ATTOUCH-WETS CONVERGENCE
AND A DIFFERENTIAL OPERATOR FOR CONVEX FUNCTIONS

GERALD BEER AND MICHEL THÉRA

(Communicated by Andrew Bruckner)

ABSTRACT. The purpose of this note is to characterize Attouch-Wets convergence for sequences of proper lower semicontinuous convex functions defined on a Banach space X in terms of the behavior of an operator Δ defined on the space of such functions with values in $X \times R \times X^*$, defined by $\Delta(f) = \{(x, f(x), y): (x, y) \in \partial f\}$. We show that $\langle f_n \rangle$ is Attouch-Wets convergent to f if and only if points of $\Delta(f)$ lying in a fixed bounded set can be uniformly approximated by points of $\Delta(f_n)$ for large n. The operator Δ is a natural carrier of the Borwein variational principle, which is a key tool in both directions of our proof.

1. INTRODUCTION

In an attempt to address the shortcomings of Mosco convergence [Mo1, Mo2] for sequences of convex sets and functions outside the reflexive setting, new topologies for convex sets and for convex functions viewed as convex sets have been intensively studied over the last few years. The two most important of these are the Attouch-Wets topology (see, e.g., [AW2, ALW, Be1, BL, AP], which is actually mentioned in passing in [Mo1], and the weaker slice topology [Be2–Be4, AB, BB], which is compatible with Mosco convergence if and only if the underlying space is reflexive, and thus constitutes a viable extension of Mosco convergence to nonreflexive spaces. There are several ways to describe the induced convergence notions. In terms of distance functionals, Attouch-Wets convergence of a sequence of closed convex sets $\langle A_n \rangle$ to a closed convex set A means the uniform convergence of the associated sequence of distance functionals to the distance functional of the limit set on bounded subsets of the underlying set X, whereas slice convergence means the pointwise convergence of distance functionals *with respect to each equivalent renorming* [Be3]. Convergence of functions, as usual, means convergence of epigraphs in $X \times R$.

The Attouch-Wets topology is highly tractable and is much better suited for estimation than the slice topology (as is often required in optimization and approximation problems), and in an arbitrary Banach space the topology is stable

with respect to duality: The Fenchel transform is a homeomorphism between $\Gamma(X)$, the proper lower semicontinuous convex functions defined on X, and $\Gamma^*(X^*)$, the proper weak* lower semicontinuous functions defined on the dual space X^* [Be1]. On the other hand, the slice topology admits several fundamental *external* characterizations in terms of duality in an arbitrary Banach space, whereas to this date none have been found for the Attouch-Wets topology (for a partial advance in this direction, see [ANT]). One striking characterization of the slice topology involves the lower semicontinuity of the multifunction $\Delta: \Gamma(X) \rightrightarrows X \times R \times X^*$ defined by $\Delta(f) = \{(x, f(x), y): (x, y) \in \partial f\}$, where ∂f is the subdifferential of the function f [Be2, Theorem 4.11]. In sequential terms, this says: a sequence $\langle f_n \rangle$ in $\Gamma(X)$ is convergent in the slice topology to $f \in \Gamma(X)$ if and only if for each point $(x_0, f(x_0), y_0) \in \Delta(f)$ there exists a sequence $\langle (x_n, f(x_n), y_n) \rangle$ (strongly) convergent to $(x_0, f(x_0), y_0)$ such that $(x_n, f(x_n), y_n) \in \Delta(f_n)$ for each $n \in Z^+$. The purpose of this paper is to show that an analogous characterization holds for the Attouch-Wets topology, strengthening the lower semicontinuity condition to "uniform lower semicontinuity on bounded sets". The tools that we use are the comprehensive Borwein variational principle, along with a characterization of Attouch-Wets convergence in $\Gamma(X)$ in terms of the convergence of Lipschitz regularizations.

2. Preliminaries

In the sequel, X will be a normed linear space—usually a Banach space—with closed unit ball U and origin θ. The closed unit ball and origin of the dual space X^* will be denoted by U^* and θ^*, respectively. We will often be forced to consider finite products of normed spaces, e.g., $X \times R$ and $X \times R \times X^*$, and in such spaces, the box norm will be understood. Thus, the unit ball in $X \times R$ is $U \times [-1, 1]$. If λ is a positive scalar, the verticle cone in $X \times R$ with slope λ will be denoted by C_λ; formally,

$$C_\lambda \equiv \{(x, \alpha): x \in X, \ \alpha \in R, \ \text{and} \ \alpha \geq \lambda\|x\|\}.$$

If A is a nonempty subset of X and $x \in X$, we write $d(x, A)$ for $\inf\{\|x - a\|: a \in A\}$, and if B is another nonempty subset of X, we define the *excess of B over A* by the formula $e(B, A) = \sup_{x \in B} d(x, A)$. We adopt the convention that $e(\varnothing, A) = 0$ whenever A is nonempty.

We represent the nonempty closed convex subsets of X by $C(X)$. As we essentially said in the introduction, the *Attouch-Wets topology* τ_{AW} on $C(X)$ is the topology that $C(X)$ inherits from $C(X, R)$, the continuous real-valued functions from X to R, equipped with the (metrizable) topology of uniform convergence on bounded subsets of X, under the identification $A \leftrightarrow d(\cdot, A)$. But this topology can be approached in a much different way as historically was the case (see, e.g., [Mo1, AW2]). For each $\rho > 0$ and A, $B \in C(X)$, write $e_\rho(B, A) = e(B \cap \rho U, A)$. Then a base for a compatible uniformity for τ_{AW} consists of all entourages of the form

$$\{(A, B): \max\{e_\rho(B, A), \ e_\rho(A, B)\} < \varepsilon\} \quad (\rho > 0, \ \varepsilon > 0).$$

It is now standard to write $\text{haus}_\rho(A, B)$ for $\max\{e_\rho(B, A), e_\rho(A, B)\}$ (it should be noted that haus_ρ is not in general a metric on $C(X)$). For a proof that this agrees with the topology of uniform convergence of distance functions on bounded sets, the reader may consult [AP].

We assume that the reader is familiar with the standard terminology and notation of convex analysis; for further details, consult [Ro]. We denote the proper lower semicontinuous convex functions on X by $\Gamma(X)$. If $f \in \Gamma(X)$, we denote its epigraph by epi f, its effective domain by dom f, and its Fenchel conjugate by f^*. Given a sequence $f, f_1, f_2, f_3, \ldots$ in $\Gamma(X)$, we will write $f = \text{AW-lim} f_n$ provided $\langle \text{epi} f_n \rangle$ is convergent in the Attouch-Wets topology to epi f. For each $x \in X$, we write $\partial f(x)$ (resp. $\partial f_\varepsilon(x)$) for the set of subgradients (resp. ε-subgradients) of f at x. The subdifferential of f is $\partial f \equiv \{(x, y) \in X \times X^* : y \in \partial f(x)\}$. It follows from the Brønsted-Rockafellar Theorem (see, e.g., [Ph, p. 51]) that when X is a Banach space, dom $\partial f \equiv \{x \in X : \exists y \in X^* \text{ such that } (x, y) \in \partial f\}$ is dense in dom f and range $\partial f \equiv \{y \in X^* : \exists x \in X \text{ such that } (x, y) \in \partial f\}$ is dense in dom f^*.

We now turn to a brief discussion of Lipschitz regularizations of proper lower semicontinuous convex functions. These have been studied by numerous authors under various names (see, e.g., [H-U1, H-U2, Az, AW1, FP, Be4]) and are special cases of epi-sums of convex functions. Given $f \in \Gamma(X)$ and $\mu > 0$, we define the *Lipschitz regularization with parameter μ of f*, $f(\mu, \cdot) : X \to [-\infty, \infty]$, by the formula

$$f(\mu, x) = \inf_{w \in X} f(w) + \mu \|x - w\|.$$

Geometrically, the epigraph of $f(\mu, \cdot)$ is obtained by taking the vector sum of epi f with the cone C_μ and then taking the closure in $X \times R$. There are two possibilities for the function $f(\mu, \cdot)$: either epi $f(\mu, \cdot)$ is all of $X \times R$, or $f(\mu, \cdot)$ is a finite-valued convex function and is the largest Lipschitz function with constant μ majorized by f [Be4, Lemma 2.1]. Simple convex algebra shows that we get a finite-valued regularization if and only if $\mu U^* \cap \text{dom} f^* \neq \varnothing$ [Az, Lemme 1.1]. For such μ, we clearly have dom $\partial f(\mu, \cdot) = X$ and range $\partial f(\mu, \cdot) \subset \mu U^*$. Furthermore, for $\mu > d(\theta^*, \text{dom} f^*)$ and x restricted to a bounded subset of X, the infimum defining $f(\mu, x)$ may be taken over a somewhat larger bounded subset of X independent of x rather than over the entire space (see, e.g., Lemma 3.4 and more generally [H-U2, Proposition 5]). Fix $x_0 \in X$; if $w \in X$ satisfies

$$f(w) + \mu \|x_0 - w\| < f(\mu, x_0) + \varepsilon,$$

then we call w an *ε-estimator* for $f(\mu, x_0)$.

For each x in X, we have $f(\mu, x) \leq f(x)$, and it can be shown that equality occurs if and only if $\partial f(\mu, x) = \partial f(x) \cap \mu U^*$ [H-U1, Proposition 2.3]. Increasing the parameter μ produces a larger function, and $\langle f(\mu, \cdot) \rangle$ converges pointwise to f as $\mu \to \infty$. Some interesting new facts regarding the relationship between $\partial f(\mu, \cdot)$ and ∂f can be found in [FP].

3. RESULTS

As announced in the introduction, we shall use the Borwein variational principle ([Bo] or [Ph, p. 50]) in the proof of our characterization theorem (actually, in both directions). We can get by with the following version, which says intuitively that if $y_0 \in \partial g_\varepsilon(x_0)$, then we can approximate $(x_0, g(x_0), y_0)$ by an element of Δg, the accuracy of the approximation being dependent on both ε and $\|y_0\|$.

Borwein variational principle. Let X be a Banach space, and let $g \in \Gamma(X)$. Suppose $\varepsilon > 0$ and $\beta > 0$ and $y_0 \in \partial g_\varepsilon(x_0)$. Then there exists $(x_\varepsilon, g(x_\varepsilon), y_\varepsilon) \in \Delta(g)$ such that:

    (1) $\|x_0 - x_\varepsilon\| \leq \sqrt{\varepsilon}$,

    (2) $|g(x_0) - g(x_\varepsilon)| \leq \sqrt{\varepsilon}(\sqrt{\varepsilon} + \beta^{-1})$, and

    (3) $\|y_0 - y_\varepsilon\| \leq \sqrt{\varepsilon}(1 + \beta\|y_0\|)$.

For the convenience of the reader, we state some elementary lemmas that will be used in the course of our proof. We leave it to the reader to establish the first two.

Lemma 3.1. *Let X be a normed linear space, and let $0 < \varepsilon < \lambda$. Suppose $x_1 \in \lambda U$ and $y_1 \in \lambda U^*$. Then whenever x, y, and w satisfy $\|x - x_1\| < \varepsilon$ and $\|y - y_1\| < \varepsilon$ and $w \in \lambda U$, we have $|\langle y_1, w - x_1 \rangle - \langle y, w - x \rangle| \leq 4\varepsilon\lambda$.*

Lemma 3.2. *Suppose X is a normed linear space and ε, σ, and μ are positive scalars. Suppose $y \in X^*$ satisfies $\|y\| \leq \mu + \varepsilon$, $\alpha \in R$, and $a(x) = \langle y, x \rangle - \alpha$. Then for each $x \in \sigma U$, we have*

$$\inf_{\|w\| \leq \sigma} a(w) + \mu\|x - w\| \geq a(x) - 2\sigma\varepsilon.$$

We now obtain some quantitative information regarding the estimators for a Lispchitz regularization $g(\mu, \cdot)$ of a function $g \in \Gamma(X)$. This information gives uniform bounds for both the arguments and the function values of g used to estimate the set of values $\{g(\mu, x): \|x\| \leq \rho\}$, where g is subject to the normalization $g(\theta) = 0$ and $\theta \in \operatorname{dom} \partial g$.

Lemma 3.3. *Let X be a normed linear space. Suppose that $g \in \Gamma(X)$ with $g(\theta) = 0$ and $\mu > \lambda > 0$. Suppose further that $\operatorname{epi} g \cap -C_\lambda = \{(\theta, 0)\}$. Then for each $x \in \rho U$, we have*

$$g(\mu, x) = \inf\{g(w) + \mu\|x - w\| : \|w\| \leq 3\mu\rho/(\mu - \lambda)\}.$$

Moreover, if w is a $\mu\rho$-estimator for $g(\mu, x)$ where $x \in \rho U$ is arbitrary, then

$$|g(w)| \leq 2\mu\rho + 6\mu^2\rho/(\mu - \lambda).$$

Proof. We can separate $\operatorname{epi} g$ from $-C_\lambda$ by the graph of a continuous linear functional with norm at most λ. Evidently, this linear functional belongs to $\partial g(\theta)$; so, $\mu > d(\theta^*, \operatorname{dom} g^*)$. In particular, $g(\mu, \cdot)$ is finite valued and μ-Lipschitz continuous. Since $\operatorname{epi} g \cap -C_\lambda = \{(\theta, 0)\}$, whenever $x \neq \theta$, we have $g(x) > -\lambda\|x\|$, and since $g(\theta) = 0$, we have $g(x, \mu) \leq \mu\|x\|$. Combining these facts, we see that for all x, $|g(x, \mu)| \leq \mu\|x\|$.

Fix $x \in \rho U$; we first claim that no point outside the ball with radius $3\mu\rho(\mu - \lambda)^{-1}$ and center θ can be a $\mu\rho$-estimator for $g(\mu, x)$. For if $\|w\| > 3\mu\rho(\mu - \lambda)^{-1}$, we have

$$g(w) + \mu\|x - w\| > -\lambda\|w\| + \mu\|w\| - \mu\|x\| > (\mu - \lambda)3\mu\rho/(\mu - \lambda) - \mu\|x\|$$

$$\geq 3\mu\rho - \mu\rho \geq \mu\|x\| + \mu\rho \geq g(\mu, x) + \mu\rho.$$

Taking an arbitrary $\mu\rho$-estimator w for $g(\mu, x)$, we obtain

$$|g(w)| < |g(\mu, x)| + \mu\|x - w\| + \mu\rho$$

$$\leq \mu\|x\| + \mu \cdot 6\mu\rho/(\mu - \lambda) + \mu\rho \leq 2\mu\rho + 6\mu^2\rho/(\mu - \lambda). \quad \square$$

Our next fact is implicit in the proof of Proposition 1.2 of [FP].

Lemma 3.4. *Let X be a normed linear space. Suppose $g \in \Gamma(X)$, $\mu > d(\theta^*, \operatorname{dom} g^*)$, and $(x_0, y_0) \in \partial g(\mu, \cdot)$. Set $a(x) = g(\mu, x_0) + \langle y_0, x - x_0 \rangle$. Then whenever x_ε is an ε-estimator of $g(\mu, x_0)$, we have $a(x_\varepsilon) > g(x_\varepsilon) - \varepsilon$.*

Proof. Suppose instead that $a(x_\varepsilon) + \varepsilon \leq g(x_\varepsilon)$. Since $g(\mu, \cdot)$ is μ-Lipschitz continuous, we have $\|y_0\| \leq \mu$, and the following inequality string is valid:

$$g(x_\varepsilon) + \mu\|x_0 - x_\varepsilon\| \geq (g(\mu, x_0) + \langle y_0, x_\varepsilon - x_0 \rangle + \varepsilon) + \mu\|x_0 - x_\varepsilon\|$$
$$\geq g(\mu, x_0) + \varepsilon + (\mu - \|y_0\|)\|x_0 - x_\varepsilon\| \geq g(\mu, x_0) + \varepsilon.$$

This contradicts the hypothesis that x_ε is an ε-estimator of $g(\mu, x_0)$. $\square$

We now come to our main result, which characterizes the Attouch-Wets convergence of $\langle f_n \rangle$ to f in $\Gamma(X)$ as follows: points of $\Delta(f)$ lying in a fixed bounded set can be uniformly approximated by points of $\Delta(f_n)$ for large n.

Theorem 3.5. *Let X be a Banach space, and let f, f_1, f_2, $f_3, \ldots$ be a sequence of proper lower semicontinuous convex functions on X. Then $f = \text{AW-lim} f_n$ if and only if for each $\rho > 0$ we have $\lim_{n \to \infty} e_\rho(\Delta(f), \Delta(f_n)) = 0$.*

Proof. Necessity. Fix $\rho > 1$ and $\varepsilon \in (0, 1)$. Write

$$\Omega \equiv \Delta(f) \cap \rho U \times [-\rho, \rho] \times \rho U^* \subset X \times R \times X^*.$$

Then for $(x, f(x), y) \in \Omega$, by the definition of subgradient, we have $f^*(y) = \langle y, x \rangle - f(x)$, so that $\|f^*(y)\| \leq \rho^2 + \rho$. Now let $\lambda = \rho^2 + \rho > \rho + 1$. Since $f = \text{AW-lim} f_n$, using the continuity of the Fenchel transform with respect to the Attouch-Wets topology [Be1], there exists $N \in Z^+$ such that for each $n \geq N$,

(1) $\operatorname{haus}_{2\lambda}(\operatorname{epi} f, \operatorname{epi} f_n) < \varepsilon$ and $\operatorname{haus}_{2\lambda}(\operatorname{epi} f^*, \operatorname{epi} f_n^*) < \varepsilon.$

Fix $n \geq N$ and $(x_0, f(x_0), y_0) \in \Omega$. By (1) and the definition of $\operatorname{haus}_{2\lambda}$, there exist points $(x_n, \alpha_n) \in \operatorname{epi} f_n$ and points $(y_n, \beta_n) \in \operatorname{epi} f_n^*$ such that $\|(x_n, \alpha_n, y_n, \beta_n)\| \leq 2\lambda$ and

(2) $|\alpha_n - f(x_0)| < \varepsilon,$ $|\beta_n - f^*(y_0)| < \varepsilon,$
           $\|x_n - x_0\| < \varepsilon,$ $\|y_n - y_0\| < \varepsilon.$

Lemma 3.1 yields $|\langle y_n, x_n \rangle - \langle y_0, x_0 \rangle| \leq 4\lambda\varepsilon$, and since $(x_0, y_0) \in \partial f$, (2) gives

$$f_n(x_n) + f_n^*(y_n) \leq \alpha_n + \beta_n \leq f(x_0) + f^*(y_0) + 2\varepsilon < \langle y_n, x_n \rangle + 5\lambda\varepsilon.$$

This shows that $y_n \in \partial_{5\lambda\varepsilon} f_n(x_n)$. It is also the case that $f_n(x_n) > f(x_0) - 5\lambda\varepsilon$, else

$$f_n(x_n) + f_n^*(y_n) < f(x_0) - 5\lambda\varepsilon + f^*(y_0) + \varepsilon < \langle y_0, x_0 \rangle - 4\lambda\varepsilon < \langle y_n, x_n \rangle,$$

which would contradict Fenchel's inequality. As a result, $|f_n(x_n) - f(x_0)| < 5\lambda\varepsilon$. We now apply the Borwein variational principle to the function f_n at $x = x_n$ with the parameters ε and β in its statement replaced by $5\lambda\varepsilon$ and 1 respectively: there exists $(w_n, f_n(w_n), z_n) \in \Delta(f_n)$ such that

$$\|w_n - x_0\| < \varepsilon + \sqrt{5\lambda\varepsilon},$$
$$|f_n(w_n) - f(x_0)| < 5\lambda\varepsilon + \sqrt{5\lambda\varepsilon}(\sqrt{5\lambda\varepsilon} + 1),$$
$$\|z_n - y_0\| < \varepsilon + \sqrt{5\lambda\varepsilon}(1 + 2\lambda).$$

This shows that $e_\rho(\Delta(f), \Delta(f_n)) \le 5\lambda\varepsilon + \sqrt{5\lambda\varepsilon}(1+3\lambda)$ whenever $n \ge N$. Since λ depends only on ρ and ε is arbitrary, this shows that $\lim_{n\to\infty} e_\rho(\Delta(f), \Delta(f_n)) = 0$.

Sufficiency. We rely on the following characterization of AW-convergence valid in an arbitrary normed linear space as established in [Be4, Theorem 4.3]: $f = \text{AW-}\lim f_n$ if and only if for each $\mu > d(\theta^*, \text{dom } f^*)$, the sequence $\langle f_n(\mu, \cdot)\rangle$ converges uniformly to $f(\mu, \cdot)$ on bounded subsets of X.

Fix $\mu > d(\theta^*, \text{dom } f^*)$. To verify uniform convergence of regularizations on bounded sets, we find it convenient to normalize the functions in a manner consistent with the statement of Lemma 3.3. Since $d(\theta^*, \text{dom } f^*) = d(\theta^*, \text{range } \partial f)$, we can find $(\overline{x}, \overline{y}) \in \partial f$ with $\|\overline{y}\| < \mu$. Choose for each $n \in Z^+$ a point $(\overline{x}_n, \overline{y}_n) \in \partial f_n$ such that

$$\lim_{n\to\infty}(\overline{x}_n, f_n(\overline{x}_n), \overline{y}_n) = (\overline{x}, f(\overline{x}), \overline{y}).$$

We now replace our initial sequence $f, f_1, f_2, f_3, \ldots$ by $g, g_1, g_2, g_3, \ldots$ defined by

$$g(x) = f(x + \overline{x}) - f(\overline{x})$$

and

$$g_n(x) = f_n(x + \overline{x}_n) - f_n(\overline{x}_n), \qquad n = 1, 2, 3, \ldots.$$

It is easy to check that for each $\rho > 0$ we have $\lim_{n\to\infty} e_\rho(\Delta(g), \Delta(g_n)) = 0$ and that the expected perturbations of Lipschitz regularizations occurs, i.e., for each $x \in X$,

$$g(\mu, x) = f(\mu, x + \overline{x}) - f(\overline{x})$$

and

$$g_n(\mu, x) = f_n(\mu, x + \overline{x}_n) - f_n(\overline{x}_n), \qquad n = 1, 2, 3, \ldots.$$

Notice that (i) $g(\theta) = g_1(\theta) = g_2(\theta) = \cdots = 0$, and (ii) $(\theta, \overline{y}) \in \partial g$ and $(\theta, \overline{y}_n) \in \partial g_n$. Since $\mu > \|\overline{y}\|$ and $\|\overline{y}_n\| \to \|\overline{y}\|$, there exists $\lambda \in (0, \mu)$ with $\|\overline{y}\| < \lambda$ and such that for all large n, $\|\overline{y}_n\| < \lambda$. Thus, by passing to a tail of $\langle g_n\rangle$, we may assume that $\text{epi } g \cap -C_\lambda = \{(\theta, 0)\}$ and $\text{epi } g_n \cap -C_\lambda = \{(\theta, 0)\}$, $n = 1, 2, 3, \ldots$. Thus, all regularizations $g(\mu, \cdot)$ and $g_n(\mu, \cdot)$ are finite valued and Lipschitz continuous with parameter μ.

If we can show that $\langle g_n(\mu, \cdot)\rangle$ converges to $g(\mu, \cdot)$ uniformly on bounded subsets of X, then clearly the same can be said for $\langle f_n(\mu, \cdot)\rangle$ with respect to $f(\mu, \cdot)$. To this end, fix $\rho > \max\{1, \mu^{-1}\}$ and $\varepsilon \in (0, 1)$. We produce $N \in Z^+$ such that $\forall n \ge N$ we have

$$\sup_{x\in\rho U} |g_n(\mu, x) - g(\mu, x)| \le \varepsilon.$$

Now let σ be the following positive constant, as suggested by Lemma 3.3:

$$\sigma = 3\mu\rho/(\mu - \lambda) + 2\mu\rho + 6\mu^2\rho/(\mu - \lambda).$$

To note for the future: Since $\rho > 1$ and $\rho\mu > 1$, the scalar σ exceeds each of the quantities μ, $3\mu\rho(\mu - \lambda)^{-1}$, and $2\mu\rho + 6\mu^2\rho(\mu - \lambda)^{-1}$ by at least 1. Now choose $N \in Z^+$ so large that $\forall n \ge N$ we have $e_\sigma(\Delta(g), \Delta(g_n)) < \varepsilon/16\sigma$. We will show that this choice of N does the job.

Fix $x_0 \in \rho U$; since $g(\mu, \cdot)$ is Lipschitz continuous with constant μ, $\partial g(\mu, x_0) \ne \varnothing$. Let $y_0 \in \partial g(\mu, x_0)$ be arbitrary, and choose $\delta \in (0, \varepsilon/2)$

such that $\sqrt{\delta}(1 + \sqrt{\sigma}) < \varepsilon/16\sigma$. Let w be a δ-estimator for $g(\mu, x_0)$. Since $\delta < \varepsilon < 1 < \mu\rho$, Lemma 3.4 yields

$$\|w\| \leq 3\mu\rho/(\mu - \lambda) \quad \text{and} \quad |g(w)| \leq 2\mu\rho + 6\mu^2\rho/(\mu - \lambda).$$

By Lemma 3.4 and the fact that $g \geq g(\mu, \cdot)$, we have $y_0 \in \partial_\delta g(w)$. Applying the Borwein variational principle with β replaced by $\sigma^{-1/2}$ and ε replaced by δ in its statement, we can find $(w', y') \in \partial g$ satisfying

(1) $\|w' - w\| \leq \sqrt{\delta} < \varepsilon/16\sigma$,
(2) $\|g(w') - g(w)\| \leq \sqrt{\delta}(1 + \sqrt{\sigma}) < \varepsilon/16\sigma$, and
(3) $\|y' - y_0\| \leq \sqrt{\delta}(1 + \sqrt{\sigma}) < \varepsilon/16\sigma$.

This implies that $\|w'\| \leq \sigma$, $|g(w')| \leq \sigma$, and $\|y'\| \leq \sigma$. Now fix $n \geq N$. By our choice of N and the triangle inequality, there exists $(w_n, g_n(w_n), y_n) \in \Delta(g_n)$ such that $\|w_n - w\| < \varepsilon/8\sigma$, $|g_n(w_n) - g(w)| < \varepsilon/8\sigma$, and $\|y_n - y_0\| < \varepsilon/8\sigma$.

Using the fact that w is a δ-estimator for $g(\mu, x_0)$ and $\mu < \sigma$, we first obtain

$$\begin{aligned}
g_n(\mu, x_0) &\leq g_n(w_n) + \mu\|x_0 - w_n\| < g(w) + \varepsilon/8\sigma + \mu(\|x_0 - w\| + \|w - w_n\|) \\
&< g(\mu, x_0) + \delta + \varepsilon/8\sigma + \mu \cdot \varepsilon/8\sigma \\
&< g(\mu, x_0) + \varepsilon/2 + \varepsilon/8 + \varepsilon/8 < g(\mu, x_0) + \varepsilon.
\end{aligned}$$

Verification of the inequality $g(\mu, x) - g_n(\mu, x) \leq \varepsilon$ is more difficult. We consider the continuous affine functional on X defined by

$$a(x) = g_n(w_n) + \langle y_n, x - w_n \rangle.$$

Of course, the graph of a is a support hyperplane to epi g_n. Since $\|y_n\| < \|y_0\| + \varepsilon/8\sigma$ and $g_n \geq a$, we have using Lemmas 3.2 and 3.3

$$\begin{aligned}
g_n(\mu, x_0) &= \inf\{g_n(w) + \mu\|x_0 - w\|: \|w\| \leq 3\mu\rho/(\mu - \lambda)\} \\
&\geq \inf\{a(w) + \mu\|x_0 - w\|: \|w\| \leq 3\mu\rho/(\mu - \lambda)\} \\
&> a(x_0) - 2\sigma \cdot \varepsilon/8\sigma = a(x_0) - \varepsilon/4.
\end{aligned}$$

But by Lemma 3.1,

$$\begin{aligned}
g(\mu, x_0) &\leq g(\mu, w) + \langle y_0, x_0 - w \rangle \leq g(w) + \langle y_0, x_0 - w \rangle \\
&\leq a(x_0) + |g_n(w_n) - g(w)| + |\langle y_0, x_0 - w \rangle - \langle y_n, x_0 - w_n \rangle| \\
&\leq a(x_0) + \varepsilon/8\sigma + 4\sigma \cdot \varepsilon/8\sigma < a(x_0) + 3\varepsilon/4.
\end{aligned}$$

Putting these last two inequalities together gives $g(\mu, x_0) < g_n(\mu, x_0) + \varepsilon$ and completes the proof of the theorem. □

We note in closing that $f = \text{AW-lim} f_n$ does not imply that for each $\rho > 0$ we have $\lim_{n \to \infty} e_\rho(\Delta(f_n), \Delta(f)) = 0$: on the line, let f be the indicator function of the ray $[0, \infty)$, and for each n define $f_n \in \Gamma(R)$ by $f_n(x) = n - n^2 x$ if $x \leq 1/n$ and $f_n(x) = 0$ otherwise. But we do not have $\lim_{n \to \infty} e_\rho(\partial f_n, \partial f) = 0$ for each $\rho > 0$, as shown in [ANT], so that the sequence of subdifferentials $\langle \partial f_n \rangle$ is convergent to ∂f in the Attouch-Wets topology.

REFERENCES

[AB]   H. Attouch and G. Beer, *On the convergence of sudifferentials of convex functions*, Sém. Anal. Convexe **21** (1991); exposé no. 8, Arch. Math. **60** (1993), 389–400.

[ALW]   H. Attouch, R. Lucchetti, and R. Wets, *The topology of the ρ-Hausdorff distance*, Ann. Mat. Pura Appl. (4) **160** (1991), 303–320.

[ANT]   H. Attouch, J. Ndoutomne, and M. Théra, *Epigraphical convergence of functions and convergence of their derivatives in Banach spaces*, Sém Anal. Convexe **20** (1990), exposé no. 9.

[AW1]   H. Attouch and R. Wets, *Isometries for the Legendre-Fenchel transform*, Trans. Amer. Math. Soc. **296** (1986), 33–60.

[AW2]   ———, *Quantitative stability of variational systems*, I. *The epigraphical distance*, Trans. Amer. Math. Soc. **328** (1991), 695–730.

[Az]   D. Azé, *Caractérisation de la convergence au sens de Mosco en terme d' approximation infconvolutive*, Ann. Fac. Sci. Toulouse Math. (5) **8** (1986–1987), 293–314.

[AP]   D. Azé and J.-P. Penot, *Operations on convergent families of sets and functions*, Optimization **21** (1990), 521–534.

[Be1]   G. Beer, *Conjugate convex functions and the epi-distance topology*, Proc. Amer. Math. Soc. **108** (1990), 117–126.

[Be2]   ———, *The slice topology: A viable alternative to Mosco convergence in nonreflexive spaces*, Sém. Anal. Convexe **21** (1991), exposé no. 3; J. Nonlinear Anal. **19** (1992), 271–290.

[Be3]   ———, *Wijsman convergence of convex sets under renorming*, J. Nonlinear Anal. (to appear).

[Be4]   ———, *Lipschitz regularization and the convergence of convex functions*, Numer. Funct. Anal. Optim. (to appear).

[BB]   G. Beer and J. Borwein, *Mosco and slice convergence of level sets and graphs of linear functionals*, J. Math. Anal. Appl. **175** (1993), 53–67.

[BL]   G. Beer and R. Lucchetti, *Convex optimization and the epi-distance topology*, Trans. Amer. Math. Soc. **327** (1991), 795–813.

[Bo]   J. Borwein, *A note on ε-subgradients and maximal monotonicity*, Pacific J. Math. **103** (1982), 307–314.

[FP]   S. Fitzpatrick and R. Phelps, *Bounded approximations to monotone operators on Banach spaces*, Anal. Non-linéaire **9** (1992), 573–595.

[H-U1] J.-B. Hiriart-Urruty, *Lipschitz r-continuity of the approximate subdifferential of a convex function*, Math. Scand. **47** (1980), 123–134.

[H-U2] ———, *Extensions of Lipschitz functions*, J. Math. Anal. Appl. **77** (1980), 539–554.

[Mo1]   U. Mosco, *Convergence of convex sets and of solutions of variational inequalities*, Adv. Math. **3** (1969), 510–585.

[Mo2]   ———, *On the continuity of the Young-Fenchel transform*, J. Math. Anal. Appl. **35** (1971), 518–535.

[Ph]   R. Phelps, *Convex functions, monotone operators, and differentiability*, Lecture Notes in Math., vol. 1364, Springer-Verlag, Berlin, 1989.

[Ro]   R. T. Rockafellar, *Convex analysis*, Princeton Univ. Press, Princeton, NJ, 1970.

DEPARTMENT OF MATHEMATICS, CALIFORNIA STATE UNIVERSITY LOS ANGELES, LOS ANGELES, CALIFORNIA 90032
E-mail address: gbeer@calstatela.edu

DÉPARTEMENT DE MATHÉMATIQUES, UNIVERSITÉ DE LIMOGES, 87060 LIMOGES, FRANCE