JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 131, 404–411 (1988)

Another Isometry for the Legendre–Fenchel Transform

HEDY ATTOUCH

Laboratoire AVAMAC, Université de Perpignan, 66025 Perpignan Cédex, France

AND

ROGER J.-B. WETS*

Mathematics Department, University of California, Davis, California 95616

Submitted by Frank H. Clarke

Received November 3, 1986

A new distance is introduced on the space of extended real-valued, lower semicontinuous convex functions defined on reflexive Banach spaces. It is shown that with this notion of distance, the Legendre–Fenchel is an isometry.

Let X be a Banach space and $\Gamma_0(X)$ the space of proper lower semicontinuous convex functions defined on X and with values in $(-\infty, +\infty]$; f is proper if its effective domain $\operatorname{dom} f := \{x \mid f(x) < \infty\}$ is nonempty. If $f \in \Gamma_0(X)$, its Legendre–Fenchel transform

$$v \mapsto f^*(v) := \sup_{x \in X} [\langle v, x\rangle - f(x)]$$

(or conjugate function) belongs to $\Gamma_0(X^*)$, where X^* is paired with X through the bilinear form $\langle \cdot, \cdot \rangle$. In [3], we introduce distance functions on $\Gamma(X)$ for which the Legendre–Fenchel transform is an isometry. Most of the results were limited to the Hilbert space case, and those that dealt with general Banach spaces were existential, by opposition to operational, in nature. The notion of distance, or more exactly of a class of distances, that we introduce here are (relatively) easy to use in an operational setting for a wide class of Banach spaces, and still allows us to conclude that Legendre–Fenchel transform is isometric.

* Supported in part by a grant of the National Science Foundation.

0022-247X/88 $3.00

The definition of the distance function relies on an a class of approximate of (convex) functions. In the context of the duality theory for convex optimization problems, this type of functions appear in the work of Rockafellar [6] and, Aubin and Ekeland [4, Chap. 4, Sect. 6].

To a function $f: X \to (-\infty, \infty]$, we associate the following approximate:

$$f_\lambda^\#(x, v) := \inf_{u \in X} \left[f(x+u) + \frac{1}{2\lambda} \|u\|^2 - \langle v, u \rangle \right],$$

where $\lambda > 0$ and $v \in X^*$. We refer to this function as the *approximate of parameter λ and slope v*; the use of the term "approximate" is justified by the fact that for λ sufficiently small and v close to $0, f_\lambda^\#(\cdot, v)$ and f should be close in value. In some cases it is useful to consider the following generalization

$$\begin{aligned} f_\lambda^{\#,p}(x, v) &:= \inf_{u \in X} \left[f(x+u) + \frac{1}{p\lambda} \|u\|^p - \langle v, u \rangle \right] \\ &= \inf_{u \in X} \left[f(u) + \frac{1}{p\lambda} \|x-u\|^p - \langle v, u \rangle \right] + \langle v, x \rangle \end{aligned}$$

that relies on the pth power, instead of the square of the norm, where $1 < p < \infty$. We refer to this function as the *approximate of order p with parameter λ and slope v*. We begin with recording some of the main properties of $f_\lambda^{\#,p}$ in the convex case, i.e., when f is convex.

PROPOSITION 1. *Suppose $f \in \Gamma_0(X)$, X is a Banach space, and $\lambda > 0$, $1 < p < \infty$. Then $(x, v) \mapsto f_\lambda^{\#,p}(x, v)$ is a convex-concave, continuous, finite-valued function. Moreover, it is bounded on bounded subsets of $X \times X^*$.*

Proof. Clearly $f_\lambda^{\#,p}$ is concave in v (as the infimum of affine functions), and convex in x, since

$$f_\lambda^{\#,p}(\cdot, v) = (f - \langle v, \cdot \rangle) \,\square\, \frac{1}{p\lambda} \|\cdot\|^p + \langle v, \cdot \rangle$$

is the sum of a linear term and the inf-convolution of two convex functions. This last expression also shows that it is finite, from which is follows that $f_\lambda^{\#,p}$ is continuous.

Now, to show that it is bounded on bounded sets, first pick any $x_0 \in \operatorname{dom} f$ and note that

$$f_\lambda^{\#,p}(x, v) \leqslant f(x_0) + \frac{1}{p\lambda} \|x - x_0\|^p - \langle v, x_0 \rangle + \langle v, x \rangle$$

and hence, if $\|x\| \leqslant \rho$ and $\|v\|_* \leqslant \rho$ we have that

$$f_\lambda^{\#,p}(x, v) \leqslant \gamma_1(p, \lambda, \rho)$$

for some constant $\gamma_1 > 0$ that depends only on ρ for fixed λ and p.

To obtain a lower bound, observe that $f \in \Gamma_0(X)$, implies the existence of a scalar α such that for all $u \in X$,

$$f(u) + \alpha(\|u\| + 1) \geqslant 0.$$

In turn, this tells us that

$$\begin{aligned} f_\lambda^{\#,p}(x, v) &\geqslant \min_u \left[-\alpha \|x + u\| + \frac{1}{p\lambda} \|u\|^p - \langle v, u \rangle \right] - \alpha \\ &\geqslant \min_u \left[-\alpha \|x\| - \alpha \|u\| + \frac{1}{p\lambda} \|u\|^p - \langle v, u \rangle \right] - \alpha \\ &\geqslant \min_u \left[-\alpha \|u\| + \frac{1}{p\lambda} \|u\|^p - \alpha\rho - \rho \|u\| \right] - \alpha \\ &= \min_{t \in \mathbb{R}} \left[\frac{1}{p\lambda} t^p - (\alpha + \rho)t - \alpha(1 + \rho) \right] \end{aligned}$$

whenever $\|x\| \leqslant \rho$ and $\|v\|_* \leqslant \rho$. This least quantity is finite (since $p > 1$) and depends only on ρ, λ, and p. And hence, for $\|x\| \leqslant \rho$ and $\|v\|_* \leqslant \rho$,

$$f_\lambda^{\#,p}(x, v) \geqslant \gamma_2(p, \lambda, \rho),$$

where γ_2 is constant that depends only on ρ for fixed λ and p. ▮

Special cases of these approximates of order p are the Moreau–Yosida approximates [3] with $p = 2$ and $v = 0$. However $f_\lambda^{\#,p}$ possess duality properties that are not shared by other approximates even by much more general classes of approximates that have been suggested in the literature [5, 7]. It is closely connected to the perturbation theory for convex optimization of Rockafellar [6] and hence with duality theory. As a point of departure, let us consider, for $f \in \Gamma_0(X)$, the problem:

$$\text{find } x \in X \text{ that minimizes } f(x).$$

We are interested in the stability of the solution—assuming that it exists—with respect to perturbations of f. Rockafellar's approach is to embed the given problem is a class of optimization problems that depend on certain parameters. The class should be rich enough to carry sufficient information and simple enough to allow for easy manipulations. One class that incorporates both classical perturbations (adding a conditioning term such as $\|\cdot\|^2$) and perturbations of the coefficients ("vertical" and/or "horizontal" perturbations) is given by

$$\mathbf{f}_{\lambda,p}(x, u) = f(x+u) + \frac{1}{p\lambda}\|u\|^p$$

for some fixed $1 < p < \infty$, where $(\lambda, u) \in (R_+ \times X)$ are the parameters. Note that $\mathbf{f}(x, 0) = f(x)$, and $\mathbf{f}(x, -x) = f(0) + (1/\lambda p)\|x\|^p$ involves only the conditioning term. The solutions and the value of

$$\text{find } x \in X \text{ that minimizes } \mathbf{f}_{\lambda,p}(x, u)$$

depend on u and λ. If perturbations can be introduced in the problem at a "cost" $-\langle v, u\rangle$, the problem is then to choose the perturbation that will allow us to reach the lowest value of $\mathbf{f}_{\lambda,p}(x, u) - \langle v, u\rangle$, i.e.

$$\text{find } x \in X \text{ that minimizes } \inf_{u \in X}\left[f(x+u) + \frac{1}{\lambda p}\|u\|^p - \langle v, u\rangle\right],$$

or, equivalently,

$$\text{find } x \in X \text{ that minimizes } \inf_{u \in X} f_\lambda^{\#,p}(x, v).$$

The interpretation to be given to the term $\langle -v, u\rangle$ depends on the application. It could correspond to the energy needed to generate these perturbations, or the cost of purchasing these perturbations, etc. These approximates are thus Lagrangian functions with v in the multipliers space. (Note that since $f_\lambda^{\#,p}$ is a Lagrangian, the fact that it is convex-concave follows directly from the convexity of $\mathbf{f}_{\lambda,p}$.)

Let us also observe that if λ converges to 0, the function $\mathbf{f}_{\lambda,p}$ epi-converges to $(x, u) \mapsto f(x) + \delta_{\{u \geqslant 0\}}(u)$ (here δ_C is the indicator function of C) which implies that $f_\lambda^{\#,p}$ epi/hypo-converges to $f(x)$, see [1, Theorem 3.2]. This, and the usual relationship between a convex function and its conjugate, suggests the following key formula.

PROPOSITION 2. *Suppose $f \in \Gamma_0(X)$ with X a reflexive Banach space, $\lambda > 0$, and $1 < p < \infty$. Then for all $x \in X$, $v \in X^*$, we have*

$$(f^*)_\lambda^{\#,p}(v, x) + f_{\lambda'}^{\#,p'}(x, v) = \langle v, x\rangle,$$

where $1/p + 1/p' = 1$ and $\lambda' = \lambda^{-p'/p}$.

Proof. We have

$$(f^*)_\lambda^{\#,p}(v,x) = \operatorname*{Min}_{v' \in X^*} \left[\operatorname*{Sup}_{x' \in X}(\langle v', x' \rangle - f(x')) + \frac{1}{p\lambda} \|v - v'\|_*^p - \langle v', x \rangle \right] + \langle v, x \rangle$$

$$= \operatorname*{Sup}_{x'} \left[-f(x') + \langle v, x' \rangle - \operatorname*{Sup}_{v'} \left[\langle v - v', x' - x \rangle - \frac{1}{p\lambda} \|v - v'\|_*^p \right] \right]$$

$$= -\operatorname*{Min}_{x' \in X} \left[f(x') + \frac{\lambda^{p'/p}}{p'} \|x - x'\|^{p'} - \langle v, x' \rangle \right]$$

$$= -[f_{\lambda'}^{\#,p'}(x,v) - \langle v, x \rangle],$$

where the second equality follows from the minimax theorem [4, Theorem 6.2.7] and the third one from the formulas

$$(\alpha g)^* = \alpha g^* \left(\frac{1}{\alpha} \cdot \right) \qquad \text{for all} \quad \alpha > 0$$

$$\left(\frac{1}{p} \|\cdot\|_*^p \right)^* = \frac{1}{p'} \|\cdot\|^{p'} \qquad \text{where} \quad \frac{1}{p} + \frac{1}{p'} = 1;$$

for this last identity see [4, Proposition 4.4.8] and use the fact that the space is reflexive. ∎

COROLLARY 3. *Suppose $\lambda > 0$, $f \in \Gamma_0(X)$, where X is a reflexive Banach space. Then*

$$(f^*)_{\lambda^{-1}}(v) = -f_\lambda^\#(0, v),$$

where $(f^*)_\lambda$ is the Moreau–Yosida approximate of f^* of parameter λ.

Proof. We know that

$$-f_\lambda^\#(0,v) = -f_\lambda^{\#,2}(0,v) = (f^*)_{\lambda^{-1}}^{\#,2}(v,0) = (f^*)_{\lambda^{-1}}^\#(v,0)$$

and, by definition

$$(f^*)_{\lambda^{-1}}^\#(v,0) = \inf_{u \in X} \left[f^*(v+u) + \frac{1}{2\lambda} \|u\|^2 \right] = (f^*)_\lambda(v). \quad ∎$$

DEFINITION 4. For $1 < p < \infty$, $\lambda > 0$, $\rho \geqslant 0$, we define the distance $d_{\lambda,\rho}^{\#;p}$ of order p, between $f, g \in \Gamma_0(X)$ by

$$d_{\lambda,\rho}^{\#;p}(f, g) := \operatorname*{Sup}_{\substack{\|x\| \leqslant \rho \\ \|v\|_* \leqslant \rho}} |f_\lambda^{\#,p}(x, v) - g_\lambda^{\#,p}(x, v)|.$$

THEOREM 5. *Suppose $f, g \in \Gamma_0(X)$ with X a reflexive Banach space $1 < p < \infty$ and $\rho > 0$. Then*

$$d_{\lambda,\rho}^{\#;p}(f, g) = d_{\lambda',\rho}^{\#;p'}(f^*, g^*).$$

This means that when $1 < p < \infty$, for all $\rho > 0$, the Legendre–Fenchel transform is an isometry for any pth order distance $d_{\lambda,\rho}^{\#;p}$ on $\Gamma_0(X)$ and $d_{\lambda,\rho}^{\#;p'}$ is the corresponding distance of order p' on $\Gamma_0(X^*)$.

Proof. From Proposition 2 it follows that

$$(f^*)_1^{\#,p}(v, x) - (g^*)_1^{\#,p}(v, x) = f_1^{\#,p'}(x, v) - g_1^{\#,p'}(x, v)$$

from which the result follows since it implies that

$$\begin{aligned}&\operatorname*{Sup}_{\substack{\|x\| \leqslant \rho \\ \|v\|_* \leqslant \rho}} |(f^*)_1^{\#,p}(v, x) - (g^*)_1^{\#,p}(v, x)| \\ &\qquad = \operatorname*{Sup}_{\substack{\|x\| \leqslant \rho \\ \|v\|_* \leqslant \rho}} |f_1^{\#,p'}(x, v) - g_1^{\#,p'}(x, v)|. \quad \blacksquare\end{aligned}$$

Remark 6. The distance functions introduced in [3] are defined in terms of the function values, whereas here we also include in addition dual quantities (slopes). We could put the accent on conjugate values, by relying on the formula of Corollary 3. Let

$$d_{\lambda,\rho}^*(f, g) := \operatorname*{Sup}_{\|v\| \leqslant \rho} |(f^*)_\lambda(v) - (g^*)_\lambda(v)|.$$

Then for all $\lambda > 0$ and $\rho > 0$, $f, g \in \Gamma_0(X)$ and X a reflexive Banach space, we have the tautological isometry:

$$d_{\lambda,\rho}^*(f, g) = d_{\lambda,\rho}(f^*, g^*),$$

where $d_{\lambda,\rho}$ is the distance function introduced in [3, Section 2]. In other words, the Legendre–Fenchel transform is an isometry

$$\text{from } (\Gamma_0(X), d_{\lambda,\rho}^*) \qquad \text{into } (\Gamma_0(X^*), d_{\lambda,\rho}).$$

In the Hilbert case, when we identify X and X^*, and observing that

$$f_1^{\#}(0, x) = f_1(x) - \tfrac{1}{2}\|x\|^2,$$

it follows directly from Corollary 3, that

$$f_1(x)+f_1^*(x)=\tfrac{1}{2}\|x\|^2,$$

and hence

$$\begin{aligned} d_{1,\rho}(f,g) &= \sup_{\|x\|\leqslant 1} |f_1(x)-g_1(x)| \\ &= d_{1,\rho}^*(f,g). \end{aligned}$$

This yields an isometry for the Hilbert space case, already found in [3, Corollary 2.22].

Remark 7. The distance function $d_{\lambda,\rho}^{\#;2}$ that introduces the term $-\langle v,u\rangle$ does not bring us anything new in the Hilbert space case with respect to the results of [3]. Indeed, we have then

$$f_\lambda^{\#}(x,v)=f_\lambda(x-v)-\tfrac{1}{2}\|v^2\|,$$

and hence

$$d_{\lambda,\rho}^{\#;2}(f,g)=\operatorname*{Sup}_{\substack{\|x\|\leqslant\rho\\ \|v\|\leqslant\rho}} |f_\lambda(x-v)-g_\lambda(x-v)|=d_{\lambda,2\rho}(f,g).$$

To conclude let us record that this convergence with respect to this distance does imply Mosco-epi-convergence, and in the finite-dimensional case it is actually equivalent to epi-convergence.

THEOREM 8. *Suppose* $1<p<\infty$, $\{f; f^\nu, \nu=1,...\}\subset\Gamma_0(X)$ *with* X *a Banach space. Then, the condition: for all* $\lambda>0$, *and* $\rho>0$,

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

implies that $f=\text{Mosco-epi-lim}_{\nu\to\infty} f^\nu$. *Moreover, if* X *is finite dimensional, the converse also holds.*

Proof. The argument is the same as in [3, Theorem 2.51] after one observes that for f, g in $\Gamma_0(X)$:

$$d_{\lambda,\rho}^{\#;p}(f,g)\geqslant \sup_{\|x\|\leqslant\rho} |f'_{\lambda,p}(x)-g'_{\lambda,p}(x)|,$$

where

$$f'_{\lambda,p}(x):=\inf_u\left\{f(x+u)+\frac{1}{p\lambda}\|u\|^p\right\},$$

and that the resolvent (of order p) is well defined and can be used in the same way as in [3] to build the sequences that converge to x. ∎

REFERENCES

1. H. ATTOUCH, D. AZE, AND R. J.-B. WETS, "Convergence of Convex-Concave Saddle Functions: Continuity Properties of the Legendre–Fenchel Transform with Applications to Convex Programming and Mechanics," *Annales de l'Institut H. Poincaré: Analyse Non Linéaire* (1988), to appear.
2. H. ATTOUCH AND R. J.-B. WETS, A convergence theory for saddle functions, *Trans. Amer. Math. Soc.* **280** (1983), 1–41.
3. H. ATTOUCH AND R. J.-B. WETS, Isometries for the Legendre–Fenchel transform, *Trans. Amer. Math. Soc.* **296** (1986), 33–60.
4. J.-P. AUBIN AND I. EKELAND, *in* "Applied Nonlinear Analysis," Wiley-Interscience, New York, 1984.
5. A. FOUGERES AND A. TRUFFERT, Regularisation s.c.i. et GAMMA-convergence. Approximation inf-convolutives associees a un referentiel, *Ann. Mat. Pura Appl.* in press.
6. R. T. ROCKAFELLAR, "Conjugate Duality and Optimization," CBMS Regional Conferences Series in Applied Mathematics Vol. 16, SIAM Philadelphia, 1974.
7. R. J.-B. WETS, Convergence of convex functions, variational inequalities and convex optimization problems, *in* "Variational Inequalities and Complementarity Problems" (R. Cottle, F. Gianessi, and J. L. Lions, Ed.), pp. 405–419, Wiley, New York, 1980.