambda\downarrow 0} \lambda\omega^* \equiv 0$. Hence, as $\text{e-}\lim_{\lambda\downarrow 0} f^* = f^*$, [23, Theorem 7.46 (b)] (or [10, Theorem 3.2]) yields

$$\text{e-}\lim_{\lambda\downarrow 0}(f^* + \lambda\omega^*) = f^*.$$

To prove the next assertion, simply note that $(f^* + \lambda\omega^*)^* = f\#\omega_\lambda$ by Proposition 3.6 b) and apply Wijsman's Theorem 3.2.

The graphical convergence of the subdifferentials is due to *Attouch's Theorem*, see [1, Theorem 3.66] or [23, Theorem 12.35].

c) If, in addition, $\omega(0) \le 0$, Lemma 3.7 tells us that the net $\{(f\#\omega_\lambda)(x)\}_{\lambda>0}$ is bounded above by $f(x)$ and monotonically increasing as $\lambda \downarrow 0$. Hence it is pointwise convergent. By part b), we have

$$f(x) \ge \lim_{\lambda\downarrow 0}(f\#\omega_\lambda)(x) \ge f(x),$$

hence $(f\#\omega_\lambda)(x) \uparrow f(x)$ as $\lambda \downarrow 0$. □

Part b) of this result was already establish in [8, Proposition 4.5 and Theorem 4.6] using a different proof technique.

In the infinite dimensional case, a similiar result was furnished by Beer in [11, Theorem 7.3.8], using a so-called *regularizing family of smoothing kernels*, see [11, Definition 7.3.5]. Beer's smoothing kernels are required to be non-negative and to take the value zero at the origin. This is slightly more restrictive than is required by Theorem 3.8 c). In addition, since, in Lemma 3.5, we establish the equivalence

$$\omega_\lambda \xrightarrow{e} \delta_{\{0\}} \quad \Longleftrightarrow \quad \omega \text{ supercoercive},$$

the smoothing kernels in [11, Theorem 7.3.8] are necessarily supercoercive in our finite-dimensional setting. Moreover, by Proposition 3.6, the supercoercivity assumption, gives the full duality correspondence

$$(f\#\omega_\lambda)^* = f + \lambda\omega^* \quad \text{and} \quad (f + \lambda\omega^*)^* = f\#\omega_\lambda,$$

a fact, which seems to be missing in the infinite dimensional setting of cite[Theorem 7.3.8]Bee 96.

It is important to note that the result [23, Theorem 7.46] on the epi-convergence of sums of epi-convergent sequences, is flawed and cannot be applied off-hand even in the convex setting. A corrected version of this result can be found in [10, Theorem 3.2], and it is this corrected version that we invoke in the proof of Theorem 3.8.

Theorem 3.8 tells us that if $0 \in \operatorname{dom}\omega$ and ω is supercoercive, then $f = \text{e-}\lim_{\lambda\downarrow 0} f\#\omega_\lambda$, or equivalently, $f^* = \text{e-}\lim_{\lambda\downarrow 0} f^* + \lambda\omega^*$. We now consider the question of wether one can relax the supercoercivity hypothesis on ω and still obtain the desired epiconvergence of $f\#\omega_\lambda$ to f. The following proposition clarifies what is possible in this regard.

Proposition 3.9 *Let $f \in \Gamma$.*

a) *If $\omega \in \Gamma$ is such that $f\#\omega_\lambda \xrightarrow{e} f$, then $\mathrm{dom}\, f^* \subset \mathrm{cl}\,(\mathrm{dom}\,\omega^*)$. In particular, if f is supercoercive, then ω must also be supercoercive.*
b) *If f is not supercoercive, then there exists $\omega \in \Gamma$ that is not supercoercive and for which $f\#\omega_\lambda \xrightarrow{e} f$.*

Proof a) If $\mathrm{dom}\, f^* \not\subset \mathrm{cl}\,\mathrm{dom}\,\omega^*$, then there is an $\bar{x} \in \mathrm{dom}\, f^* \setminus \mathrm{cl}\,\mathrm{dom}\,\omega^*$. Hence, for every sequence $\{x^k\} \to \bar{x}$, we know that $x^k \notin \mathrm{dom}\,\omega^*$ for all k sufficiently large. Therefore, $\limsup_{k\to\infty} f^*(x^k) + \lambda^k\omega^*(x^k) = +\infty > f(\bar{x})$. Consequently, by (2), $f^* + \lambda\omega$ cannot epiconverge to f^*, or equivalently, $f\#\omega_\lambda$ does not epiconverge to f. Finally, note that if ω is supercoercive, then, by Lemma 3.5, $\mathrm{dom}\,\omega^* = \mathbb{R}^n$, and so trivially $\mathrm{dom}\, f^* \subset \mathrm{cl}\,(\mathrm{dom}\,\omega^*)$.

b) If f is not supercoercive, $\mathrm{dom}\, f^* \subsetneq \mathbb{R}^n$. Define $\omega := \sigma_{\mathrm{dom}\, f^*}$. Then $\omega^* = \delta_{\overline{\mathrm{conv}}\,(\mathrm{dom}\, f^*)} = \delta_{\mathrm{cl}\,(\mathrm{dom}\, f^*)}$, and, hence, ω is not supercoercive (as $\mathrm{dom}\,\omega^* \neq \mathbb{R}^n$). Yet, we have

$$f^* + \lambda\omega^* = f^* \xrightarrow{e} f^*,$$

which proves the assertion. □

The following theorem is the main result on infimal-convolution smoothing. It resembles a result obtained by Strömberg [24, Theorem 5.10] (see the discussion following the proof). However, due in part to finite dimensionality, the proof we give here is self-contained, whereas, the technique of proof in [24] makes use of results for *viscosity functions* obtained by Attouch in [2, 3].

Theorem 3.10 *Let $f, \omega \in \Gamma$ be such that ω is differentiable on $\mathbb{R}^n$ (and consequently continuously differentiable by* [23, Corollary 9.20]*) and supercoercive. Then the following hold:*

a) *ω^* is strictly convex and supercoercive with $\mathrm{dom}\,\omega^* = \mathbb{R}^n$.*
b) *For all $\bar{x} \in \mathbb{R}^n$ the points*

$$u_\lambda(\bar{x}) \in \underset{u\in\mathbb{R}^n}{\mathrm{argmin}} \left\{ f(u) + \lambda\omega\Big(\frac{x-u}{\lambda}\Big)\right\} \quad (\lambda > 0)$$

are characterized by

$$\nabla\omega\left(\frac{\bar{x} - u_\lambda(\bar{x})}{\lambda}\right) \in \partial f\,(u_\lambda(\bar{x})) \quad (\lambda > 0).$$

c) *$f\#\omega_\lambda$ is continuously differentiable for all $\lambda > 0$ with*

$$\nabla(f\#\omega_\lambda)(\bar{x}) = \nabla\omega\left(\frac{\bar{x} - u_\lambda(\bar{x})}{\lambda}\right).$$

d) *We have*

$$f\#\omega_\lambda \xrightarrow{e} f \quad \textit{and} \quad \mathrm{gph}\,\nabla(f\#\omega_\lambda) \to \mathrm{gph}\,\partial f \quad (\lambda \downarrow 0);$$

hence, in particular

$$\underset{\lambda\downarrow 0, x\to\bar{x}}{\mathit{Lim\,sup}}\, \nabla(f\#\omega_\lambda)(x) = \partial f(\bar{x}) \quad \forall \bar{x} \in \mathbb{R}^n.$$

e) *For all* $x \in \mathbb{R}^n$ *and all* $\lambda > 0$*, we have*

$$\underset{y\in\mathbb{R}^n}{argmin}\{f^*(y) + \lambda\omega^*(y) - \bar{x}^T y\} = \{\nabla(f\#\omega_\lambda)(\bar{x})\} \quad \forall \bar{x} \in \mathbb{R}^n.$$

f) *For all* $\bar{x} \in \operatorname{dom} \partial f$*, we have*

$$\omega^*(\nabla(f\#\omega_\lambda)(\bar{x})) \le \min\{\omega^*(y) \mid y \in \partial f(\bar{x})\} \quad \forall \lambda > 0.$$

g) *If* $\omega(0) \le 0$*, then*

$$(f\#\omega_\lambda)(x) \uparrow f(x) \quad (\lambda \downarrow 0) \quad \forall x \in \mathbb{R}.$$

In particular,

$$f\#\omega_\lambda \xrightarrow{p} f \quad (\lambda \downarrow 0).$$

h) *If* $\omega(0) \le 0$*, then, for all* $\bar{x} \in \operatorname{dom} \partial f$*, we have*

$$\lim_{\lambda\downarrow 0} \nabla(f\#\omega_\lambda)(\bar{x}) = argmin\{\omega^*(y) \mid y \in \partial f(\bar{x})\},$$

and, in particular,

$$\lim_{\lambda\downarrow 0} \omega^*(\nabla(f\#\omega_\lambda)(\bar{x})) = \min\{\omega^*(y) \mid y \in \partial f(\bar{x})\}.$$

Proof a) By [23, Theorem 11.1], $\omega = \omega^{**}$; hence, since $\operatorname{dom} \omega = \mathbb{R}^n$, Lemma 3.5 tells us that ω^* is supercoercive. In addition, since ω is differentiable on $\mathbb{R}^n$, [23, Theorem 11.13] tells us that ω^* is strictly convex on every convex subset of $\operatorname{dom} \partial\omega^*$. But, again by Lemma 3.5, $\operatorname{dom} w^* = \mathbb{R}^n$, so ω^* is strictly convex on $\mathbb{R}^n$. We use these facts freely in the remainder of the proof.

b) Since

$$\min_{u\in\mathbb{R}^n} \left\{ f(u) + \lambda\omega\left(\frac{x-u}{\lambda}\right) \right\} \quad (\lambda > 0),$$

is an unconstrained convex optimization problem, the condition

$$0 \in \partial f(u) - \nabla\omega\left(\frac{x-u}{\lambda}\right)$$

characterizes its solutions, which exist by Proposition 3.6 c).

c) As in Proposition 3.6 c), $f\#\omega_\lambda$ is exact, so differentiability and the derivative formula follow from [5, Proposition 18.7 /Corollary 18.8]. Continuous differentiability follows from [23, Corollary 9.20].

d) This follows from Theorem 3.8 b), since trivially $0 \in \operatorname{dom} \omega = \mathbb{R}^n$.

e) We have

$$\begin{aligned}
\underset{y\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ f^*(y) + \lambda\omega^*(y) - \bar{x}^T y \right\} &= \underset{y\in\mathbb{R}^n}{\operatorname{argmax}} \left\{ x^T y - f^*(y) - \lambda\omega^*(y) \right\} \\
&= \underset{y\in\mathbb{R}^n}{\operatorname{argmax}} \left\{ x^T y - (f\#\omega_\lambda)^*(y) \right\} \\
&= \partial(f\#\omega_\lambda)(\bar{x}) \\
&= \{\nabla(f\#\omega_\lambda)(\bar{x})\},
\end{aligned}$$

where the third equality follows from, e.g., [23, Proposition 11.3].

f) For $\lambda > 0$, set $y_\lambda := \nabla(f\#\omega_\lambda)(\bar{x})$. By part d),

$$f^*(y_\lambda) + \lambda\omega^*(y_\lambda) - \bar{x}^T y_\lambda \le f^*(y) - x^T y + \lambda\omega^*(y) \quad \forall y \in \mathbb{R}^n.$$

Hence,

$$\lambda w^*(y_\lambda) \le f^*(y) - f^*(y_\lambda) - \bar{x}^T(y - y_\lambda) + \lambda\omega^*(y) \quad \forall y \in \mathbb{R}^n.$$

Now if $y \in \partial f(\bar{x})$, or equivalently $\bar{x} \in \partial f^*(y)$, we have

$$f^*(y) - f^*(y_\lambda) - x^T(y - y_\lambda) \le 0.$$

Consequently,

$$\omega^*(y_\lambda) \le \omega^*(y) \quad \forall y \in \partial f(\bar{x}).$$

This proves the assertion, since the minimum in fact exists, due to the (1-)coercivity and lower semicontinuity of ω^* as well as the closedness of $\partial f(\bar{x})$.

g) See Theorem 3.8 c).

h) Again, for $\lambda > 0$, set $y_\lambda := \nabla(f\#\omega_\lambda)(\bar{x})$. By part e),

$$\omega^*(y_\lambda) \le \min\{\omega^*(y) \mid y \in \partial f(\bar{x})\}, \tag{8}$$

and so the (1-)coercivity of ω^* implies that the set $\{y_\lambda : \lambda > 0\}$ is necessarily bounded. Let $\bar{y}$ be an accumulation point of $\{y_\lambda : \lambda > 0\}$. By parts a) and b),

$$y_\lambda \in \partial f(u_\lambda(\bar{x})) \quad \forall \lambda > 0,$$

and, by Proposition 3.6 d), $u_\lambda(\bar{x}) \to \bar{x}$. Hence, by the outer semicontinuity of ∂f, $\bar{y} \in \partial f(\bar{x})$. Thus, in view of (8) and the lower semi-continuity of ω^*,

$$\bar{y} \in \operatorname{argmin}\{\omega^*(y) \mid y \in \partial f(\bar{x})\}.$$

On the other hand, as ω^* is strictly convex and $\partial f(\bar{x})$ is (closed and) convex, the set $\operatorname{argmin}\{\omega^*(y) \mid y \in \partial f(\bar{x})\}$ is a singleton. Therefore, the bounded set $\{y_\lambda \mid \lambda > 0\}$ has exactly one accumulation point $\bar{y} = \operatorname{argmin}\{\omega^*(y) \mid y \in \partial f(\bar{x})\}$, and hence $y_\lambda \to \bar{y}$ as $\lambda \downarrow 0$. □

We point out that due to item b) and c) of the above result, the functions $f\#\omega_\lambda$ constitute *gradient consistent epi-smoothing functions* for f in the sense of [8].

As previously stated, Theorem 3.10 is closely related to Strömberg's result [24, Theorem 5.1]. While the assertions are basically the same, the assumptions differ. However some of these differences only stem from the fact that in [24] general normed spaces are considered, while here we focus our attention on finite-dimensional spaces.

In order to compare the assumptions of Theorem 3.10 and [24, Theorem 5.1], we first need to recall that ω is differentiable and supercoercive if and only if ω^* is strictly convex, supercoercive and $\operatorname{dom}\omega^* = \mathbb{R}^n$. Hence, the coercivity and differentiabililty assumptions essentially coincide in both Theorems. However, we do not impose the assumptions that ω be nonnegative and $\omega(0) = 0$, but merely demand $\omega(0) \le 0$. In addition, while we assume ω^* to be strictly convex (by assuming ω to be differentiable), [24, Theorem 5.1] uses *locally uniform convexity* of ω^* in order to obtain the Fréchet differentiability of $f\#\omega_t$ [24, Theorem 3.8 (b)]. However, if one assumes that ω^* is strictly convex, then $f\#\omega_t$ is Gâteaux differentiable [24, Theorem 3.8(a)], which, in finite-dimensions, is equivalent to Fréchet differentiability.

Theorem 3.10 shows that if ω is differentiable on $\mathbb{R}^n$ and supercoercive, then $f\#\omega_\lambda$ is continuously differentiable and epi-converges to f. The following very elementary result shows that, without further knowledge of f, these hypotheses on ω are required for such a result.

Proposition 3.11 *Let $\omega \in \Gamma$. Then $f\#\omega_\lambda$ is continuously differentiable and epi-converges to f for all $f \in \Gamma$ if and only if ω is differentiable on $\mathbb{R}^n$ and supercoercive.*

Proof One direction has already been established in Theorem 3.10. The reverse direction follows by taking $f = \delta_{\{0\}}$. Indeed, in this case $f\#\omega_\lambda = \omega_\lambda$, so we must have that ω is differentiable on $\mathbb{R}^n$. In addition, $\omega_\lambda = f\#\omega_\lambda \xrightarrow{e} f = \delta_{\{0\}}$, so, by Lemma 3.5, ω is supercoercive. □

The following corollary, which was partly covered in [8, Theorem 4.6], shows that the function $f\#\omega_\lambda$ $(\lambda > 0)$ has Lipschitz gradient if the smoothing kernel ω has a Lipschitz gradient. In what follows, we call a function $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *strongly convex* [23, Exercise 12.59] if

$$\exists \sigma > 0 : \ \phi - \frac{\sigma}{2}\|\cdot\|^2 \text{ is convex.}$$

Corollary 3.12 *Let the assumptions of Theorem 3.10 hold, and let ω have Lipschitz gradient (with Lipschitz modulus $L > 0$). Then, in addition to all the properties from Theorem 3.10, $\nabla(f\#\omega_\lambda)$ is also Lipschitz (with modulus $\frac{L}{\lambda}$) for all $\lambda > 0$.*

Proof We have the following chain of implications

$$\begin{aligned}
& \omega \text{ is differentiable with } \nabla\omega \text{ Lipschitz (with modulus } L)\\
\Longleftrightarrow\ & \nabla\omega_\lambda \text{ is Lipschitz (with modulus } \tfrac{L}{\lambda})\\
\Longleftrightarrow\ & (\omega_\lambda)^* = \lambda\omega^* \text{ is strongly convex (with modulus } \tfrac{\lambda}{L})\\
\Longrightarrow\ & f^* + \lambda\omega^* \text{ is strongly convex (with modulus } \tfrac{\lambda}{L})\\
\Longleftrightarrow\ & (f^*+\lambda\omega^*)^* = \nabla f\#\omega_\lambda \text{ is differentiable with } \nabla(f^*+\lambda\omega^*)^* \text{ Lipschitz (with modulus } \tfrac{L}{\lambda}).
\end{aligned}$$

Here the first and the last equivalence employ [23, Proposition 12.60]. □

The question immediately arises as to whether further smoothness properties of the kernel ω can increase the smoothness of $f\#\omega_\lambda$. However, in general, this is not the case as the following example illustrates.

Example 3.13 Let $C \in \mathbb{R}^n$ be a closed convex set. With $f := \delta_C$ and $\omega := \frac{1}{2}\|\cdot\|^2$, we have that ω

$$\nabla(\delta_C\#\omega_\lambda) = \frac{1}{\lambda}(\mathrm{id} - \Pi_C),$$

cf., [23], where Π_C is the Euclidean projector onto C, which is Lipschitz (with modulus $L = 1$), but not differentiable in general [23, Exercise 2.25].

4 Supercoercive Kernels

The results of the previous section establish supercoercivity as an essential hypothesis in obtaining the epi-convergence and smoothness of the the family $f\#\omega_\lambda$. Hence, for the purposes of applications, it is important to understand this class of functions. The following proposition is provided in order to give some insight into the nature and richness of this class.

Proposition 4.1 *Let $\omega \in \Gamma$. Then the following hold:*

a) *If ω is supercoercive, then*

i) $\forall \phi \in \Gamma: \ \phi + \omega \ (\in \Gamma)$ *is supercoercive;*
ii) $\forall \alpha > 0: \ \alpha\omega$ *is supercoercive.*

b) *If ω is strongly convex, then ω is supercoercive.*
c) *Suppose ω is supercoercive and $\theta : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is a proper, lsc, and nondecreasing convex function that is nonconstant on its domain. If there exists $\hat{x} \in \operatorname{dom} \omega$ such that $\omega(\hat{x}) \in \operatorname{dom} \theta$, then $\theta \circ \omega$ is a supercoercive element of Γ (under the convention that $\theta(+\infty) = \infty$).*

Proof a) A proof of i) can be found in [5, Proposition 11.13] and follows immediately from the fact that a convex function is minorized by an affine minorant, while ii) is obvious.

b) As ω is strongly convex, there exists $\sigma > 0$ such that $\omega - \frac{\sigma}{2}\|\cdot\|^2 \in \Gamma$. Moreover, trivially, $\frac{\sigma}{2}\|\cdot\|^2$ is supercoercive. Hence, the assertion follows from a), as

$$\omega = (\omega - \frac{\sigma}{2}\|\cdot\|^2 \in \Gamma) + \frac{\sigma}{2}\|\cdot\|^2.$$

c) It is well known, see, e.g., [23, Exercise 2.20 (b)], that $\theta \circ \omega$ is lsc, proper, and convex (under the convention that $\theta(+\infty) = \infty$). Since θ is nondecreasing and proper, there is a $t_0 \in \mathbb{R}$ such that $(-\infty, t_0] \subset \operatorname{dom} \theta$. The result will follow once we have shown that for all t sufficiently large $\theta(t)$ is bounded below by a linear function with positive slope. To this end, note that for every $t_1, t_2, t_3 \in \operatorname{dom} \theta$ with $t_1 < t_2 < t_3$, we have

$$t_2 = \frac{t_2 - t_1}{t_3 - t_1} t_3 + \frac{t_3 - t_2}{t_3 - t_1} t_1.$$

Hence

$$\theta(t_2) \leq \frac{t_2 - t_1}{t_3 - t_1} \theta(t_3) + \frac{t_3 - t_2}{t_3 - t_1} \theta(t_1),$$

or equivalently,

$$[(t_3 - t_2) + (t_2 - t_1)]\theta(t_2) \leq (t_2 - t_1)\theta(t_3) + (t_3 - t_2)\theta(t_1).$$

By rearranging, we obtain

$$(t_3 - t_2)[\theta(t_2) - \theta(t_1)] \leq (t_2 - t_1)[\theta(t_3) - \theta(t_2)]$$

Consequently,

$$\frac{\theta(t_2) - \theta(t_1)}{t_2 - t_1} \leq \frac{\theta(t_3) - \theta(t_2)}{t_3 - t_2} \tag{9}$$

for all $t_1, t_2, t_3 \in \operatorname{dom} \theta$ with $t_1 < t_2 < t_3$.

Since θ is nonconstant on its domain, there are $t_1, t_2 \in \operatorname{dom} \theta$ such that $\theta(t_1) < \theta(t_2)$. Set $0 < \gamma = (\theta(t_2) - \theta(t_1))/(t_2 - t_1)$. By (9),

$$\theta(t_2) + \gamma(t - t_2) \leq \theta(t) \quad \forall\, t \geq t_2,$$

which establishes the result.

□

Clearly, e.g. from Proposition 4.1 b), for all $A \succ 0$, $b \in \mathbb{R}^n$, $\gamma \in \mathbb{R}$, the function

$$x \in \mathbb{R}^n \mapsto \frac{1}{2} x^T A x + b^T x + \gamma \in \mathbb{R}.$$

is supercoercive.

5 Application to Convex-Composite Functions

One of the major contributions of [8] is the construction of epi-convergent smoothing functions for so-called *convex-composite* functions, i.e., functions of the type $\phi := f \circ G : \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ where $f \in \Gamma$ and $G : \mathbb{R}^p \to \mathbb{R}^n$ is continuously differentiable. Note that, in general, such functions ϕ are lsc but not necessarily convex.

While supercoercivity will yield epi-convergence, the key assumption to obtain (set-)convergence of the gradients of the infimal convolutions to the subdifferential of the underlying convex-composite at the respective point is the *basic constraint qualification (BCQ)*, which is arguably due to Rockafellar and given as follows:

Definition 5.1 Let $\phi := f \circ G$ be a convex-composite, i.e., $f \in \Gamma$ and $G : \mathbb{R}^p \to \mathbb{R}^n$ continuously differentiable. Then the *basic constraint qualification (BCQ)* is satisfied (for ϕ) at $\bar{x} \in \operatorname{dom}\phi$ if

$$N_{\operatorname{dom} f}(G(\bar{x})) \cap \ker G'(\bar{x}) = \{0\}.$$

Note that, since f is convex, its domain is a convex set and $\bar{x} \in \operatorname{dom}\phi$ if and only if $G(\bar{x}) \in \operatorname{dom} f$. Moreover, note that BCQ for the convex-composite function $\phi = f \circ G$ is everywhere satisfied if f is finite-valued since in this case $N_{\operatorname{dom} f}(x) = \{0\}$ for all x. Clearly, it is also satisfied at any point where $G'(\bar{x})$ has full column rank, i.e., $\ker G'(\bar{x}) = \{0\}$.

In our analysis of the graph convergence of the subgradient mapping for convex-composite functions we also require that the mapping G be $C^{1,1}$ at the point of interest $\bar{x}$, i.e. there exist $\varepsilon > 0$ and $L > 0$ such that $\|G'(x) - G'(y)\| \leq L\,\|x - y\|$ whenever $x, y \in B_\varepsilon(\bar{x})$. We say that G is locally $C^{1,1}$ on a set $X \subset \mathbb{R}^p$ if it is $C^{1,1}$ at every point of X. If G is $C^{1,1}$ at a point $\bar{x}$, then we have the following well-known bound on the accuracy of the linear approximation to G near $\bar{x}$:

$$\left\|G(y) - (G(x) + G'(x)(y - x))\right\| \leq \frac{L}{2}\,\|y - x\|^2 \quad \forall\, x, y \in B_\varepsilon(\bar{x}), \tag{10}$$

where L is the Lipschitz constant for G' on $B_\varepsilon(\bar{x})$. Moreover, we make use of the well-known fact that if G is locally $C^{1,1}$ on X, then G' is (globally) Lipschitz on every compact subset of X.

Some other useful consequences of BCQ and the $C^{1,1}$ hypothesis are subsumed below.

Proposition 5.2 *Let $\phi := f \circ G$ be a convex-composite, i.e., $f \in \Gamma$ and $G : \mathbb{R}^p \to \mathbb{R}^n$ continuously differentiable, and let $\bar{x} \in \operatorname{dom}\phi$. Then the following hold:*

a) *If BCQ holds at $\bar{x}$, there exists a neighborhood N of $\bar{x}$ such that BCQ (for ϕ) holds at all $x \in N \cap \operatorname{dom}\phi$.*
b) *If BCQ holds at $\bar{x}$, then $\partial\phi(\bar{x}) = \hat{\partial}\phi(\bar{x}) = G'(\bar{x})^T \partial f(G(\bar{x}))$.*
c) *If G is $C^{1,1}$ at $\bar{x}$, we have $\partial_p\phi(\bar{x}) \supset G'(\bar{x})^T \partial f(G(\bar{x}))$. Consequently, under BCQ at $\bar{x}$, we get $\partial\phi(\bar{x}) = \hat{\partial}\phi(\bar{x}) = \partial_p\phi(\bar{x}) = G'(\bar{x})^T \partial f(G(\bar{x}))$.*

Proof a) See [23, Exercise 10.25 (b)].
b) See [23, Theorem 10.6].

c) If $\partial f(G(\bar{x})) = \emptyset$, we are done. Otherwise, let $v \in \partial f(G(\bar{x}))$. Then, by convexity of f and (10), we obtain

$$\begin{aligned}\phi(x) &\geq \phi(\bar{x}) + \langle v,\ G(x) - G(\bar{x})\rangle \\ &= \phi(\bar{x}) + \left\langle G'(\bar{x})^T v,\ x - \bar{x}\right\rangle + \left\langle v,\ G(x) - (G(\bar{x}) + G'(\bar{x})(x - \bar{x})\right\rangle \\ &\geq \phi(\bar{x}) + \left\langle G'(\bar{x})^T v,\ x - \bar{x}\right\rangle - \frac{L}{2}\|v\| \cdot \|x - \bar{x}\|^2,\end{aligned}$$

where L is the local Lipschitz constant for G' at $\bar{x}$. Hence, $G'(\bar{x})^T v \in \partial_p \phi(\bar{x})$, which gives the first assertion. The remainder follows from this and part b).

□

The following result is a refinement of the results from [8] in the sense that the smoothing kernel ω only needs to be continuously differentiable, as opposed to also having Lipschitz gradient. Furthermore, pointwise convergence is established for the smooth approximations.

Theorem 5.3 *Let $G : \mathbb{R}^p \to \mathbb{R}^n$ be continuously differentiable and $f, \omega \in \Gamma$ such that ω is supercoercive and differentiable with $\omega(0) \leq 0$. Then the following hold:*

a) *The function $\phi_\lambda := (f \# \omega_\lambda) \circ G$ is continuously differentiable for all $\lambda > 0$.*
b) *We have*

$$\phi_\lambda \xrightarrow{e} \phi := f \circ G \quad (\lambda \downarrow 0).$$

c) *For all $x \in \mathbb{R}$, we have*

$$\phi_\lambda(x) \uparrow \phi(x) \quad (\lambda \downarrow 0).$$

Hence, in particular, $\phi_\lambda \xrightarrow{p} \phi$ as $\lambda \downarrow 0$.
d) *If BCQ is satisfied at $\bar{x} \in \operatorname{dom}\phi$, we have*

$$\operatorname*{Lim\,sup}_{\lambda\downarrow 0, x\to\bar{x}} \nabla\phi_\lambda(x) = G'(\bar{x})^T \partial f(G(\bar{x})) = \partial\phi(\bar{x}).$$

Proof a) Follows immediately from Theorem 3.10 b) and the fact that G is continuously differentiable by assumption.
b) Follows from Theorem 3.10 b)-c) and Lemma 3.7 in combination with [8, Proposition 5.2].
c) Follows from Theorem 3.10 f).
d) Follows from a)-b) in combination with [8, Theorem 5.7].

□

Part d) of Theorem 5.3 establishes conditions under which the smoothing functions ϕ_λ satisfy gradient consistency at a point $\bar{x}$. This is significantly weaker than the local graph convergence of $\nabla\phi_\lambda$ to $\partial(f \circ G)$ that one might expect due to part d) of Theorem 3.10. However, establishing graph convergence is highly nontrivial as it requires an extension of *Attouch's Theorem* ([1, Theorem 3.66] or [23, Theorem 12.35]) to convex-composite functions. We now show that the local graph convergence of the subdifferentials occurs if G is assumed to be in the class $C^{1,1}$ at $\bar{x}$. This type of graph convergence result was already proven by Poliquin in [22, Theorem 2.1 and Proposition 2.3] in much greater generality than is required here, but for the case that $G \in C^2$. The proof makes use of the notion of *primal lower-nice* functions developed by Poliquin in [21, 22].

Definition 5.4 Let $f : \mathbb{R}^n \to \mathbb{R}$ be closed and proper. We say that f is primal lower-nice at $\bar{x} \in \operatorname{dom} f$ if there exists $\varepsilon > 0$, $c > 0$, and $T > 0$, such that

$$\begin{aligned} \langle y_1 - y_2,\, x_1 - x_2 \rangle &\geq -t\, \|x_1 - x_2\|^2 \\ \forall\, t \geq T,\; x_i \in B_\varepsilon(\bar{x}),\; y_i &\in \partial_p f(x_i),\; \|y_i\| \leq ct,\; i = 1, 2. \end{aligned} \tag{11}$$

A family of closed proper functions $\{f_\nu \mid \nu \in I\}$ is said to be equi primal lower-nice at $\bar{x} \in \cap_{\nu \in I} \operatorname{dom} f_\nu$ if $\varepsilon > 0$, $c > 0$, and $T > 0$ can be chosen so that (11) holds uniformly in $\nu \in I$.

Remark 5.5 In [22, Proposition 2.2], Poliquin shows that the condition (11) is equivalent to the condition that there exists $\hat{\varepsilon} > 0$, $\hat{c} > 0$, and $\hat{T} > 0$ such that

$$\begin{aligned} f(x') &\geq f(x) + \left\langle u,\, x' - x \right\rangle - (\hat{t}/2) \left\| x - x' \right\|^2 \\ \forall\, t \geq \hat{T},\; x \in B_\varepsilon(\bar{x}),\; y &\in \partial_p f(x),\; \|y\| \leq \hat{c}t,\; x' \in x + \hat{\varepsilon}\mathbb{B}. \end{aligned}$$

By using the bound from (10), it is easily shown that if the BCQ is satisfied at $\bar{x}$ (Definition 5.1) and G is $C^{1,1}$ at $\bar{x}$, then ϕ is primal lower-nice at $\bar{x}$. Indeed, this has already been observed in [14, Theorem 2.1] in the Banach space setting.

Theorem 5.6 [14, Theorem 2.1] *Let ϕ, f and G be as in Definition 5.1, and let $\bar{x} \in \operatorname{dom} \phi$ be such that the BCQ is satisfied at $\bar{x}$ and G is $C^{1,1}$ at $\bar{x}$. Then ϕ is primal lower-nice at $\bar{x}$.*

However, to obtain the graph convergence of the subdifferential we need to establish that the the family of functions ϕ_λ is equi primal lower-nice. The first step toward proving this is the following revised version of [21, Lemma 5.2].

Lemma 5.7 *Let $G : \mathbb{R}^p \to \mathbb{R}^n$ be locally $C^{1,1}$ on a compact set $X \subset \mathbb{R}^p$. Then there is a Lipschitz constant L for G' over X such that for all $f \in \Gamma$ and $t > 0$*

$$\left.\begin{array}{r} x_i \in X,\, v_i \in \partial f(G(x_i)) \\ \|v_i\| \leq \sigma t,\;\; i = 1, 2 \end{array}\right\} \Longrightarrow \left\langle G'(x_1)^T v_1 - G'(x_2)^T v_2,\, x_1 - x_2 \right\rangle \geq -t\, \|x_1 - x_2\|^2, \tag{12}$$

where $\sigma := (2L)^{-1}$ is independent of the choice of the function f.

Proof As previously observed, the hypotheses imply that G' has a uniform Lipschitz constant L over X. Hence, the constant $\sigma := (2L)^{-1}$ is well defined. Let $x_1, x_2 \in X$. Given $t > 0$ note that (12) holds trivially if either $\partial f(G(x_1)) \cap \sigma t\mathbb{B}$ or $\partial f(G(x_2)) \cap \sigma t\mathbb{B}$ is empty. So we may assume that $G(x_1), G(x_2) \in \operatorname{dom} \partial f$ with $x_1 \neq x_2$ and $\partial f(G(x_1)) \cap \sigma t\mathbb{B}$ and $\partial f(G(x_2)) \cap \sigma t\mathbb{B}$ both nonempty.

Let $v_i \in \partial f(G(x_i))$ $i = 1, 2$ be such that $\|v_i\| \leq \sigma t$. Monotonicity of the subdifferential, cf. [23, Theorem 12.17], gives

$$\langle v_1 - v_2,\, G(x_1) - G(x_2) \rangle \geq 0,$$

and so

$$\begin{aligned} \left\langle G'(x_2)^T v_1 - G'(x_2)^T v_2,\, x_1 - x_2 \right\rangle &= \left\langle v_1 - v_2,\, G'(x_2)(x_1 - x_2) \right\rangle \\ &\geq \left\langle v_1 - v_2,\, G(x_2) + G'(x_2)(x_1 - x_2) - G(x_1) \right\rangle \\ &\geq \frac{-L}{2}\, \|v_1 - v_2\|\, \|x_1 - x_2\|^2 . \end{aligned}$$

Here, the first inequality is due to the inequaltiy from above and the last one follows from the Cauchy-Schwarz inequality in combination with (10) applied to X. In addition,

$$|\left\langle G'(x_1)^T v_1 - G'(x_2)^T v_1,\ x_1 - x_2\right\rangle| \leq L\,\|v_1\|\,\|x_1 - x_2\|^2\,,$$

and so

$$\begin{aligned}\left\langle G'(x_1)^T v_1 - G'(x_2)^T v_2,\ x_1 - x_2\right\rangle &\geq -L(\|v_1\| + (1/2)\,\|v_1 - v_2\|)\,\|x_1 - v_2\|^2\\ &\geq -2\sigma t L\,\|x_1 - x_2\|^2\,,\end{aligned}$$

which establishes (12). □

We now show that if $\phi = f \circ G$ satisfies the BCQ at a point $\bar{x} \in \operatorname{dom}\phi$, then the the family $\phi_\lambda := (f\#\omega_\lambda) \circ G$ for $\lambda > 0$ as given in Theorem 5.3 is equi primal lower-nice at $\bar{x}$. The proof follows the pattern established in [22, Proposition 2.3].

Proposition 5.8 *Let $f, \omega \in \Gamma$ such that ω is supercoercive and differentiable such that $\omega(0) \leq 0$. As in Theorem 5.3 define $\phi := f \circ G$ and $\phi_\lambda := (f\#\omega_\lambda) \circ G$ for $\lambda > 0$. Let $\bar{x} \in \operatorname{dom}\phi$ be such that BCQ is satisfied at $\bar{x}$ and G is $C^{1,1}$ at $\bar{x}$. Then there exists $\bar{\lambda} > 0$ such that the family of functions $\left\{\phi_\lambda \;\middle|\; \lambda \in (0, \bar{\lambda}]\right\}$ is equi primal lower-nice at $\bar{x}$.*

Proof Due to Theorem 3.8,

$$(f\#\omega_\lambda) \to f \;\text{ and }\; \operatorname{gph}\nabla(f\#\omega_\lambda) \to \operatorname{gph}\partial f. \tag{13}$$

By Theorem 5.3, we know that $\phi_\lambda \xrightarrow{e} \phi$. In addition, the formula for the subdifferential of ϕ and the fact that BCQ is a local property follows from Proposition 5.2. Let $\varepsilon > 0$ be chosen to that G' is Lipschitz on $B_{2\varepsilon}(\bar{x})$ and the BCQ is satisfied at all points in $\operatorname{dom}\phi \cap B_{2\varepsilon}(\bar{x})$. In addition, since all of the functions ϕ_λ are finite-valued for $\lambda > 0$, they all satisfy the BCQ on all of $\mathbb{R}^n$ with $\nabla\phi_\lambda(x) = G'(x)^T\nabla(f\#\omega_\lambda)(G(x))$ for all $\lambda > 0$.

Set $\sigma := (2L)^{-1}$ where L is a Lipschitz constant for G' on $B_{2\varepsilon}(\bar{x})$. By Lemma 5.7, for all $t > 0$ and $\lambda > 0$,

$$\left.\begin{array}{c} x_i \in B_{2\varepsilon}(\bar{x}),\\ v_i = \nabla(f\#\omega_\lambda)(G(x_i)),\\ \|v_i\| \leq \sigma t,\ i = 1, 2\end{array}\right\} \Longrightarrow \left\langle G'(x_1)^T v_1 - G'(x_2)^T v_2,\ x_1 - x_2\right\rangle \geq -t\,\|x_1 - x_2\|^2. \tag{14}$$

We now claim that (14) and the BCQ implies that there exist $\bar{\lambda} > 0,\ T > 0$, and $\mu > 0$ such that

$$\left.\begin{array}{c} x_i \in B_\varepsilon(\bar{x}),\ \lambda \in (0, \bar{\lambda}]\\ u_i = \nabla\phi_\lambda(x_i),\\ t \geq T,\ \|u_i\| \leq \mu t,\ i = 1, 2\end{array}\right\} \Longrightarrow \langle u_1 - u_2,\ x_1 - x_2\rangle \geq -t\,\|x_1 - x_2\|^2. \tag{15}$$

In view of the fact that $\nabla\phi_\lambda(x) = G'(x)^T\nabla(f\#\omega_\lambda)(G(x))$ for all $\lambda > 0$, the implication (14) tells us that this claim will follow if it can be shown that $\bar{\lambda} > 0,\ T > 0$, and $\mu > 0$ can be chosen so that

$$\left.\begin{array}{c} x \in B_\varepsilon(\bar{x}),\ \lambda \in (0, \bar{\lambda}],\\ v = \nabla(f\#\omega_\lambda)(G(x)),\\ t \geq T,\ \left\|G'(x)^T v\right\| \leq \mu t\end{array}\right\} \Longrightarrow \|v\| \leq \sigma t.$$

To this end, suppose that this were not the case, then there exist sequences $\{x_k\} \subset B_\varepsilon(\bar{x})$, $\lambda_k \downarrow 0$, $t_k \uparrow \infty$, $\mu_k \downarrow 0$, and $v_k \in \nabla(f\#\omega_{\lambda_k})(G(x_k))$ such that

$$\left\| G'(x_k)^T v_k \right\| \le \mu_k t_k \quad \text{while} \quad \|v_k\| > \sigma t_k.$$

Due to compactness and (13), we may assume with no loss in generality that there exist vectors $\hat{x} \in B_{2\varepsilon}(\bar{x})$ and $w \in \partial f^\infty(G(\hat{x})) = N_{\operatorname{dom} f}(G(\hat{x}))$ such that $x_k \to \hat{x}$ and $v_k/\|v_k\| \to w$. But then

$$\frac{\left\| G'(x_k)^T v_k \right\|}{\|v_k\|} \le \frac{\mu_k t_k}{\sigma t_k} = \frac{\mu_k}{\sigma} \to 0,$$

which implies that $\left\| G'(\hat{x})^T w \right\| = 0$ contradicting the fact that the BCQ holds at every point of $B_{2\varepsilon}(\bar{x})$. □

We are now able to establish the graph convergence of the subdifferential $\partial\phi_\lambda$ in a neighborhood of a point satisfying the BCQ.

Theorem 5.9 [22, Theorem 2.1 and Proposition 2.3] *Let $f, \omega \in \Gamma$ such that ω is supercoercive and differentiable such that $\omega(0) \le 0$. As in Theorem 5.3 define $\phi := f \circ G$ and $\phi_\lambda := (f\#\omega_\lambda) \circ G$ for $\lambda > 0$. Let $\bar{x} \in \operatorname{dom}\phi$ be such that BCQ is satisfied at $\bar{x}$ and G is $C^{1,1}$ at $\bar{x}$. Then there exists $\varepsilon > 0$ such that*

$$\operatorname{gph} \nabla\phi_\lambda \to \operatorname{gph} \partial\phi \quad \textit{on} \quad B_\varepsilon(\bar{x}),$$

with $\partial\phi(x) = G'(x)^T \partial f(G(x))$ for all $x \in B_\varepsilon(\bar{x})$.

Proof The hypotheses imply that the consequences of Proposition 5.8 hold. Let $\bar{\lambda} > 0$ be such that the family $\Phi_{\bar{\lambda}} := \left\{\phi_\lambda \;\middle|\; \phi \in (0, \bar{\lambda}]\right\}$ is equi primal lower-nice at $\bar{x}$. Let $\eta > 0$. By continuity, $\phi_{\bar{\lambda}}$ is lower bounded on every set of the form $B_\eta(\bar{x})$. By Theorem 5.3, $\phi_\lambda \uparrow \phi$, and so the family $\Phi_{\bar{\lambda}}$ is uniformly bounded below on $B_\eta(\bar{x})$. Define $B_\eta := \operatorname{cl} B_\eta(\bar{x})$, $\hat{\phi} := \phi + \delta_{B_\eta}$, and $\hat{\phi}_\lambda := \phi_\lambda + \delta_{B_\eta}$ for all $\lambda > 0$. Then the family $\hat{\Phi}_{\bar{\lambda}} := \left\{\hat{\phi}_\lambda \;\middle|\; \phi \in (0, \bar{\lambda}]\right\}$ is equi primal lower-nice at $\bar{x}$ and bounded below by a constant. Apply [22, Theorem 2.1 (a)] to the family $\hat{\Phi}_{\bar{\lambda}}$ to obtain the result. □

6 Final Remarks

In this paper we studied smoothing techniques based on infimal convolution in order to construct epi-convergent and gradient consistent smooth approximations to nonsmooth convex and convex-composite functions. The early contributions by Strömberg in the development of these ideas is highlighted and combined with our earlier findings as well as new results. The main tool for proving epi-convergence is Wijsman's Theorem, which facilitates a unified and self-contained presentation.

In Lemma 3.5 we provide characterizations of supercoercivity that guide the basic epi-convergence results in Proposition 3.6 and Theorem 3.8 for convex functions. Theorem 3.8 also reveals the role of the condition $\omega(0) \le 0$ on the smoothing kernel ω to establish the monotonicity of the approximations $f\#\omega_\lambda$. The limitations to relaxing the supercoercivity hypothesis while maintaining epi-convergence are revealed in Proposition 3.9. In Theorem 3.10 we present our main result on gradient consistency in epi-smoothing for convex functions. This result requires the obvious additional hypothesis that the smoothing kernel be differentiable, and a representation for the gradient $\nabla(f\#\omega_\lambda)$ is given. The essential nature of the combined hypotheses of supercoercivity and differentiability on the kernel ω is illustrated in Proposition 3.11.

Due to the central role played by supercoercivity, we present in Section 4 a brief description of the calculus of supercoercive functions before moving on to convex-composite functions.

The key to extending the smoothing results for convex functions to convex composite functions is Rockafellar's basic constraint qualification for convex-composite functions, BCQ (Definition 5.1). The relationship between the subdifferential calculus and the BCQ is presented in Proposition 5.2, and the main result on epi-smoothing , gradient consistency for convex-composite functions is presented in Theorem 5.3.

The final results of the paper are devoted to the local graph convergence of the gradients $\nabla(\phi_\lambda)$ to the subdifferential of the convex-composite function $\phi = f \circ G$. Here we appealed to the approach developed by Poliquin [22] based on primal lower-nice functions and equi primal lower-nice families.

Acknowledgments The authors thank two anonymous referees and the associate editor for their helpful remarks and corrections.

References

1. Attouch, H.: Variational Convergence of Functions and Operators. Applicable Mathematics Series, Pitman Advanced Publishing Program, Boston (1984)
2. Attouch, H.: Viscosity solutions of minimization problems. Epi-convergence and scaling: Séminaire d'Analyse Convexe **22**, 8.1–8.46 (1992)
3. Attouch, H.: Viscosity solutions of minimization problems. SIAM J. Optim. **6**(3), 769–806 (1996)
4. Attouch, H., Azé, D.: Approximation and regularization in Hilbert spaces by Lasry-Lions method. Annales de l'Institute Henri Poincaré (Section C) **10**(3), 289–312 (1993)
5. Bauschke, H.H., Combettes, P.L.: Convex analysis and Monotone Operator Theory in Hilbert Spaces. CMS Books in Mathematics. Springer, New York (2011)
6. Beck, A., Teboulle, M.: Smoothing and first order methods: A unified framework. SIAM J. Optim. **22**(2), 557–580 (2012)
7. Bougeard, M., Penot, J.-P., Pommellet, A.: Towards minimal assumptions for the infimal convolution regularization. Journal of Approximation Theory **64**(3), 245–270 (1991)
8. Burke, J.V., Hoheisel, T.: Epi-smoothing with applications to convex composite functions. SIAM J. Optim. **23**(3), 1457–1479 (2013)
9. Burke, J.V., Hoheisel, T., Kanzow, C.: Gradient consistency for integral-convolution smoothing functions. Set-valued and Variational Analysis **21**(2), 359–376 (2013)
10. Burke, J.V., Hoheisel, T.: A note on epi-convergene of sums under the inf-addition rule. Optimization Online 2015
11. Beer, G.: Topologies on Closed and Closed Convex Sets. Kluwer, Dordrecht (1993)
12. Chen, X.: Smoothing methods for nonsmooth, nonconvex minimization. Math. Program. **134**, 71–99 (2012)
13. Chen, C., Mangasarian, O.L.: A class of smoothing functions for nonlinear and mixed complementarity problems. Comput. Optim. Appl. **5**, 97–113 (1996)
14. Combari, C., Alaoui, A.E., Levy, A., Poliquin, R., Thibault, L.: Convex composite functions in Banach spaces and the primal lower-nice property. Proc. Am. Math. Soc. **126**, 3701–3708 (1998)
15. Goebel, R.: Self-dual smoothing of convex and saddle functions. Journal of Convex Analysis **15**, 179–190 (2008)
16. Hiriart-Urrruty, J.-B., Lemaréchal, C.: Fundamentals of Convex Analysis. Grundlehren Text Editions. Springer, Berlin (2001)
17. Jongen, H.T., Stein, O.: Smoothing by mollifiers. Part I: Semi-infinite optimization. J. Glob. Optim. **41**, 319–334 (2008)

18. Jongen, H.T., Stein, O.: Smoothing by mollifiers. Part II: Nonlinear Optimization. J. Glob. Optim. **41**, 335–350 (2008)
19. Lasry, J.M., Lions, P.L.: A remark on regularization in Hilbert spaces. Israel Journal of Mathematics **55**(3), 257–266 (1986)
20. Moreau, J.J.: Proximité et dualité dans un espace Hilbertien. Bulletin de la Société Mathématique de France **93**, 273–299 (1965)
21. Poliquin, R.: Integration of subdifferentials of nonconvex functions. Nonlinear Analysis, Theory, Methods, and Applications **17**, 385–398 (1991)
22. Poliquin, R.: An extension of Attouch's theorem and its application to second-order epi-differentiation of convexly composite functions. Trans. Am. Math. Soc. **332**, 861–874 (1992)
23. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. A Series of Comprehensive Studies in Mathematics, vol. 317. Springer, Berlin (1998)
24. Strömberg, T.: The operation of infimal convolution. Dissertationes Mathematicae (Rozprawy Matematyczne) **352**, 58 (1996)
25. Strömberg, T.: On regularization in Banach spaces. Arkiv för Matematik **34**(2), 383–406 (1996)
26. Wijsman, R.A.: Convergence of sequences of convex sets, cones and functions. Bull. Am. Math. Soc. **70**, 186–188 (1964)
27. Wijsman, R.A.: Convergence of sequences of convex sets, cones and functions II. Trans. Am. Math. Soc. **123**, 32–45 (1966)