TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 332, Number 2, August 1992

AN EXTENSION OF ATTOUCH'S THEOREM AND ITS APPLICATION TO SECOND-ORDER EPI-DIFFERENTIATION OF CONVEXLY COMPOSITE FUNCTIONS

RENÉ A. POLIQUIN

ABSTRACT. In 1977, Hedy Attouch established that a sequence of (closed proper) convex functions epi-converges to a convex function if and only if the graphs of the subdifferentials converge (in the Mosco sense) to the subdifferential of the limiting function and (roughly speaking) there is a condition that fixes the constant of integration. We show that the theorem is valid if instead one considers functions that are the composition of a closed proper convex function with a twice continuously differentiable mapping (in addition a constraint qualification is imposed). Using Attouch's Theorem, Rockafellar showed that second-order epi-differentiation of a convex function and proto-differentiability of the subdifferential set-valued mapping are equivalent, moreover the subdifferential of one-half the second-order epi-derivative is the proto-derivative of the subdifferential mapping; we will extend this result to the convexly composite setting.

1. INTRODUCTION

In (finite dimensional) optimization theory, epi-convergence is quickly emerging as the natural and "correct" concept of (function) convergence. To recall the concept of epi-convergence we must first discuss set convergence.

For a sequence $\{C_\nu\}_{\nu\in\mathbb{N}}$, of nonempty sets in $\mathbb{R}^n$, the $\liminf_{\nu\to\infty} C_\nu$ consists of all limits points of sequences $\{\omega_\nu\}_{\nu\in\mathbb{N}}$ selected with $\omega_\nu \in C_\nu$, while the $\limsup_{\nu\to\infty} C_\nu$ consists of all cluster points of such sequences. We say that C_ν converge, in the *Painlevé-Kuratowski* sense, to C if $\limsup_{\nu\to\infty} C_\nu = \liminf_{\nu\to\infty} C_\nu = C$ (for convergence of sets that depend on a continuous parameter one makes the obvious extension). In finite dimensions, which is the setting of this paper, Painlevé-Kuratowski convergence is the same as Mosco convergence and Attouch-Wets convergence; see [1, 2, 9, 26 and 27].

A sequence $\{f_\nu\}_{\nu\in\mathbb{N}}$ of extended real-valued functions on $\mathbb{R}^n$ (i.e., they may take the value $+\infty$) is said to *epi-converge* to f, denoted by $f_\nu \xrightarrow{e} f$, if the epigraphs $\operatorname{epi} f_\nu = \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : f_\nu(x) \le \alpha\}$ converge (in the above sense) to epi f. We say that f_ν epi-converges to f on $S \subset \mathbb{R}^n$, if $f_\nu + \delta_S$

Received by the editors June 26, 1990.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 49A52, 58C06; Secondary 58C20, 49A50.
Key words and phrases. Attouch's Theorem, convex analysis, proto-differentiation, epi-differentiation, composite functions, quadratic-conjugate, proximal subgradients.
This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) under grant OGP41983.

epi-converges to $f + \delta_S$, where δ_S is the indicator of S, i.e. the function with value 0 on S, infinity otherwise. For more on epi-convergence and its importance in optimization see [1 and 27].

In 1977, Hedy Attouch [1] showed that for convex functions, epi-convergence is the appropriate concept of convergence if one is interested in convergence of their subdifferentials. For a convex function f the subdifferential to f at $\bar{x}$, is the set $\partial f(\bar{x}) = \{u: f(x) \geq f(\bar{x}) + \langle u, x - \bar{x}\rangle \text{ for all } x\}$, where $\langle\cdot, \cdot\rangle$ is the usual dot product; for an arbitrary function, $\partial f(x)$ will denote the set of (Clarke) *generalized subgradients* to f at $\bar{x}$ (in the convex case, the subdifferential is the set of generalized subgradients), cf. [3 and 20].

Attouch showed that a sequence $\{f_\nu\}_{\nu\in\mathbb{N}}$ of lower semicontinuous proper convex functions epi-converge to f if and only if the sets $\{\text{gph } \partial f_\nu\}_{\nu\in\mathbb{N}}$ converge to gph ∂f (where gph $\partial g = \{(x, u): u \in \partial g(x)\}$) and there exist $\{(x_\nu, u_\nu)\}_{\nu\in\mathbb{N}}$ in gph ∂f_ν converging to (x, u) with $u \in \partial f(x)$ and $f_\nu(x_\nu)$ converging to $f(x)$. The pointwise convergence condition in Attouch's Theorem can be interpreted as a condition that "fixes" the constant of integration. Note that Attouch's Theorem implicitly states that a convex function is determined, up to an additive constant, by its subdifferential (a fact that was well known at the time; see [19]). For a proof of this theorem in the finite dimensional setting we refer to [27, Theorem 3], in the case of a reflexive Banach space we refer to [1, Theorem 3.66].

Since a convex function f attains its minima on $\mathbb{R}^n$ at $\bar{x}$ if and only if $0 \in \partial f(\bar{x})$, one senses the importance of convergence of subdifferentials (and therefore of epi-convergence) when one is interested in convergence of minima. For an arbitrary function f, $0 \in \partial f(\bar{x})$ represents a first-order necessary condition for optimality.

The extension of Attouch's Theorem that we present in this paper involves *equi primal lower-nice* functions; these functions were first introduced in [11], but we postpone the definition until §2. The main example of a primal lower-nice function is the composition of a lower semicontinuous proper convex function with a twice continuously differentiable mapping, in addition a *basic constraint qualification* (cf. Definition 2.2) must be verified. It was pointed out by Rockafellar in [17] that most common problems of optimization that arise in practice can be reformulated using convexly composite functions, this fact serves to establish the importance of an extension of Attouch's Theorem that covers the convexly composite functions. In Proposition 2.3 we give conditions under which a family of convexly composite functions is equi primal-lower-nice.

As in the convex case, the primal-lower-nice functions are determined, up to an additive constant, by their (Clarke) generalized subgradients; see [11]. For these functions the generalized subgradients and the *proximal* subgradients agree, enabling us to employ techniques from proximal analysis. In particular, the quadratic conjugate function introduced in [10] as a tool for studying proximal subgradients will be used extensively in this paper; see §2. For a discussion of proximal subgradients, see [4, 6, 10, 21], and §2.

Our extension of Attouch's Theorem (Theorem 2.1) contains a subtle difference compared to the original theorem for convex functions (as stated in [1, Theorem 3.66]). In Theorem 2.1, we do not assume that gph ∂f_ν converges to the graph of a subgradient set-valued mapping, we merely assume that the sequence of graphs converges, we then show that the limit is the graph of a subgradient set-valued mapping (Attouch's Theorem could also be stated in this manner, because the limit of subdifferential mappings is again a subdifferential mapping; see [1, Corollary 3.65]). The limiting subgradients and the pointwise convergence condition then determine the epi-limit of the sequence of functions. This extra degree of generality in our extension is motivated by the study of epi-limits of second-order difference quotients which we now discuss.

For the main application of Theorem 2.1, we turn to second-order epi-differentiation of functions and proto-differentiation of set-valued mappings. We adumbrate here these two concepts: A function f is *twice epi-differentiable* at x relative to v if the second-order difference quotients

$$\varphi_{x,v,t}(\xi) = \frac{f(x+t\xi) - f(x) - t\langle v, \xi\rangle}{(\frac{1}{2})t^2}, \quad \text{for } t > 0$$

epi-converge as t goes to 0 to a proper function. The epi-limit of these second-order difference quotients is the *second-order epi-derivative* and is denoted by $f''_{x,v}$; see [17]. Second-order epi-derivatives have been used by Rockafellar to establish necessary and sufficient conditions for optimality that mimic the classical ones; see [18] for details. In [17] it is shown that when the basic constraint qualification holds, the composition of a *piecewise linear-quadratic* convex function with a twice continuously differentiable mapping is twice epi-differentiable relative to any subgradient; this fact can be used to gauge the importance of second-order epi-differentiation in optimization theory because most optimization problems that arise in practice can be reformulated using these special convexly composite functions. For more on second-order epi-differentiation we refer to [5, 17, 18], and to Poliquin and Rockafellar [15], where a calculus of second-order epi-derivatives is presented for convexly composite functions. We now turn to proto-differentiation.

A set-valued mapping $\Gamma\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *proto-differentiable* at x relative to v in $\Gamma(x)$ if Γ_t, where $\Gamma_t = t^{-1}(\operatorname{gph}\Gamma - (x, v))$, converge as t goes to 0. The *proto-derivative* of Γ at x relative to v, denoted by $\Gamma'_{x,v}$, is the set-valued mapping whose graph is the limiting set; see [13, 14, 22 and 23]. As in the case of the derivative of a function, estimates of set-valued mappings can be obtained using proto-derivatives; see [22]. Many of the set-valued encountered in optimization, e.g. set-valued mappings expressing feasibility or optimality, are proto-differentiable; see [22].

When we apply the proto-derivative theory to a subgradient set-valued mapping one obtains a generalized second-order differentiation theory. In the convex case the link between the second-order epi-derivative and the proto-derivative of the subdifferential mapping is well understood. Rockafellar [24] established that a lower semicontinuous proper convex function f on $\mathbb{R}^n$ is twice epi-differentiable at x relative to v, a subgradient to f at x, if and only if the subdifferential mapping is proto-differentiable at x relative to v. Moreover the subdifferential of one-half the second-order epi-derivative is the proto-derivative of the subdifferential mapping at x relative to v; see [24]. The main tool in the proof of this result is Attouch's Theorem. This type of formula (relating the subgradients of the second-order epi-derivative to the proto-derivative of the subgradient mapping) has tremendous applications to the study of perturbed optimal solutions in parametric optimization; for a discussion on this subject we refer to [14, 16, and 23].

Our extension of Attouch's Theorem, enables us to deduce that for convexly composite functions, there is equivalence between second-order epi-differentiation of the function and proto-differentiation of the subgradients, and the same formula holds; this is carried out in Theorem 2.2. In [12], the above result was presented for the special piecewise linear-quadratic case; at that time this extension of Attouch's Theorem was not available, and the specific formula for the second-order epi-derivative had to be used.

2. Main results

The proof of our extension of Attouch's Theorem relies heavily on the *quadratic conjugate function* introduced in [10].

Recall that for an extended real-valued function on $\mathbb{R}^n$ (i.e., $f\colon \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$), proper (the effective domain $\operatorname{dom} f := \{x : f(x) < \infty\}$ is nonempty) and lower semicontinuous ($\operatorname{epi} f$ is a closed set), the quadratic conjugate of f, with parameter $t > 0$, is the function defined by

$$h_f(z, t) = \sup_{x\in\mathbb{R}^n}\{\langle z, x\rangle - (t/2)\|x\|^2 - f(x)\}, \tag{2.1}$$

where $\langle z, x\rangle$ is the usual dot product, and $\|\cdot\|$ the Euclidean norm. The set of points where the supremum is attained will be denoted by $\operatorname{argmax} h_f(z, t)$ i.e.

$$\operatorname{argmax} h_f(z, t) = \{x : h_f(z, t) = \langle z, x\rangle - (t/2)\|x\|^2 - f(x)\}. \tag{2.2}$$

For all z and t with $h_f(z, t)$ finite, we have

$$f(x) \geq \langle z, x\rangle - (t/2)\|x\|^2 - h_f(z, t) \tag{2.3}$$

hence $h_f(z, t)$ is the smallest α such that $\langle z, x\rangle - (t/2)\|x\|^2 - \alpha \leq f(x)$ for all x. This is reminiscent of the convex conjugate function and hence the name; recall that for a function g, the convex conjugate

$$g^*(z) = \sup_{x\in\mathbb{R}^n}\{\langle z, x\rangle - g(x)\}$$

is the smallest β such that $g(x) \geq \langle z, x\rangle - \beta$ for all x. This quadratic conjugate function is closely related to the study of *proximal subgradients.*

Recall that a vector u is a proximal subgradient to f at x, denoted by $u \in \partial_p f(x)$, if for some $t > 0$

$$f(x') \geq f(x) + \langle u, x' - x\rangle - (t/2)\|x' - x\|^2 \tag{2.4}$$

in a neighborhood of x ($\partial_p f(x) \subset \partial f(x)$). If we assume that f is bounded below by a quadratic function, then we can replace sup by max in (2.1), we can assume that (2.4) holds for all x', and that for some $T > 0$, $\operatorname{argmax} h_f(z, t)$ is nonempty for all $t \geq T$.

We summarize, from [10, §3], some important facts illustrating the relationship between the quadratic conjugate and proximal subgradients; these properties essentially mimic the relationship between a convex function and its convex conjugate.

(2.5) If $x \in \operatorname{argmax} h_f(z, t)$, then $z - tx \in \partial_p f(x)$.

(2.6) If $u \in \partial_p f(x)$, then for t big enough, $\operatorname{argmax} h_f(u + tx, t) = \{x\}$.

If f is bounded below by a quadratic then there exist $T > 0$ such that for all $t > T$

$$\partial h_f(z, t) = \operatorname{co}\left\{\left(x, -\frac{\|x\|^2}{2}\right) : x \in \operatorname{argmax} h_f(z, t)\right\}, \tag{2.7}$$

moreover, and this is the main reason for employing a quadratic function of two variables,

$$\text{if } (x, -(1/2)\|x\|^2) \in \partial h_f(z, t), \text{ then } z - tx \in \partial_p f(x). \tag{2.8}$$

Just as a convex function can be recovered by taking the conjugate of its conjugate the same (general result) is true for the quadratic conjugate, i.e.

$$f(x) = \sup_{\substack{z \in \mathbf{R}^n \\ t \geq T}} \{\langle z, x\rangle - (t/2)\|x\|^2 - h(z, t)\}. \tag{2.9}$$

The quadratic conjugate is closely related to the Moreau-Yosida approximates; this fact is crucial in the proofs of Proposition 2.1, and Theorem 2.1. The Moreau-Yosida approximate of f, with parameter $\lambda > 0$, is given by

$$f^\lambda(x) = \inf_{\xi \in \mathbf{R}^n} \left\{ f(\xi) + \left(\frac{1}{2\lambda}\right) \|x - \xi\|^2 \right\}. \tag{2.10}$$

In the terminology of convex analysis, f^λ is the infimal convolution of f with $(\frac{1}{\lambda})J(x)$, where $J(x) = \frac{\|x\|^2}{2}$; see [19]. The connection between the quadratic conjugate and the Moreau-Yosida approximates is the following:

$$\begin{aligned} -f^\lambda(x) &= \sup_\xi \left\{ -f(\xi) - \left(\frac{1}{2\lambda}\right)\|x\|^2 + \left(\frac{1}{\lambda}\right)\langle x, \xi\rangle - \frac{1}{2\lambda}\|\xi\|^2 \right\} \\ &= \sup_\xi \left\{ \langle \lambda^{-1}x, \xi\rangle - \left(\frac{1}{2\lambda}\right)\|\xi\|^2 - f(\xi) \right\} - \left(\frac{1}{2\lambda}\right)\|x\|^2 \\ &= h_f(\lambda^{-1}x, \lambda^{-1}) - \left(\frac{1}{\lambda}\right)\left(\frac{\|x\|^2}{2}\right). \end{aligned} \tag{2.11}$$

In the next proposition we record the fact that epi-convergence of a sequence of lower semicontinuous functions, uniformly minorized by a quadratic, is equivalent to epi-convergence of the sequence of quadratic conjugates. This proposition is essentially obtained by combining [1, Corollary 2.67 and 2, Proposition 4.2]. In the proposition we denote by h_ν and h the quadratic conjugates of f_ν and f with parameter t.

Proposition 2.1. *Suppose f_ν, $\nu = 1, 2, \ldots$, are lower semicontinuous extended real-valued functions on $\mathbb{R}^n$ uniformly minorized by a quadratic. Then, there exists $T > 0$ such that the following are equivalent*

(a) $f_\nu \xrightarrow{e} f$ *(epi-convergence),*

(b) $h_\nu \xrightarrow{e} h$ *or* $\mathbb{R}^n \times [T, \infty)$,

(c) $\forall t \geq T$, $h_\nu(\cdot, t) \xrightarrow{pt\omega} h(\cdot, t)$,

(d) $\forall t \geq T$, $h_\nu(\cdot, t) \xrightarrow{e} h(\cdot, t)$.

Proof. Since f_ν, $\nu = 1, 2, \ldots$, are uniformly minorized by a quadratic, there exists $T > 0$, such that $\tilde{f}_\nu = f_\nu + (\frac{T}{2})\|x\|^2$, $\nu = 1, 2, \ldots$, are nonnegative. Let $\tilde{f} = f + (\frac{T}{2})\|x\|^2$. It is easy to see that

$$f_\nu \xrightarrow{e} f \Leftrightarrow \tilde{f}_\nu \xrightarrow{e} \tilde{f}. \tag{2.12}$$

By [2, Proposition 4.2], $\tilde{f}_\nu \xrightarrow{e} \tilde{f} \Rightarrow \forall \lambda > 0$, $\tilde{f}_\nu^\lambda \xrightarrow{e} \tilde{f}^\lambda$ (in [2, Proposition 4.2] let $g^\nu = \tilde{f}_\nu$ and $f^\nu(\cdot) = (1/2\lambda)\|\cdot\|^2$). Hence, by (2.11) and (2.12), $f_\nu \xrightarrow{e} f \Rightarrow \forall t \geq T$, $h_\nu(\cdot, t) \xrightarrow{e} h(\cdot, t)$, this shows that (a) $\Rightarrow$ (d).

To prove (c) $\Rightarrow$ (a) simply invoke [1, Corollary 2.65]. Indeed, under assumption (c), for every $\lambda > 0$ small enough, the sequence of Moreau-Yosida approximate of parameter λ converges. By [1, (2.166)] we have the desired result.

To complete the proof of the proposition use the fact that for finite convex functions on $\mathbb{R}^n$ (each $h_f(\cdot, t)$ is a finite convex function) epi-convergence is equivalent to pointwise convergence (cf. [27, Corollaries 4 and 5]); this gives (c) $\Leftrightarrow$ (d). By, [27, Corollary 4], we also have (c) $\Rightarrow$ (b), because the effective domain of h is nonempty. By [27, Corollary 5] (b) $\Rightarrow$ (c) on $\mathbb{R}^n \times (T, \infty)$, but by taking T bigger if necessary we have the desired result. □

The extension of Attouch's Theorem that we present involves *primal lower-nice* functions; these functions were first introduced in [11, §3]. For technical reasons we will use throughout this paper a different characterization than the one found in [11, Definition 3.1] (we show in Proposition 2.2 that they are equivalent).

Definition 2.1. A lower semicontinuous function $f: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is said to be primal lower-nice at $\bar{x}$ if there exist $\sigma > 0$, $\mu > 0$ and $T > 0$ such that if $t \geq T$, $\|u\| \leq \mu t$, $\|x - \bar{x}\| < \sigma$ and $u \in \partial_p f(x)$ then

$$f(x') \geq f(x) + \langle u, x' - x \rangle - (t/2)\|x - x'\|^2 \quad \text{if } \|x' - x\| < \sigma. \tag{2.13}$$

The advantage of dealing with primal lower-nice functions is that given a proximal subgradient u one knows the "steepness" of the quadratic needed to "realize" u. Another advantage of dealing with primal lower-nice functions is that *all generalized subgradients are proximal subgradients;* see [11, Proposition 3.5]. We now give an equivalent characterization of primal lower-nice functions: we show that the subgradients of primal lower-nice functions are "t-monotone" i.e. $\partial f + tI$ is a monotone set-valued mapping (Γ is monotone if whenever $u_i \in \Gamma(x_i)$, $i = 1, 2$, then $\langle u_1 - u_2, x_1 - x_2 \rangle \geq 0$).

Proposition 2.2. *Assume that $f: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is lower semicontinuous and bounded below by a quadratic. The following are equivalent.*

(a) *The function f is primal lower-nice at $\bar{x}$.*

(b) *(t-monotonicity) There exist $\tilde{\sigma} > 0$, $\tilde{\mu} > 0$ and $\widetilde{T} > 0$ such that if $t \geq \widetilde{T}$, $\|u_i\| \leq \tilde{\mu} t$, $\|x_i - \bar{x}\| < \tilde{\sigma}$ and $u_i \in \partial_p f(x_i)$, $i = 1, 2$, then*

$$\langle u_1 - u_2, \ x_1 - x_2 \rangle \geq -t\|x_1 - x_2\|^2,$$

i.e. $\langle u_1 + tx_1 - (u_2 + tx_2), x_1 - x_2 \rangle \geq 0.$

Proof. It follows from [11, Corollary 3.4] that (b) $\Rightarrow$ (a).

Assume that f is primal lower-nice at $\bar{x}$, with constants σ, μ and T given by Definition 2.1. Let $t \geq T$, $\|u_i\| \leq \mu t$, $\|x_i - \bar{x}\| < \sigma$ and $u_i \in \partial_p f(x_i)$, $i = 1, 2$. Since f is primal lower-nice at $\bar{x}$ and bounded below by a quadratic we may assume from (2.7) (by choosing T bigger if necessary) that

$$\nabla h_f(z_i, t) = (x_i, -(1/2)\|x_i\|^2), \qquad i = 1, 2,$$

where $z_i = u_i + tx_i$. Because the subdifferential of a convex function is monotone (cf. [19]) it follows that

$$\langle ((z_1, t) - (z_2, t)), (x_1 - x_2, -(1/2)\|x_1\|^2 - (-(1/2)\|x_2\|^2)))\rangle \geq 0.$$

By simplifying we obtain the desired result, i.e. $\langle z_1 - z_2, x_1 - x_2\rangle \geq 0$. □

Obviously, convex functions are primal lower-nice (the subdifferential of a convex function is monotone).

To provide other examples of primal lower-nice functions, we need to first recall the definition of the basic constraint qualification (b.c.q.).

Definition 2.2. The basic constraint qualification (b.c.q.) holds at $\bar{x} \in \operatorname{dom}(g \circ F)$, where $g : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ is a proper lower semicontinuous convex and $F : \mathbb{R}^n \to \mathbb{R}^m$ is of class C^2, if the only $y \in N(F(\bar{x})|\operatorname{dom} g)$ with $\nabla F(\bar{x})^* y = 0$ is $y = 0$, where $N(F(\bar{x})|\operatorname{dom} g)$ is the normal cone to $\operatorname{dom} g$ at $F(\bar{x})$; see [19].

In [11] we showed that when the basic constraint qualification holds at $\bar{x}$, then the composition of a lower semicontinuous proper convex function $g: \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ with a twice continuously differentiable mapping $F: \mathbb{R}^n \to \mathbb{R}^m$ is primal lower-nice at $\bar{x}$.

We are now ready to state our main theorem. In the theorem, we employ a family of *equi* primal-lower-nice functions at $\bar{x}$, i.e. the σ, μ and T in Definition 2.1 work for all functions.

Theorem 2.1 (Extension of Attouch's Theorem).

Assume $f_\nu : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $\nu = 1, 2, \dots$, are lower semicontinuous uniformly minorized by a quadratic and equi primal-lower-nice at $\bar{x}$ with constants $\sigma > 0$, $\mu > 0$ and $T > 0$ (see Definition 2.1).

(a) *If $f_\nu \xrightarrow{e} f$ on $\|x-\bar{x}\| < \sigma$, then $\operatorname{gph} \partial f_\nu \to \operatorname{gph} \partial f$ on $\|x-\bar{x}\| < \sigma$ and there exist $\operatorname{gph} \partial f_\nu \ni (x_\nu, u_\nu) \to (\tilde{x}, \tilde{u}) \in \operatorname{gph} \partial f$ with $f_\nu(x_\nu) \to f(\tilde{x})$ and $\|\tilde{x} - \bar{x}\| < \sigma$.*

(b) *If*

(b′) *$\operatorname{gph} \partial f_\nu$ converges on $\|x - \bar{x}\| < \sigma$ and*

(b″) *there exist $\operatorname{gph} \partial f_\nu \ni (x_\nu, u_\nu) \to (\tilde{x}, \tilde{u})$ with $\|\tilde{x} - \bar{x}\| < \sigma$ and $f_\nu(x_\nu)$ converging,*

then there exist $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ such that $\operatorname{gph} \partial f_\nu \to \operatorname{gph} \partial f$, $f_\nu \xrightarrow{e} f$ on $\|x - \bar{x}\| < \sigma$, $(\tilde{x}, \tilde{u}) \in \operatorname{gph} \partial f$ and $f_\nu(x_\nu) \to f(\tilde{x})$.

Moreover, in both (a) *and* (b) *the function f is primal lower-nice at $\bar{x}$ with constants $\sigma > 0$, $\mu > 0$ and $T > 0$.*

To prove Theorem 2.1 we will make use of the following lemma; the proof of which was provided by Professor R. T. Rockafellar.

Lemma 2.1. *Let A_ν, $\nu = 1, 2, \dots$, and A be set-valued mapping on $\mathbb{R}^n$. Assume that $\operatorname{gph} A = \liminf_{\nu\to\infty} \operatorname{gph} A_\nu$, and that A_ν are maximal monotone. Under these assumptions, A has convex values i.e. for each x, the set $A(x)$ is convex.*

Proof. First notice that if a sequence D_ν of maximal monotone set-valued mappings graph converges to D (i.e. their graphs converge), with the domain of D nonempty (i.e. for some x, $D(x)$ is nonempty), then D is maximal monotone. Indeed D is obviously monotone, and graph convergence of D_ν to D is equivalent to graph convergence of $(I+D_\nu)^{-1}$ to $(I+D)^{-1}$, where I is the identity. Because D_ν and D are monotone, the mappings $(I+D_\nu)^{-1}$ and $(I+D)^{-1}$, are Lipschitz with modulus 1 (in particular single-valued), and therefore graph convergence is equivalent to pointwise convergence. By Minty's Theorem (see [1, Theorem 3.54]), the maximality of a monotone mapping D is equivalent to $(I+D)^{-1}$ having the whole space as its domain. Because $(I+D_\nu)^{-1}$ are Lipschitz of modulus one with domain the whole space, the domain of $(I+D)^{-1}$, is either the whole space or is empty; by our assumptions D is therefore maximal monotone.

To complete the proof of the lemma we use the fact that for any sets C_ν, $\liminf_{\nu\to\infty} C_\nu$ is the intersection of all sets B, such that some subsequence converges to B; see [25]. Using this fact, we have that $\operatorname{gph} A$ is the intersection of sets $\operatorname{gph} B$ such that a subsequence of A_ν graph converges to B. Because B is either maximal monotone or has empty domain, we conclude that for any x, $B(x)$ is a convex set. The set $A(x)$ is then convex because it is the intersection of convex sets. □

Proof of Theorem 2.1. We will assume without loss of generality that $\bar{x}=0$, that f_ν, $\nu=1,2,\ldots$, are nonnegative and that $f_\nu(x)=+\infty$ if $\|x\|\geq\sigma$. To do so, just translate to the origin, and add a C^∞ function ψ that has effective domain $\|x\|<\sigma$ and that "blows-up" on the boundary of its effective domain. It follows that $\partial(f_\nu+\psi)(x)=\partial f_\nu(x)+\nabla\psi(x)$, for all x with $\|x\|<\sigma$ and that $f_\nu \xrightarrow{e} f$ on $\|x\|<\sigma$ if and only if $(f_\nu+\psi)\xrightarrow{e}(f+\psi)$.

As in Proposition 2.1, let h_ν be the quadratic conjugate function of f_ν, $\nu=1,2,\ldots$.

Proof of part (a). Let h be the quadratic conjugate function of f.

By choosing T bigger if necessary, we have by Proposition 2.1 that

$$f_\nu \xrightarrow{e} f \Leftrightarrow h_\nu \xrightarrow{e} h \quad \text{on } \mathbb{R}^n\times[T,\infty). \tag{2.14}$$

By Attouch's Theorem the following are equivalent:

(A1) $h_\nu \xrightarrow{e} h$ on $\{(z,t): t\geq T\}$.

(A2) $\operatorname{gph}\partial h_\nu \to \operatorname{gph}\partial h$ on $\{(z,t): t\geq T\}$ and there exist $(z_\nu,t_\nu)\to(\tilde{z},\tilde{t})$, $(x_\nu,a_\nu)\to(\tilde{x},\tilde{a})$ with $\|\tilde{x}\|<\sigma$, $\tilde{t}\geq T$, $(x_\nu,a_\nu)\in\partial h_\nu(z_\nu,t_\nu)$, $h(z_\nu,t_\nu)\to h(\tilde{z},\tilde{t})$, and $(\tilde{x},\tilde{a})\in\partial h(\tilde{z},\tilde{t})$.

We first show that $\limsup_{\nu\to\infty}\operatorname{gph}\partial f_\nu\subset\liminf_{\nu\to\infty}\operatorname{gph}\partial f_\nu$. To do so let $(x,u)\in\limsup_{\nu\to\infty}\operatorname{gph}\partial f_\nu$ i.e. $\exists x_{\nu_i}\to x$, $u_{\nu_i}\to u$ with $\nu_i\to\infty$ and $u_{\nu_i}\in\partial f_{\nu_i}(x_{\nu_i})$.

By the equi primal-lower-nice property, there exist $t\geq T$ such that

$$f_{\nu_i}(x')\geq f_{\nu_i}(x_{\nu_i})+\langle u_{\nu_i}, x'-x_{\nu_i}\rangle-(t/2)\|x'-x_{\nu_i}\|^2,$$

i.e. $(x_{\nu_i}, -(1/2)\|x_{\nu_i}\|^2)\in\partial h_{\nu_i}(u_{\nu_i}+tx_{\nu_i}, t)$; see (2.7). This implies that

$$\begin{aligned}\big((u+tx,t),(x,-(1/2)\|x\|^2)\big)\in{}&\limsup_{\nu\to\infty}\operatorname{gph}\partial h_\nu\\ &=\liminf_{\nu\to\infty}\operatorname{gph}\partial h_\nu\quad\text{on }\{(z,t):t\geq T\}\\ &=\operatorname{gph}\partial h\quad\text{on }\{(z,t):t\geq T\}.\end{aligned}$$

By (2.8) we conclude that $u \in \partial_p f(x)$, i.e. we have shown that

$$\limsup_{\nu\to\infty} \operatorname{gph} \partial f_\nu \subset \operatorname{gph} \partial_p f.$$

By the nature of the lim inf, we have

$$\forall \nu_n \to \infty, \ \exists (z_{\nu_n}, t_{\nu_n}) \to (u + tx, t), \quad (x_{\nu_n}, a_{\nu_n}) \to \big(x, -(1/2)\|x\|^2\big)$$

with $(x_{\nu_n}, a_{\nu_n}) \in \partial h_{\nu_n}(z_{\nu_n}, t_{\nu_n}) = \operatorname{co}\{(y, -(1/2)\|y\|^2) : y \in \operatorname{argmax} h_{\nu_n}(z_{\nu_n}, t_{\nu_n})\}$. Since

$$(x_{\nu_n}, a_{\nu_n}) \to (x, -(1/2)\|x\|^2)$$

we conclude that there exist $y_{\nu_n} \in \operatorname{argmax} h_{\nu_n}(z_{\nu_n}, t_{\nu_n})$ with $\big(y_{\nu_n}, -(1/2)\|y_{\nu_n}\|^2\big) \to \big(x, -(1/2)\|x\|^2\big)$; see [10, Lemma 3.3].

In other words, $\forall \nu_n \to \infty$, $\exists y_{\nu_n} \to x$ with

$$(z_{\nu_n} - ty_{\nu_n}) \in \partial f_{\nu_n}(y_{\nu_n}) \quad \text{and} \quad (z_{\nu_n} - ty_{\nu_n}) \to (u + tx) - tx = u;$$

hence $(x, u) \in \liminf_{\nu\to\infty} \operatorname{gph} \partial f_\nu$. Notice that we have also shown that $\operatorname{gph} \partial_p f \subset \liminf_{\nu\to\infty} \operatorname{gph} \partial f_\nu$. Because we have shown that $\operatorname{gph} \partial f_\nu \to \operatorname{gph} \partial_p f$ we conclude that f is primal lower-nice at the origin, and that $\operatorname{gph} \partial_p f = \operatorname{gph} \partial f$.

To complete the proof of part (a) we need only show that the pointwise convergence property holds. In (A2), because of (2.7) and [10, Lemma 3.3], we may assume without loss of generality that $\tilde{a} = -(1/2)\|\tilde{x}\|^2$ and that $a_\nu = -(1/2)\|x_\nu\|^2$. This implies that $\operatorname{gph} \partial f_\nu \ni (x_\nu, z_\nu - t_\nu x_\nu) \to (\tilde{x}, \tilde{z} - \tilde{t}\tilde{x}) \in \operatorname{gph} \partial f$ and $f_\nu(x_\nu) = \langle z_\nu, x_\nu\rangle - (t_\nu/2)\|x_\nu\|^2 - h_\nu(z_\nu, t_\nu)$ converges to $\langle \tilde{z}, \tilde{x}\rangle - (\tilde{t}/2)\|\tilde{x}\|^2 - h(\tilde{z}, \tilde{t})$. By (2.8) the previous quantity equals $f(\tilde{x})$. This completes the proof of (a).

Proof of part (b). Let $0 < \bar{\sigma} < \sigma$ such that $2\bar{\sigma} < \mu$. We will show that f_ν epi-converge to f on this $\bar{\sigma}$ neighborhood of the origin (if we can prove (b) on some neighborhood of the origin then we will have shown that (b) holds on $\|x\| < \sigma$, this is because f_ν, $\nu = 1, 2, \ldots$, are uniformly equi primal-lower-nice at all x with $\|x\| < \sigma$).

We first show that $\operatorname{gph} \partial h_\nu$ converges on $\{(z, t), t \geq T, \|z\| \leq \bar{\sigma} t\}$.

Consider

$$\big((z, t), (x, a)\big) \in \limsup_{\nu\to\infty} \operatorname{gph} \partial h_\nu, \quad \text{with } t \geq T, \ \|z\| < \bar{\sigma} t.$$

There exist

$$\big(z_{\nu_n}, t_{\nu_n}\big) \to (z, t), \ \big(x_{\nu_n}, a_{\nu_n}\big) \to (x, a) \quad \text{with } \big(x_{\nu_n}, a_{\nu_n}\big) \in \partial h_{\nu_n}\big(z_{\nu_n}, t_{\nu_n}\big).$$

By Carathéodory's Theorem (cf. [19]), there exist $y^i_{\nu_n} \in \operatorname{argmax} h_{\nu_n}\big(z_{\nu_n}, t_n\big)$, $\lambda^i_{\nu_n} \geq 0$, $i = 1, 2, \ldots, n+2$, with $\sum \lambda^i_{\nu_n} = 1$ and

$$\big(x_{\nu_n}, a_{\nu_n}\big) = \sum_1^{n+2} \lambda^i_{\nu_n}\big(y^i_{\nu_n}, -(1/2)\|y^i_{\nu_n}\|^2\big).$$

We may assume that $\lambda^i_{\nu_n} \to \lambda_i$ and that $y^i_{\nu_n} \to y_i$. It follows that $(x, a) = \sum_1^{n+2} \lambda_i\big(y_i, -(1/2)\|y_i\|^2\big)$. We have $\big(z_{\nu_n} - t_{\nu_n} y^i_{\nu_n}\big) \in \partial f_{\nu_n}\big(y^i_{\nu_n}\big)$, $i = 1, 2, \ldots, n+2$. Let $u_i = z - ty_i$, we have

$$(y_i, u_i) \in \limsup_{\nu\to\infty} \operatorname{gph} \partial f_\nu = \liminf_{\nu\to\infty} \operatorname{gph} \partial f_\nu = \operatorname{gph} \partial f.$$

Since $\|u_i\| = \|z - ty_i\| \leq \|z\| + \bar{\sigma}t \leq 2\bar{\sigma}t < \mu t$ we have, by the equi primal-lower-nice property, that $\big((z, t), (y_i, -(1/2)\|y_i\|^2)\big) \in \liminf_{\nu\to\infty} \operatorname{gph} \partial h_\nu$.

To summarize, given

$$\big((z, t), (x, a)\big) \in \limsup_{\nu\to\infty} \operatorname{gph} \partial h_\nu, \quad \text{with } t \geq T, \ \|z\| < \bar{\sigma}t.$$

there exist (y_i, λ_i), $i = 1, 2, \dots$, such that $\sum \lambda_i = 1$, $\lambda_i \geq 0$,

$$\big((z, t), (x, a)\big) = \sum_1^{n+2} \lambda_i \big((z, t), (y_i, -(1/2)\|y_i\|^2)\big) \tag{2.15}$$

with $\big((z, t), (y_i, -(1/2)\|y_i\|^2)\big) \in \liminf_{\nu\to\infty} \operatorname{gph} \partial h_\nu$. By Lemma 2.1 we conclude that $\big((z, t), (x, a)\big) \in \liminf_{\nu\to\infty} \operatorname{gph} \partial h_\nu$.

Choose t big enough so that

$$f_\nu(x_\nu) = \langle u_\nu + tx_\nu, t\rangle - (t/2)\|x_\nu\|^2 - h_\nu(u_\nu + tx_\nu, t)$$

i.e.

$$h_\nu(u_\nu + tx_\nu, t) = \langle u_\nu + tx_\nu, t\rangle - (t/2)\|x_\nu\|^2 - f_\nu(x_\nu). \tag{2.16}$$

By Attouch's Theorem there exists h convex lower semicontinous such that $h_\nu + \delta_S \xrightarrow{e} h$, where $S = \{(z, t) | t \geq T, \|z\| \leq \bar{\sigma}t\}$, and we can choose h such that $h(\tilde{u} + t\tilde{x}, t)$ is the limit on the left-hand side of (2.16). Because f_ν are uniformly bounded below by a quadratic, the values $h_\nu(z, t)$ are bounded above (see (2.1)), for all $(z, t) \in S$; this implies that the effective domain of h is S (i.e. h is finite valued on S).

Claim 1. If $\nabla h(z, t) = (x, a)$ where $\|z\| < \bar{\sigma}t$, , then $a = -(1/2)\|x\|^2$.

Proof of Claim 1. Since $h_\nu \xrightarrow{e} h$ it follows by Attouch's Theorem that there exist $(z_\nu, t_\nu) \to (z, t)$ and $(x_\nu, a_\nu) \to (x, a)$ with $(x_\nu, a_\nu) \in \partial h_\nu(z_\nu, t_\nu)$. Choose $(y_\nu, b_\nu) \in \partial h_\nu(z_\nu, t_\nu)$. If (y_ν, b_ν) has an accumulation point different from (x, a), then h would not be differentiable at (z, t); see [19]. Therefore $\partial h_\nu(z_\nu, t_\nu)$ has an unique accumulation point which by [10, Lemma 3.3] can only be of the desired form. This concludes the proof of Claim 1.

Define

$$f(x) = \sup_{\substack{\|z\|\leq\bar{\sigma}t \\ t\geq T}} \{\langle z, x\rangle - (t/2)\|x\|^2 - h(z, t)\}.$$

Claim 2. Assume h is differentiable at (z_0, t_0) with $t_0 > T$ and $\|z_0\| < \bar{\sigma}t$, and $\nabla h(z_0, t_0) = (x_0, -(t_0/2)\|x_0\|^2)$ then $h(z_0, t_0) = h_f(z_0, t_0)$ and $f(x_0) = \langle z_0, x_0\rangle - (t_0/2)\|x_0\|^2 - h(z_0, t_0)$.

Proof of Claim 2. Consider $L_{x_0}(z, t) = \langle z, x_0\rangle - (t/2)\|x_0\|^2 - h(z, t)$. The function L_{x_0} is concave with $\nabla L_{x_0}(z_0, t_0) = (0, 0)$. Therefore L_{x_0} attains a global maximum at (z_0, t_0). This means that $f(x_0) = \langle z_0, x_0\rangle - (t_0/2)\|x_0\|^2 - h(z_0, t_0)$. It follows that for all x, $h(z_0, t_0) \geq \langle z_0, x\rangle - (t_0/2)\|x\|^2 - f(x)$, therefore $h(z_0, t_0) \geq h_f(z_0, t_0)$. But, $h_f(z_0, t_0) \geq \langle z_0, x_0\rangle - (t_0/2)\|x_0\|^2 = (z_0, t_0)$. Hence, $h(z_0, t_0) = h_f(z_0, t_0)$, and this completes the proof of Claim 2.

Since h is differentiable on a dense subset of $\{(z, t) : \|z\| < \bar{\sigma}t, t < T\}$, and continuous on the same set (a convex function is continuous on the interior of its effective domain; see [19]) we conclude by Claims 2 and 3 that $h(z, t) = h_f(z, t)$ for all such (z, t) with $\|z\| \le \bar{\sigma}t$ and $t \ge T$ (a lower semicontinous convex function is uniquely determined by the values it assumes on the relative interior of its effective domain). By Proposition 2.1, $f_\nu \xrightarrow{e} f$ on $\|x\| \le \bar{\sigma}$ (this is because if we restrict convergence of the quadratic conjugates to $\{(z, t)\colon \|z\| \le \bar{\sigma}t, t \ge T\}$, then by (2.11) we obtain convergence of the functions on $\{x : \|x\| \le \bar{\sigma}\}$). By the very nature of f and h, the pointwise convergence property in (b) is verified. Finally, apply part (a) to conlude that f is primal lower-nice at $\bar{x} = 0$ and that $\operatorname{gph} \partial f_\nu \to \operatorname{gph} \partial f$.

This concludes the proof of Theorem 2.1. □

For an application of Theorem 2.1 we turn to convergence of convex composite functions.

Proposition 2.3. *Let $g, g_\nu : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$, $\nu = 1, 2, \dots$, be lower semicontinuous proper convex functions and $F, F_\nu : \mathbb{R}^n \to \mathbb{R}^m$, $\nu = 1, 2, \dots$, be C^2 mappings. Assume that:*

(a) $g_\nu \xrightarrow{e} g$.
(b) ∇F_ν *converges uniformly to* ∇F.
(c) $F_\nu(\bar{x}) \to F(\bar{x})$.
(d) $F_\nu(\bar{x}) \in \operatorname{dom} g_\nu$.
(e) *The basic constraint qualification holds* (*for* $g \circ F$) *at* $\bar{x} \in \operatorname{dom}(g \circ F)$.

Under these assumptions, $f_\nu = g_\nu \circ F_\nu$ is equi primal-lower-nice at $\bar{x}$.

Proof. Since $g_\nu \xrightarrow{e} g$ one can easily show that there exist $\tilde{\sigma} > 0$ such that for all x, with $\|x - \bar{x}\| \le \tilde{\sigma}$, the basic constraint qualification holds at $x \in \operatorname{dom}(g_\nu \circ F_\nu)$, for all ν large enough.

According to (b) and [11, Lemma 5.2] there exist $c > 0$ such that if $t > 0$, $\|x_i - \bar{x}\| \le \tilde{\sigma}$, $v_\nu^i \in \partial g_\nu(F_\nu(x_i))$, $\|v_\nu^i\| \le ct$, $i = 1, 2$, then

$$\langle \nabla F_\nu(x_1)^* v_\nu^1 - \nabla F_\nu(x_2)^* v_\nu^2, \, x_1 - x_2 \rangle \ge -t\|x_1 - x_2\|^2. \tag{2.17}$$

Claim. There exist N, $T > 0$, and $\mu > 0$ such that if $\nu > N$, $u_\nu^i \in \partial f_\nu(x_i)$, $\|x_i - \bar{x}\| < \tilde{\sigma}$, $t > T$, $\|u_\nu^i\| \le \mu t$, $i = 1, 2$, then $\langle u_\nu^1 - u_\nu^2, x_1 - x_2\rangle \ge -t\|x_1 - x_2\|^2$.

Proof of Claim. Since the b.c.q. holds for all ν large enough, there exists $\tilde{N}$ such that if $\nu \ge \tilde{N}$, then $\partial f_\nu(x) = \nabla F_\nu(x)^* \partial g_\nu(F_\nu(x))$ for all x with $\|x - \bar{x}\| \le \tilde{\sigma}$; see [11]. Therefore according to (2.17), to establish the claim we need only show that there exist N, T, and μ such that if $\nu \ge N$, $t \ge T$, $v_\nu \in \partial g_\nu(F_\nu(x))$ with $\|x - \bar{x}\| \le \tilde{\sigma}$ and $\|\nabla F_\nu(x)^* v_\nu\| \le \mu t$, then $\|v_\nu\| \le ct$.

Suppose not, then there exist $\mu_\nu \downarrow 0$, $x_\nu \to x$, with $\|x_\nu - \bar{x}\| \le \tilde{\sigma}$, $t_\nu \uparrow \infty$ and v_ν in $\partial g_\nu(F_\nu(x))$, such that $\|\nabla F_\nu(x_\nu)^* v_\nu\| \le \mu_\nu t_\nu$ and $\|v_\nu\| > ct_\nu$. We may assume that $v_\nu/\|v_\nu\| \to v$, with $\|v\| = 1$. The following holds:

$$\frac{\|\nabla F_\nu(x_\nu)^* v_\nu\|}{\|v_\nu\|} \le \frac{\mu_\nu t_\nu}{ct_\nu} = \frac{\mu_\nu}{c}.$$

Since μ_ν decreases to 0,

$$\nabla F(x)^* v = \lim_{\nu \to \infty} \frac{\nabla F_\nu(x_\nu)^* v_\nu}{\|v_\nu\|} = 0.$$

But one can easily show that $v \in N(F(x)|\text{dom } g)$ with $\|v\| = 1$. This contradicts the b.c.q. at x and completes the proof of the claim.

By the Claim, $\{f_\nu\}$ is "equi t-monotone", it then follows from [11, Corollary 3.4] that f_ν is equi primal-lower-nice at $\bar{x}$ (The proof of [11, Corollary 3.4] relies on [11, Proposition 3.3 and Lemma 3.2]. The size of T in [11, Lemma 3.2] is proportional to the infimum on $\|x - \bar{x}\| \leq \tilde{\sigma}$. By assumptions (a), (b), and (c) the infima of f_ν on $\|x - \bar{x}\| \leq \tilde{\sigma}$ are bounded below.) □

For the main application of Theorem 2.1 we turn to second-order epi-differentiability.

Theorem 2.2. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be primal lower-nice at $\bar{x}$ and $\bar{v} \in \partial f(\bar{x})$. Then f is twice epi-differentiable at $\bar{x}$ relative to $\bar{v}$ $\Leftrightarrow$ ∂f is proto-differentiable at $(\bar{x}, \bar{v})$. Moreover $\partial(\frac{1}{2} f''_{\bar{x},\bar{v}})(\xi) = (\partial f)'_{\bar{x},\bar{v}}(\xi)$, for all ξ.*

Proof. For $t > 0$, let

$$k_t(\xi) = \frac{f(\bar{x} + t\xi) - f(\bar{x}) - t\langle \bar{v}, \xi\rangle}{t^2}\ ; \tag{2.18}$$

note that $k_t(0) = 0$, for all $t > 0$.

We have $\partial(k_t)(\xi) = \frac{\partial f(\bar{x}+t\xi) - \bar{v}}{t}$; note that

$$0 \in \partial(k_t)(0), \quad \text{for all } t > 0. \tag{2.19}$$

We wish to apply Theorem 2.1 to this family of functions indexed by t, as t goes to 0. To do so we need to verify that this family of functions is equi primal-lower-nice at 0 and uniformly minorized by a quadratic.

Since f is primal lower-nice at $\bar{x}$ there exist $\mu > 0$, $R > 0$ and $\sigma > 0$ such that if $r \geq R$, $\|y\| \leq \mu r$, $\|x - \bar{x}\| < \sigma$ and $y \in \partial f(x)$ then $f(x') \geq f(x) + \langle y, x' - x\rangle - \frac{r}{2}\|x - x'\|^2$ for all x'. Hence if $\|t\xi\| \leq \sigma$, $y \in \partial f(\bar{x} + t\xi)$, $\|y\| \leq \mu r$ then

$$\frac{1}{t^2} f(\bar{x} + t\xi') \geq \frac{1}{t^2}\left[f(\bar{x} + t\xi) + \langle y, t(\xi' - \xi)\rangle - \frac{r}{2}\|t(\xi' - \xi)\|^2\right]$$

for all ξ'. Since $k_t(\xi') = (\frac{1}{t^2})[f(\bar{x}+t\xi') - f(\bar{x}) - t\langle \bar{v}, \xi'\rangle]$, we have that $k_t(\xi') \geq k_t(\xi) + \langle (\frac{1}{t})(y - \bar{v}), \xi' - \xi\rangle - \frac{r}{2}\|\xi' - \xi\|^2$, for all ξ'; this shows that k_t are equi primal-lower-nice at 0 (if $\omega \in \partial k_t(\xi)$, $\omega = \frac{y - \bar{v}}{t}$ for $y \in \partial f(\bar{x}+t\xi)$, $\exists \lambda > 0$ s.t. if $\|\omega\| \leq \lambda r$ then $\|y\| \leq \mu r$, since $\|y\| \leq (t\lambda r + \|\bar{v}\|)$).

Since f is bounded below by a quadratic, so is $f(\cdot) - f(\bar{x}) - \langle \bar{v}, \cdot - \bar{x}\rangle$, i.e. there exist $r > 0$ s.t. $f(x) - f(\bar{x}) - \langle \bar{v}, x\rangle \geq -\frac{r}{2}\|x - \bar{x}\|^2$ for all x. Therefore

$$f(\bar{x} + t\xi) - f(\bar{x}) - \langle \bar{v}, t\xi\rangle \geq -\tfrac{r}{2}\|t\xi\|^2.$$

This implies that

$$k_t(\xi) \geq -\tfrac{r}{4}\|\xi\|^2 \quad \text{for all } \xi.$$

If f is twice epi-differentiable at $\bar{x}$ relative to $\bar{v}$, i.e. $k_t \xrightarrow{e} ((1/2)f''_{\bar{x},\bar{v}})$, then, according to Theorem 2.1 (note that there is no need for the pointwise convergence condition by (2.18) and (2.19)), $\text{gph}\,\partial k_t \to \text{gph}\,\partial((1/2)f''_{\bar{x},\bar{v}})$, but by definition $\text{gph}\,\partial k_t$ converges to $\text{gph}(\partial f)'_{\bar{x},\bar{v}}$.

If ∂f is proto-differentiable at $\bar{x}$ relative to $\bar{v}$, i.e. $\text{gph}\,\partial k_t$ converge, then, according to Theorem 2.1, there exits g such that $k_t \xrightarrow{e} g$ and $\text{gph}\,\partial k_t \to$

gph ∂g. By definition $g = ((1/2)f''_{x,v})$, and $\partial((1/2)f''_{x,v})(\xi) = (\partial f)'_{x,v}(\xi)$, for all ξ. □

Acknowledgment

I am greatly indebted to Professors R. T. Rockafellar and R. J.-B. Wets for their encouragement and their many helpful comments and suggestions. I am also indebted to the anonymous referee for suggestions that have done much to bring this paper to its present form.

References

1. H. Attouch, *Variational convergence of functions and operators*, Pitman, London, 1984.
2. H. Attouch and R. J.-B. Wets, *Epigraphical analysis*, Ann. Inst. Henri Poincaré Analyse Non-Linéaire **6** (1989), 73–100.
3. F. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc. **205** (1975), 247–262.
4. ——, *Optimization and nonsmooth analysis*, Centre de Recherches Mathématiques, Université de Montréal, 1983.
5. R. Cominetti, *On pseudo-differentiability*, Trans. Amer. Math. Soc. **322** (1991), 843–865.
6. P. Loewen, *The proximal normal formula in Hilbert spaces*, Nonlinear Analysis **11** (1987), 979–995.
7. ——, *The proximal normal formula in Banach space*, Canad. Math. Bull. **31** (1988), 353–361.
8. J. L. Ndoutoume, *Conditions nécessaires d'optimalité du premier ordre pour des problèmes de contrôle optimal d'inéquations variationnelles*, Publications AVAMAC, Université Perpignan, Mathématiques, 66025 Cedex, 1987.
9. U. Mosco, *Convergences of convex sets and solutions to variational inequalities*, Adv. in Math. **3** (1969), 510–585.
10. R. Poliquin, *Subgradient monotonicity and convex functions*, Nonlinear Analysis **14** (1990), 305–317.
11. ——, *Integration of subdifferentials of nonconvex functions*, Nonlinear Analysis **17** (1991), 385–398.
12. ——, *Proto-differentiation of subgradient set-valued mappings*, Canad. J. Math. (3) **42** (1990), 522–534.
13. ——, *Proto-differentiation and integration of proximal subgradients*, Dissertation, Univ. of Washington, June 1988.
14. R. Poliquin and R.T. Rockafellar, *Proto-derivative of solution mappings in parametric optimization*, manuscript.
15. ——, *A calculus of epi-derivatives applicable to optimization*, manuscript.
16. ——, *Amenable functions in optimization*, Proc. Internat. School of Math. G. Stampacchia, 10th course: Nonsmooth Optimization: Methods and Applications (to appear).
17. R.T. Rockafellar, *First and second-order epi-differentiation*, Trans. Amer. Math. Soc. **307** (1988), 75–108.
18. ——, *Second-order optimality conditions in mathematical programming obtained by way of epi-derivatives*, Math. Oper. Res. **14** (1989), 462–484.
19. ——, *Convex analysis*, Princeton Univ. Press, 1970.
20. ——, *La théorie des sous-gradients et ses applications à l'optimisation, fonctions convexes et non convexes*, Les Presses de l'Université de Montréal, 1979.
21. ——, *Proximal subgradients, marginal values and augmented Lagrangians*, Math. Oper. Res. **6** (1981), 424–436.

22. ——, *Proto-differentiability of set-valued mappings and its applications in optimization*, Ann. Inst. Henri Poincaré Analyse Non-Linéaire **6** (1989), 449–482.

23. ——, *Perturbation of Kuhn-Tucker points in finite-dimensional optimization*, Nonsmooth Optimization and Related Topics, (F. Giannessi, Ed.), Plenum, 1989.

24. ——, *Generalized second derivatives of convex functions and saddle functions*, Trans. Amer. Math. Soc. **320** (1990), 810–822.

25. R. T. Rockafellar and R. J.-B. Wets, *Variational analysis* (to appear).

26. G. Salinetti and R. J.-B. Wets, *On the convergence of sequences of convex functions in finite dimensions*, SIAM Rev. **21** (1979), 18–33.

27. R. J.-B. Wets, *Convergence of convex functions, variational inequalities and convex optimization problems*, Variational Inequalities and Complementary Problems, (R. Cottle, F. Giannessi and J.-L. Lions, Eds.), Wiley, 1980, pp. 376–403.

Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

E-mail address: rpoliqui@vm.ucs.ualberta.ca