TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 296, Number 1, July 1986

# ISOMETRIES FOR THE LEGENDRE-FENCHEL TRANSFORM

BY

HEDY ATTOUCH AND ROGER J. B. WETS[1]

ABSTRACT. It is shown that on the space of lower semicontinuous convex functions defined on $R^n$, the conjugation map—the Legendre-Fenchel transform—is an isometry with respect to some metrics consistent with the epi-topology. We also obtain isometries for the infinite dimensional case (Hilbert space and reflexive Banach space), but this time they correspond to topologies finer than the Mosco-epi-topology.

**1. Introduction.** Initially, the study of the epi-topology for the space of lower semicontinuous functions was motivated by the fact that on the subspace of convex functions the Legendre-Fenchel transform, i.e. the conjugation map, is bicontinuous. Actually, it is to state this result, which he proved for functions defined on $R^n$, that Wijsman [**1**] was led to introduce the concept of epi-convergence. Mosco [**2**], and also Joly [**3**], generalized this theorem to functions defined on a reflexive Banach space by considering an epi-topology generated by both the weak and the strong topology on the underlying space. We refer to it today as the Mosco-epi-topology. A further extension to the nonreflexive Banach case has been obtained recently by Back [**4**].

Walkup and Wets [**5**] obtained a related result, namely that on $\mathscr{C}$ the hyperspace of closed convex cones, subsets of a reflexive Banach space, the polar map is an isometry when the distance between two cones $P$ and $Q$ is measured in terms of the Hausdorff distance between $P \cap B$ and $Q \cap B$ with $B$ the unit ball. In finite dimensions this isometry implies the bicontinuity of the Legendre-Fenchel transform on the space of lower semicontinuous convex functions equipped with the epi-topology; details are worked out in [**6**]. In §4, we refine this result and show that the isometry of the polar map yields an isometry for the Legendre-Fenchel transform, provided the notion of distance between two functions is defined in terms of a suitable metric.

For infinite dimensions, however, the Walkup-Wets result is not immediately transferable to the functional setting, at least not in an operational form. This can be traced back to the fact that the unit ball is not compact. In §2 we exhibit new

Received by the editors November 25, 1984.

1980 *Mathematics Subject Classification*. Primary 49A50, 58E30; Secondary 49A29, 47H05, 52A05, 90C31.

*Key words and phrases*. Convexity, epi-convergence, variational convergence, duality, conjugation, Legendre-Fenchel transform, isometry, polarity.

[1]Supported in part by a Fellowship of the Centre National de la Recherche Scientifique and a grant of the National Science Foundation.

isometries for the Legendre-Fenchel transform relying on Moreau-Yosida approximates. We obtain one isometry in terms of the approximated functionals and another one in terms of the resolvents and subdifferentials. We then explore the important relationship between these two isometries. Still another isometry is brought to the fore in §3 involving the indicator functions and the support functions of convex sets. It also relies on approximates of the original functions, to which we refer as Wijsman approximates in recognition of the fact that these were the tools used by Wijsman in his derivation of the bicontinuity result refered to earlier.

**2. Isometries for Moreau-Yosida approximates.** Let $X$ be a Hilbert space, identified with its dual, with norm $\|\cdot\|$ and inner product $\langle\cdot,\cdot\rangle$. Let $f\colon X \to ]-\infty, \infty]$ be an extended real valued function, finite valued for at least some $x$ in $X$. Such a function is said to be *proper*; it is the only type of function that appears in this paper. For every $\lambda > 0$

$$f_\lambda(x) := \left(f \,\square\, \frac{1}{2\lambda}\|\cdot\|^2\right)(x) = \inf_{y \in X}\left[f(y) + \frac{1}{2\lambda}\|x - y\|^2\right] \tag{2.1}$$

is the *Moreau-Yosida approximate of $f$ of parameter* $\lambda$; here $\square$ denotes *inf-convolution*. These approximates play an important role in the analysis of variational limit problems, basically because a sequence of functions $\{f^\nu\colon X \to ]-\infty, \infty],\ \nu = 1, \ldots\}$ epi-converges (with respect to the strong topology of $X$) to the lower semicontinuous function $f$ if and only if

$$f = \sup_{\lambda > 0} \limsup_{\nu \to \infty} f_\lambda^\nu = \sup_{\lambda > 0} \liminf_{\nu \to \infty} f_\lambda^\nu, \tag{2.2}$$

provided the $f^\nu$ are (quadratically) minorized, i.e. there exists $x_0 \in X$ and $\beta > 0$ such that for all $\nu = 1, \ldots$

$$f^\nu(x) \geqslant -\beta\left(\|x - x_0\|^2 + 1\right),$$

[7, Theorem 5.37]. Recall that, given $(X, \tau)$ a first countable topological space, a sequence $\{f^\nu\colon X \to \overline{R},\ \nu = 1, \ldots\}$ *epi-converges* to $f$ (with respect to the topology $\tau$), if for all $x$ in $X$

$$\liminf_{\nu \to \infty} f^\nu(x^\nu) \geqslant f(x), \quad \text{whenever } x = \lim_{\nu \to \infty} x^\nu, \tag{2.3}$$

and for some sequence $\{x^\nu,\ \nu = 1, \ldots\}$ with $x = \lim_{\nu \to \infty} x^\nu$

$$\limsup_{\nu \to \infty} f^\nu(x^\nu) \leqslant f(x). \tag{2.4}$$

A short review of the properties of epi-convergence can be found in [7, §2] (for more details consult the monograph [8]).

The Moreau-Yosida approximates possess a number of properties that make them well suited for the analysis of the limit of sequences of functions. For example, with $f$ (quadratically) minorized, we have that $f_\lambda$ is locally Lipschitz with the Lipschitz constant depending only on the parameters $\lambda$, $x_0$ and $\beta$ [7, Theorem 5.8]. Thus, if a collection $\{f^\nu,\ \nu = 1, \ldots\}$ is minorized with the same quadratic form, the $\{f_\lambda^\nu,\ \nu = 1, \ldots\}$ are locally equi-Lipschitz. Another property [8, Proposition 2.67] needed later on is: For any $\lambda > 0$ and $\mu > 0$, we have

$$(f_\lambda)_\mu = f_{\lambda+\mu}, \tag{2.5}$$

which is called the *resolvent equation*. Indeed

$$\begin{aligned}(f_\lambda)_\mu(x) &= \inf_y\left[\frac{1}{2\mu}\|x-y\|^2 + \inf_z\left(f(z) + \frac{1}{2\lambda}\|y-z\|^2\right)\right]\\ &= \inf_z\left[f(z) + \inf_y\left(\frac{1}{2\mu}\|x-y\|^2 + \frac{1}{2\lambda}\|y-z\|^2\right)\right]\\ &= \inf_z\left[f(z) + \frac{1}{2\mu}\left\|x - \frac{1}{\lambda+\mu}(\mu z + \lambda x)\right\|^2 + \frac{1}{2\lambda}\left\|\frac{1}{\lambda+\mu}(\mu z + \lambda x) - z\right\|^2\right]\\ &= \inf_z\left[f(z) + \frac{1}{2(\lambda+\mu)}\|x-z\|^2\right].\end{aligned}$$

Note that this identity remains valid if $X$ is a Banach space.

The analysis of the limit properties of sequences of convex functions via their Moreau-Yosida approximates highlights the full potential of this technique. Instead of (2.2), we have that a sequence of convex functions $\{f^\nu\colon X \to ]-\infty,\infty],\ \nu = 1,\dots\}$ Mosco-epi-converges to the (necessarily convex and lower semicontinuous) function $f$ if and only if for all $\lambda > 0$

$$f_\lambda(x) = \lim_{\nu\to\infty} f_\lambda^\nu(x) \quad \text{for all } x \in X, \tag{2.6}$$

provided only that $f$ be proper [**9**, Théorème 1.2]. Recall that Mosco-epi-convergence is epiconvergence for *both* the strong and weak topologies of $X$, which means that in (2.3) one considers all weakly converging sequences while in (2.4) one requires the sequence to be strongly converging. Also

$$(f_\lambda)^* = \left(f \,\square\, \frac{1}{2\lambda}\|\cdot\|^2\right)^* = f^* + \frac{\lambda}{2}\|\cdot\|^2 \tag{2.7}$$

since $(\|\cdot\|^2/2\lambda)^* = \lambda\|\cdot\|^2/2$. As usual, $f^*\colon X \to \overline{\mathbf{R}}$, the *conjugate* of $f$, is defined by

$$f^*(y) := \sup_{x\in X}\left[\langle y, x\rangle - f(x)\right]. \tag{2.8}$$

The map $f \mapsto f^*$ is the *Legendre-Fenchel transform*.

An important relationship between the Moreau-Yosida approximates of $f$ and its conjugate $f^*$ is highlighted by the next theorem. This identity was already known to Moreau [**10**] in the case $\lambda = 1$, see also [**11**, §31]. It is the key to a number of isometries.

2.9. THEOREM. *Suppose $X$ is a Hilbert space and $f\colon X \to \overline{R}$ is a proper convex function. Then for any $\lambda > 0$*

$$(f^*)_\lambda(\lambda x) = \frac{\lambda}{2}\|x\|^2 - f_{\lambda^{-1}}(x). \tag{2.10}$$

PROOF. We have

$$\begin{aligned}(f^*)_\lambda(\lambda x) &= \inf_u \left[ f^*(u) + \frac{1}{2\lambda}\|\lambda x - u\|^2 \right] \\ &= \inf_u \sup_y \left[ \langle y, u\rangle - f(y) + \frac{1}{2\lambda}\|\lambda x - u\|^2 \right] \\ &= \sup_y \left[ -f(y) + \min_u \left( \langle y, u\rangle + \frac{1}{2\lambda}\|\lambda x - u\|^2 \right) \right] \\ &= \sup_y \left[ -f(y) + \langle y, \lambda x - \lambda y\rangle + \frac{1}{2\lambda}\|\lambda x - (\lambda x - \lambda y)\|^2 \right] \\ &= -\inf_y \left[ f(y) + \frac{\lambda}{2}\|x - y\|^2 \right] + \frac{\lambda}{2}\|x\|^2\end{aligned}$$

which yields (2.10). The interchange of inf and sup can be justified as follows: Define

$$g(y) = \min_u \left( \langle y, u\rangle + \frac{1}{2\lambda}\|\lambda x - u\|^2 \right) = \frac{\lambda}{2}\left( \|x\|^2 - \|x - y\|^2 \right).$$

The function $g$ is concave, finite and continuous on $X$. Thus, by Fenchel's duality theorem [**12**] we have

$$\sup_y [g(y) - f(y)] = \inf_u [f^*(u) - g_*(u)],$$

where $g_*$ is the concave conjugate of $g$, i.e.

$$g_*(u) = \inf_y [\langle u, y\rangle - g(y)],$$

and a straightforward calculation shows that

$$g_*(u) = -\frac{1}{2\lambda}\|\lambda x - u\|^2. \qquad \square$$

As a corollary to this theorem, we obtain an interesting identity between the gradients of $(f^*)_\lambda$ and the resolvents of parameter $\lambda^{-1}$ associated with $f$. Let $\partial f(x)$ denote the set of subgradients of the convex function $f$ at $x$, i.e.

$$\partial f(x) := \{ v \mid f(y) \geq f(x) + \langle v, y - x\rangle \text{ for all } y \in X \}. \tag{2.11}$$

The convex function $f_\lambda$ is differentiable, and

$$\nabla f_\lambda(x) = \lambda^{-1}(x - J_\lambda x), \tag{2.12}$$

where $J_\lambda$ for $\lambda > 0$ is the *resolvent of parameter* $\lambda$ associated to $f$, i.e. the operator from $X$ into $X$ defined by

$$J_\lambda x := (I + \lambda \partial f)^{-1}(x). \tag{2.13}$$

Note that

$$f_\lambda(x) = f(J_\lambda x) + \frac{1}{2\lambda}\|x - J_\lambda x\|^2.$$

To obtain (2.12), observe that for any $\lambda > 0$

$$y \in \arg\min\left[f + \frac{1}{2\lambda}\|x - \cdot\|^2\right]$$

if and only if

$$(\partial f + \lambda^{-1}I)(y) - \lambda^{-1}x = 0$$

or equivalently, if and only if

$$y = (I + \lambda\partial f)^{-1}(x) = J_\lambda x.$$

Since

$$\partial f_\lambda(x) = \left\{ v|(v,0) \in \partial_{(x,y)}\left(f(y) + \frac{1}{2\lambda}\|x - y\|^2\right)\right\},$$

we have

$$\partial f_\lambda(x) = \left\{ v|v = \lambda^{-1}(x - J_\lambda x)\right\},$$

which also means that the set of subgradients is a singleton and, $f_\lambda$ being convex, it is thus Fréchet differentiable with its gradient $\nabla f_\lambda$ given by (2.12). The resolvent $J_\lambda$ is a contraction and the gradient $x \mapsto \nabla f_\lambda(x)$ is Lipschitz with constant $\lambda^{-1}$ (for more about resolvents and the properties of Yosida approximates $(I + \lambda\partial f)^{-1}$ of the monotone operator $\partial f$, consult [**13**]).

Combining (2.12) with (2.10), we obtain

2.14. COROLLARY. *Suppose $f$ is a proper closed convex function defined on a separable Hilbert space $X$. Then for any $\lambda > 0$*

$$\nabla(f^*)_\lambda(\lambda x) = J_{\lambda^{-1}}x = x - \lambda^{-1}\nabla f_{\lambda^{-1}}(x). \tag{2.15}$$

Now let SCC($X$) be *the cone of proper lower semicontinuous convex functions defined on $X$*, here a Hilbert space. For every $\lambda > 0$ and $\rho \geqslant 0$ we define on SCC($X$) the distance functions

$$d_{\lambda,\rho}(f,g) = \sup_{\|x\|\leqslant\rho} |f_\lambda(x) - g_\lambda(x)| \tag{2.16}$$

and

$$d^J_{\lambda,\rho}(f,g) = \sup_{\|x\|\leqslant\rho} \|J^f_\lambda x - J^g_\lambda x\|, \tag{2.17}$$

where $f$ and $g$ are two elements of SCC($X$) and $J^f_\lambda$ and $J^g_\lambda$ are the resolvents of parameter $\lambda$ associated with $f$ and $g$ respectively. Note that in view of (2.12) we could also define $d^J_{\lambda,\rho}$ as follows:

$$d^J_{\lambda,\rho}(f,g) = \lambda \sup_{\|x\|\leqslant\rho} \|\nabla f_\lambda(x) - \nabla g_\lambda(x)\|. \tag{2.18}$$

Recall that if $f \in \mathrm{SCC}(X)$, so does $f^*$, and $f^{**} = f$. This suggests comparing distances between any two functions and their conjugates. This leads us to

2.19. THEOREM. *Suppose $X$ is a Hilbert space. Then for any $f,\ g \in \mathrm{SCC}(X)$ and any $\lambda > 0$ and $\rho \geqslant 0$*

$$d_{\lambda,\rho}(f,g) = d_{\lambda^{-1},\rho\lambda^{-1}}(f^*,g^*), \tag{2.20}$$

*and*

$$d^J_{\lambda,\rho}(f,g) = \lambda d^J_{\lambda^{-1},\rho\lambda^{-1}}(f^*,g^*). \tag{2.21}$$

PROOF. From (2.10) it follows that

$$|f_\lambda(x) - g_\lambda(x)| = |(f^*)_{\lambda^{-1}}(\lambda^{-1}x) - (g^*)_{\lambda^{-1}}(\lambda^{-1}x)|$$

and hence

$$\sup_{\|x\|\leqslant\rho} |f_\lambda(x) - g_\lambda(x)| = \sup_{\|y\|\leqslant\lambda^{-1}\rho} |(f^*)_{\lambda^{-1}}(y) - (g^*)_{\lambda^{-1}}(y)|$$

which gives us (2.20).

From (2.15) and (2.12) we see that

$$\|J^f_\lambda x - J^g_\lambda x\| = \lambda \|J^{f^*}_{\lambda^{-1}}(\lambda^{-1}x) - J^{g^*}_{\lambda^{-1}}(\lambda^{-1}x)\|$$

which implies that

$$\sup_{\|x\|\leqslant\rho} \|J^f_\lambda x - J^g_\lambda x\| = \lambda \sup_{\|y\|\leqslant\lambda^{-1}\rho} \|J^{f^*}_{\lambda^{-1}}(y) - J^{g^*}_{\lambda^{-1}}(y)\|$$

and this yields (2.21). □

Of course with $\lambda = 1$, as a direct consequence of the above and (2.18), we obtain

2.22. COROLLARY (ISOMETRIES). *For every $\rho \geqslant 0$, the Legendre-Fenchel transform on $\mathrm{SCC}(X)$ is an isometry for $d_{1,\rho}$ and $d^J_{1,\rho}$, i.e. for all $f,\ g$ in $\mathrm{SCC}(X)$:*

$$d_{1,\rho}(f,g) = d_{1,\rho}(f^*,g^*) \tag{2.23}$$

*and*

$$d^J_{1,\rho}(f,g) = d^J_{1,\rho}(f^*,g^*). \tag{2.24}$$

*Note also that for $\lambda = 1$,*

$$d^J_{1,\rho}(f,g) = \sup_{\|x\|\leqslant\rho} \|\nabla f_1(x) - \nabla g_1(x)\|.$$

The distance function $d_{\lambda,\rho}$ is calculated in terms of the function values, whereas $d^J_{\lambda,\rho}$ is in terms of slopes (2.18) or resolvents (2.17). Both distance functions generate Hausdorff metrics, for example:

$$\mathrm{dist}(f,g) := \sum_{k=1}^{\infty} 2^{-k} \frac{d_{1,\rho_k}(f,g)}{1 + d_{1,\rho_k}(f,g)} \tag{2.25}$$

with the $\{\rho_k,\ k = 1, \ldots\}$ a sequence of positive real numbers increasing to $+\infty$, and

$$\text{(2.26)} \qquad \operatorname{dist}^J(f, g) := \sum_{k=1}^{\infty} \left[ \left(2^{-k} d^J_{1,\rho_k}(f, g)\right) / \left(1 + d^J_{1,\rho_k}(f, g)\right) \right].$$

Let us note that $\operatorname{dist}(f, g) = 0$ implies $f_1 \equiv g_1$ and hence $f = g$. This last implication, which might seem at first surprising, relies on the fact that the functions $f$ and $g$ are closed and convex. Indeed if for such functions and some $\lambda_0 > 0$ one has $f_{\lambda_0} = g_{\lambda_0}$, then $f = g$. To see this, just take the conjugates of $f_{\lambda_0}$ and $g_{\lambda_0}$. From (2.7)

$$f^* + \frac{\lambda_0}{2}\|\cdot\|^2 = g^* + \frac{\lambda_0}{2}\|\cdot\|^2$$

which implies

$$f^* = g^*.$$

It follows that $f = g$ since the functions $f$ and $g$ are closed and convex. Let us stress the fact that in order to recover general closed functions $f$ from their Moreau-Yosida approximates one needs all approximates $f_\lambda$ (or at least a sequence $f_{\lambda_n}$ with $\lambda_n \to 0$ as $n \to +\infty$).

Similarly, $\operatorname{dist}^J(f, g) = 0$ implies $\partial f = \partial g$ and hence $f = g$ after some normalization of the functions.

From (2.23) and (2.24), one gets the following

2.27. COROLLARY (ISOMETRY). *Suppose $X$ is a Hilbert space. The Legendre-Fenchel transform on* SCC($X$) *is an isometry for the Hausdorff metrics* dist *and* $\operatorname{dist}^J$ *defined by* (2.25) *and* (2.26). *In particular, we have*

$$\text{(2.28)} \qquad \operatorname{dist}(f, g) = \operatorname{dist}(f^*, g^*)$$

*and*

$$\text{(2.29)} \qquad \operatorname{dist}^J(f, g) = \operatorname{dist}^J(f^*, g^*).$$

It should be emphasized that convergence for a sequence of convex functions $\{f^\nu\colon X \to \overline{R},\ \nu = 1, \ldots\}$ to a limit function $f$ can of course be defined in terms of these distance functions. As can be surmised from our earlier comments, there is a close connection between epi-convergence in SCC($X$) and the convergence generated by the metrics introduced earlier. To study these relationships, we begin with comparing the uniform structures associated to the distance functions $\{d_{\lambda,\rho};\ \lambda > 0,\ \rho \geqslant 0\}$ and $\{d^J_{\lambda,\rho};\ \lambda > 0,\ \rho \geqslant 0\}$.

2.30. PROPOSITION. *Suppose $f$ and $g$ are proper closed convex functions defined on a Hilbert space $X$. For any $\lambda > 0$ and $\rho \geqslant 0$ let $d_{\lambda,\rho}$ and $d^J_{\lambda,\rho}$ be the distance functions defined by* (2.16) *and* (2.17) *respectively. Then*

$$\text{(2.31)} \qquad d_{\lambda,\rho}(f, g) \leqslant \lambda^{-1}\rho\, d^J_{\lambda,\rho}(f, g) + \alpha_\lambda(f, g),$$

*where*

$$\alpha_\lambda(f,g) = |f_\lambda(0) - g_\lambda(0)| = d_{\lambda,0}(f,g). \tag{2.32}$$

PROOF. We have that

$$f_\lambda(x) = f_\lambda(0) + \int_0^1 \langle \nabla f_\lambda(\tau x), x\rangle\, d\tau,$$

since $f_\lambda$ is finite everywhere and differentiable, see (2.12). The same holds with $g$ and thus

$$f_\lambda(x) - g_\lambda(x) = f_\lambda(0) - g_\lambda(0) + \int_0^1 \langle \nabla f_\lambda(\tau x) - \nabla g_\lambda(\tau x), x\rangle\, d\tau.$$

This yields

$$\begin{aligned}|f_\lambda(x) - g_\lambda(x)| &\leqslant |f_\lambda(0) - g_\lambda(0)| + \int_0^1 |\langle \nabla f_\lambda(\tau x) - \nabla g_\lambda(\tau x), x\rangle|\, d\tau \\ &\leqslant \alpha_\lambda(f,g) + \|x\| \cdot \int_0^1 \|\nabla f_\lambda(\tau x) - \nabla g_\lambda(\tau x)\|\, d\tau,\end{aligned}$$

where the last inequality follows from the Cauchy-Schwarz inequality. Taking the supremum on both sides with $x$ restricted to the closed ball of radius $\rho$, we have

$$d_{\lambda,\rho}(f,g) \leqslant \alpha_\lambda(f,g) + \rho\int_0^1 \lambda^{-1} d^J_{\lambda,\rho}(f,g)\, d\tau,$$

utilizing here the relation (2.18). And this in turn gives (2.31). □

2.33. THEOREM. *Suppose $f$ and $g$ are proper closed convex functions defined on a Hilbert space $X$. For any $\lambda > 0$ and $\rho \geqslant 0$ let $d_{\lambda,\rho}$ and $d^J_{\lambda,\rho}$ be the distance functions defined by* (2.16) *and* (2.17) *respectively. Then*

$$d^J_{\lambda,\rho}(f,g) \leqslant (1+\lambda)\big(2 d_{\lambda,\rho_0}(f,g)\big)^{1/2} \tag{2.34}$$

*for any $\rho_0$ such that*

$$\rho_0 \geqslant (1+\lambda^{-1})\rho + \lambda^{-1}\theta_\lambda(f,g), \tag{2.35}$$

*where*

$$\theta_\lambda(f,g) = \|J^f_\lambda 0\| + \|J^g_\lambda 0\|. \tag{2.36}$$

PROOF. Since $f_\lambda$ and $g_\lambda$ are convex, finite everywhere, and differentiable with gradients $\nabla f_\lambda$ and $\nabla g_\lambda$, we have that for any $x$ and $y$ in $X$:

$$\begin{aligned} f_\lambda(y) - f_\lambda(x) &\geqslant \langle \nabla f_\lambda(x), y - x\rangle, \\ g_\lambda(x) - g_\lambda(y) &\geqslant \langle \nabla g_\lambda(y), x - y\rangle. \end{aligned}$$

Adding them up, these inequalities yield for all $x$ and $y$ in $X$:

$$[f_\lambda(y) - g_\lambda(y)] - [f_\lambda(x) - g_\lambda(x)] \geqslant \langle \nabla f_\lambda(x) - \nabla g_\lambda(y), y - x\rangle. \tag{2.37}$$

Now fix $x$ and choose $y$ such that

$$y := (I + \nabla g_\lambda)^{-1}[(I + \nabla f_\lambda)(x)] = \frac{1}{\lambda + 1}(\lambda I + J^g_{\lambda+1})[(I + \nabla f_\lambda)(x)]; \tag{2.38}$$

the last equality comes from (2.5), which implies that

$$J_1^{g_\lambda}x = (\lambda+1)^{-1}(\lambda x + J_{\lambda+1}^g x) = (I + \nabla g_\lambda)^{-1}x.$$

Thus for all $x \in X$ and $y$ as above, the inequality (2.37) becomes

$$\|\nabla f_\lambda(x) - \nabla g_\lambda(y)\| \leqslant |(f_\lambda(y) - g_\lambda(y)) - (f_\lambda(x) - g_\lambda(x))|^{1/2}. \tag{2.39}$$

Next, we obtain a lower bound for the left-hand side term of (2.39) in terms of $\|\nabla f_\lambda(x) - \nabla g_\lambda(x)\|$. Indeed we have

$$\begin{aligned}\|\nabla f_\lambda(x) - \nabla g_\lambda(x)\| &\leqslant \|\nabla f_\lambda(x) - \nabla g_\lambda(y)\| + \|\nabla g_\lambda(y) - \nabla g_\lambda(x)\|\\ &\leqslant \|\nabla f_\lambda(x) - \nabla g_\lambda(y)\| + \lambda^{-1}\|x - y\|\\ &\leqslant (1+\lambda^{-1})\|\nabla f_\lambda(x) - \nabla g_\lambda(y)\|;\end{aligned}$$

the second inequality follows from the fact that $\nabla g_\lambda$ is Lipschitz with constant $\lambda^{-1}$, and the third inequality from the definition (2.38) of $y$ which implies that

$$x - y = \nabla g_\lambda(y) - \nabla f_\lambda(x).$$

This last inequality with (2.39) implies

$$\|\nabla f_\lambda(x) - \nabla g_\lambda(x)\| \leqslant (1+\lambda^{-1})\big(|f_\lambda(y) - g_\lambda(y)| + |f_\lambda(x) - g_\lambda(x)|\big)^{1/2}. \tag{2.40}$$

Now since $J_{\lambda+1}^g$ is a contraction, so is

$$D_\lambda := \frac{1}{\lambda+1}J_{\lambda+1}^g + \frac{\lambda}{\lambda+1}I$$

and thus for any $z$

$$\|D_\lambda z\| \leqslant \|D_\lambda z - D_\lambda 0\| + \|D_\lambda 0\| \leqslant \|z\| + \frac{1}{\lambda+1}\|J_{\lambda+1}^g 0\|.$$

Also,

$$\begin{aligned}\|(I + \nabla f_\lambda)(x)\| &\leqslant \|x\| + \|\nabla f_\lambda(x) - \nabla f_\lambda(0)\| + \|\nabla f_\lambda(0)\|\\ &\leqslant \frac{\lambda+1}{\lambda}\|x\| + \frac{1}{\lambda}\|J_\lambda^f 0\|,\end{aligned}$$

where to obtain the last inequality we have used the facts that $\nabla f_\lambda$ is Lipschitz with constant $\lambda^{-1}$ and, as follows from (2.12), that

$$\nabla f_\lambda(0) = \frac{1}{\lambda}(0 - J_\lambda^f 0).$$

Thus for $y$, as defined by (2.38), we have

$$\|y\| \leqslant \frac{\lambda+1}{\lambda}\|x\| + \frac{1}{\lambda}\|J_\lambda^f 0\| + \frac{1}{\lambda+1}\|J_{\lambda+1}^g 0\|. \tag{2.41}$$

Noting that for every $x \in X$, where $\lambda \mapsto \|\nabla g_\lambda x\|$ increases as $\lambda$ decreases to zero, we obtain

$$\|\nabla g_{\lambda+1} 0\| \leqslant \|\nabla g_\lambda 0\|,$$

that is

$$\frac{1}{\lambda+1}\|J^g_{\lambda+1}0\| \leqslant \frac{1}{\lambda}\|J^g_\lambda 0\|.$$

Returning to (2.41) we get

$$\|y\| \leqslant \frac{\lambda+1}{\lambda}\|x\| + \lambda^{-1}\big(\|J^f_\lambda 0\| + \|J^g_\lambda 0\|\big).$$

Taking the supremum on both sides of (2.40) with $\|x\| \leqslant \rho$ and appealing to (2.18) and the above inequality, we obtain

$$\lambda^{-1}d^J_{\lambda,\rho}(f,g) \leqslant (1+\lambda^{-1})\Big[\sup_{\|y\|\leqslant\rho_0}|f_\lambda(y)-g_\lambda(y)| + \sup_{\|x\|\leqslant\rho}|f_\lambda(x)-g_\lambda(x)|\Big]^{1/2}$$

with

$$\rho_0 = (1+\lambda^{-1})\rho + \lambda^{-1}\big(\|J^f_\lambda 0\| + \|J^g_\lambda 0\|\big).$$

Since $\rho \leqslant \rho_0$, this yields (2.34). □

2.42. COROLLARY. *Suppose f and g are proper closed convex functions defined on a Hilbert space X such that*

$$f(0) = \inf f = 0 = \inf g = g(0).$$

*Then for any* $\lambda > 0$ *and* $\rho \geqslant 0$ *we have*

(2.43) $$\lambda d_{\lambda,\rho}(f,g) \leqslant \rho d^J_{\lambda,\rho}(f,g) \leqslant \rho(1+\lambda)\big[2d_{\lambda,(1+\lambda^{-1})\rho}(f,g)\big]^{1/2}.$$

*In particular, when* $\lambda = 1$, *this implies*

(2.44) $$d_{1,\rho}(f,g) \leqslant \rho d^J_{1,\rho}(f,g) \leqslant 3\rho\big(d_{1,2\rho}(f,g)\big)^{1/2}.$$

The inequalities (2.31) and (2.34), summarized in (2.43) in the "normalized" case ($f(0) = 0 = \inf f$), make explicit the relationship between the uniform structures induced by $d_{\lambda,\rho}$ and $d^J_{\lambda,\rho}$ on SCC($X$). If we restrict ourselves to the convex cone of "normalized" functions in SCC($X$), i.e. such that $f(0) = 0 = \inf f$, then

(2.45) $$d_{\lambda,\rho}(f,g) \leqslant \rho\lambda^{-1}d^J_{\lambda,\rho}(f,g) \leqslant \rho(1+\lambda^{-1})\big[2d_{\lambda,\rho(1+\lambda^{-1})}(f,g)\big]^{1/2}.$$

Let $\{f^\nu\colon X \to \overline{R},\ \nu = 1,\ldots\}$ be a sequence of proper convex functions, and suppose they are "normalized" as defined above. Then (2.45) allows us to compare the convergence "rate" of the functions, or at least of their Moreau-Yosida approximates of parameter $\lambda$, and that of their resolvents. And either one of these then give us a measure of the convergence rate of a sequence of epi-convergent functions as we show next. Let us begin with a couple of preliminary lemmas that are of independent interest in various applications of these results.

2.46. LEMMA. *Suppose X is a Hilbert space, with f and g proper, lower semicontinuous convex functions defined on X. Then for all* $\lambda > 0$, $\mu > 0$ *and* $\rho \geqslant 0$

$$d_{\lambda+\mu,\rho}(f,g) \leqslant d_{\lambda,\rho_0}(f,g)$$

*for any* $\rho_0$ *such that*

$$\rho_0 \geqslant \rho + \mu\lambda^{-1}\max\big\{\|J^f_\lambda 0\|, \|J^g_\lambda 0\|\big\}.$$

PROOF. Let $\lambda > 0$ and $\mu > 0$ and recall that, in view of (2.5), $(f_\lambda)_\mu = f_{\lambda+\mu}$. Hence

$$f_{\lambda+\mu}(x) = \inf_{v\in X}\left\{ f_\lambda(v) + \frac{1}{2\mu}\|x - v\|^2\right\}$$

and

$$g_{\lambda+\mu}(x) = \inf_{v\in X}\left\{ g_\lambda(v) + \frac{1}{2\mu}\|x - v\|^2\right\}.$$

Suppose

$$v_\lambda^f := \arg\min\left( f_\lambda(\cdot) + \frac{1}{2\mu}\|x - \cdot\|^2\right).$$

Then

$$0 = \nabla f_\lambda(v_\lambda^f) + \frac{1}{\mu}(v_\lambda^f - x)$$

which yields

$$v_\lambda^f = (I + \mu\nabla f_\lambda)^{-1}(x).$$

We also have that

$$f_{\lambda+\mu}(x) = f_\lambda(v_\lambda^f) + \frac{1}{2\mu}\|x - v_\lambda^f\|^2$$

and, of course,

$$g_{\lambda+\mu}(x) \leqslant g_\lambda(v_\lambda^f) + \frac{1}{2\mu}\|x - v_\lambda^f\|^2,$$

and these two relations imply

$$g_{\lambda+\mu}(x) - f_{\lambda+\mu}(x) \leqslant g_\lambda(v_\lambda^f) - f_\lambda(v_\lambda^f). \tag{2.47}$$

Since $x \mapsto J_\mu^{f_\lambda}x = v_\lambda^f$ is a contraction,

$$\|v_\lambda^f\| \leqslant \|v_\lambda^f - J_\mu^{f_\lambda}0\| + \|J_\mu^{f_\lambda}0\| \leqslant \|x - 0\| + \|J_\mu^{f_\lambda}0\|.$$

On the other hand, from the equality $\nabla(f_\lambda)_\mu = \nabla f_{\lambda+\mu}$, it follows that

$$J_\mu^{f_\lambda}x = (\lambda+\mu)^{-1}\left[\lambda x + \mu J_{\lambda+\mu}^f x\right],$$
$$J_\mu^{f_\lambda}0 = (\lambda+\mu)^{-1}\mu J_{\lambda+\mu}^f 0,$$

and thus

$$\|v_\lambda^f\| \leqslant \|x\| + \mu(\lambda+\mu)^{-1}\|J_{\lambda+\mu}^f 0\|.$$

Returning to (2.47), for any $\rho \geqslant 0$

$$\sup_{\|x\|\leqslant\rho}\left(g_{\lambda+\mu}(x) - f_{\lambda+\mu}(x)\right) \leqslant \sup_{\|x\|\leqslant\rho_0}\left(g_\lambda(x) - f_\lambda(x)\right)$$

with

$$\rho_0 \geqslant \rho + \mu(\lambda+\mu)^{-1}\max\left\{\|J_{\lambda+\mu}^f 0\|, \|J_{\lambda+\mu}^g 0\|\right\}.$$

Observe that $\lambda \mapsto \|\nabla_\lambda^f x\|$ and $\lambda \mapsto \|\nabla_\lambda^g x\|$ are monotonically decreasing functions, and thus the above inequality is satisfied if

$$\rho_0 \geqslant \rho + \mu\lambda^{-1} \max\left\{\|J_\lambda^f 0\|, \|J_\lambda^g 0\|\right\}.$$

Repeating the same argument, but interchanging the role of $f$ and $g$, and using the definition of $d_{\lambda,\rho}$ yields

$$d_{\lambda+\mu,\rho}(f,g) \leqslant d_{\lambda,\rho_0}(f,g). \qquad \square$$

2.48. Lemma. *Suppose $X$ is a Hilbert space, with $f$ and $g$ proper convex lower semicontinuous functions defined on $X$. Then for any $\lambda > 0$, $\mu > 0$ and $\rho \geqslant 0$*

$$d_{\lambda,\rho}(f,g) \leqslant d_{\lambda+\mu,\rho'}(f,g),$$

*where*

$$\rho' := (1 + \mu\lambda^{-1})\rho + \mu\lambda^{-1} \max\left\{\|J_\lambda^f 0\|, \|J_\lambda^g 0\|\right\}.$$

Proof. With the same construction as in the proof of Lemma 2.46 we have, from (2.47), that

$$f_\lambda(v_\lambda^f) - g_\lambda(v_\lambda^f) \leqslant f_{\lambda+\mu}(x) - g_{\lambda+\mu}(x).$$

Observe that $(I + \mu\nabla f_\lambda)(v_\lambda^f) = x$. Given any $y \in X$, taking $x = (I + \mu\nabla f_\lambda)(y)$ we obtain $v_\lambda^f(x) = y$ and thus

$$f_\lambda(y) - g_\lambda(y) \leqslant f_{\lambda+\mu}(w_\lambda^f) - g_{\lambda+\mu}(w_\lambda^f) \tag{2.49}$$

with $w_\lambda^f := (I + \mu\nabla f_\lambda)(y)$. Note that

$$\|w_\lambda^f\| \leqslant \|y\| + \mu\lambda^{-1}\left(\|y\| + \|J_\lambda^f 0\|\right).$$

When $\|y\| \leqslant \rho$, it follows that

$$\|w_\lambda^f\| \leqslant (1 + \mu\lambda^{-1})\rho + \mu\lambda^{-1}\|J_\lambda^f 0\| \leqslant \rho'.$$

Taking suprema on both sides of (2.49) with $\|y\| \leqslant \rho$ yields

$$\sup_{\|y\|\leqslant\rho} \left(f_\lambda(y) - g_\lambda(y)\right) \leqslant \sup_{\|y\|\leqslant\rho'} \left(f_{\lambda+\mu}(y) - g_{\lambda+\mu}(y)\right).$$

The same holds when the roles of $f$ and $g$ are interchanged, and this completes the proof. $\square$

As a direct consequence of Lemmas 2.46 and 2.48 we obtain

2.50. Corollary. *Suppose $X$ is a Hilbert space with $f$ and $g$ proper lower semicontinuous convex functions defined on $X$ such that $f(0) = g(0) = \inf f = \inf g$. Then for all $\lambda > 0$, $\mu > 0$ and $\rho \geqslant 0$*

$$d_{\lambda+\mu,\rho}(f,g) \leqslant d_{\lambda,\rho}(f,g) \leqslant d_{(\lambda+\mu),\rho(1+\lambda^{-1}\mu)}(f,g).$$

2.51. Theorem. *Suppose $X$ is a Hilbert space, and $\{f^\nu\colon X \to \overline{R},\ \nu = 1,\dots\}$ and $f\colon X \to \overline{R}$ are a collection of proper closed convex functions, such that for some $\lambda > 0$ and all $\rho \geqslant 0$,*

$$\lim_{\nu\to\infty} d_{\lambda,\rho}(f, f^\nu) = 0.$$

*Then*

$$f = \text{Mosco-epi-lim}_{\nu\to\infty} f^\nu.$$

*Moreover, if X is finite dimensional, then the reverse implication is also valid.*

PROOF. In view of [**9**, Théorème 1.2] a sequence of proper convex functions $\{f^\nu, \nu = 1,\dots\}$ Mosco-epi-converges to $f$ if and only if for all $\lambda > 0$, the Moreau-Yosida approximates of parameter $\lambda$ converge pointwise to $f_\lambda$, i.e. for all $x \in X$ and all $\lambda > 0$,

$$\lim_{\nu\to\infty} f_\lambda^\nu(x) = f_\lambda(x).$$

Now, if

$$\lim_{\nu\to\infty} d_{\lambda,\rho}(f, f^\nu) = 0 \tag{2.52}$$

for some $\lambda = \lambda_0 > 0$ and all $\rho \geqslant 0$, it follows from Lemmas 2.46 and 2.48 that the above holds for all $\lambda > 0$. The first one of these lemmas gives the convergence for all $\lambda \geqslant \lambda_0$, and the second one for all $0 < \lambda < \lambda_0$, since it implies that

$$d_{\lambda,\rho}(f, g) \leqslant d_{\lambda_0,\rho_0}(f, g),$$

where

$$\rho_0 = \left(1 + (\lambda_0 - \lambda)\lambda^{-1}\right)\rho + (\lambda_0 - \lambda)\lambda^{-1}\left(\max\left[\|J_\lambda^f 0\|, \|J_\lambda^g 0\|\right]\right).$$

Since (2.52) implies the uniform (pointwise) convergence on all balls of radius $\rho$ of the $f_\lambda^\nu$ to $f_\lambda$, and this for all $\lambda > 0$, we have that $f$ is the Mosco-epi-limit of the $f^\nu$. To obtain the converse observe that the convex functions $\{f_\lambda^\nu, \nu = 1,\dots\}$ and $f_\lambda$ are equi-locally Lipschitz, and this combined with pointwise convergence implies, by the Arzelà-Ascoli Theorem, the uniform convergence on compact sets, which in finite dimensions are the closed bounded sets. □

Note that we used Lemmas 2.46 and 2.48 to pass from requiring that $d_{\lambda,\rho}(f, f^\nu)$ goes to 0 for *some* $\lambda > 0$ instead of for *all* $\lambda > 0$. This also confirms that the epi-convergence engendered by the convergence of the distance functions $d_{\lambda,\rho}$ is strictly stronger (in infinite dimensions) than the Mosco-epi-convergence, since having

$$f_\lambda = \text{pointwise-lim} f_\lambda^\nu$$

for some $\lambda > 0$ is not sufficient to ensure that $f$ is the Mosco-epi-limit of the $f^\nu$. It is also easy to see that we could not obtain the epi-convergence of the sequence $\{f^\nu, \nu = 1,\dots\}$ to $f$ by requiring that $d_{\lambda,\rho}(f, f^\nu)$ goes to 0 for all $\lambda > 0$ and some $\rho_0 > 0$. So, we may feel that Theorem 2.51 is in this setting the best result possible.

Let us finally notice that Theorem 2.51 is certainly valid in a reflexive Banach space. One has to extend Lemmas 2.46 and 2.48 to this setting and rely on [**8**, Theorem 3.26].

The next result could be obtained from the equivalence of Mosco-epi-convergence and the convergence of the resolvents [**9**, Théorème 1.2(c),(d)]. However, it is easier to obtain it here as a corollary of the previous theorem, Proposition 2.30 and Theorem 2.33.

2.53. Corollary. *Suppose $X$ is a Hilbert space, and $\{f^\nu\colon X \to \overline{R}, \nu = 1, \dots\}$ and $f\colon X \to \overline{R}$ are proper convex lower semicontinuous functions. If for some $\lambda > 0$ and all $\rho \geqslant 0$*

$$\lim_{\nu\to\infty} d^J_{\lambda,\rho}(f, f^\nu) = 0 \quad \textit{and} \quad \lim_{\nu\to\infty} d_{\lambda,0}(f, f^\nu) = 0,$$

*then*

$$f = \operatorname*{Mosco\text{-}epi\text{-}lim}_{\nu\to\infty} f^\nu.$$

*Moreover, if $X$ is finite dimensional, then the converse is also true.*

Proof. It really suffices to observe that the hypotheses of the corollary imply that for all $\lambda > 0$ and $\rho \geqslant 0$

$$\lim_{\nu\to\infty} d_{\lambda,\rho}(f, f^\nu) = 0, \tag{2.54}$$

as follows from (2.31), and then appeal to the theorem to complete the proof of the first claim. In the other direction, we first rely on Theorem 2.51 to obtain (2.54) from which it follows (2.34) that

$$\lim_{\nu\to\infty} d^J_{\lambda,\rho}(f, f^\nu) = 0$$

for all $\rho \geqslant 0$, since (2.54) also implies that $\lim_{\nu\to\infty} d_{\lambda,0}(f, f^\nu) = 0$. □

To conclude this section let us describe a situation that covers a number of important applications, where the metrics $d_{\lambda,\rho}$ and $d^J_{\lambda,\rho}$ appear as the appropriate concepts for measuring distance between convex functions.

We write

$$f = \tau_X\text{-}\operatorname*{epi\text{-}lim}_{\nu\to\infty} f^\nu$$

if epi-convergence is with respect to the $\tau$-topology on $X$, i.e. for all $x$ in $X$, (2.3) must hold for all sequences $\{x^\nu, \nu = 1, \dots\}$ $\tau$-converging to $x$ and (2.4) for some $\tau$-converging sequence. Recall that a collection of functions $\{f^\alpha\colon X \to \overline{R}, \alpha \in A\}$ is said to be *equi-coercive* if there exist a function $\theta\colon R_+ \to [0, \infty]$ with $\lim_{r\to\infty}\theta(r) = \infty$ such that for all $\alpha \in A$

$$f^\alpha(x) \geqslant \theta(\|x\|_X) \quad \text{for all } x \in X.$$

2.55. Theorem. *Suppose $X$ and $H$ are two Hilbert spaces and $X \hookrightarrow H$ is a continuous compact embedding. Then, for any collection $\{f; f^\nu, \nu = 1, \dots\}$ of proper, equi-coercive, lower semicontinuous, convex functions defined on $X$, the following four assertions are equivalent*:

(i) $f = \text{weak}_X\text{-epi-lim}_{\nu\to\infty} f^\nu$;

(ii) $f = \text{Mosco-epi-lim}_{\nu\to\infty} f^\nu$ *on $H$*;

(iii) *for all* $\rho \geqslant 0$, $\lim_{\nu\to\infty} d_{1,\rho}(f, f^\nu) = 0$;

(iv) *for all* $\rho \geqslant 0$, $\lim_{\nu\to\infty} d^J_{1,\rho}(f, f^\nu) = 0$ *and* $\lim_{\nu\to\infty} d_{\lambda,0}(f, f^\nu) = 0$;

*where $d_{\lambda,\rho}$ and $d^J_{\lambda,\rho}$ are defined in terms of the norm $\|\cdot\|_H$ on $H$.*

PROOF. Because of the coercivity assumption, each function $\tilde{f}^\nu$ on $H$ defined by

$$\tilde{f}^\nu(x) = \begin{cases} f^\nu(x) & \text{if } x \in X, \\ +\infty & \text{if } x \in H \setminus X \end{cases}$$

is also a proper, lower semicontinuous convex function. We simply write $f^\nu$ when no ambiguity is possible.

(i) → (ii). We begin by verifying the appropriate version of (2.4) for the Mosco-epi-convergence, i.e. to every $x \in H$ there corresponds a strongly convergent sequence $\{x^\nu \in H,\ \nu = 1, \dots\}$ such that $\limsup_{\nu\to\infty} f^\nu(x^\nu) \leqslant f(x)$. There is nothing to prove if $f(x) = \infty$. If $f(x) < \infty$, then $x \in X$ and from (i) it follows that there exists a weakly convergent sequence $\{x^\nu \in X,\ \nu = 1, \dots\}$ that gives the desired inequality. This sequence is strongly converging in $H$ since $\hookrightarrow$ is a compact injection from $X$ into $H$.

Now to establish (2.3), pick any sequence $\{x^\nu \in H,\ \nu = 1, \dots\}$ weakly converging to $x$ in $H$. We have to show that

$$\liminf_{\nu\to\infty} f^\nu(x^\nu) \geqslant f(x).$$

If $\liminf_{\nu\to\infty} f^\nu(x^\nu) = \infty$, the inequality is clearly satisfied. Otherwise, passing to a subsequence if necessary, we may assume that

$$\liminf_{\nu\to\infty} f^\nu(x^\nu) = \lim_{\nu\to\infty} f^\nu(x^\nu) < \infty.$$

From the equi-coercivity of the sequence $\{f^\nu,\ \nu = 1, \dots\}$, and the compact embedding $X \hookrightarrow H$, it follows that for $\nu$ sufficiently large the $x^\nu$ are in $X$ and the sequence is weakly converging *in* $X$ to $x$. The desired inequality now follows from (2.3) itself since $f$ is the weak$_X$-epi-limit of the $f^\nu$.

(ii) → (i). The argument is similar to the preceding one, simplified by the fact that we are now going from convergence in $H$ to convergence in $X$.

(ii) → (iv). We show that to every $\rho \geqslant 0$ and $\varepsilon > 0$ there corresponds $\nu_\varepsilon$ such that for all $\nu \geqslant \nu_\varepsilon$ and $x \in H$ with $\|x\|_H \leqslant \rho$,

$$\|J_1^\nu x - J_1 x\|_H \leqslant \varepsilon,$$

where $J_1^\nu = J_1^{f^\nu}$ and $J_1 = J_1^f$. We argue by contradiction. Assume that for some $\rho_0 > 0$ and $\varepsilon_0 > 0$ there exists a sequence $\{\nu(k),\ k = 1, \dots\}$ that goes to $+\infty$, and a sequence $\{x^k \in H,\ k = 1, \dots\}$ bounded in norm by $\rho_0$ such that for all $k$:

$$\|J_1^{\nu(k)} x^k - J_1 x^k\| > \varepsilon_0. \tag{2.56}$$

Since $f = \text{Mosco-epi-lim}_{\nu\to\infty} f^\nu$ (on $H$) we have that for all $x$

$$f_1(x) = \lim_{\nu\to\infty} f_1^\nu(x),$$

cf. (2.6), and

$$J_1 x = \text{strong}_H\text{-}\lim_{\nu\to\infty} J_1^\nu x. \tag{2.57}$$

From the definition (2.13) of $J_1^\nu x$, we have

$$f_1^\nu(x) = f^\nu(J_1^\nu x) + \tfrac{1}{2}\|x - J_1^\nu x\|_H^2,$$

which implies

$$f_1^\nu(x) \geqslant f^\nu(J_1^\nu x).$$

Pick $x^0 \in \operatorname{dom} f$. Then again $f = \text{Mosco-epi-lim}_{\nu\to\infty} f^\nu$ implies the existence of a sequence $\{x^{0\nu},\ \nu = 1,\dots\}$ strongly converging in $H$ to $x^0$ such that

$$f(x^0) = \lim_{\nu\to\infty} f^\nu(x^{0\nu});$$

using here conditions (2.4) and (2.3). By definition of $f_1^\nu$

$$f_1^\nu(x) \leqslant f^\nu(x^{0\nu}) + \tfrac{1}{2}\|x - x^{0\nu}\|_H,$$

and hence from the above, for $k = 1,\dots$

$$f^{\nu(k)}\left(J_1^{\nu(k)}x^k\right) \leqslant f^{\nu(k)}\left(x^{0,\nu(k)}\right) + \tfrac{1}{2}\|x^k - x^{0,\nu(k)}\|_H.$$

Now, the $\{x^{0,\nu(k)},\ k = 1,\dots\}$ converge to $x^0$ and the $\{x^k,\ k = 1,\dots\}$ are norm-bounded (by $\rho_0$). Thus

$$\sup_k f^{\nu(k)}\left(J_1^{\nu(k)}x^k\right) < \infty,$$

which implies that

$$\sup_k \|J_1^{\nu(k)}x^k\| < \infty$$

since the collection $\{f^{\nu(k)},\ k = 1,\dots\}$ is equi-coercive. This means that the sequence $\{J_1^{\nu(k)}x^k,\ k = 1,\dots\}$ is (strongly) relatively compact in $H$ since $\hookrightarrow$ is a compact embedding of $X$ into $H$. Similarly, the sequence $\{J_1x^k,\ k = 1,\dots\}$ is relatively compact in $H$. We can thus extract subsequences that we still denote by $\{x^k,\ k = 1,\dots\}$ and $\{J_1^{\nu(k)},\ k = 1,\dots\}$ such that

$$(2.58)\qquad \begin{aligned} x &= \text{weak}_H\text{-}\lim_{k\to\infty} x^k,\\ u &= \text{strong}_H\text{-}\lim_{k\to\infty} J_1x^k,\\ v &= \text{strong}_H\text{-}\lim_{k\to\infty} J_1^{\nu(k)}x^k \end{aligned} \quad ,$$

for some $u$ and $v$ in $H$.

For any $y$ in $H$, from the monotonicity of the operator $\partial f^\nu$—the functions $f^\nu$ are convex—and (2.13) the definition of $J_\lambda x$ which implies $x - J_1^\nu x \in \partial f^\nu(J_1^\nu x)$ and $y - J_1^\nu y \in \partial f^\nu(J_1^\nu y)$, we have

$$\left\langle J_1^\nu y - J_1^\nu x, (y - J_1^\nu y) - (x - J_1^\nu x)\right\rangle \geqslant 0,$$

that is

$$\left\langle J_1^\nu y - J_1^\nu x,\ y - x\right\rangle \geqslant \|J_1^\nu y - J_1^\nu x\|_H^2.$$

From the continuity of $x \mapsto J_1^\nu x$ it follows that these operators are maximal monotone [**13**] from $H$ into $H$. Moreover, the sequence of operators $\{J_1^\nu,\ \nu = 1,\dots\}$ is graph convergent to $J_1$ as follows from (2.57), which implies the following closure property, see [**9**]:

(2.59) *whenever* $x = \text{weak}_H\text{-}\lim_{\nu\to\infty} x^\nu$; *and* $y = \text{strong}_H\text{-}\lim_{k\to\infty} y^\nu$, *and* $y^\nu = J_1^\nu x^\nu$, *it follows that* $y = J_1x$.

Using this result, with the sequences identified by (2.58), it yields

$$J_1 x = \text{strong}_H\text{-}\lim_{k\to\infty} J_1 x^k; \qquad J_1 x = \text{strong}_H\text{-}\lim_{k\to\infty} J_1^{\nu(k)} x^k;$$

which clearly contradicts (2.56).

(iii) → (iv) → (ii). This follows from Proposition 2.30 and Theorem 2.51. □

**3. Isometry for Wijsman-approximates.** In [**6**, §6] it is shown that the study of the epi-convergence of a sequence $\{f^\nu, \nu = 1, \dots\}$ to a limit function $f$ can be reduced to the study of the convergence of a parametrized family of functions $\{f^\nu(\lambda; \cdot);$ $\nu = 1, \dots, \lambda > 0\}$ obtained from the $f^\nu$ by inf-convoluting them with a collection $\{g(\lambda, \cdot), \lambda > 0\}$, called a *cast* in [**6**]. One is allowed to choose this collection $\{g(\lambda, \cdot), \lambda > 0\}$ so as to endow the "regularized" functions $\{f^\nu(\lambda; \cdot); \nu = 1, \dots,$ $\lambda > 0\}$ with some desired properties, provided naturally that as $\lambda \downarrow 0$ the $f^\nu(\lambda; \cdot)$ converge (pointwise) to $f^\nu$, and that if $f = \text{epi-lim}_{\nu\to\infty} f^\nu$, then a formula of the type

$$\lim_{\lambda\downarrow 0} \lim_{\nu\to\infty} f^\nu(\lambda, \cdot) = f$$

holds. One possibility is to choose for $\lambda > 0$

$$g(\lambda, \cdot) = \frac{1}{2\lambda}\|\cdot\|^2.$$

This leads to the Moreau-Yosida approximates $\{f_\lambda^\nu, \nu = 1, \dots\}$, introduced in §2, that have played such an important role in the theory and the applications of epi-convergence beginning with [**9**]; for a more recent account consult [**8**]. Another possibility is to choose the $g(\lambda, \cdot)$ so that they are adapted to the sequence in question such as done by Fougères and Truffert [**14**] in their work on lower semicontinuous regularizations, in particular of integral functionals, by a reference function. Each cast $\{g(\lambda, \cdot), \lambda > 0\}$ is potentially the source of new isometries for the Legendre-Fenchel transform. In this section, we work with $g(\lambda, \cdot) = \lambda^{-1}\|\cdot\|$ to construct for given $f$, the "regularized" function:

$$f_{[\lambda]}(x) := (f \,\square\, \lambda^{-1}\|\cdot\|)(x) = \inf_y [f(y) + \lambda^{-1}\|x - y\|]. \tag{3.1}$$

We refer to $f_{[\lambda]}$ as the *Wijsman-approximate of $f$ of parameter* $\lambda$ in recognition of the role played by this type of function in the seminal work of Wijsman [**1**]. Of course, we have that

$$\text{epi}\, f_{[\lambda]} = \{(x, \alpha = \inf \alpha') \,|\, (x, \alpha') \in \text{epi}\, f + \text{epi}\, \lambda^{-1}\|\cdot\|\}$$

and if $f$ is a proper lower semicontinuous convex function, then so is $f_{[\lambda]}$ for all $\lambda > 0$. Moreover, $f_{[\lambda]}$ is Lipschitz assuming only that there exists a constant $\gamma \geqslant 0$ such that $f(x) + \gamma(\|x\| + 1) \geqslant 0$ for all $x$ (see [**14**, Theorem 3.2]). Here, unless specified, we let $X$ be a Banach space with norm $\|\cdot\|$ paired with its dual $X^*$ through the bilinear form $\langle \cdot, \cdot \rangle$. The norm on $X^*$ is denoted by $\|\cdot\|_*$. Let $f$: $X \to \overline{R}$ be a proper lower semicontinuous convex function. Then

$$(f_{[\lambda]})^*(v) = f^*(v) + \psi_{\lambda^{-1}B_*}(v), \tag{3.2}$$

where $\psi_C$ is the *indicator function* of the set $C$, i.e.

$$\psi_C(x) = \begin{bmatrix} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise,} \end{bmatrix} \tag{3.3}$$

and $B_* = \{v \mid \|v\|_* \leqslant 1\}$ is the unit ball of $X^*$. Also

3.4. LEMMA. *Suppose $X$ is a reflexive Banach space and $f$ is a proper lower semicontinuous convex function defined on $X$. Then*

$$f_{[\lambda]}(x) = \sup_{\|v\|_* \leqslant \lambda^{-1}} \left[\langle v, x\rangle - f^*(v)\right]. \tag{3.5}$$

PROOF. We have

$$\begin{aligned} f_{[\lambda]}(x) &= \inf_y \left[f(x-y) + \lambda^{-1}\|y\|\right] \\ &= \inf_y \sup_v \left[\langle v, x-y\rangle - f^*(v) + \lambda^{-1}\|y\|\right] \\ &= \sup_v \left[-f^*(v) + \langle v, x\rangle - \sup_y \left(\langle v, y\rangle - \lambda^{-1}\|y\|\right)\right] \\ &= \sup_v \left[\langle v, x\rangle - f^*(v) - \psi_{\lambda^{-1}B_*}(v)\right], \end{aligned}$$

where the interchange of $\inf_y$ and $\sup_v$ can be justified by the same arguments as those used in the proof of Theorem 2.9; we have also used the fact that $\|\cdot\|^* = \psi_{B_*}$. □

To begin with we exhibit a relationship between indicator and support functions. We need a couple of lemmas, whose proofs we include for easy reference. The second one is due to Hörmander [**15**], whereas the first one is, or should be, part of the folklore. Recall that if $C$ and $D$ are two nonempty subsets of $X$, the Hausdorff distance between $C$ and $D$ is given by

$$\operatorname{haus}(C, D) := \sup\left[\sup_{x \in D} \operatorname{dist}(x, C),\ \sup_{x \in C} \operatorname{dist}(x, D)\right], \tag{3.6}$$

where $\operatorname{dist}(x, C) := \inf_{y \in C}\|x - y\|$.

3.7. LEMMA. *Suppose $C$ and $D$ are nonempty subsets of a Banach space $X$ with norm $\|\cdot\|$ such that* haus$(C, D)$ *is finite. Then*

$$\operatorname{haus}(C, D) = \sup_{x \in X} |\operatorname{dist}(x, C) - \operatorname{dist}(x, D)|. \tag{3.8}$$

PROOF. Given any nonempty set $S$, and any pair $(y, z)$ in $X$, we always have

$$\|y - z\| + \operatorname{dist}(y, S) \geqslant \operatorname{dist}(z, S). \tag{3.9}$$

Indeed, for any $\varepsilon > 0$, let

$$x_\varepsilon \in \varepsilon\text{-arg min}\left(\|y - \cdot\| + \psi_S\right) := \left\{ x \in S \mid \|y - x\| - \varepsilon \leqslant \operatorname{dist}(y, S)\right\},$$

Then

$$\|y - z\| + \operatorname{dist}(y, S) + \varepsilon \geqslant \|y - z\| + \|y - x_\varepsilon\| \geqslant \|z - x_\varepsilon\| \geqslant \operatorname{dist}(z, S)$$

which yields (3.9) since the above holds for all $\varepsilon > 0$. Now take any point $x \in X$, then for any $\varepsilon > 0$ there always exists $y \in C$ such that

$$|\operatorname{dist}(x, C) - \operatorname{dist}(x, D)| \leqslant |\operatorname{dist}(y, D)| + \varepsilon, \tag{3.10}$$

or a point $z \in D$ such that

$$|\operatorname{dist}(x, C) - \operatorname{dist}(x, D)| \leqslant |\operatorname{dist}(z, D)| + \varepsilon. \tag{3.11}$$

If $\operatorname{dist}(x, C) = \operatorname{dist}(x, D)$ there is nothing to prove. Suppose that

$$\alpha := \operatorname{dist}(x, D) - \operatorname{dist}(x, C) > 0.$$

For $\nu = 1, \ldots,$ let

$$y^\nu \in (1/\nu)\text{-arg min}\big(\|x - \cdot\| + \psi_C\big).$$

Then by (3.9), for $\nu = 1, \ldots$

$$\operatorname{dist}(y^\nu, D) \geqslant \operatorname{dist}(x, D) - \|x - y^\nu\| \geqslant \operatorname{dist}(x, D) - \operatorname{dist}(x, C) - \nu^{-1}$$

which gives (3.10). We have to rely on (3.11) if $\alpha < 0$. This means that

$$\sup_{x \in X} |\operatorname{dist}(x, C) - \operatorname{dist}(x, D)| = \sup_{x \in C \cup D} |\operatorname{dist}(x, C) - \operatorname{dist}(x, D)|,$$

which in view of (3.10) and (3.11) can also be written:

$$\sup\left[\sup_{x \in C} \operatorname{dist}(x, D),\ \sup_{x \in D} \operatorname{dist}(x, C)\right]$$

and this is the definition of the Hausdorff distance. □

The conjugate of the indicator function $\psi_S$ of a set $S$ is the *support function* of $S$ denoted by $\psi_S^*$, i.e.

$$\psi_S^*(v) = \sup\big[\langle v, x\rangle \,|\, x \in S\big].$$

3.12. Lemma [15]. *Suppose $C$ and $D$ are nonempty closed convex subsets of a reflexive Banach space $X$ with norm $\|\cdot\|$ such that* $\operatorname{haus}(C, D)$ *is finite. Then*

$$\operatorname{haus}(C, D) = \sup_{\|v\|_* \leqslant 1} |\psi_C^*(v) - \psi_D^*(v)|. \tag{3.13}$$

Proof. First, observe that for $C$ nonempty convex we have

$$\begin{aligned}\operatorname{dist}(x, C) &= \big(\|\cdot\| \,\square\, \psi_C\big)(x) \\ &= \sup_v \Big[\langle x, v\rangle - \big(\|\cdot\| \,\square\, \psi_C\big)^*(v)\Big] \\ &= \sup_v \Big[\langle x, v\rangle - \big(\psi_{\|v\|_* \leqslant 1}(v) + \psi_C^*(v)\big)\Big] \\ &= \sup_{\|v\|_* \leqslant 1} \big[\langle x, v\rangle - \psi_C^*(v)\big] = (\psi_C)_{[1]}(x);\end{aligned}$$

compare with (3.5). Now

$$\operatorname{haus}(C, D) = \inf\left[\theta \,\middle|\, \theta \geqslant \sup_{x \in C} \operatorname{dist}(x, D),\ \theta \geqslant \sup_{x \in D} \operatorname{dist}(x, C)\right]$$

as follows from the definition of the Hausdorff distance. Since

$$\sup_{x\in D} \operatorname{dist}(x,C) = \sup_{x\in D} \sup_{\|v\|_*\leqslant 1} \left[\langle x,v\rangle - \psi_C^*(v)\right]$$

$$= \sup_{\|v\|_*\leqslant 1} \left[-\psi_C^*(v) + \sup_{x\in D}\langle x,v\rangle\right] = \sup_{\|v\|_*\leqslant 1}\left[\psi_D^*(v) - \psi_C^*(v)\right]$$

we have that

$$\operatorname{haus}(C,D) = \inf\left[\theta|\theta \geqslant \sup_{\|v\|_*\leqslant 1}\left(\psi_C^*(v) - \psi_D^*(v)\right),\ \theta \geqslant \sup_{\|v\|_*\leqslant 1}\left(\psi_D^*(v) - \psi_C^*(v)\right)\right]$$

which is just another version of (3.13). □

Equipped with these two formulas, we are now ready to state the main result of this section. For any pair $(f, g)$ of proper functions defined on $X$, we define the Wijsman *distance* $d_{[\lambda],\rho}$ between $f$ and $g$ as follows:

$$d_{[\lambda],\rho}(f,g) = \sup_{\|x\|\leqslant\rho} \left|f_{[\lambda]}(x) - g_{[\lambda]}(x)\right|. \tag{3.14}$$

If the functions are defined on the dual space $X^*$ then we write $d^*_{[\lambda],\rho}$ to emphasize the fact that the dual norm has been used in the definition of the Wijsman distance.

3.15. THEOREM. *Suppose $X$ is a Hilbert space and $C$, $D$ are closed convex subsets of $X$ such that $0 \in C \cap D$. Then, for all $\lambda > 0$ and $\rho > 0$ we have*

$$d_{[\lambda],\rho}(\psi_C,\psi_D) = d_{[\rho^{-1}],\lambda^{-1}}(\psi_C^*,\psi_D^*). \tag{3.16}$$

PROOF. By definition (3.14) of $d_{[\lambda],\rho}$

$$d_{[\lambda],\rho}(\psi_C,\psi_D) = \sup_{\|x\|\leqslant\rho}\left|(\psi_C)_{[\lambda]}(x) - (\psi_D)_{[\lambda]}(x)\right|$$

$$= \lambda^{-1}\sup_{\|x\|\leqslant\rho}\left|\operatorname{dist}(x,C) - \operatorname{dist}(x,D)\right|$$

$$= \lambda^{-1}\sup_{\|x\|\leqslant\rho}\left|\operatorname{dist}(x,C\cap\rho B) - \operatorname{dist}(x,D\cap\rho B)\right|,$$

where $B = \{x|\|x\| \leqslant 1\}$ is the unit ball in $X$. The second equality follows from the definition of $(\psi_C)_{[\lambda]}$, and the third one from the fact that, whenever $C$ is a closed convex set containing the origin and $x \in X$ satisfies $\|x\| \leqslant \rho$, then

$$\operatorname{dist}(x,C) = \operatorname{dist}(x,C\cap\rho B). \tag{3.17}$$

Let us first notice that $\operatorname{dist}(x,C) \leqslant \operatorname{dist}(x,C\cap\rho B)$. On the other hand since $X$ is a Hilbert space, the mapping $y \mapsto \operatorname{proj}_C y$ is a contraction. From $0 \in C$ it follows that

$$\|\operatorname{proj}_C x\| \leqslant \|x\| \leqslant \rho.$$

Consequently, $\operatorname{proj}_C x$ belongs to $C \cap \rho B$ and

$$\operatorname{dist}(x,C) = \|x - \operatorname{proj}_C x\| \geqslant \operatorname{dist}(x,C\cap\rho B)$$

which combined with the opposite inequality yields (3.17).

Let us return to the computation of $d_{[\lambda],\rho}(\psi_C, \psi_D)$ and note that from (3.10) and (3.11)

$$\sup_{x\in X} |\mathrm{dist}(x, C\cap \rho B) - \mathrm{dist}(x, D\cap \rho B)| = \sup_{\|x\|\leqslant\rho} |\mathrm{dist}(x, C\cap \rho B) - \mathrm{dist}(x, D\cap \rho B)|,$$

the supremum being achieved (up to an arbitrary small quantity) on the set $(C\cup D)\cap \rho B$. From Lemma 3.7 we obtain

$$d_{[\lambda],\rho}(\psi_C, \psi_D) = \lambda^{-1}\,\mathrm{haus}(C\cap\rho B, D\cap\rho B).$$

Lemma 3.12 yields the dual formulation of haus($C\cap\rho B$, $D\cap\rho B$), that is,

$$\mathrm{haus}(C\cap\rho B,\ D\cap\rho B) = \sup_{\|v\|_*\leqslant 1} \left|(\psi_{C\cap\rho B})^*(v) - (\psi_{D\cap\rho B})^*(v)\right|.$$

Let us observe that

$$(\psi_{C\cap\rho B})^* = (\psi_C + \psi_{\rho B})^* = \psi_C^* \,\square\, \rho\|\cdot\|_* = (\psi_C^*)_{[\rho^{-1}]}.$$

Combining these last equalities and using the positive homogeneity of support functions, we finally obtain

$$d_{[\lambda],\rho}(\psi_C, \psi_D) = \sup_{\|v\|_*\leqslant\lambda^{-1}} \left|(\psi_C^*)_{[\rho^{-1}]}(v) - (\psi_D)^*_{[\rho^{-1}]}(v)\right| = d_{[\rho^{-1}],\lambda^{-1}}(\psi_C^*, \psi_D^*). \qquad \square$$

3.18. Corollary (Isometry). *Suppose $X$ is a Hilbert space and $C$, $D$ are closed convex subsets of $X$ such that $0\in C$ and $0\in D$. Then*

$$d_{[1],1}(\psi_C, \psi_D) = d_{[1],1}(\psi_C^*, \psi_D^*), \tag{3.19}$$

*i.e. the Legendre-Fenchel transform is an isometry as a map between the space of indicator functions of convex sets and the space of support functions of convex sets when the distance is defined in terms of $d_{[1],1}$.*

We recover in the Hilbert case the Walkup-Wets result [5] as a direct consequence of this corollary. Indeed if $C$ and $D$ are nonempty closed convex cones, then both $C$ and $D$ contain the origin and

$$\psi_C^* = \psi_{\mathrm{pol}\,C},$$

where $\mathrm{pol}\,C = \{v\in X^* | \langle v, x\rangle \leqslant 0 \text{ for all } x\in C\}$ is the *polar cone* of $C$. Then

$$d_{[1],1}(\psi_C, \psi_D) = \mathrm{haus}(C\cap B,\ D\cap B)$$

and

$$d_{[1],1}(\psi_{\mathrm{pol}\,C}, \psi_{\mathrm{pol}\,D}) = \mathrm{haus}(\mathrm{pol}\,C\cap B, \mathrm{pol}\,D\cap B).$$

Thus

3.20. Corollary (Isometry [5]). *Suppose $\mathscr{C}$ is the space of nonempty closed convex cones included in a Hilbert space $X$. Then* pol: $\mathscr{C}\to\mathscr{C}$ *is an isometry in the following sense: Given $P$ and $Q$ in $\mathscr{C}$*

$$\mathrm{haus}(P\cap B, Q\cap B) = \mathrm{haus}(\mathrm{pol}\,P\cap B, \mathrm{pol}\,Q\cap B).$$

Given $\{C^\nu,\ \nu = 1,\dots\}$ a sequence of subsets of $X$, the formula (3.16) allows us to define a convergence rate for the epi-convergence of their support and indicator functions. Indeed we have

3.21. THEOREM. *Suppose $X$ is a reflexive Banach space, and $\{f^\nu\colon X \to \overline{R},\ \nu = 1,\dots\}$ and $f\colon X \to \overline{R}$ are a collection of proper lower semicontinuous convex functions uniformly minorized by the quadratic form $-\alpha(\|x\| + 1)$ for some $\alpha > 0$, such that for all $\lambda > 0$ and all $\rho \geqslant 0$*

$$\lim_{\nu \to \infty} d_{[\lambda],\rho}(f, f^\nu) = 0.$$

*Then*

$$f = \operatorname*{Mosco\text{-}epi\text{-}lim}_{\nu \to +\infty} f^\nu.$$

*Moreover, when $X$ is finite dimensional, the reverse implication is also valid.*

PROOF. Let us first verify that given any sequence $\{x^\nu,\ \nu = 1,\dots\}$ weakly converging to $x$, the inequality (2.3) holds, i.e.

$$\liminf_{\nu} f^\nu(x^\nu) \geqslant f(x). \tag{3.22}$$

By definition of $f^\nu_{[\lambda]}$, for every $\lambda > 0$

$$f^\nu(x^\nu) \geqslant f^\nu_{[\lambda]}(x^\nu).$$

The sequence $\{x^\nu,\ \nu = 1,\dots\}$, being weakly convergent, is contained in a fixed ball $B_\rho$ of $X$. By definition of $d_{[\lambda],\rho}$

$$\left|f^\nu_{[\lambda]}(x^\nu) - f_{[\lambda]}(x^\nu)\right| \leqslant d_{[\lambda],\rho}(f^\nu, f).$$

Therefore

$$f^\nu(x^\nu) \geqslant f_{[\lambda]}(x_\nu) - d_{[\lambda],\rho}(f^\nu, f).$$

Using the assumption that $d_{[\lambda],\rho}(f^\nu, f)$ goes to 0 as $\nu$ goes to $\infty$, and the weak lower semicontinuity of the convex continuous function $f_{[\lambda]}$, this inequality yields

$$\liminf_{\nu \to \infty} f^\nu(x_\nu) \geqslant f_{[\lambda]}(x).$$

Since $f = \sup_{\lambda > 0} f_{[\lambda]}$ (see [**6**, §6]), we obtain (3.22).

Let us now verify the second assertion (2.4) of the Mosco-epi-convergence definition, i.e., the existence of a sequence $\{x^\nu,\ \nu = 1,\dots\}$ for every $x \in X$ strongly converging to $x$ such that

$$f(x) \geqslant \limsup_{\nu} f^\nu(x^\nu). \tag{3.23}$$

For $\lambda > 0$, and $\nu = 1, 2, \dots,$ let

$$J^\nu_{[\lambda]}x \in \operatorname{argmin}_{y \in Y}\left\{f^\nu(y) + \frac{1}{\lambda}\|x - y\|\right\}, \tag{3.24}$$

which means that

$$f^\nu_{[\lambda]}(x) = f^\nu(J^\nu_{[\lambda]}x) + \frac{1}{\lambda}\|x - J^\nu_{[\lambda]}x\|. \tag{3.25}$$

Note that $J^\nu_{[\lambda]}x$ is not necessarily unique. Since $f \geqslant f_{[\lambda]}$, and $f_{[\lambda]} = \lim_{\nu\to\infty} f^\nu_{[\lambda]}$, it follows that for every $x \in X$,

$$f(x) \geqslant \limsup_{\lambda\downarrow 0}\ \limsup_{\nu\to\infty} f^\nu_{[\lambda]}(x),$$

which with (3.24) yields

$$f(x) \geqslant \limsup_{\lambda\downarrow 0}\ \limsup_{\nu\to\infty}\left[f^\nu\left(J^\nu_{[\lambda]}x\right) + \frac{1}{\lambda}\left\|x - J^\nu_{[\lambda]}x\right\|\right]. \tag{3.26}$$

We can now rely on a diagonalization process [**8**], to choose a sequence $\{\lambda_\nu,\ \nu \in N\}$ decreasing to zero such that

$$f(x) \geqslant \limsup_{\lambda\downarrow 0}\left[f^\nu\left(J^\nu_{[\lambda_\nu]}x\right) + \frac{1}{\lambda_\nu}\left\|x - J^\nu_{[\lambda_\nu]}x\right\|\right]. \tag{3.27}$$

If $f(x) = \infty$, there is nothing to prove, the inequality (3.23) is trivially satisfied by any sequence, converging to $x$. So let us assume that $f(x) < \infty$. Since the functions, $\{f^\nu,\ \nu = 1,\dots\}$ are uniformly minorized by $x \mapsto -\alpha(\|x\| + 1)$, we have that

$$f^\nu\left(J^\nu_{[\lambda_\nu]}x\right) \leqslant -\alpha\left(\left\|J^\nu_{[\lambda_\nu]}x\right\| + 1\right)$$

and hence for $\nu$ sufficiently large

$$f(x) + 1 \geqslant \left[-\alpha\left(\left\|J^\nu_{[\lambda_\nu]}x\right\| + 1\right) + \frac{1}{\lambda_\nu}\left\|x - J^\nu_{[\lambda_\nu]}x\right\|\right]. \tag{3.28}$$

This in turn implies that

$$\left\|x - J^\nu_{[\lambda_\nu]}x\right\| \leqslant \frac{\lambda_\nu}{1 - \alpha\lambda_\nu}\left(f(x) + \alpha\|x\| + \alpha + 1\right). \tag{3.29}$$

Since $\lambda_\nu \downarrow 0$, it follows that with $x^\nu := J^\nu_{[\lambda_\nu]}x$ that

$$x = \text{strong-}\lim_{\nu\to\infty} x^\nu,$$

which with (3.27) yields (3.23).

For the converse observe that for all $\lambda > 0$, the functions $\{f^\nu_{[\lambda]},\ \nu = 1,\dots\}$ and $f_{[\lambda]}$ are equi-Lipschitz and this combined with pointwise convergence [**6**, Theorem 5] implies, by the Arzela-Ascoli Theorem, the uniform convergence on compact sets, which in finite dimensions correspond to the closed bounded sets, i.e. for all $\lambda > 0$ and $\rho \geqslant 0$

$$0 = \lim_{\nu\to\infty}\ \sup_{\|x\|\leqslant\rho}\left|f_{[\lambda]}(x) - f^\nu_{[\lambda]}(x)\right| = \lim_{\nu\to\infty} d_{[\lambda],\rho(f,f^\nu)}. \qquad \square$$

3.30. Corollary [**16**, p. 523; **17**, §4]. *Suppose $X$ is a reflexive Banach space and $\{C;\ C^\nu,\ \nu = 1,\dots\}$ is a collection of closed nonempty convex subsets of $X$ and for all $\rho \geqslant 0$,*

$$\lim_{\nu\to\infty} \operatorname{haus}(C \cap \rho B,\ D \cap \rho B) = 0.$$

*Then $C = \text{Mosco-lim}_{\nu\to\infty} C^\nu$, i.e. $\psi_C = \text{Mosco-epi-lim}_{\nu\to\infty}\psi_{C^\nu}$. Moreover, if $X$ is a finite dimensional Euclidean space, the reverse implication is also valid.*

PROOF. Returning to the beginning of the proof of Theorem 3.15, we see that in general

$$\begin{aligned} d_{[\lambda],\rho}(\psi_C, \psi_D) &= \lambda^{-1} \sup_{\|x\|\leqslant\rho} |\operatorname{dist}(x, C\cap\rho B) - \operatorname{dist}(x, D\cap\rho B)| \\ &\leqslant \lambda^{-1} \sup_{\|x\|\leqslant\rho} |\operatorname{dist}(x, C\cap\rho B) - \operatorname{dist}(x, D\cap\rho B)| \\ &= \lambda^{-1} \operatorname{haus}(C\cap\rho B,\ D\cap\rho B), \end{aligned}$$

with equality if $X$ is a Hilbert space. Thus, $\lim_{\nu\to\infty} \operatorname{haus}(C\cap\rho B,\ C^\nu\cap\rho B) = 0$ implies in the reflexive Banach case that $\lim_{\nu\to\infty} d_{[\lambda],\rho}(\psi_C, \psi_{C^\nu}) = 0$, and is equivalent to that condition when $X$ is the finite dimensional Euclidean space. It now suffices to apply Theorem 3.21 with $f = \psi_C$ and $f^\nu = \psi_{C^\nu}$. □

Note finally that as a consequence of this corollary, (3.16), and (3.13), in finite dimension, we have that

$$\psi_C^* = \operatorname*{epi-lim}_{\nu\to\infty} \psi_{C^\nu}^*$$

if and only if

$$\lim_{\nu\to\infty}\left[\sup_{\|v\|\leqslant 1} \left|(\psi_{C^\nu}^*)_{[\lambda]}(v) - (\psi_C^*)_{[\lambda]}(v)\right|\right] = 0. \tag{3.31}$$

It is not known, as is the case for Moreau-Yosida approximates, if the pointwise convergence of *all* the Wijsman approximates does actually imply Mosco-convergence, although it is natural to conjecture that it does.

As with the distance functions $d_{\lambda,\rho}$ and $d_{\lambda,\rho}^J$ generated by the Moreau-Yosida approximates (see (2.25) and (2.26)), we can define a Hausdorff metric on the spaces of indicator and support functions, equivalently on the space of closed convex sets, for which the Legendre-Fenchel transform is an isometry. We would refer to it as the *Wijsman metric*. We can also go one step further in using the preceding results to build a distance function on the space $\mathrm{SCC}(X)$ of proper lower semicontinuous convex functions defined on $X$. One way to achieve this is detailed in §4; here we record another possibility. Let $f, g \in \mathrm{SCC}(X)$ and $\rho > 0$ such that $C := \operatorname{epi} f \cap (\rho B \times R)$ and $D := \operatorname{epi} g \cap (\rho B \times R)$ are nonempty.

Then, by (3.13),

$$\operatorname{haus}(\operatorname{epi} f\cap(\rho B\times R), \operatorname{epi} g\cap(\rho B\times R)) = \sup_{\|v,\beta\|_*\leqslant 1} |\psi_C^*(v,\beta) - \psi_D^*(v,\beta)| \tag{3.32}$$

since here the Hausdorff distance is finite. By straightforward convex calculus we have

$$\begin{aligned} \psi^*_{\operatorname{epi} f\cap(\rho B\times R)}(v,\beta) &= \sup_{\|x\|\leqslant\rho} \left[\langle v, x\rangle + \alpha\beta \mid \alpha \geqslant f(x)\right] \\ &= \begin{cases} +\infty & \text{if } \beta > 0, \\ \sup_{\|x\|\leqslant\rho} \langle v, x\rangle & \text{if } \beta = 0, \\ \sup_{\|x\|\leqslant\rho} \left[\langle v, x\rangle - (-\beta) f(x)\right] & \text{if } \beta < 0. \end{cases} \end{aligned}$$

For $h \in \mathrm{SCC}(X)$ and $\theta \in R_+$, the *epigraphical multiplication* $\theta * h$ is defined by

$$(\theta * h)(x) = \begin{bmatrix} 0 & \text{if } \theta = 0,\ x = 0, \\ \theta h(\theta^{-1}x) & \text{if } \theta > 0. \end{bmatrix}$$

For $\theta \geqslant 0$, we have

$$(\theta * h)^*(v) = \theta h^*(v).$$

Thus for $\beta < 0$

$$\begin{aligned} \sup_{\|x\| \leqslant \rho} \left[\langle v, x\rangle - (-\beta) f(x)\right] &= \sup_{\|x\| \leqslant \rho} \left[\langle v, x\rangle - ((-\beta) * f^*)^*(x)\right] \\ &= ((-\beta) * f^*)_{[\rho^{-1}]}(v) \end{aligned}$$

as follows from (3.5). Also for $\beta = 0$ we have a similar formula and thus

$$\psi^*_{\operatorname{epi} f \cap (\rho B \times R)}(v, \beta) = ((-\beta) * f^*)_{[\rho^{-1}]}(v).$$

Substituting this in (3.32), with an obvious change of variable and $\operatorname{epi}_{[\rho]} f = \{(x, \alpha) | f(x) \leqslant \alpha,\ \|x\| \leqslant \rho\}$, we get

$$\operatorname{haus}(\operatorname{epi}_{[\rho]} f, \operatorname{epi}_{[\rho]} g) = \sup_{\|v, \rho\|_* \leqslant 1} \left|(\beta * f^*)_{[\rho^{-1}]}(v) - (\beta * g^*)_{[\rho^{-1}]}(v)\right|.$$

From this relation we could extract a notion of distance between $f$ and $g$ and their conjugates.

**4. The cosmic distance.** In [**18**, §1.F] Rockafellar and Wets introduce the notion of *extended real vector space* by adjoining to the finite dimensional space $R^n$ its *horizon*, $\operatorname{hor} R^n$, consisting of all *direction points*; each direction point corresponding to an equivalence class determined by the congruence relation that identifies parallel closed half-lines. In contrast to a 1-point compactification of $R^n$, the compactification by direction points allows us to discriminate between different directions of unboundedness of sets and sequences. Geometrically, one can identify this extended space with the surface of an $n$-dimensional hemisphere with the rim (the horizon) representing the direction points, and the open half-sphere the points of $R^n$, cf. Figure A. Given any normed linear space $(X, \|\cdot\|)$, the same construction enables one to identify the extended real vector space $X \cup \operatorname{hor} X$, where $\operatorname{hor} X$ consists of the direction points of $X$, with a closed hemisphere in $X \times R_-$; the unit ball in $X \times R$ is $\{(x, \eta) | (\|x\|^2 + \eta^2)^{1/2} \leqslant 1\}$. In general this is not a compactification of $X$ but it suggests defining a metric on $X$ that makes it a bounded space. The cosmic metric was introduced in [**18**, Definition 1F3] when $X$ is a finite dimensional Euclidean space. Here we extend its definition to a more general setting.

4.1. DEFINITION. *The cosmic distance between two points $x$ and $y$ of an extended normed linear space $X \cup \operatorname{hor} X$, denoted by $\operatorname{dist}^c(x, y)$ is the geodesic distance between the corresponding elements of the hemisphere $H$, i.e. the distance along the great circle joining the elements in question.*

We shall not pursue here a detailed study of the cosmic metric. This is done in the Euclidean case in [**18**, §§1F and 3B]. We only use it to exhibit an isometry for the Legendre-Fenchel transform that does not rely on approximates.

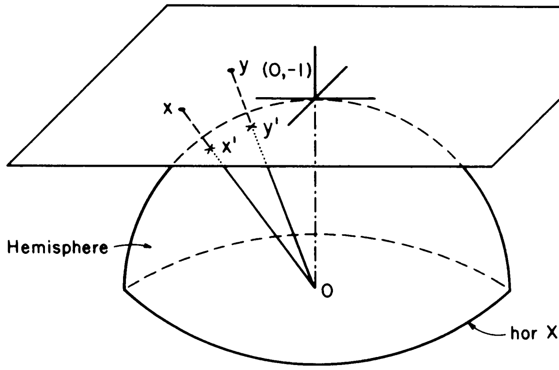


FIGURE A. The cosmic view of $X$

Every extended-real-valued function $f$ defined on $X$, a reflexive Banach space, is completely determined by its epigraph, epi $f$, a subset of $X \times R$. Given any subset of $(X \times R)$, in particular epi $f$, we can identify it as in **[6]** with a closed convex cone in $(X \times R) \times R_-$, viz.:

$$\mathrm{cl}\{\lambda(x, \alpha, -1) | (x, \alpha) \in \mathrm{epi}\, f,\ \lambda \geqslant 0\} =: \mathrm{cl\ cn}(\mathrm{epi}\, f \times \{-1\}),$$

or equivalently with a closed subset of the hemisphere $H$ in $(X \times R) \times R_-$. Given two proper functions $f$ and $g$ we can use as a measure of the distance between them the Hausdorff distance between the (bounded) subsets of the hemisphere determined by the preceding construction or equivalently, in view of Definition 4.1, the cosmic-Huasdorff distance, denoted by $\mathrm{haus}^c$, between the (unbounded) subsets epi $f$ and epi $g$ of $X \times R$,

(4.2) $\mathrm{haus}^c(\mathrm{epi}\, f, \mathrm{epi}\, g)$

$$:= \max\left[\sup_{(x,\alpha)\in \mathrm{epi}\, f} \mathrm{dist}^c((x, \alpha), \mathrm{epi}\, g),\ \sup_{(y,\beta)\in \mathrm{epi}\, g} \mathrm{dist}^c((y, \beta), \mathrm{epi}\, f)\right]$$

and

$$\mathrm{dist}^c((x, \alpha), \mathrm{epi}\, g) := \inf_{(y,\beta)\in \mathrm{epi}\, g} \mathrm{dist}^c((x, \alpha), (y, \beta)).$$

If $f$ and $g$ are convex functions on a Hilbert space $X$, then the closed cones they generate in $(X \times R) \times R_-$ are also convex. For closed convex cones we already have an isometry for the polar map in terms of the Hausdorff distance between their intersections with the unit ball, or equivalently between their intersections with the unit sphere. Since, by construction and definition of the cosmic distance,

(4.3)

$$\mathrm{haus}^c(\mathrm{epi}\, f, \mathrm{epi}\, g) = \mathrm{haus}(\mathrm{cl\ cn}(\mathrm{epi}\, f \times \{-1\}) \cap H, \mathrm{cl\ cn}(\mathrm{epi}\, g \times \{-1\}) \cap H),$$

which by Corollary 3.20 equals

$$\mathrm{haus}[(\mathrm{pol\ cn}(\mathrm{epi}\, f \times \{-1\})) \cap H, (\mathrm{pol\ cn}(\mathrm{epi}\, g \times \{-1\})) \cap H],$$

which again by Definition 4.1 is equal to

$$\mathrm{haus}^c(E_f^*, E_g^*),$$

where

$$E_f^* := \{ (v, \beta) | (v, -1, \beta) \in \text{pol cn}(\text{epi} f \times \{-1\}) \}$$

and $E_g^*$ is defined similarly. Now observe that $(v, \beta)$ belongs to $E_f^*$ if and only if

$$\langle v, x \rangle - \alpha - \beta \leqslant 0 \quad \text{for all } (x, \alpha) \in \text{epi} f,$$

or equivalently if $(v, \beta) \in \text{epi} f^*$. And thus

$$\text{haus}^c(\text{epi} f, \text{epi} g) = \text{haus}^c(\text{epi} f^*, \text{epi} g^*). \tag{4.4}$$

Let us denote by $d^c(f, g)$ the *cosmic distance* between two proper functions $f$ and $g$, defined by

$$d^c(f, g) = \text{haus}^c(\text{epi} f, \text{epi} g). \tag{4.5}$$

We thus obtain the theorem below that completes a similar result to Rockafellar and Wets [**18**, Chapter 3]:

4.6. THEOREM. *Suppose $X$ is a Hilbert space and $f$ and $g$ are proper lower semicontinuous convex functions defined on $X$. Then*

$$d^c(f, g) = d^c(f^*, g^*). \tag{4.7}$$

Given a collection of convex functions $\{f^\nu, \nu = 1, \dots\}$ defined on $X$, a reflexive Banach space, we can introduce a notion of convergence in terms of the cosmic distance. We say that $f$ is the *cosmic-epi-limit* of the sequence $\{f^\nu, \nu = 1, \dots\}$ which we write

$$f = \text{epi}^c\text{-}\lim_{\nu \to \infty} f^\nu, \quad \text{if } \lim_{\nu \to \infty} d^c(f^\nu, f) = 0.$$

The next theorem justifies referring to $f$ as an epi-limit.

4.8. THEOREM. *Suppose $X$ is a reflexive Banach space and $\{f; f^\nu, \nu = 1, \dots\}$ is a collection of proper lower semicontinuous convex functions defined on $X$ such that $f = \text{epi}^c\text{-}\lim_{\nu \to \infty} f^\nu$. Then*

$$f = \text{Mosco-epi-}\lim_{\nu \to \infty} f^\nu.$$

*Moreover, if $X$ is a finite dimensional Euclidean space, then $f = \text{epi-}\lim_{\nu \to \infty} f^\nu$ if and only if $f = \text{epi}^c\text{-}\lim_{\nu \to \infty} f^\nu$.*

PROOF. We know that

$$\lim_{\nu \to \infty} \text{haus}^c(\text{epi} f^\nu, \text{epi} f) = 0$$

if and only if

$$\lim_{\nu \to \infty} \text{haus}([\text{cl cn}(\text{epi} f^\nu \times \{-1\})] \cap B, [\text{cl cn}(\text{epi} f \times \{-1\})] \cap B) = 0.$$

The last equality holds if and only if the same holds with $B$ replaced by $\rho B$ with any $\rho > 0$, the sets involved being closed convex cones. We now apply Corollary 3.30 and we see that the above implies that

$$\text{cl cn}(\text{epi} f \times \{-1\}) = \text{Mosco-}\lim_{\nu \to \infty} \text{cl cn}(\text{epi} f^\nu \times \{-1\}),$$

and this in turn implies that

$$\operatorname{epi} f = \operatorname*{Mosco-lim}_{\nu \to \infty} \operatorname{epi} f^{\nu}.$$

If $X$ is the Euclidean $n$-space, to obtain the second assertion we argue as above except that we rely on the second part of Corollary 3.30:

$$\lim_{\nu \to \infty} \operatorname{haus}^{c}(\operatorname{epi} f^{\nu}, \operatorname{epi} f) = 0$$

if and only if

$$\operatorname{cl} \operatorname{cn}(\operatorname{epi} f \times \{-1\}) = \lim_{\nu \to \infty} \operatorname{cl} \operatorname{cn}(\operatorname{epi} f^{\nu} \times \{-1\}).$$

By [6, §4] this occurs if and only if $\operatorname{epi} f = \lim_{\nu \to \infty} \operatorname{epi} f^{\nu}$, or equivalently, $f = \operatorname{epi-lim}_{\nu \to \infty} f^{\nu}$. □

## References

1. R. Wijsman, *Convergence of sequences of convex sets, cones and functions*. II, Trans. Amer. Math. Soc. **123** (166), 32–45.

2. U. Mosco, *On the continuity of the Young-Fenchel transform*, J. Math. Anal. Appl. **35** (1971), 518–535.

3. J. L. Joly, *Une famille de topologies sur l'ensemble des fonctions convexes pour lesquelles la polarité est bicontinue*, J. Math. Pures Appl. **52** (1973), 421–441.

4. K. Back, *Continuity of the Fenchel transform of convex functions*, Tech. Report, Northwestern University, November 1983.

5. D. Walkup and R. Wets, *Continuity of some convex-cone valued mappings*, Proc. Amer. Math. Soc. **18** (1967), 229–235.

6. R. Wets, *Convergence of convex functions, variational inequalities and convex optimization problems*, Variational Inequalities and Complementarity Problems (R. Cottle, F. Gianessi and J. L. Lions, eds.), Wiley, New York, 1980, pp. 375–403.

7. H. Attouch and R. Wets, *A convergence theory for saddle functions*, Trans. Amer. Math. Soc. **280** (1983), 1–41.

8. H. Attouch, *Variational convergences for functions and operators*, Pitman Research Notes in Math., London, 1984.

9. ______, *Familles d'opérateurs maximaux monotones et mesurabilité*, Ann. Mat. Pura Appl. **120** (1979), 35–111.

10. J. J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France **93** (1965), 273–299.

11. R. T. Rockafellar, *Convex analysis*, Princeton Univ. Press, Princeton, N. J., 1969.

12. ______, *Extension of Fenchel's duality theorem for convex functions*. Duke Math. J. **33** (1966), 81–89.

13. H. Brezis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Lecture Notes, vol. 5, North-Holland, Amsterdam, 1972.

14. A. Fougères and A. Truffert, *Régularisation-s.c.i. et epi-convergence: approximations inf-convolutives associées à un référentiel*, AVAMAC 84-08, Université Perpignan, 1984.

15. L. Hörmander, *Sur la fonction d'appui des ensembles convexes dans un espace localement convexe*, Ark. Mat. **3** (1954), 181–186.

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

17. G. Salinetti and R. Wets, *On the convergence of sequences of convex sets in finite dimensions*, SIAM Rev. **21** (1979), 18–33.

18. R. T. Rockafellar and R. Wets, *Extended real analysis* (to appear).

Department of Mathematics, Universite de Perpignan, Perpignan, France

Department of Mathematics, Universite de Paris, Dauphine, Paris, France

*Current address* (R. J. B. Wets): Department of Mathematics, University of California, Davis, California 95616