FERMAT Days 85: Mathematics for Optimization
J.-B. Hiriart-Urruty (editor)

ON CONTINUITY PROPERTIES OF THE PARTIAL LEGENDRE–FENCHEL TRANSFORM : CONVERGENCE OF SEQUENCES OF AUGMENTED LAGRANGIAN FUNCTIONS, MOREAU–YOSIDA APPROXIMATES AND SUBDIFFERENTIAL OPERATORS

H. ATTOUCH
A.V.A.M.A.C Département de Mathématiques
Université
PERPIGNAN (France)

D. AZÉ
A.V.A.M.A.C Département de Mathématiques
Université
PERPIGNAN (France)

R. WETS
University of California
DAVIS (U.S.A.)

Abstract. In this article we consider the continuity properties of the partial Legendre-Fenchel transform which associates, with a bivariate convex function $F: X \times Y \to \mathbf{R} \cup \{+\infty\}$, its partial conjugate

$$L : X \times Y^* \to \overline{\mathbf{R}}, \text{ i.e. } L(x,y^*) = \inf y \in Y\{F(x,y) - \langle y^* \mid y\rangle\}.$$

Following [3] where this tranformation has been proved to be bicontinuous when convex functions F are equipped with the Mosco-epi-convergence, and convex concave Lagrangian functions L with the Mosco-epi/hypo-convergence, we now investigate the corresponding convergence notions for augmented Lagrangians, Moreau-Yosida approximates and subdifferential operators.

1. INTRODUCTION

In [4], [5] the authors have introduced a new concept of convergence for bivariate functions specifically designed to study the convergence of sequences of saddle value problems, called *epi/hypo-convergence.*

A main feature of this convergence notion is, in the convex setting, to make the *partial Legendre–Fenchel transform bicontinuous.* We recall that, given a convex function $F: X \times Y \to \overline{\mathbf{R}}$ its partial Legendre–Fenchel transform is the convex-concave function $L: X \times Y^* \to \overline{\mathbf{R}}$

$$L(x,y^*) = \inf_{y\in Y}\{F(x,y) - \langle y^* \mid y\rangle\}. \tag{1.1}$$

The transformation $F \mapsto L$ is one-to-one bicontinuous when convex functions are equipped with epi-convergence and closed convex-concave functions (in the sense of R.T. Rockafellar [37] with epi/hypo convergence (see [5], [3])).

When, following the classical duality scheme, functions F^n are perturbation functions attached to the primal problems

$$\inf_{x\in X} F^n(x,0),$$

the above continuity property, combined with the variational properties of epi/hypo-convergence, is a key tool in order to study the convergence of the saddle points (that is of primal and dual solutions) of the corresponding Lagrangian functions $\{L^n ; n \in \mathbf{N}\}$. The reduced problem is the study of epi-convergence of the sequence of perturbations functions $\{F^n ; n \in \mathbf{N}\}$. This approach has been successfully applied to various situations in Convex Analysis (in Convex Programming see D. Azé [8], for convergence problems in Mechanics like homogenization of composite materials or reinforcement by thin structures see [9], H. Chabi [17], ...).

Indeed there are many other mathematical objects attached to this classical duality scheme. Our main purpose in this article is to study for each of them the corresponding convergence notion.

Particular attention is paid to the so-called *augmented Lagrangian* (especially quadratic augmented) whose definition is (compare with (1.1))

$$L_r(x,y^*) = \inf_{y\in Y} \left\{F(x,y) + \frac{r}{2}\|y\|^2 - \langle y^* \mid y\rangle\right\} \tag{1.2}$$

and which can be viewed as an "augmented "partial Legendre–Fenchel transform. In theorem 4.2 we prove the equivalence between Mosco epi/hypo-convergence of Lagrangian functions L^n and

$$\begin{aligned}&\text{for every } r > 0 \text{ and } y^* \in Y^*, \text{ the sequence}\\ &\text{of convex functions } \{L^n_r(\cdot,y^*);\ n \in \mathbf{N}\}\\ &\text{Mosco epi-converges to } L_r(\cdot,y^*).\end{aligned} \tag{1.3}$$

By the way, since L_r can be written as an inf-convolution

$$-L_r = (-L)\,\nabla \frac{1}{2r}\,\|\cdot\|^2_{Y^*}, \tag{1.4}$$

we are led to study the two following basic properties of the inf-convolution operation, which explains the practical importance (especially from a numerical point of view) of the augmented Lagrangian :

- regularization effect;
- conservation of the infima and minimizing elements. This is considered in Propositions 3.1 and 3.2 for general convolution Kernels, see also M. Bougeard and J.P. Penot [14], M. Bougeard [13].

Iterating this regularization process but, now, on the x–variable, we obtain the so called *Moreau-Yosida approximate*

$$L_{\lambda,\mu}(x,y^*) = \inf_{\xi\in X}\sup_{\eta\in Y^*}\left\{L(\xi,\eta)) + \frac{1}{2\lambda}\|x-\xi\|^2 - \frac{1}{2\mu}\|y^*-\eta\|^2\right\}, \quad (1.5)$$

the inf-sup being equal to the sup-inf (for closed convex–concave functions (theorem 5.1 d)) and the Mosco epi/hypo-convergence of L^n to L is equivalent to the pointwise convergence of the associated Moreau–Yosida approximates (Theorem 5.2). Moreover $L_{\lambda,\mu}$ has the same saddle elements as L! (Theorem 5.1 b).

Finally we characterize in terms of graph convergence of subdifferential operators

$$\partial L^n \xrightarrow{G} \partial L$$

the above notions (Theorem 6.1), and summarize in a diagram all these equivalent convergence properties.

2. CONVERGENCE OF CONVEX–CONCAVE SADDLE FUNCTIONS AND CONTINUITY OF THE PARTIAL LEGENDRE–FENCHEL TRANSFORMATION

2.1. Duality scheme

Let us first briefly review the main feature of Rockafellar's duality scheme (cf. [37], [38], [39]). Let X, Y, X^*, Y^* be linear spaces such that X (resp. Y) is in separate duality with X^* (resp. Y^*) via pairings denoted by $\langle\cdot\mid\cdot\rangle$.

Let us consider

$$L: X \times Y^* \to \overline{\mathbf{R}}$$

which is

convex in the x variable,

concave in the y^* variable.

Let us define

$$F: X \times Y \to \overline{\mathbf{R}}$$

$$G: X^* \times Y^* \to \overline{\mathbf{R}}$$

by:

$$F(x,y) = \sup_{y^* \in Y^*} \{L(x,y^*) + \langle y^* \mid y \rangle\}, \tag{2.1}$$

$$G(x^*,y^*) = \inf_{x \in X} \{L(x,y^*) - \langle x^* \mid x \rangle\}. \tag{2.2}$$

F (resp. G) is the convex (resp.concave) parent of the convex–concave function L.

Two convex–concave functions are said to be equivalent if they have the same parents. A function L is said to be closed if its parents are conjugate to each other, i.e.,

$$-G = F^* \text{ and } (-G)^* = F. \tag{2.3}$$

For closed convex–concave functions L, the associated equivalence class is an interval, denoted by $[\underline{L}, \overline{L}]$ with

$$\underline{L}(x,y^*) = \sup_{x^* \in X^*} \{G(x^*,y^*) + \langle x^* \mid x \rangle\}, \tag{2.4}$$

$$\overline{L}(x,y^*) = \inf_{y \in Y} \{F(x,y) - \langle y^* \mid y \rangle\}. \tag{2.5}$$

Let us observe that

$$\underline{L} = (-G)^{*x^*} \tag{2.6}$$

$$-\overline{L} = F^{*y}, \tag{2.7}$$

where $*y$ (resp. $*x^*$) denotes the partial conjugation with respect to the y (resp. x^*) variable.

If we denote by $\Gamma(X \times Y)$ the class of all convex l.s.c. functions defined on $X \times Y$ with value in $\overline{\mathbf{R}}$, we have the following ([37]).

Proposition 2.1. *The map $K \mapsto F$ establishes a one-to-one correspondence between closed convex-concave equivalence classes and $\Gamma(X \times Y)$.*

In the sequel, closed convex-concave functions will be assumed to be proper, i.e., convex parent F is neither the function $\equiv +\infty$ nor the function $\equiv -\infty$.

In the classical theory of convex duality (see [19], [39]) the Lagrangian associated with the proper closed convex perturbation function F is the convex-concave function $\overline{L}$ defined in (2.5). The research for a primal and dual solution is then equivalent to that of a saddle point for the equivalence class which contains $\overline{L}$.

2.2. Mosco epi-convergence

For further results see [1], [25], [34].

Definition 2.2. *Let X be a reflexive Banach space. A sequence $\{F^n : X \to \overline{\mathbf{R}}\}$ is said to be Mosco-epi-convergent to $F: X \to \overline{\mathbf{R}}$ if*

(i) for every $x \in X$, for every $x_n \xrightarrow{w} x$,

$$\liminf_n F^n(x_n) \geq F(x),$$

(ii) for every $x \in X$, there exists $x_n \xrightarrow{s} x$,

$$\limsup_n F^n(x_n) \leq F(x),$$

where w and s denote the weak and the strong topology of X respectively.

We then write

$$F = M - \lim_e F^n. \tag{2.8}$$

A basic property of Mosco-convergence is the following (cf. [33])

Theorem 2.2. *Let X be a reflexive Banach space and*

$$\{F^n;\ F: X \to \mathbf{R} \cup \{+\infty\}\}$$

a collection of closed convex proper functions. Then

$$F = M - \lim_e F^n \Leftrightarrow F^* = M - \lim_e (F^n)^*.$$

Comment. The above results establishes that the conjugacy operation is bicontinuous with respect to Mosco-convergence. In fact, as proved in [6], this operation is an isometry for suitable choice of metrics on $\Gamma_0(X)$ and $\Gamma_0(X^*)$.

2.3. Extended Mosco-epi/hypo-convergence

Let (E,τ) and (F,σ) be topological spaces and $\{L^n: E \times F \to \overline{\mathbf{R}}\}$ be a sequence of bivariate functions; we define, for every $(x,y) \in E \times F$:

$$(e_\tau/h_\sigma - \text{ls } L^n)(x,y) = \sup_{y_n \xrightarrow{\sigma} y} \inf_{x_n \xrightarrow{\tau} x} (\limsup_n L^n(x_n, y_n)), \tag{2.9}$$

$$(h_\sigma/e_\tau - \text{li } L^n)(x,y) = \inf_{x_n \xrightarrow{\tau} x} \sup_{y_n \xrightarrow{\sigma} y} (\liminf_n L^n(x_n, y_n)). \tag{2.10}$$

Definition 2.3 *(see [4], [5], [3]). Let X and Y be reflexive Banach spaces and $\{L^n,\ L: X \times Y^* \to \overline{\mathbf{R}}\}$ a collection of bivariate functions. We say that L^n Mosco epi/hypo-converges to L in the extended sense if*

$$\underline{\text{cl}}\,(e_s/h_w - \text{ls } L^n) \le L \le \overline{\text{cl}}\,(h_s/e_w - \text{li } L^n), \tag{2.11}$$

where $\underline{cl}$ and $\overline{\text{cl}}$ respectively denote the extended lower closure and the extended upper closure, i.e., for any function $F: (X,\tau) \to \overline{\mathbf{R}}$

$$\underline{cl}\, F = \begin{cases} \text{cl } F & \text{if cl } F > -\infty, \\ -\infty & \text{otherwise,} \end{cases}$$

cl F denoting the l.s.c. regularization of F, and $\overline{\text{cl}}\ F = \underline{cl}\,(-F)$.

For a convex function, it is well known that $\underline{\mathrm{cl}}\ F = F^{**}$ (let us observe that if $e_s/h_w - \mathrm{ls}\ L^n$ is convex in X and $h_s/e_w - \mathrm{li}\ L^n$ is concave in y^*, then in definition (2.11) the extended closure operations reduce to biconjugation).

The following result ([3]) establishes that the partial conjugation defined in (2.4) and (2.5) is bicontinuous when $\Gamma_0(X \times Y)$ is endowed with Mosco convergence and the classes of closed convex-concave functions is endowed with Mosco epi/hypo-convergence.

Theorem 2.4 *([3], Theorem 3.2). Let us consider X and Y, reflexive Banach spaces, and $\{F^n,\ F: X \times Y \to \mathbf{R}\}$ a collection of closed proper convex functions with associated equivalence classes of closed convex-concave functions denoted by L^n, L. Then, the following are equivalent :*

(i) $F^n \xrightarrow{M} F$,

(ii) $L^n \xrightarrow{M-e/h} L$ *(extended Mosco epi/hypo-convergence).*

The extended Mosco epi/hypo-convergence is variational convergence in a sense made precise by

Theorem 2.5 *([3], Theorem 2.6). Let us consider (X, τ) and (Y, σ) two general topological spaces and $\{K^n,\ K: X \times Y \to \overline{\mathbf{R}}\}$ a sequence of bivariate functions such that*

$$\underline{\mathrm{cl}}(e_\tau/h_\sigma - \mathrm{ls}\ K^n) \le K \le \overline{\mathrm{cl}}(h_\sigma/e_\tau - \mathrm{li}\ K^n),$$

$$\begin{cases} (\overline{x}_k, \overline{y}_k) \text{ is a saddle point of } K^{n_k} \text{ for all } k \in \mathbf{N}, \\ \overline{x}_k \xrightarrow{\tau} \overline{x} \text{ and } \overline{y}_k \xrightarrow{\sigma} \overline{y}. \end{cases} \tag{2.12}$$

Then

$$\begin{aligned} &(\overline{x}, \overline{y}) \text{ is a saddle point of } K \text{ and} \\ &K(\overline{x}, y) = \lim_{k \to +\infty} K^{n_k}(\overline{x}_k, \overline{y}_k) \end{aligned} \tag{2.13}$$

3. FURTHER PROPERTIES OF INFIMAL CONVOLUTION: REGULARIZATION EFFECTS, CONSERVATION OF INFIMAL VALUE AND MINIMIZING ELEMENTS

In preceding section 2.3 the partial Legendre-Fenchel transform

$$F \xrightarrow{*y} L = -(F^{*y}),$$
$$L(x,y^*) = \inf_{y\in Y}\{F(x,y) - \langle y^* \mid y\rangle\},$$

has been introduced and its continuity properties have been briefly reviewed. In the case of Convex Programming

$$\begin{cases} \min_{x\in X} f_0(x) \\ \text{subject to } f_i(x) \le 0 \quad i = 1,2,\dots,m \end{cases}$$

The Lagrangian function L attached to the classical perturbation function F is given by

$$L(x,y^*) = \begin{cases} f_0(x) - \sum_{i=1}^m y_i^* f_i(x) & \text{if } y^* \le 0, \\ -\infty & \text{otherwise.} \end{cases}$$

A major technical difficulty which arises when using directly this Lagrangian comes from the fact that the value $-\infty$ is taken on. A natural idea is to replace it by some smoother function either by approximation (penalization of the constraint $y^* \le 0$) or even better relying on the approximation-regularization by *infimal convolution* (with respect to the perturbation variable y^*). This last approach gives rise to the so-called *augmented Lagrangian*, for example

$$L_r(x,y^*) = L(x,\cdot)\,\Delta\frac{1}{2r}\,\|\cdot\|_{y^*}^2.$$

(where Δ denotes the sup-convolution) is the "quadratic" augmented Lagrangian).

In the next section we shall study the correspondance $F \mapsto L_r$ which can be viewed as a "generalized" partial duality transform and shall describe its continuity properties.

In this paragraph we study two main features of the inf-convolution operation which enlight the practical importance of augmented Lagrangian functions:

$$\text{The inf-convolution by a smooth kernel has a smoothing effect.} \tag{3.0}$$

$$\begin{array}{c}\text{The inf-convolution preserves the infimal value}\\ \text{and the set of minimizing elements.}\end{array} \tag{3.1}$$

In the above setting it follows that the Lagrangian and corresponding augmented Lagrangian functions have exactly the same saddle elements.

The following propositions, which are related to some results obtained simultaneously by M. Bougeard and J.P. Penot [14] (see also [13]) allow us to select well-behaved convolution kernels for which the two above basic properties (3.0) and (3.1) hold.

Proposition 3.1. *Let* (X, d) *be a general metric space,* $F: X \to \overline{\mathbf{R}}$ *a real extended valued function and* $k: \mathbf{R}^+ \to \mathbf{R}^+$ *a positive function such that*

$$k(0) = 0. \tag{3.2}$$

Let us define, for every x *belonging to* X,

$$F_k(x) = \inf_{y \in X} \{F(y) + k(d(x, y))\}.$$

Then

(a) $\inf_{x \in X} F_k(x) = \inf_{x \in X} F(x)$

(b) $\arg\min\,(\text{cl } F) \subset \arg\min\,(\text{cl } F_k)$.

Moreover, if we assume that

$$\inf F > -\infty \tag{3.3}$$

and

$$k(t) \to 0 \text{ implies } t \to 0, \tag{3.4}$$

then

(c) *arg min* (cl F) = *arg min* (cl F_k), *where* cl($\cdot$) *denotes the lower-semicontinuous regularization operation (with respect to the topology induced by d!).*

Proof

(a)

$$\begin{aligned}\inf_{x\in X} F_k(x) &= \inf_{x\in X}\left[\inf_{y\in Y}(F(y)+kd(x,y))\right]\\ &= \inf_{y\in Y}\left[\inf_{x\in X}(F(y)+kd(x,y))\right]\\ &= \inf_{y\in Y} F(y)\end{aligned}$$

since $k(0) = 0$.

(b) Let us now consider $\overline{x} \in$ arg min (cl F), that means

$$\text{cl } F(\overline{x}) = \inf_X(\text{cl } F) = \inf_X F = \inf_X F_k = \inf_X \text{cl } F_k,$$

thus we derive since $F_k \leq F$,

$$\text{cl } F_k(\overline{x}) \leq \text{cl } F(\overline{x}) = \inf_X \text{cl } F_k$$

and (b) follows.

(c) If $F \equiv +\infty$, there is nothing to prove, so, we can assume that F is proper.

Let us consider $x^{\#} \in$ arg min (cl F_k), that is

$$\text{cl } F_k(x^{\#}) = \inf_X \text{cl } F_k = \inf_X F.$$

For every $\epsilon > 0$, by definition of cl F_k, there exists $\xi_\epsilon \in X$ which satisfies

$$\begin{aligned}&d(\xi_\epsilon, x^{\#}) \leq \epsilon\\ &F_k(\xi_\epsilon) < \inf_X F + \epsilon\end{aligned} \tag{3.5}$$

(let us recall that $\inf_X F$ is finite thanks to (3.3) and the properness of F). Using now the definition of F_k, we derive the existence of $y_\epsilon \in X$ such that

$$F(y_\epsilon) + k(d(y_\epsilon, \xi_\epsilon)) < \inf_X F + \epsilon. \tag{3.6}$$

Since $F(y_\epsilon) \geq \inf_X F$, we obtain

$$k(d(y_\epsilon, \xi_\epsilon)) < \epsilon,$$

which ensures $y_\epsilon \to x^\#$, thanks to (3.4) and (3.5).

Passing to the limit inferior on both sides of (3.6), using the fact that $k \geq 0$, we derive

$$\begin{aligned} \operatorname{cl} F(x^\#) &\leq \liminf_{\epsilon \to 0} F(y_\epsilon) \\ &\leq \inf_X F, \end{aligned}$$

that is, $x^\#$ minimizes cl F. ∎

The next proposition deals with regularity of the approximates. Roughly speaking , F_k inherits the Lipschitz regularity of k. For technical reasons, we shall distinguish the Lipschitz case and the locally Lipschitz one which is surprisingly more involved.

Proposition 3.2. *Let us assume that the function* $k: \mathbf{R}^+ \to \mathbf{R}^+$ *satisfies* $k(0) = 0$ *and let us consider a proper function* $F: X \to \overline{\mathbf{R}}$ *which satisfies the growth condition*

$$\begin{aligned} &\text{for every } x \in X, \text{ there exists } c(x) \in \mathbf{R} \text{ such that} \\ &F(y) \geq -k(d(x,y)) + c(x), \text{ for every } y \in X. \end{aligned} \tag{3.7}$$

(a) If $k(\cdot)$ *is Lipschitz on* $\mathbf{R}^+$ *then*

$$F_k \text{ is Lipschitz on } X. \tag{3.8}$$

(b) If $k(\cdot)$ *is locally Lipschitz and verifies*

$$\begin{aligned} &F(\cdot) + k(d(\cdot, x)) \text{ is uniformly coercive} \\ &\text{when } x \text{ ranges over a bounded set,} \end{aligned} \tag{3.9}$$

(this means that $F(y) + k(d(y,x)) \leq M$ *with* x *in a bounded set implies that* y *ranges over a bounded set), then*

$$F_k \text{ is locally Lipschitz on } X. \tag{3.10}$$

Proof. Let us observe, thanks to the growth condition (3.7) and the properness of F, that F_k is everywhere finite.

(a) Let $x_1 \in X$, $x_2 \in X$, $\epsilon > 0$ and $\xi_{1,\epsilon} \in X$ such that

$$F(\xi_{1,\epsilon}) + k(d(\xi_{1,\epsilon}, x_1)) \le F_k(x_1) + \epsilon.$$

From the definition of $F_k(x_2)$, we derive

$$F_k(x_2) \le F(\xi_{1,\epsilon}) + k(d(\xi_{1,\epsilon}, x_2)).$$

Adding the two last inequalities and keeping in mind that $F(\xi_{1,\epsilon})$ is finite, we obtain

$$F_k(x_2) - F_k(x_1) \le k(d(\xi_{1,\epsilon}, x_2) - k(d(\xi_{1,\epsilon}, x_1)) + \epsilon. \tag{3.11}$$

Assuming $k(\cdot)$ to be Lipschitz, there exists $L > 0$ such that

$$|k(s) - k(t)| \le L|s - t|$$

for every $s \ge 0$, $t \ge 0$. Using the triangle inequality in (3.11), we derive

$$F_k(x_2) - F_k(x_1) \le Ld(x_1, x_2) + \epsilon.$$

Letting $\epsilon \downarrow 0$, we obtain that F_k is Lipschitz.

(b) We claim that

$$F_k \text{ is bounded from above on bounded subsets of } X. \tag{3.12}$$

Indeed

$$F_k(x) \le F(x_0) + k(d(x_0, x)),$$

where $x_0 \in X$ is such that $F(x_0) < +\infty$

Therefore (3.12) follows from the continuity of $k(\cdot)$. Let us consider a bounded set $B \subset X$, $(x_1, x_2) \in B \times B$ and $\xi_{1,\epsilon}$ defined as above. By definition of $\xi_{1,\epsilon}$

$$F(\xi_{1,\epsilon}) + k(d(\xi_{1,\epsilon}, x_1)) \le F_k(x_1) + \epsilon \le M$$

when x_1 ranges over B (see (3.12)) and $0 < \epsilon < \epsilon_0$. Using (3.9), we know that $\xi_{1,\epsilon}$ remains bounded. On other hand, let us recall that

$$F_k(x_2) - F_k(x_1) \leq k(d(\xi_{1,\epsilon}, x_2) - k(d(\xi_{1,\epsilon}, x_1)) + \epsilon. \tag{3.13}$$

Using the fact that $\xi_{1,\epsilon}$ is bounded, we derive the existence of $M > 0$ such that

$$(d(\xi_{1,\epsilon}, x_2) \leq M, \quad (d(\xi_{1,\epsilon}, x_1) \leq M$$

for every $(x_1, x_2) \in B \times B$ and $0 < \epsilon \leq \epsilon_0$.

The function $k(\cdot)$ being locally Lipschitz, is Lipschitz on $[0, M]$, so, there exists $L > 0$ such that

$$|k(s) - k(t)| \leq L|s - t|, \text{ for every } (s,t) \in [0, M] \times [0, M].$$

From (3.13) we derive

$$\begin{aligned} F_k(x_2) - F_k(x_1) &\leq L|d(\xi_{1,\epsilon}, x_2) - d(\xi_{1,\epsilon}, x_1)) + \epsilon \\ &\leq Ld(x_1, x_2) + \epsilon \end{aligned}$$

for every $(x_1, x_2) \in B \times B$.

Letting $\epsilon \downarrow 0$ archieves the proof of (3.10). ∎

Comments

(1) A sufficient conditions which guarantees the growth condition (3.7) and the coerciveness assumptions (3.9) is the following:

$$\left.\begin{aligned} &\text{for every } B \subset X \text{ bounded,} \\ &\text{there exists } \alpha < 1 \text{ and } C \in \mathbf{R} \\ &\text{such that } F(y) \geq -\alpha k(d(x,y)) - C \\ &\text{for every } x \in B \text{ and } y \in X \end{aligned}\right. \tag{3.14a}$$

and

$$k(\cdot) \text{ is coercive } (\lim_{t \to +\infty} k(t) = +\infty) \tag{3.14b}.$$

Indeed by taking $B = \{x\}$ for every $x \in X$ the growth condition (3.7) is fulfilled. Moreover if B is a bounded subset and if $F(y) + k(d(x,y)) \leq M$ with $x \in B$, we derive, using $\alpha < 1$ and $C \in \mathbf{R}$ defined in (3.14),

$$-C + (1-\alpha)k(d(x,y)) \leq M \text{ and}$$

$$k(d(x,y)) \leq \frac{M+C}{1-\alpha} \text{for every } x \in B.$$

From the coerciveness of $k(\cdot)$ and the boundness of B, (3.9) follows.

(2) Take X a Banach space, for the following possible choices of $k(\cdot)$, we have

$k(r) = \frac{r^2}{2}$	Moreau-Yosida approximate
$k(r) = r$	Baire-Wijsman approximate
$k(r) = \frac{1}{2}r^2 + \epsilon r$	Gauvin approximate

(3) We stress the fact that, as far one is only concerned by the minimization problem, one can replace any function F by a smoother Lipschitzian function which has exactly same minima and same minimization set as the original one.

This feature has been already exploited by the authors ([6]) when defining rate of convergence for sequences of convex functions.

A major difficulty in this kind of question is that the domains of the functions may also vary. By using the above device (note that the regularized functions F_k are everywhere defined and locally Lipschitz) one can define for every $\rho \geq 0$ the following distance

$$d_\rho(F,G) = \sup_{\|x\| \leq \rho} |F_k(x) - G_k(x)|$$

which allows us to derive convergence rates for the solutions of the corresponding minimization problems.

Indeed, in the convex case, and $k(r) = \frac{r^2}{2}$ (that is the Moreau-Yosida approximate), the whole function F is determined by one of its approximates.

Just notice that

$$F_k = F \nabla \frac{|\cdot|^2}{2k},$$

hence

$$F_k^* = F^* + k\frac{|\cdot|^2}{2} \quad (F_k^* \text{ stands for } (F_k)^*),$$

and if F is closed convex

$$F = F^{**} = \left(F_k^* - k\frac{|\cdot|^2}{2}\right)^* . \tag{3.15}$$

At this stage a natural question is: what is the largest class of functions for which the correspondance $F \mapsto F_k$ is one-to-one? (i.e., F uniquely determined by one approximate). The class of closed convex function by the preceding argument does satisfy this property. Indeed one can exhibit a larger class, namely functions which are convex up to the square of the norm, for which this property still holds (further results concerning this class of functions can be found in M. Bougeard [13]). This is made precise in the following

Proposition 3.3. *Let H be a Hilbert space; for any proper function $F: H \to \mathbf{R} \cup \{+\infty\}$ and $\lambda > 0$ let us denote*

$$F_\lambda(x) = \inf_{y \in X} \left\{ F(y) + \frac{1}{2\lambda} \|x - y\|^2 \right\} \tag{3.16}$$

the Moreau-Yosida approximate of index λ of F. Let us denote by Γ_k the class of functions $F: H \to \mathbf{R} \cup \{+\infty\}$ such that $F + k\|\cdot\|^2$ is proper, closed and convex. Then, for every $\lambda > 0$, $k \geq 0$ such that $\frac{1}{2\lambda} > k$ the correspondence

$$F \in \Gamma_k \mapsto F_\lambda$$

is one-to-one, i.e., $F \in \Gamma_k$ is uniquely determined by one of its approximates. Moreover F_λ is C^1.

Proof. Let us first notice that F satisfies a growth condition

$$F(x) \geq -k\|x\|^2 - c(\|x\| + 1)$$

since $F + k\left\|\cdot\right\|^2$ is closed convex and proper.

Hence for every $\lambda > 0$, $k \geq 0$ such that $\frac{1}{2\lambda} > k$, F satisfies conditions of Proposition 3.2 and F_λ is everywhere defined and locally Lipschitz. Introducing φ, a closed convex function such that $F = \varphi - k\left\|\cdot\right\|^2$, we have

$$\begin{aligned} F_\lambda(x) &= \inf_{y\in X}\left\{\varphi(y) - k\|y\|^2 + \frac{1}{2\lambda}\|x-y\|^2\right\} \\ &= \inf_{y\in X}\left\{\varphi(y) + \left(\frac{1}{2\lambda} - k\right)\|y\|^2 - \frac{1}{2\lambda}\|y\|^2 + \frac{1}{2\lambda}\|x-y\|^2\right\}. \end{aligned}$$

Simplifying the last expression, we obtain

$$\begin{aligned} F_\lambda(x) &= \inf_{y\in X}\left\{\varphi(y) + \left(\frac{1}{2\lambda} - k\right)\|y\|^2 - \frac{1}{\lambda}\langle x \mid y\rangle\right\} + \frac{1}{2\lambda}\|x\|^2 \\ &= -\left[\varphi + \left(\frac{1}{2\lambda} - k\right)\|\cdot\|^2\right]^*\left(\frac{x}{\lambda}\right) + \frac{1}{2\lambda}\|x\|^2 \\ &= \frac{1}{2\lambda}\|x\|^2 - \left(\varphi^* \nabla \frac{1}{2\left(\frac{1}{\lambda} - 2k\right)}\|\cdot\|^2\right)\left(\frac{x}{\lambda}\right), \end{aligned}$$

and finally

$$F_\lambda(x) = \frac{1}{2\lambda}\|x\|^2 - (\varphi^*)_{\frac{1}{\lambda}-2k}\left(\frac{x}{\lambda}\right). \tag{3.17}$$

From this last expression we easily derive the conclusions of Proposition 3.3, we first notice that given F_λ, (3.17) uniquely determines $(\varphi^*)_{\frac{1}{\lambda}-2k}$, and from the above argument in the convex case, φ^* is uniquely determined. The function φ being closed and convex is again uniquely determined by its conjugate and so is F.

Moreover from classical properties of the Moreau-Yosida approximation for closed convex functions (cf. [15], [1]) $(\varphi^*)_{\frac{1}{\lambda}-2k}$ is a C^1 function and from (3.17) so is F_λ. ∎

Remark. Without geometric assumptions on F, the Moreau-Yosida transform $F \to F_\lambda$ fails to be a one-to-one mapping. Take for instance $H = \mathbf{R}$, $\lambda = 1/2$ and

$$F(x) = \begin{cases} -1 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0. \end{cases}$$

A quite elementary computation shows that all functions Q such that $F \geq Q \geq G$ with

$$G(x) = \begin{cases} -1 & \text{if } x \leq 0, \\ 1 - (\sqrt{2} - x)^2 & \text{if } 0 \leq x \leq \sqrt{2}, \\ 1 & \text{if } x \geq \sqrt{2}, \end{cases}$$

verify

$$Q_{1/2}(x) = F_{1/2}(x) = \begin{cases} -1 & \text{if } x \leq 0, \\ x^2 - 1 & \text{if } 0 \leq x \leq \sqrt{2}, \\ 1 & \text{if } x \geq \sqrt{2}. \end{cases}$$

4. CONVERGENCE OF AUGMENTED LAGRANGIANS AND CONTINUITY OF THE "AUGMENTED" PARTIAL LEGENDRE-FENCHEL TRANSFORM

From now on we assume that

X, X^*, Y, Y^* are reflexive Banach spaces
equipped with strictly convex norms
and satisfy the following property: (4.1)
weak convergence and convergence of the norms
imply strong convergence.

As far as one is only concerned with topological properties it is not a restrictive assumption since a theorem of S. Trojanski and E. Asplund asserts that every reflexive Banach space can be renormed in order to verify (4.1). When this is done, the norm is Frechet-differentiable (except at the origin!) and one can define

$$\forall x \in X,\ H(x) = D\left(\frac{1}{2}\|\cdot\|^2\right)(x) \in X^*. \tag{4.2}$$

The map $H: X \to X^*$ is called the duality map and characterized by

$$\begin{aligned} &H(x) \text{ is the unique element } x^* \in X^* \text{ which satisfies} \\ &\|x^*\|_* = \|x\| \text{ and } \langle x^* \mid x\rangle = \|x\|^2). \end{aligned} \tag{4.3}$$

The duality map is then a homeomorphism between X and X^* and verifies

$$H(\lambda x) = \lambda H(x)$$
$$H^{-1}(x^*) = D\left(\frac{1}{2}\left\|\cdot\right\|_*^2\right)(x^*)$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

From preceding results (section 2 and 3) it follows that the "augmented" partial Legendre-Fenchel transform $F \mapsto L_r$ is one-to-one correspondence, where

$F: X \times Y \to \mathbf{R} \cup \{+\infty\}$ is a closed convex proper function,

$L: X \times Y^* \to \overline{\mathbf{R}}$ is an element of the class of closed proper convex-concave functions associated with F by (2.4) and (2.5).

$$L: X \times Y^* \to \mathbf{R} \cup \{+\infty\},\ r > 0,$$

is the classical "quadratic " augmented Lagrangian (see [11], [20], [35], [36]):

$$\begin{aligned} L_r(x, y^*) &= \left(L(x, \cdot)\Delta \frac{1}{2r}\left\|\cdot\right\|_Y^2 *\right)(y^*) \\ &= \sup_{\eta \in Y^*}\left\{L(x,\eta) - \frac{1}{2r}\left\|y^* - \eta\right\|_*^2\right\} \\ &= \text{for every } L \in [\underline{L}, \overline{L}]. \end{aligned}$$

The terminology is justified by the following equivalent formulation of L_r, obtained by taking $L = \overline{L}$,

$$\begin{aligned} -L_r &= F^{*y} \nabla \frac{1}{2r}\left\|\cdot\right\|_{Y^*}^2 \\ &= \left(F + \frac{r}{2}\left\|\cdot\right\|_Y^2\right)^{*y}; \end{aligned} \tag{4.4}$$

thus

$$L_r(x, y^*) = \inf_{y \in Y}\left\{F(x, y) + \frac{r}{2}\left\|y\right\|^2 - \langle y^* \mid y\rangle\right\},$$

which amounts to replacing F by $F + \frac{r}{2}\left\|\cdot\right\|_y^2$ where Y is the perturbation space.

In the case of Convex Programming the quadratic augmented Lagrangian is given by the following formula

$$L_r(x, y^*) = f_0(x) + \sum_{i=1}^{m} \psi_r(f_i(x), y_i^*) \forall x \in X, \ \forall y^* \in \mathbf{R}^m,$$

where

$$\psi_r(s,t) = \begin{cases} \frac{rs^2}{2} - ts & \text{if } s \geq \frac{t}{r}, \\ -\frac{1}{2r} t^2 & \text{if } s \leq \frac{t}{r}. \end{cases}$$

The following proposition guarantees that the saddle points and saddle values are preserved when replacing L by L_r, in the general (metric) setting.

Proposition 4.1. *L and L_r have same saddle value, and every saddle point of L is a saddle point of L_r.*

Proof. Take $(\overline{x}, \overline{y}^*)$ a saddle point of L; it is characterized by the following inequality:

$$\sup_{y^* \in Y^*} L(\overline{x}, y^*) \leq \inf_{x \in X} L(x, \overline{y}^*).$$

By Proposition 3.1

$$\sup_{y^* \in Y^*} L_r(\overline{x}, y^*) = \sup_{y^* \in Y^*} L(\overline{x}, y^*).$$

Noticing that L_r is greater than or equal to L,

$$\inf_{x \in X} L(x, \overline{y}^*) \leq \inf_{x \in X} L_r(x, \overline{y}^*).$$

Combining the preceding inequalities induces

$$\sup_{y^* \in Y^*} L_r(\overline{x}, y^*) = \sup_{y^* \in Y^*} L(\overline{x}, y^*) \leq \inf_{x \in X} L(x, \overline{y}^*) \leq \inf_{x \in X} L_r(x, \overline{y}^*),$$

that is

$$(\overline{x}, \overline{y}^*) \text{ is also a saddle point of } L_r,$$

and

$$L(\overline{x}, \overline{y}^*) = L_r(\overline{x}, \overline{y}^*)$$

i.e., L and L_r have same saddle value. ∎

Remark. The preceding conclusions still hold when instead of quadratic augmented Lagrangian, one considers augmented Lagrangian obtained through inf-convolution by a kernel $k(\cdot)$ satisfying assumptions of Proposition 3.1.

In the convex-concave setting, a more precise result can be obtained.

> **Proposition 4.1'** . *Let $L: X \times Y^* \to \overline{\mathbf{R}}$ be a closed convex-concave function where X, Y are reflexive Banach spaces verifying 4.1. Then L and L_r have the same saddle points and saddle values.*

Proof. In the lines of R.T. Rockafellar (see [37], [38]), we consider, for a closed convex-concave function L, its subdifferential

$$\partial L(x, y^*) = \partial_1 L(x, y^*) \times (-\partial_2(-L(x, y^*))),$$

where $\partial_1 L$ and $\partial_2(-L)$ denote the convex subdifferential with respect to the first and the second variable. It is well known that

$$(u^*, v) \in \partial L(x, y^*) \Leftrightarrow (u^*, y^*) \in \partial F(x, -v), \tag{4.5}$$

and

$$(\overline{x}, \overline{y}^*) \text{ is a saddle point of } L \Leftrightarrow (0,0) \in \partial L(\overline{x}, \overline{y}^*). \tag{4.6}$$

Moreover, when $(\overline{x}, \overline{y}^*)$ is a saddle point of L,

$$L(\overline{x}, \overline{y}^*) = F(\overline{x}, 0) = G(0, \overline{y}^*),$$

where F and G are respectively the convex and concave parent of L (see (2.1), (2.2)).

Let us now return to the proof of Proposition (4.1').

$$\begin{aligned}(\overline{x}, \overline{y}^*) \text{ is a saddle point of } L_r &\Leftrightarrow (0,0) \in \partial L_r(\overline{x}, \overline{y}^*) \\ &\Leftrightarrow (0, \overline{y}^*) \in \partial \left(F + \frac{r}{2} \|\cdot\|_Y^2\right)(\overline{x}, 0).\end{aligned}$$

Let us observe that

$$\partial\left(F+\frac{r}{2}\|\cdot\|_Y^2\right)(x,y)=\partial F(x,y)+(O,rH(y))$$

since the function $\frac{r}{2}\|\cdot\|_Y^2$ is continuous. ($H(y)$ is the duality map defined in 4.3). Using the fact that $H(0)=0$, we derive

$$\begin{aligned}(\overline{x},\overline{y}^*) \text{ is a saddle point of } L_r &\Leftrightarrow (0,\overline{y}^*)\in\partial F(\overline{x},0)\\ &\Leftrightarrow (0,0)\in\partial L(\overline{x},\overline{y}^*) \quad \text{(from 4.5)}\\ &\Leftrightarrow (\overline{x},\overline{y}^*) \text{ is a saddle point of } L.\end{aligned}$$

Moreover the saddle values verify

$$\begin{aligned}L_r(\overline{x},\overline{y}^*)&=\left(F+\frac{r}{2}\|\cdot\|_Y^2\right)(\overline{x},0)\\ &=F(\overline{x},0)=L(\overline{x},\overline{y}^*),\end{aligned}$$

which ends the proof of Proposition 4.1'. ∎

Comment. The conclusions of Proposition 4.1' still hold when replacing L_r by $L_k(x,y^*)=\sup_{\eta^*\in Y^*}\{L(x,\eta^*)-k(\|y^*-\eta^*\|)\}$, where $k:\mathbf{R}\to\mathbf{R}$ is an even convex function such that k^* is derivable at the origin and verifies $k^*(0)=0$ and $(k^*)'(0)=0$.

Indeed, in this setting, L_k is the closed convex-concave function associated with the convex parent

$$\Phi(x;y)=F(x,y)+k^*(\|y\|).$$

We can now give the main result of this section.

Theorem 4.2. *Let X, Y be reflexive Banach spaces renormed as in (4.1). There is equivalence between*

(i) $F^n \xrightarrow{M} F$;

(ii) $L^n \xrightarrow{M-e/h} L$;

(iii) $L_r^n \xrightarrow{M-e/h} L_r$ *for every (resp. some)* $r>0$,

(iv) $L_r^n(\cdot, y^*) \xrightarrow{M} L_r(\cdot, y^*)$ *for every* $r > 0$ *and* $y^* \in Y^*$.

Before detailing the proof of Theorem 4.2, we recall the key facts used in this proof:

$$F^n \xrightarrow{M} F \Leftrightarrow (F^n)^* \xrightarrow{M} F^*; \tag{4.7}$$

$$F^n \xrightarrow{M} F \Leftrightarrow \forall \lambda > 0,\ \forall x \in X,\ \lim_{n \to +\infty} (F^n)_\lambda(x) = F_\lambda(x), \tag{4.8}$$

where $\{F^n,\ F: X \to \mathbf{R}\}$ is a collection of convex closed proper functions and

$$F_\lambda(x) = \inf_{y \in X} \left\{ F(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\} \tag{4.9}$$

is the Moreau-Yosida approximate of parameter λ of F. Equivalence (4.7) was proved by U. Mosco in [33], and (4.8) is Theorem 3.26 of [1].

Proof of Theorem 4.2

$(i) \Leftrightarrow (ii)$ is Theorem 2.4.

$(i) \Rightarrow (iii)$. $F^n \to F \Rightarrow F^n + \frac{r}{2} \|\cdot\|_Y^2 \xrightarrow{M} F + \frac{r}{2} \|\cdot\|_Y^2$ and (iii) follows from Theorem 2.4 and formula (4.4).

$(iii) \Rightarrow (i)$. Assuming (iii) holds for some $r > 0$, we derive, from Theorem 2.4 that

$$F^n + \frac{r}{2} \|\cdot\|_Y^2 \xrightarrow{M} F + \frac{r}{2} \|\cdot\|_Y^2$$

and then

$$F^n + \frac{r}{2} \left(\|\cdot\|_Y^2 + \|\cdot\|_Y^2 \right) \xrightarrow{M} F + \frac{r}{2} \left(\|\cdot\|_Y^2 + \|\cdot\|_Y^2 \right).$$

From (4.7), we derive that

$$(F^n)_r^* \xrightarrow{M} (F^*)_r, \text{ for some } r > 0. \tag{4.10}$$

Using the resolvent equation $(\varphi_r)_s = \varphi_{r+s}$, we obtain from (4.8) and (4.10) that

$$(F^n)_\rho^*(x^*) \xrightarrow{n \to +\infty} (F^*)_\rho(x^*) \text{ for every } \rho > 0 \text{ and } x^* \in X^*.$$

Using again (4.8), in fact a slightly weakened version (see [1]), we derive

$$(F^n)^* \xrightarrow{M} (F^*)$$

and by (4.7)

$$F^n \xrightarrow{M} F.$$

$(i) \Leftrightarrow (iv)$. We observe that the convex function

$$\psi(x) = L_(x, y^*) = \inf_{y \in Y} \{F(x, y) + \frac{r}{2} \|y\|^2 - \langle y^* \mid y \rangle\},$$

which is not identically equal to $+\infty$ since F is proper, does not take on the value $-\infty$, and is l.s.c. since, for every $x \in X$, the function

$$y \mapsto F(x, y) + \frac{r}{2} \|y\|^2 - \langle y^* \mid y \rangle$$

is uniformly coercive when x remains bounded. Let us define (in the following argument y^* is fixed)

$$\Psi^n(x) = L_r^n(x, y^*),$$

$$\Psi(x) = L_r(x, y^*),$$

and observe that

$$(\Psi^n)(x^*) = \left(F^n + \frac{r}{2} \|\cdot\|_Y^2\right)^* (x^*, y^*) \text{ for every } y^* \in Y^*$$
$$= \inf_{\eta^* \in Y^*} \left\{ (F^n)^*(x^*, \eta^*) + \frac{1}{2r} \|y^* - \eta^*\|^2 \right\}.$$

Let us now consider $\rho > 0$ and $(\Psi^n)^*_\rho$ the Moreau-Yosida approximate of $(\Psi^n)^*$ of parameter ρ. We derive

$$(\Psi^n)^*_\rho(x^*) = \inf_{\xi^* \in X^*} \left\{ (\Psi^n)^*(\xi^*) + \frac{1}{2\rho} \|x^* - \xi^*\|^2 \right\} \text{ for every } x \in X^*;$$

$$(\Psi^n)^*_\rho(x^*) = \inf_{(\xi^*, \eta^*) \in X^* \times Y^*} \left\{ (F^n)^*(\xi^*, \eta^*) + \frac{1}{2\rho} \|x^* - \xi^*\|^2 + \frac{1}{2r} \|y^* - \eta^*\|^2 \right\}. \tag{4.11}$$

The same calculation holds for Ψ and we obtain

$$(\Psi^n)^*_\rho(x^*) = \inf_{(\xi^*, \eta^*) \in X^* \times Y^*} \left\{ F^*(\xi^*, \eta^*) + \frac{1}{2\rho} \|x^* - \xi^*\|^2 + \frac{1}{2r} \|y^* - \eta^*\|^2 \right\}. \tag{4.12}$$

Let us return to the proof of the equivalence $(i) \Leftrightarrow (iv)$. By definitions of Ψ and Ψ^n

$$
\begin{aligned}
\textbf{(iv)} &\Leftrightarrow \forall r > 0,\ \forall y^* \in Y^*,\ \Psi^n \xrightarrow{M} \Psi \\
&\Leftrightarrow \forall r > 0,\ \forall y^* \in Y^*,\ (\Psi^n)^* \xrightarrow{M} \Psi^* \\
&\Leftrightarrow \forall r > 0,\ \forall \rho > 0,\ \forall (x^*, y^*) \in X^* \times Y^*, \\
&\qquad\qquad \lim_{n \to +\infty} (\Psi^n)^*_\rho(x^*) = \Psi^*_\rho(x^*) \\
&\Leftrightarrow \forall r > 0,\ \forall (x^*, y^*) \in X^* \times Y^*, \\
&\qquad\qquad \lim_{n \to +\infty} (F^n)^*_r(x^*, y^*) = F^*_r(x^*, y^*) \\
&\Leftrightarrow (F^n)^* \xrightarrow{M} F^* \\
&\Leftrightarrow F^n \xrightarrow{M} F,
\end{aligned}
$$

which ends the proof of Theorem 4.2. ∎

Comments

(1) Theorem 4.2 can be viewed as a continuity result of the generalized partial duality transform

$$-L_r(x, y^*) = \sup_{y \in Y} (y^*, y) - F(x, y),$$

where (y^*, y) denotes the non-bilinear coupling

$$(y^*, y) =: \langle y^* \mid y \rangle - \frac{r}{2} \|y\|^2$$

(cf. the papers of S. Dolecki [18] and M. Volle [41]).

(2) One can give an equivalent expression of the augmented Lagrangian in the Hilbert spaces by using Theorem 2.9 of [6]

$$
\begin{aligned}
-L_r(x, y^*) &= \left(F^{*y} \nabla \frac{1}{2r} \|\cdot\|^2 \right)(x, y^*), \\
&= (F^{*y})_r(x, y^*), \\
&= \frac{1}{2r} \|y^*\|^2 - F(x, \cdot)_{1/r}(y^*/r). \ \blacksquare
\end{aligned}
$$

5. Moreau-Yosida approximates of closed convex-concave functions. Equivalence between extended Mosco epi/hypo convergence and pointwise limit of Moreau-Yosida approximates

In [4], H. Attouch and R. Wets have defined the upper and lower Moreau-Yosida approximates of general bevariate functions L by means of the following formula

$$L^{\uparrow}_{\lambda,\mu} = \inf_{\xi\in X}\sup_{\eta^*\in Y^*}\left\{L(\xi,\eta^*) + \frac{1}{2\lambda}\|x-\xi\|^2 - \frac{1}{2\mu}\|y^*-\eta^*\|^2\right\}$$

$$L^{\downarrow}_{\lambda,\mu} = \sup_{\eta^*\in Y^*}\inf_{\xi\in X}\left\{L(\xi,\eta^*) + \frac{1}{2\lambda}\|x-\xi\|^2 - \frac{1}{2\mu}\|y^*-\eta^*\|^2\right\}.$$

When L is a closed convex-concave function, these two quantities prove to be equal as made precise by the following

Theorem 5.1. *Let X, Y be reflexive Banach spaces (renormed as in (4.1)) and*

$$L: X\times Y^* \to \overline{\mathbf{R}}$$

a closed convex-concave function.

(a) Then, for all $\lambda>0$, $\mu>0$

$$L^{\downarrow}_{\lambda,\mu} = L^{\uparrow}_{\lambda,\mu} := L_{\lambda,\mu}. \tag{5.1}$$

$L_{\lambda,\mu}$ is called the Moreau-Yosida approximate of index λ, μ of L.

(b) L and $L_{\lambda,\mu}$ have same saddle value and saddle points.

(c) For all $(x,y^*)\in X\times Y^*$ the function

$$L(\xi,\eta^*) = L(\xi,\eta^*) + \frac{1}{2\lambda}\|x-\xi\|^2 - \frac{1}{2\mu}\|y^*-\eta^*\|^2$$

has a unique saddle point $\left(x_{\lambda,\mu}, y^*_{\lambda,\mu}\right)$ characterized by

$$\left(H\left(\frac{x-x_{\lambda,\mu}}{\lambda}\right), -H^*\left(\frac{y^*-y^*_{\lambda,\mu}}{\mu}\right)\right) \in \partial L\left(x_{\lambda,\mu}, y^*_{\lambda,\mu}\right), \tag{5.2}$$

where $H: X \to X^*$ and $H^*: Y^* \to Y$ are the duality maps defined in (4.3) and $\partial L = \partial_1 L \times \left(-\partial_2(-L)\right)$.

(d) $L_{\lambda,\mu}$ is locally Lipschitz convex-concave function of class C^1 on $X \times Y^*$, with derivative

$$DL_{\lambda,\mu}(x, y^*) = \left(H\left(\frac{x - x_{\lambda,\mu}}{\lambda}\right), -H^*\left(\frac{y^* - y^*_{\lambda,\mu}}{\mu}\right)\right).$$

Proof

(a) and (c). We shall use the inf-sup theorem of J.J. Moreau [30]; let us recall this result. Under the assumptions

$$\begin{aligned}
&U,\ V \text{ are locally convex t.v.s.}\\
&K: U \times V \to \overline{\mathbf{R}} \text{ is convex-concave}\\
&K(\cdot, v) \in \Gamma(U) \text{ for all } v \in V,\\
&\text{there exists } v_0 \in V,\ k_0 > \inf\nolimits_{u\in U} K(u, v_0) \text{ such that}\\
&\{u \in U : K(u, v_0 \le k_0\} \text{ is weakly compact.}
\end{aligned} \tag{5.3}$$

Then

$$\inf_{u\in U} \sup_{v\in V} K(u,v) = \sup_{v\in V} \inf_{u\in U} K(u,v). \tag{5.4}$$

Moreover

$$\inf_{u\in U} \sup_{v\in V} K(u,v) = \min_{u\in U} \sup_{v\in V} K(u,v). \tag{5.5}$$

Let us define

$$K(\xi, \eta^*) = L(\xi, \eta^*) + \frac{1}{2\lambda} \|\xi - x\|^2 - \frac{1}{2\mu} \|\eta^* - y^*\|^2.$$

K is closed convex-concave function such that

$$\overline{K}(\xi, \eta^*) = \overline{L}(\xi, \eta^*) + \frac{1}{2\lambda} \|\xi - x\|^2 - \frac{1}{2\mu} \|\eta^* - y^*\|^2,$$

$$\underline{K}(\xi, \eta^*) = \underline{L}(\xi, \eta^*) + \frac{1}{2\lambda} \|\xi - x\|^2 - \frac{1}{2\mu} \|\eta^* - y^*\|^2.$$

It is clear that $\underline{K}$ verifies the assumptions (5.3); we derive that

$$\begin{aligned}\sup_{\eta^*\in Y^*}\inf_{\xi\in X} K(\xi,\eta^*) &= \sup_{\eta^*\in Y^*}\inf_{\xi\in X}\underline{K}(\xi,\eta^*),\\ &= \min_{\xi\in X}\sup_{\eta^*\in Y^*}\underline{K}(\xi,\eta^*),\ (\text{ from } (5.4),(5.5)),\\ &= \min_{\xi\in X}\sup_{\eta^*\in Y^*}\overline{K}(\xi,\eta^*),\\ &= \min_{\xi\in X}\sup_{\eta^*\in Y^*}K(\xi,\eta^*).\end{aligned}$$

The same argument applied to $(-\overline{K})$ shows that

$$\inf_{\xi\in X}\sup_{\eta^*\in Y^*}K(\xi,\eta^*) = \max_{\eta^*\in Y^*}\inf_{\xi\in X}K(\xi,\eta^*).$$

It follows that

$$\max_{\eta^*\in Y^*}\inf_{\xi\in X}K(\xi,\eta^*) = \min_{\xi\in X}\sup_{\eta^*\in Y^*}K(\xi,\eta^*),$$

which ensures the existence of a saddle point which is unique thanks to the strict convexity-concavity of K; the characterization (5.2) of this saddle point is then straightforward.

(b) Let us consider the quadratic augmented Lagrangian

$$L_\mu(x,y^*) = \sup_{\eta^*\in Y^*}\left\{L(x,\eta^*) - \frac{1}{2\mu}\|y^*-\eta^*\|^2\right\}.$$

From Proposition 4.1', L_μ and L have same saddle values and saddle points. Exchanging the role played by the variables and taking the augmented Lagrangian of parameter λ of $(-L_\mu)$, we obtain the closed concave -convex function K defined by

$$\begin{aligned}K(x,y^*) &= \sup_{\xi\in X}\inf_{\eta^*\in Y^*}\left\{L(\xi,\eta^*) + \frac{1}{2\mu}\|y^*-\eta^*\|^2 - \frac{1}{2\lambda}\|x-\xi\|^2\right\},\\ &= -L^{\dagger}_{\lambda,\mu}(x,y^*),\\ &= -L_{\lambda,\mu}(x,y^*).\end{aligned}$$

Using again Proposition 4.1', part (b) of Theorem 5.1 follows.

(d) We claim that the operator

$$J_{\lambda,\mu}(x, y^*) = (x_{\lambda,\mu},\, y^*_{\lambda,\mu}) \tag{5.6}$$

is strongly continuous and bounded on bounded sets. Indeed, let us consider $x_0 \in X$ and $y^*_0 \in Y^*$ such that

$$\overline{L}(x_0, \cdot) \neq +\infty, \quad \underline{L}(\cdot, y^*_0) \neq -\infty. \tag{5.7}$$

We deduce the existence of a positive constant c such that

$$\begin{aligned} &\underline{L}(x, y^*_0) \geq -c(\|x\| + 1) \text{ for every } x \in X \\ &\overline{L}(x_0, y^*) \leq c(\|x\| + 1) \text{ for every } y^* \in Y^*. \end{aligned} \tag{5.8}$$

Using the fact that

$$\left(H\left(\frac{x - x_{\lambda,\mu}}{\lambda}\right), -H^*\left(\frac{y^* - y^*_{\lambda,\mu}}{\mu}\right)\right) \in \partial L\left(x_{\lambda,\mu}, y^*_{\lambda,\mu}\right),$$

we derive

$$\left\langle H\left(\frac{x_{\lambda,\mu} - x}{\lambda}\right) \mid x_{\lambda,\mu} - x_0\right\rangle \leq L\left(x_0, y^*_{\lambda,\mu}\right) - L\left(x_{\lambda,\mu}, y^*_{\lambda,\mu}\right),$$

$$\left\langle H^*\left(\frac{y^*_{\lambda,\mu} - y^*}{\mu}\right) \mid y^*_{\lambda,\mu} - y^*_0\right\rangle \leq L\left(x_{\lambda,\mu}, y^*_{\lambda,\mu}\right) - L\left(x_{\lambda,\mu}, y^*_0\right).$$

Adding these two inequalities, we obtain

$$\begin{aligned} &\langle \mu H\left(x_{\lambda,\mu} - x\right) \mid \left(x_{\lambda,\mu} - x\right) + \left(x - x_0\right)\rangle + \\ &\qquad \langle \mu H^*\left(y^*_{\lambda,\mu} - y^*\right) \mid \left(y^*_{\lambda,\mu} - y^*\right) + \left(y^* - y^*_0\right)\rangle \\ &\qquad\qquad \leq \lambda\mu\left[L\left(x_0, y^*_{\lambda,\mu}\right) - L\left(x_{\lambda,\mu}, y^*_0\right)\right]. \end{aligned}$$

Using (5.8) and properties of H and H^*, it follows

$$\begin{aligned} \mu\left\|x_{\lambda,\mu} - x\right\|^2 + \lambda\left\|y^*_{\lambda,\mu} - y^*\right\|^2 \leq{} & \mu\left\|x_{\lambda,\mu} - x\right\|\left\|x - x_0\right\| + \\ & \lambda\left\|y^*_{\lambda,\mu} - y^*\right\|\left\|y^* - y^*_0\right\| + \\ & \lambda\mu c\left(\left\|x_{\lambda,\mu}\right\| + \left\|y^*_{\lambda,\mu}\right\| + 2\right). \end{aligned}$$

From this it follows that

$$\mu \|x_{\lambda,\mu}\| + \lambda \|y^*_{\lambda,\mu}\| \le M(1 + \|x\| + \|y^*\|), \tag{5.9}$$

with M depending only on $(\|x_0\|, \|y_0^*\|, c, \lambda, \mu)$, and that the operator $J_{\lambda,\mu}$ is bounded on bounded sets.

Let us now prove the continuity of $J_{\lambda,\mu}$. Consider $x^n \xrightarrow{s} x$ and $y^{*n} \xrightarrow{s} y^*$; we claim that $x^n_{\lambda,\mu} \xrightarrow{s} x_{\lambda,\mu}$ and $y^{*n}_{\lambda,\mu} \xrightarrow{s} y^*_{\lambda,\mu}$. Indeed, define

$$K^n(\xi, \eta^*) = L(\xi, \eta^*) + \frac{1}{2\lambda} \|x^n - \xi\|^2 - \frac{1}{2\mu} \|y^{*n} - \eta^*\|^2 .$$

It is clear that

$$K^n \overset{M-e/h}{\longrightarrow} K,$$

where

$$K(\xi, \eta^*) = L(\xi, \eta^*) + \frac{1}{2\lambda} \|x - \xi\|^2 - \frac{1}{2\mu} \|y^* - \eta^*\|^2 .$$

Indeed

$$\forall \xi_n \xrightarrow{w} \xi,\ \liminf_n \underline{K}^n(\xi_n, \eta^*) \ge \underline{K}(\xi, \eta^*)$$

and

$$\forall \eta^*_n \xrightarrow{w} \eta^*,\ \overline{K}(\xi, \eta^*) \ge \limsup_n \overline{K}^n(\xi, \eta^*_n).$$

From Theorem 2.5, it follows that the sequence $\left(x^n_{\lambda,\mu}, y^{*n}_{\lambda,\mu}\right)$, which is bounded, converges weakly to $\left(x_{\lambda,\mu}, y^*_{\lambda,\mu}\right)$ since the saddle point of K is unique. Moreover we obtain, for every $K^n \in [\underline{K}^n, \overline{K}^n]$ and $K \in [\underline{K}, \overline{K}]$,

$$\lim_{n \to +\infty} K^n\left(x^n_{\lambda,\mu}, y^{*n}_{\lambda,\mu}\right) = K\left(x_{\lambda,\mu}, y^*_{\lambda,\mu}\right), \tag{5.10}$$

that is

$$\begin{aligned} &\overline{L}\left(x^n_{\lambda,\mu}, y^{*n}_{\lambda,\mu}\right) + \frac{1}{2\lambda} \left\|x^n_{\lambda,\mu} - x^n\right\|^2 - \frac{1}{2\mu} \left\|y^{*n}_{\lambda,\mu} - y^*_n\right\|^2 \\ &\quad \overset{n \to \infty}{\longrightarrow} \overline{L}\left(x_{\lambda,\mu}, y^*_{\lambda,\mu}\right) + \frac{1}{2\lambda} \|x_{\lambda,\mu} - x\|^2 - \frac{1}{2\mu} \left\|y^*_{\lambda,\mu} - y^*\right\|^2 . \end{aligned} \tag{5.11}$$

From the saddle point property of $\left(x^n_{\lambda,\mu}, y^{*n}_{\lambda,\mu}\right)$, we derive

$$\overline{L}\left(x^n_{\lambda,\mu}, y^{*n}_{\lambda,\mu}\right) + \frac{1}{2\lambda}\left\|x^n_{\lambda,\mu} - x^n\right\|^2 \leq \overline{L}\left(x_{\lambda,\mu}, y^{*n}_{\lambda,\mu}\right) + \frac{1}{2\lambda}\left\|x_{\lambda,\mu} - x\right\|^2 .$$

Passing to the limit superior in the above inequality and using (5.11) we derive

$$\limsup_n \left\|y^{*n}_{\lambda,\mu} - y^*_n\right\|^2 \leq \left\|y^*_{\lambda,\mu} - y^*\right\|^2 .$$

So, $y^{*n}_{\lambda,\mu} \xrightarrow{s} y^*_n$ since weak-convergence and convergence of the norms imply strong convergence thanks to assmption (4.1).

The strong convergence $x^n_{\lambda,\mu} \xrightarrow{s} x_{\lambda,\mu}$ is then obtain by using a similar method.

Let us now calculate the Frechet derivative of $L_{\lambda,\mu}$. Let L_μ be the quadratic augmented Lagrangian defined in (4.4); its convex parent F_μ is the function $F_\mu(x,y) = F(x,y) + \frac{\mu}{2}\|y\|^2$. Using formula (4.5) we derive

$$\begin{aligned}(u^*, v) \in \partial L_\mu(x, y^*) &\Leftrightarrow (u^*, y^*) \in \partial F_\mu(x, -v),\\ &\Leftrightarrow (u^*, y^*) \in \partial F(x, -v) + \left(0, -\mu H^{*\,-1}(v)\right),\\ &\Leftrightarrow (u^*, y^*) \in \partial L\left(x, y^* + \mu H^{*\,-1}(v)\right),\end{aligned}$$

An analogous calculation after a regularization of parameter λ on the first variable provides

$$(u^*, v) \in \partial L_{\lambda,\mu}(x, y^*) \Leftrightarrow (u^*, y^*) \in \partial L\left(x - \lambda H^{*\,-1}(u^*),\, y^* + \mu H^{*\,-1}(v)\right),$$

and then

$$\partial L_{\lambda,\mu}(x, y^*) = \left(H\left(\frac{x - x_{\lambda,\mu}}{\lambda}\right), -H^*\left(\frac{y^* - y^*_{\lambda,\mu}}{\mu}\right)\right)$$

thanks to the unicity of $\left(x_{\lambda,\mu}, y^*_{\lambda,\mu}\right)$.

Let us fix $y^* \in Y^*$. The function $L_{\lambda,\mu}(\cdot, y^*)$ is then convex, continuous (in fact locally Lipschitz) and its subdifferential taken in x reduces to

$$H\left(\frac{x - x_{\lambda,\mu}}{\lambda}\right).$$

It follows ([19] Chap. I, Proposition 5.3) that $L_{\lambda,\mu}(\cdot, y^*)$ is Gâteaux differentiable in x and then Frechet differentiable since its derivative is continuous as seen above. In the same way, the function $L_{\lambda,\mu}(x,\cdot)$ has a continuous Frechet derivative namely

$$-H^*\left(\frac{y^* - y^*_{\lambda,\mu}}{\mu}\right).$$

It follows that $L_{\lambda,\mu}$ is a C^1 function and is locally Lipschitz since its derivative is bounded on bounded subsets of $X \times Y^*$, which ends the proof of Theorem 5.1. ∎

We can now prove

Theorem 5.2. *There is equivalence between*

(i) $L^n \xrightarrow{M-e/h} L$,

(ii) $\left| \begin{array}{l} \forall (x,y^*) \in X \times Y^*,\ \forall \lambda > 0,\ \forall \mu > 0 \\ \lim_{n\to+\infty} L^n_{\lambda,\mu}(x,y^*) = L_{\lambda,\mu}(x,y^*). \end{array} \right.$

Proof. Let us consider the augmented Lagrangians L^n_μ and L_μ; we define

$$\psi^n(x) = L^n_\mu(x,y^*), \quad \psi(x) = L_\mu(x,y^*).$$

From Theorem 3.2 we know that

$$F^n \xrightarrow{M} F \Leftrightarrow \psi^n \xrightarrow{M} \psi \text{ for all } y^* \in Y^*.$$

Let us compute, for $\lambda > 0$, the Moreau-Yosida approximation ψ^n_λ:

$$\begin{aligned} \psi^n_\lambda(x) &= \inf_{\xi\in X}\left(L^n_\mu(\xi,y^*) + \frac{1}{2\lambda}\|\xi - x\|^2\right) \\ \psi^n_\lambda(x) &= \inf_{\xi\in X}\sup_{\eta\in Y^*}\left(L^n(\xi,\eta) - \frac{1}{2\mu}\|\eta - y^*\|^2 + \frac{1}{2\lambda}\|\xi - x\|^2\right) \\ &= L^n_{\lambda,\mu}(x,y^*). \end{aligned}$$

Using the characterization of Mosco convergence in terms of pointwise convergence of the Moreau-Yosida approximates (4.8), we derive that (ii) is equivalent to (i). ∎

Comments

1) In the Hilbert case, an easy computation based on the formula

$$(\varphi^*)_\lambda(\cdot) = \frac{|\cdot|^2}{2\lambda} - \varphi_{1/\lambda}\left(\frac{\cdot}{\lambda}\right) \text{ for } \varphi \in \Gamma_0(X)$$

(see [6] Theorem 2.9) shows that

$$L_{\lambda,1/\lambda}(x,y) = F_\lambda(x,\lambda y) - \frac{\lambda}{2}|y|^2.$$

It follows in an evident way, that

$$L^n \overset{M-e/h}{\longrightarrow} L \Leftrightarrow \forall\lambda > 0,\ \forall(x,y) \in X \times Y, \quad \lim_{n\to+\infty} L^n_{\lambda,1/\lambda}(x,y) = L_{\lambda,1/\lambda}(x,y),$$

which is stronger than the equivalence (i) $\Leftrightarrow$ (ii) in Theorem (5.2).

2) An interesting open question is to know whether the equivalence

$$L^n \overset{M-e/h}{\longrightarrow} L \Leftrightarrow \forall\lambda > 0,\ \forall y^* \in Y^*, \quad \lim_{n\to+\infty} L^n_{\lambda,\lambda}(x,y^*) = L_{\lambda,\lambda}(x,y^*) \tag{5.12}$$

is true or not. If this were the case, the class of maximal monotone operators (see [38] or [22])

$$A(x,y^*) := \partial_1 L(x,y^*) \times \partial_2(-L)(x,y^*)$$

associated with closed proper convex-concave functions L would verify

$$\left|\begin{array}{l} \text{for every sequence } A^n, \text{ for every } (x,y^*) \in X \times Y^* \\ A^n_\lambda(x,y^*) \overset{w}{\longrightarrow} A_\lambda(x,y^*) \text{ implies } A^n_\lambda(x,y^*) \overset{s}{\longrightarrow} A_\lambda(x,y^*), \end{array}\right. \tag{5.13}$$

where A^n_λ and A_λ are the Yosida approximates of the operators A^n and A. In [1], remark 3.30, H. Attouch has proved that (5.13) is true for the subdifferentials of convex functions.

6. Equivalence between Mosco-epi/hypo convergence of closed convex-concave saddle functions and graph convergence of their subdifferentials

In [1] H. Attouch has established the following equivalence for sequences of closed convex proper functions defined on a reflexive Banach space with value in $\mathbf{R} \cup +\infty$, (see also [29]):

$$F^n \xrightarrow{M} F \tag{6.1}$$

is equivalent to

$$\partial F^n \xrightarrow{G} \partial F + \text{ some normalization condition,} \tag{6.2}$$

where ∂F is the subdifferential of the closed convex proper function F.

The normalisation condition comes from the fact that F is determined by ∂F up to an additive constant and is described below

$$\text{(N.C)} \quad \left| \begin{array}{l} \exists (x, x^*) \in \partial F,\ \exists (x_n, x_n^*) \in \partial F^n \text{ for every } n \in \mathbf{N} \\ \text{such that } x_n \xrightarrow{s} x,\ x_n^* \xrightarrow{s} x^* \text{ and } F^n(x_n) \to F(x). \end{array} \right.$$

The code letter G means graph convergence that is:

(i) $\forall (x, x^*) \in \partial F\ \exists x_n \xrightarrow{s} x,\ \ x_n^* \xrightarrow{s} x^*$ with (x_n, x_n^*) belonging to ∂F^n for every $n \in \mathbf{N}$;

(ii) for every sequence $(x_k, x_k^*) \in \partial F^{n_k}$ such that $x_k \xrightarrow{s} x$, $x_k^* \xrightarrow{s} x^*$, we have $(x, x^*) \in \partial F$.

In fact (ii) is implied by (i) thanks to the maximal monotonicity of the subdifferential operator. Moreover (ii) can be replaced by a weaker assumption in which one of the two strong limits is in fact a weak limit (see [1], 3.7).

Let us return to convex-concave functions. In [38], R.T. Rockafellar has introduced the notion of subdifferential of closed convex-concave function L by the formula

$$\partial L(x, y^*) = \partial_1 L(x, y^*) \times (-\partial_2(-L)(x, y^*)) \tag{6.3}$$

where $\partial_1 L$ and $\partial_2(-L)$ denote the partial convex subdifferentials with respect to the first and second variable. He proved that dom (∂L) and ∂L itself is independent of $L \in [\underline{L}, \overline{L}]$ and that the graph of ∂L is related to the graph of the subdifferential ∂F of the convex parent F via the relation

$$(u^*, v) \in \partial L(x, y^*) \Leftrightarrow (u^*, y^*) \in \partial F(x, -v). \tag{6.4}$$

It is clear that

$$(x, y^*) \text{ is a saddle point of } L \Leftrightarrow (0,0) \in \partial L(x, y^*). \tag{6.5}$$

From (6.4), and the definition of graph convergence, it follows

$$\partial F^n \xrightarrow{G} \partial F \Leftrightarrow \partial L^n \xrightarrow{G} \partial L. \tag{6.6}$$

Putting together (6.6), the equivalence between (6.1) and (6.2) and Theorem 2.4 provides

Theorem 6.1. *Let $\{L^n : X \times Y^* \to \overline{\mathbf{R}}\}$ be a sequence of closed convex-concave functions (X and Y are reflexive Banach spaces) whose convex parents (F^n) verify (N.C). Then the following are equivalent:*

(i) $L^n \xrightarrow{M-e/h} L$,

(ii) $\partial L^n \xrightarrow{G} \partial L$.

Theorem 6.1 points out the fact that extended Mosco epi/hypo-convergence is the notion of convergence for classes of closed convex-concave functions associated with graph convergence of subdifferentials. This graph convergence makes precise the variational properties of extended Mosco epi/hypo-convergence in order to obtain strong stability of saddle points (see [42] and [8] for applications in Convex Programming).

Theorem 6.2. *Let $\{L^n,\ L : X \times Y^* \to \overline{\mathbf{R}}\}$ be a collection of closed convex-concave functions (X and Y are reflexive Banach spaces) such that*

$$L^n \xrightarrow{M-e/h} L.$$

Then

$$\begin{aligned}&\textit{for every sequence } (x_n, y_n^*)\\ &\textit{of saddle points of } L^n,\\ &\textit{which is } (w \times w) \textit{ relatively compact,}\\ &\textit{each } (w \times w) \textit{ limit } (x, y^*) \textit{ of a subsequence}\\ &\textit{is a saddle point of } L;\end{aligned} \tag{6.7}$$

$$\begin{aligned}&\textit{for every saddle point } (x, y^*) \textit{ of } L, \textit{ there exist sequences}\\ &x_n \xrightarrow{s} x,\ y_n^* \xrightarrow{s} y^*,\ u_n^* \xrightarrow{s} 0,\ v_n \xrightarrow{s} 0,\\ &\textit{such that } (u_n^*, v_n) \in \partial L^n(x_n, y_n^*).\end{aligned} \tag{6.8}$$

The sequence (x_n, y_n^*) is then a saddle point of the convex-concave functions

$$K^n(x, y^*) = L^n(x, y) - \langle u_n^* \mid x \rangle - \langle v_n \mid y^* \rangle ,$$

whose convex parent is

$$\Phi^n(x, y) = F^n(x, y - v_n) - \langle u_n^* \mid x \rangle .$$

Proof. Since $L^n \overset{M-e/h}{\longrightarrow} L$, we derive, from Theorem 6.1 that

$$\partial L^n \xrightarrow{G} \partial L.$$

If (x, y^*) is a saddle point of L, we derive $(0, 0) \in \partial L(x, y^*)$.

Using the definition of graph convergence it follows (6.8), the calculation of K^n and Φ^n being straightforward.

In order to prove (6.7), let us consider $(\xi, \eta) \in X \times Y$ and $\xi_n \xrightarrow{s} \xi$, $\eta_n \xrightarrow{s} \eta$ such that $F(\xi, \eta) = \lim_{n \to +\infty} F^n(\xi_n, \eta_n)$; such sequences exist since $F^n \xrightarrow{M} F$, thanks to the assumption that $L^n \overset{M-e/h}{\longrightarrow} L$ (see Theorem 2.4).

From the fact that (x_n, y_n^*) is a saddle point of L^n, we derive

$$\begin{aligned}&(0, 0) \in \partial L^n(x_n, y_n^*),\\ &(0, y_n^*) \in \partial F^n(x_n, 0) \quad \text{(see (6.4) and (6.5))}.\end{aligned}$$

It follows that

$$F^n(\xi_n,\eta_n) - F^n(x_n,0) \geq \langle y_n^* \mid \eta_n\rangle .$$

Taking the lim sup on both sides and still denoting by x_n and y_n^* the sequences such that $x_n \xrightarrow{w} x$ and $y_n^* \xrightarrow{w} y^*$, we obtain

$$F(\xi,\eta) - F(x,0) \geq \langle y^* \mid \eta\rangle \text{ ,that is } (0,y^*) \in \partial F(x,0),$$

and (x,y^*) is a saddle point of L, which proves 6.7. ∎

Let us conclude this work by giving another characterization of the extended Mosco epi/hypo-convergence in terms of the resolvents and Yosida approximates of the maximal monotone operator

$$A(x,y^*) = \{(u^*,v) \in X^* \times Y \ ; \ (u^*,-v) \in \partial L(x,y^*)\}$$

associated with every closed convex-concave function L.

Let us return to (4.2) and consider $x_{\lambda,\lambda}$ and $y^*_{\lambda,\lambda}$ defined in (4.2); it follows

$$\left(H\left(\frac{x - x_{\lambda,\lambda}}{\lambda}\right), H^*\left(\frac{y^* - y^*_{\lambda,\lambda}}{\lambda}\right)\right) \in A(x,y^*),$$

which yields the following formulae:

$$(x_{\lambda,\lambda},\, y^*_{\lambda,\lambda}) = J_\lambda^A(x,y^*) \quad \text{resolvent of index } \lambda),$$

and

$$\left(H\left(\frac{x - x_{\lambda,\lambda}}{\lambda}\right), H^*\left(\frac{y^* - y^*_{\lambda,\lambda}}{\lambda}\right)\right) = A_\lambda(x,y^*)$$

(Yosida approximate of index λ).

Theorem 6.3. *Let X, Y be two reflexive Banach spaces which verify (4.1), and $\{L^n,\, L: X \times Y^* \to \overline{\mathbf{R}}\}$ be a collection of closed convex-concave functions. Then are equivalent*

(i) $L^n \xrightarrow{M-e/h} L$,

(ii) $J_\lambda^{A^n}(x, y^*) \overset{n\to+\infty}{\longrightarrow} J_\lambda^A(x, y^*)$ *strongly, for every* $\lambda > 0$,

(iii) $A_\lambda^n(x, y^*) \overset{n\to+\infty}{\longrightarrow} A_\lambda^n(x, y^*)$ *strongly, for every* $\lambda > 0$, $(x, y^*) \in X \times Y^*$.

Proof. From Theorem 6.1, we obtain

$$L^n \overset{M-e/h}{\longrightarrow} L \Leftrightarrow \partial L^n \overset{G}{\longrightarrow} \partial L,$$

and from the definitions of A^n and A, we derive

$$\partial L^n \overset{G}{\longrightarrow} \partial L \Leftrightarrow A^n \overset{G}{\longrightarrow} A.$$

Then apply Proposition 3.60 of [1], and Theorem 6.3 follows. ∎

Let us summarize the preceding results with the following diagram

Left-hand side:

$$\begin{array}{ccccccc}
F_\lambda^n & \overset{\text{pointwise}}{\longrightarrow} & F_\lambda & \Longleftrightarrow & L^n_{\lambda,\mu} & \overset{\text{pointwise}}{\longrightarrow} & L_{\lambda,\mu} \\
 & \Updownarrow & & & & \Updownarrow \text{Thm. 5.2} & \\
F^n & \overset{\text{Mosco-epi}}{\longrightarrow} & F & \overset{\text{Thm. 2.4}}{\Longleftrightarrow} & L^n & \overset{\text{Mosco-epi/hypo}}{\longrightarrow} & L \\
 & \Updownarrow & & & & \Updownarrow \text{Thm. 6.1} & \\
\partial F^n & \overset{\text{Graph conv.}}{\longrightarrow} & \partial F & \Longleftrightarrow & \partial L^n & \overset{\text{Graph conv.}}{\longrightarrow} & \partial L \\
 & & & & & \Updownarrow \text{Thm. 6.3} & \\
 & & & & J_\lambda^n & \overset{\text{pointwise}}{\longrightarrow} & J_\lambda
\end{array}$$

Right-hand side:

$$\begin{array}{ccccc}
\overset{\text{pointwise}}{\longrightarrow} & L_{\lambda,\mu} & & & \\
\Updownarrow \text{Thm. 5.2} & & & & \\
\overset{\text{Mosco-epi/hypo}}{\longrightarrow} & L & \overset{\text{Thm. 4.2}}{\Longleftrightarrow} & L_r^n(\cdot, y^*) & \overset{\text{Mosco-epi}}{\longrightarrow} \quad L_r(\cdot, y^*)_{\forall y^* \in Y^*} \\
\Updownarrow \text{Thm. 6.1} & & & & \\
\overset{\text{Graph conv.}}{\longrightarrow} & \partial L & & & \\
\Updownarrow \text{Thm. 6.3} & & & & \\
\overset{\text{pointwise}}{\longrightarrow} & J_\lambda & & &
\end{array}$$

REFERENCES

[1] H. ATTOUCH, **Variational convergence for functions and operators**, Applicable Mathematics series, Pitman (1984).

[2] A. AUSLENDER, **Optimisation – Méthodes numériques**, Masson, Paris (1976).

[3] H. ATTOUCH, D. AZÉ, R. WETS, *Convergence of convex-concave saddle functions; continuity properties of the Legendre-Fenchel transform with applications to convex programming and mechanics*, Techn. Report A.V.A.M.A.C. Perpignan n°85-08, submitted (1985).

[4] H. ATTOUCH, R. WETS, *A convergence theory for saddle functions*, Tra,s. A.M.S., 280, n°1 (1983), 1–41.

[5] H. ATTOUCH, R. WETS, *A convergence for bivariate functions aimed at the convergence of saddle values*, **Mathematical theories of optimization**, ed. by J.P. Cecconi and T. Zolezzi, Springer-Verlag Lectures Notes in Mathematics, n°979 (1981).

[6] H. ATTOUCH, R. WETS, *Isometries for the Legendre-Fenchel transform*, to appear in Trans. of the A.M.S.

[7] H. ATTOUCH, R. WETS, *Approximation and convergence in nonlinear optimization*, **Nonlinear Programming 4**, Academic Press (1981), 367–394.

[8] D. AZÉ, *Stability results in convex programming*, Techn. Report n°85-04, A.V.A.M.A.C., Perpignan (1985).

[9] D. AZÉ, *Convergence des variables duales dans des problèmes de transmission à travers des couches minces par des méthodes d'epi-convergence*, Techn. Report n°85-06, A.V.A.M.A.C., Perpignan, submitted (1985).

[10] B. BANK, J. GUDDAT, D. KLATTE, B. KUMMER, K. TAMMER, **Nonlinear parametric optimization**, Birkhaüser Verlag (1983).

[11] D.P. BERTSEKAS, **Constrained optimization and Lagrange multipliers methods**, Academic Press (1982).

[12] D.P. BERTSEKAS, *Convexification procedures and decomposition methods for nonconvex optimization problems*, J.O.T.A., 29, n°2 (1979), 169–197.

[13] M. BOUGEARD, *Contribution à la théorie de Morse*, Cahiers du CEREMADE, Univ. Paris IX Dauphine, n°7911 (1979).

[14] M. BOUGEARD, J.P. PENOT, *Approximation and decomposition properties of some classes of locally d.c. functions*, to appear.

[15] H. BREZIS, **Opérateurs maximaux monotones et semi-groupes de contraction dans les espaces de Hilbert**, North-Holland (1973).

[16] E. CAVAZUTTI, *Γ-convergenza multipla, convergenza di punti di sella e di max-min*, Boll. Un. Mat. Ital., 6, 1-B (1982), 251–274.

[17] E.H. CHABI, *Convergence des variables duales dans des problèmes d'homogénéisation et de renforcement*, Techn. report, A.V.A.M.A.C., Perpignan, to appear (1985).

[18] S. DOLECKI, *On perspective of abstract Lagrangien*, Proc. of the meeting on Generalized Lagrangians in systems and economic theory, I.I.A.S.A., Laxenburg, Austria (1979).

[19] I. EKELAND, R. TEMAM, **Analyse convexe et problèmes variationnels**, Dunod (1974).

[20] M. FORTIN, *Leçons sur l'analyse convexe et l'approximation des problèmes de point-selle*, Publications mathématiques d'Orsay, n°78.03 (1978).

[21] A. FOUGERES, A. TRUFFERT, *Régularisation s.c.i. et Γ-convergence. Approximations inf-convolutive associées à un référentiel*, to appear in Ann. di Mat. Pura. ed. Appl. (1986).

[22] J.P. GOSSEZ, *On the subdifferential of a saddle function*, Jour. of Functional Anal., 11 (1972), 220–230.

[23] J.-B. HIRIART- URRUTY, *Extension of Lipschitz functions*, J. Math. Anal. Appl., 77 (1980), 539–554.

[24] R. JANIN, **Sur la dualité et la sensibilité dans les problèmes des programmes mathématiques**, Thèse d'Etat, Paris VI (1974).

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

[26] P.J. LAURENT, **Approximation et Optimisation**, Hermann (1972).

[27] L. Mc LINDEN, *Dual operation on saddle functions*, Trans. A.M.S., 179 (1973), 363–381.

[28] L. Mc LINDEN, *An extension of Fenchel duality theorem to saddle functions and dual minimax problems*, Pacific Jour. of Maths., 50 (1974), 135–158.

[29] M. MATZEU, *Su un tipo di continuita dell'operatore subdifferenziale*, Boll. Un. Mat. Ital., 5, 14-B (1977),480–490.

[30] J.J. MOREAU, *Théorèmes "inf-sup"*, C.R.A.S., Paris, 258 (1964), 2720–2722.

[31] J.J. MOREAU, *Inf-convolution de fonctions numériques sur un espace vectoriel*, C.R.A.S., Paris, 256 (1963), 5047–5049.

[32] J.J. MOREAU, **Fonctionnelles convexes**, Séminaire du Collège de France (1966).

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

[34] U. MOSCO, *Convergence of convex sets and of solutions of variational inequalities*, Advances in Maths., 3 (1969), 510–585.

[35] R.T. ROCKAFELLAR, *Augmented Lagrangian and applications of the proximal point algorithms in convex programming*, Math. of Operations Research, 1, n°2 (1976), 97–116.

[36] R.T. ROCKAFELLAR, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, SIAM Jour. on Control, 12, n°2 (1974), 268–285.

[37] R.T. ROCKAFELLAR, *A general correspondence between dual minimax problems and convex programs*, Pacific Jour. of Maths., 25, n°3 (1968), 597–611.

[38] R.T. ROCKAFELLAR, *Monotone operators associated with saddle functions and minimax problems*, Proc. symp. pure Maths. A.M.S., 18 (1970), 241–250.

[39] R.T. ROCKAFELLAR, **Conjugate duality and optimization**, regional conference series in Applied Mathematics 16, SIAM Publications, Philadelphia (1974).

[40] Y. SONNTAG, **Convergence au sens de Mosco; théorie et applications à l'approximation des solutions d'inéquations**, Thèse d'Etat, Marseille (1982).

[41] M. VOLLE, *Conjugaison par tranches*, Ann. di. Mat. Pura ed Appl. IV, vol. CXXXIX (1985), 279–312.

[42] T. ZOLEZZI, *On stability analysis in mathematical programming*, Math. Programming study, 21 (1984), 227–242.