LIPSCHITZIAN STABILITY OF ε-APPROXIMATE SOLUTIONS IN CONVEX OPTIMIZATION

Hedy Attouch
*Roger J.-B. Wets**
*Partially supported by a grant of the National Science Foundation

March 1987
WP-87-25

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

One component in the validation of mathematical models is related to their stability properties under data perturbation. The authors obtain a very strong (Lipschitz) stability result for the nearly optimal solutions when the data perturbation is measured in terms of a new distance function introduced here.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

ABSTRACT

We prove that the ε-optimal solutions are Lipschitzian with respect to data perturbations when these are measured in terms of the epigraphical distance.

CONTENTS

LIPSCHITZIAN STABILITY OF ε-APPROXIMATE SOLUTIONS IN CONVEX OPTIMIZATION

Hedy Attouch[1] *and Roger J.-B. Wets* [2]

[1]AVAMAC, Mathématiques, Université de Perpignan, Perpignan
[2]Mathematics, University of California, Davis

1. INTRODUCTION

Our main objective is the study of the stability of the solutions of convex optimization problems from a quantitative viewpoint. More precisely, mainly for numerical reasons, we like to be able to estimate the distance between the solutions of two convex minimization problems. We aim at the study of the rate of convergence of the solutions of approximation schemes, or the quantitative study of the way errors in, or perturbations of the data affect the solution.

Our approach is based on the introduction of a distance on $\Gamma_0(X)$ the set of closed convex proper functions from the normed linear space X into $\mathbf{R} \cup \{+\infty\}$. At this point, there are (at least!) two ways of attacking the problem:

(a) Assuming the functions to be strongly convex, for example, working with $\Gamma_\alpha(X) = \Gamma_0(X) + \alpha/2\|\cdot\|^2$ when X is an Hilbert space, one can prove the following *Hölder continuity* result:

$$\|\arg\min F - \arg\min G\| \leq C \cdot d(F, G)^{1/2} \ . \tag{1.1}$$

This approach, which we follow in Attouch and Wets [4], leads to the introduction of conditioning in convex optimization. The distance d in this theory

$$d_{\lambda,\rho}(F, G) = \sup_{\|x\|\leq\rho} |F_\lambda(x) - G_\lambda(x)| \tag{1.2}$$

is obtained by using F_λ, G_λ the *Moreau-Yosida approximation* of F and G defined by

$$F_\lambda(x) = inf_{u\in X}\left[F(u) + \frac{1}{2\lambda}\|x - u\|^2\right]$$

and similarly for G_λ. Let us note that the exponent 1/2 in (1.1) is optimal (it is related to the fact we work with a Hilbert structure). The distance $d_{\lambda,\varphi}$ has ultimate connection with Mosco epi-convergence which is the good topological concept [4].

(b) In this paper we present an alternative way to study the stability question that relies on the utilization of the concept of ε-approximating solution. A major advantage is that the set

$$\varepsilon\text{-arg min } F = \{x \in X; F(x) \leq \inf F + \varepsilon\}$$

is non void. Moreover this notion of solution might be considered as more natural from a numerical point of view, refer to the systematic studies of Hiriart-Urruty [12], R.T. Rockafellar [19], for example. However, ε-arg min F is a (convex) set while with conditioning arg min F is reduced to a single element. Thanks to the flexibility provided by this notion of approximate solution we are able to prove a *Lipschitz* dependence of the ε-approximate solution on the data, i.e.

$$\text{haus}\,(\varepsilon\text{-arg min } F,\ \varepsilon\text{-arg min } G) \leq C_\varepsilon \cdot d\,(F, G)$$

where haus stands for the Hausdorff metric and d is again a distance on $\Gamma_0(X)$ that induces Mosco-epi-convergence. We give practical criteria (Kenmochi's condition [13]) allowing the computation or estimation (from above) of this distance. In the process we obtain as a byproduct of the theory the Lipschitzian behavior of the ε-subdifferential mapping, see Nurminski [17], Hiriart-Urruty [12].

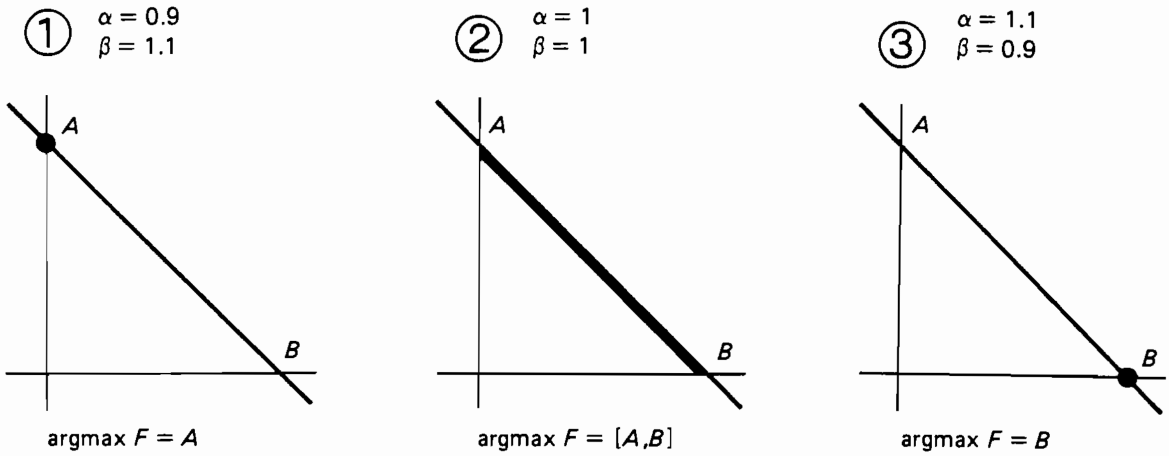
Let us describe on an elementary example coming from scenario analysis how the two above approaches permit to obtain quantitative statements about stability in convex optimization.

Take $X = \mathbf{R}^2$ and consider the following linear programming problem (we take maximization instead of minization, one can convert easily each formulation into the other)

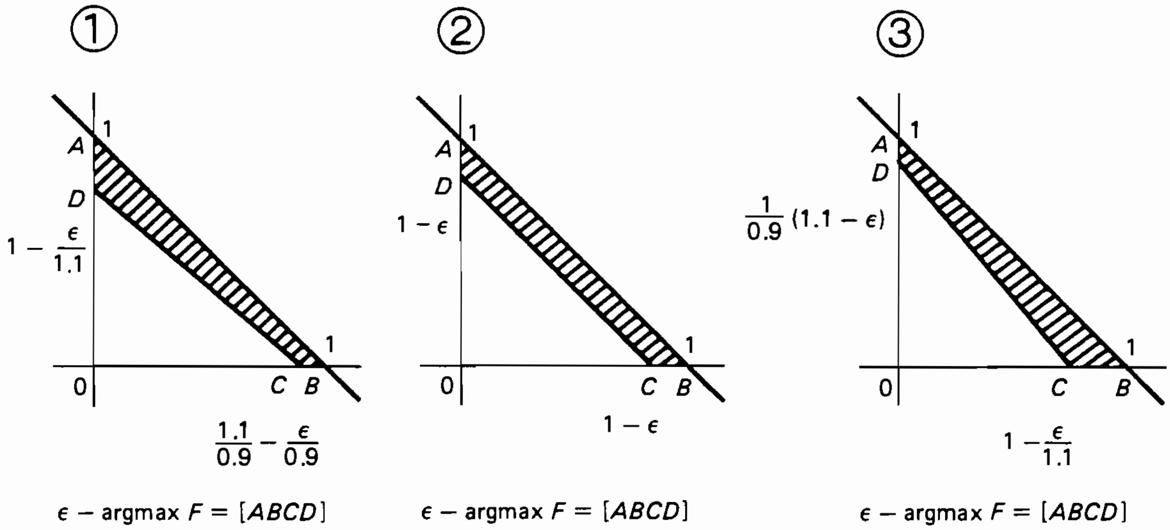
$$\begin{cases} \max\ \alpha x + \beta y \\ \text{subject to the constraint } \{x \geq 0,\ y \geq 0,\ x + y \leq 1\} \end{cases}$$

where αx is the profit which results from an investment x in a first product (oil production for example) βy is the profit which results from an investment y in a second product (nuclear energy plant for example), and the total available capital is 1. Let us consider the following possible scenarios ($\alpha = 0.9$, $\beta = 1.1$), ($\alpha = 1$, $\beta = 1$), ($\alpha = 1.1$, $\beta = 0.9$) (note that there are only slight differences between these scenarios and in all cases $\alpha + \beta = 2$) and describe in each case

- the solution set arg max F where $F(x, y) = \alpha x + \beta y - \delta_{\{\xi+\eta \leq 1\}}(x, y)$
- the ε-approximated solution set ε-arg max F
- the solution of the ε-well conditioned problem arg max $(F - \varepsilon/2|\cdot|^2)$

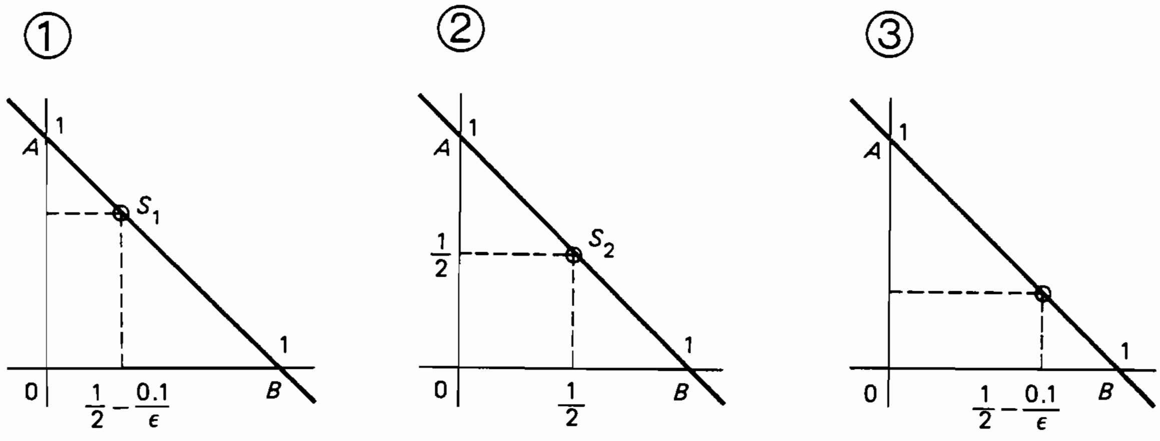


We can observe that a small change in the data (scenario 1 and 3 are "close" to each other) has, as a consequence an important modification in the solution set: one pass from solution A to solution B. This means a completely different strategy is to be followed in case of scenarios 1 and 3. Let us observe how the introduction of some flexibility via ε-solution corrects this phenomena ($\varepsilon \geq 0.2$):



Now the Hausdorff distance between the ε- arg max solution sets in scenario 1 and 3 is less or equal than 3d, where d is the perturbation on the data, here $d = 0.2$. Let us finally examine the effect of the addition of a ε-well conditioning term, which corresponds here to subtraction of $-\varepsilon/2\,(x^2+y^2)$:

$$\text{maximize } \{\alpha x + \beta y - \frac{\varepsilon}{2}(x^2+y^2) \quad \text{such that } x \geq 0,\ y \geq 0,\ x+y \leq 1\}$$



argmax $F - \frac{\epsilon}{2}|\cdot|^2 = S_1$ argmax $F - \frac{\epsilon}{2}|\cdot|^2 = S_2$ argmax $F - \frac{\epsilon}{2}|\cdot|^2 = S_3$

We observe that the variation of the solution is of order $(1/\varepsilon)d$. (It is a quite favorable situation, one might end up with a variation of order $(1/\varepsilon)\sqrt{d}$!).

The article is organized as follows. In Section 2, we begin with the study of the projection of a point on a convex set, the major advantage being, that in this idealized situation, the variation of the data, (the convex set) is naturally taken in the sense of the Hausdorff distance. In Section 3, we introduce a distance on the set of convex lower semicontinuous (lsc) proper functions which allows us to extend the preceding Lipschitz continuity property of the ε-minimizing set to this general situation. As by product, we obtain the Lipschitz property of the ε-subdifferential of a convex function $\partial_\varepsilon F: X \to 2^{X^*}$, see Hiriart-Urruty [12]. We finally study the properties of the epigraphic distance which is introduced in Section 3 and relate it to the distance $d_{\lambda,\rho}$ based on the Moreau-Yosida approximates introduced in [4], and the so called Kenmochi condition, see Kenmochi [13], Attouch and Damlamian [2].

2. PROJECTION ON A CONVEX SET-STABILITY OF ε-PROJECTIONS

Let X be a normed space. Given C a convex subset of X, $\varepsilon > 0$ and $x_0 \in X$ we denote by

$$d(x_0, C) = \inf\{\|x_0 - x\|;\ x \in C\}$$

$$\varepsilon\text{-prj}(x_0, C) = \{x \in C;\ \|x_0 - x\| \le d(x_0, C) + \varepsilon\} \tag{2.1}$$

For any pair of nonempty subsets C and D of X we write

$$\text{haus}(C, D) = \sup\{e(C, D);\ e(D, C)\}$$

where

$$e(C, D) = sup\,\{d(x, D);\ x \in C\}\ .$$

THEOREM 2.1 *Let C and D be two convex subsets of a normed linear space X. Given any $\varepsilon > 0$ and $x_0 \in X$, the following estimation holds*

$$\text{haus}(\varepsilon\text{-prj}(x_0, C);\ \varepsilon\text{-prj}(x_0, D)) \le \rho_\varepsilon(\|x_0\|)\,\text{haus}(C; D) \tag{2.2}$$

with

$$\rho_\varepsilon(\|x_0\|) = 3 + \frac{4}{\varepsilon}[d(x_0, D) + d(x_0, C)]\ . \tag{2.3}$$

PROOF Let us pick up an arbitrary point x belonging to $\varepsilon\text{-}prj\ (x_0, C)$ and prove that

$$d(x, \varepsilon\text{-}prj\,(x_0, D)) \le \rho_\varepsilon(\|x_0\|)\,\text{haus}(C, D)\ . \tag{2.4}$$

Since the sets C and D play a symmetric role, this will prove the theorem.

By definition (2.1) of ε-prj (x_0, C) we have

$$\|x_0 - x\| \le d(x_0, C) + \varepsilon\ .$$

Now, we notice that

$$d(x_0, C) \le d(x_0, D) + \text{haus}(C, D)$$

hence

$$\|x_0 - x\| \le d(x_0, D) + \text{haus}(C, D) + \varepsilon\ . \tag{2.5}$$

Moreover since x belongs to C, by definition of haus (C, D) for every $\mu > 0$ there exists some $y_\mu \in D$ satisfying

$$\|x - y_\mu\| \le \text{haus}(C, D) + \mu\ . \tag{2.6}$$

From (2.5) and (2.6) and by the triangle inequality

$$\begin{cases} \|x_0 - y_\mu\| \le d(x_0, D) + \varepsilon + 2\,\text{haus}(C, D) + \mu \\ y_\mu \in D \end{cases} \tag{2.7}$$

that is $y_\mu \in (\varepsilon + 2h + \mu)\text{-prj}(x_0, D)$, where $h = \text{haus}(C, D)$.

From the triangle inequality, it follows that

$$d(x, \varepsilon\text{-prj}(x_0, D) \le d(y_\mu, \varepsilon\text{-prj}(x_0, D)) + \|x - y_\mu\|$$

which with (2.6) yields

$$d(x, \varepsilon\text{-prj}(x_0, D)) \le (2h + \mu)\left[1 + \frac{2}{\varepsilon + 2h} d(x_0, D)\right] + h + \mu$$

$$\le (h + \mu)\left[3 + \frac{4}{\varepsilon + 2h} d(x_0, D)\right]$$

– which yields the desired inequality with ρ_ε as defined by (2.3), let $\mu \to 0$ in the above inequality, provided we have that

$$d(y_\mu, \varepsilon\text{-prj}(x_0, D) \le (2h + \mu)\left[1 + \frac{2}{\varepsilon + 2h} d(x_0, D)\right] .$$

And that actually follows from the next Lemma. □

LEMMA 2.2 *Let K be a convex set in a normed linear space X, $x_0 \in X$ and $y \in K$ which satisfies*

$$\|x_0 - y\| \le d(x_0, K) + \varepsilon + h$$

where $\varepsilon > 0$ and $h \ge 0$. Then, for each $\mu \in]0, 1]$ there exists some $z \in K$ such that

$$\begin{cases} \|x_0 - z\| \le d(x_0, K) + \varepsilon , \\ \|y - z\| \le h\left[1 + 2\dfrac{d(x_0, K) + \mu\varepsilon}{h + (1 - \mu)\varepsilon}\right] , \end{cases} \tag{2.8}$$

Letting $\mu \to 0$ in (2.8), it follows

$$d(y, \varepsilon\text{-prj}(x_0, K)) \le h\left[1 + \frac{2}{\varepsilon + h} d(x_0, K)\right] , \tag{2.9}$$

and

$$\text{haus}((\varepsilon + h)\text{-prj}(x_0, K); \varepsilon\text{-prj}(x_0, K)) \le h\left[1 + \frac{2}{\varepsilon + h} d(x_0, K)\right] . \tag{2.10}$$

PROOF For any $0 < \mu \leq 1$ let us introduce x_μ belonging to K such that

$$\|x_0 - x_\mu\| \leq d(x_0, K) + \mu\varepsilon .$$

Then for any $0 \leq \lambda \leq 1$ let us define

$$\xi_{\lambda,\mu} = \lambda y + (1 - \lambda)x_\mu$$

which still belongs to the convex set K.

Let us estimate

$$\begin{aligned}\|x_0 - \xi_{\lambda,\mu}\| &= \|\lambda(x_0 - y) + (1-\lambda)(x_0 - x_\mu)\| \\ &\leq \lambda\|x_0 - y\| + (1-\lambda)\|x_0 - x_\mu\| \\ &\leq \lambda(d(x_0, K) + \varepsilon + h) + (1-\lambda)(d(x_0, K) + \mu\varepsilon) \\ &\leq d(x_0, K) + \mu\varepsilon + \lambda[(1-\mu)\varepsilon + h] . \qquad (2.11)\end{aligned}$$

In order to have the second member of (2.11) less or equal than $d(x_0, K) + \varepsilon$, take

$$\lambda = \frac{(1-\mu)\varepsilon}{(1-\mu)\varepsilon + h} . \qquad (2.12)$$

Then

$$\|y - \xi_{\lambda,\mu}\| = (1-\lambda)\|y - x_\mu\|$$

$$\|y - \xi_{\lambda,\mu}\| \leq \frac{h}{h + (1-\mu)\varepsilon}[2d(x_0, K) + (1+\mu)\varepsilon + h] . \qquad (2.13)$$

Since $\xi_{\lambda,\mu}$ belongs to ε-prj(x_0, K)

$$d(y, \varepsilon\text{-prj}(x_0, K)) \leq \frac{h}{h + (1-\mu)\varepsilon}[2d(x_0, K) + (1+\mu)\varepsilon + h] .$$

which yields (2.8). This inequality being true for any μ belonging to the interval $]0, 1]$, by letting μ tend to zero it follows

$$\begin{aligned}d(y, \varepsilon\text{-prj}(x_0, K)) &\leq \frac{h}{h+\varepsilon}[2d(x_0, K) + \varepsilon + h] \\ &\leq h\left[1 + \frac{2}{\varepsilon + h} d(x_0, K)\right] .\end{aligned}$$

This being true for any y belonging to $(\varepsilon + h)$-prj(x_0, K) and noticing that ε-prj(x_0, K) is contained in $(\varepsilon + h)$-prj(x_0, K), formula (2.10) follows. □

3. STABILITY OF ε-APPROXIMATE SOLUTIONS OF CONVEX MINIMIZATION PROBLEMS

Let us now examine the general situation. We would like, as suggested by the mathematical model involving just projections studied in the preceding section, to prove a Lipschitzian dependence of the ε-approximate solution set on the data. In order to prove a Lipschitz property for the map

$$F \mapsto \varepsilon\text{-argmin}\, F$$

we need to introduce a distance on the set of convex functions. At this stage, we may decide to work with the distances $d_{\lambda,\rho}$, $d_{[\lambda],\rho}$, or $d_{\lambda,\rho}^{\xi,p}$ see (1.2), which have been introduced in [4], [5]. Indeed, these distances are well fitted to the study of strictly convex minimization problems in reflexive Banach spaces. But, when considering ε-solutions in general normed spaces they don't play such a natural role.

Relying on the following elementary geometric considerations,

a) epiconvergence is equivalent to set-convergence of the epigraphs

b) ε-argmin F is obtained by taking the projection on the space X of the intersection of the epigraph of F with the ball of radius $(\inf F + \varepsilon)$, we are naturally led to introduce the following distances on the set of convex functions (cf. the r-distance for convex sets introduced by Salinetti and Wets [21], Mosco [24]).

DEFINITION 3.1 *Let X be a normed space.*

a) *Given any $\rho > 0$, and any pair C and D of convex subsets of X*

$$h_\rho(C, D) = \sup\{e(C_\rho, D);\ e(D_\rho, C)\}$$

where

$$C_\rho = C \cap B_X(\rho)$$

and

$$e(C_\rho, D) = \sup\{d(x, D);\ x \in C_\rho\}$$

b) *Given any $\rho > 0$, and any pair F and G of convex functions from X into $\mathbf{R} \cup \{+\infty\}$,*

$$d_\rho(F, G) = h_\rho(epi\, F,\ epi\, G) \tag{3.1}$$

where epi F and epi G are two convex subsets of $X \times \mathbf{R}$ and

$$B_{X\times\mathbf{R}}(\rho) = \{(x, \lambda) \in X \times \mathbf{R};\ \|x\| \leq \rho \text{ and } |\lambda| < \rho\} \ .$$

$$B_{X\times\mathbf{R}}(\rho) = \{(x, \lambda) \in X \times \mathbf{R};\ \|x\| \leq \rho \quad \text{and } |\lambda| \leq \rho\} \ .$$

THEOREM 3.2 *Let $F, G : X \rightarrow \mathbf{R} \cup \{+\infty\}$ be two convex functions on a normed linear space such that*

$$\inf F \quad \text{and } \inf G \textit{ are finite} \ . \tag{3.2}$$

(i) *If there exists some $\rho_0 > 0$ such that for every $\varepsilon > 0$*

$$\varepsilon\text{-argmin} F \cap B_X(\rho_0) \neq \phi,\ \varepsilon\text{-argmin}\, G \cap B_X(\rho_0) \neq \phi \ , \tag{3.3}$$

then

$$|\inf F - \inf G| \leq d_{\rho_1}(F, G) \tag{3.4}$$

with

$$\rho_1 = \sup\{\rho_0; |\inf F| + 1; |\inf G| + 1\} \ . \tag{3.5}$$

(ii) *Under the stronger assumption*

$$\varepsilon\text{-argmin} F \cup \varepsilon\text{-argmin}\, G \subset B_X(\rho_0) \tag{3.6}$$

the following Lipschitz continuity property of the ε-solutions holds

$$\text{haus}\,(\varepsilon\text{-argmin} F,\ \varepsilon\text{-argmin}\, G) \leq d_{\rho_1}(F, G)\left[1 + \frac{4\rho_0 + 2d_{\rho_1}(F, G)}{\varepsilon + 2d_{\rho_1}(F, G)}\right] \tag{3.7}$$

PROOF i) Take for each $1 > \varepsilon > 0$ some $x_\varepsilon \in B_X(\rho_0)$ such that

$$F(x_\varepsilon) \leq \inf F + \varepsilon \ .$$

$$(x_\varepsilon, \inf F + \varepsilon) \in \text{epi} F \cap B_X(\rho_1)$$

where

$$\rho_1 = \sup\{\rho_0; \sup\{|\inf F|; |\inf G|\} + 1\} \ .$$

By definition of d_ρ, see (3.1), for every $\mu > 0$ there exists $(\xi_\mu, \lambda_\mu) \in \text{epi}\, G$ such that

$$\|x_\varepsilon - \xi_\mu\| \le d_{\rho_1}(F, G) + \mu$$

$$|\inf F + \varepsilon - \lambda_\mu| \le d_{\rho_1}(F, G) + \mu \ .$$

Thus

$$\lambda_\mu \le \inf F + \varepsilon + d_{\rho_1}(F, G) + \mu \ .$$

Since $\lambda_\mu \ge G(\xi_\mu) \ge \inf G$, it follows

$$\inf G - \inf F \le d_{\rho_1}(F, G) + \varepsilon + \mu \ .$$

This inequality being true for any $\varepsilon > 0$ and $\mu > 0$

$$\inf G - \inf F \le d_{\rho_1}(F, G)$$

and by exchanging the role of F and G we finally obtain (3.4).

ii) Let us take some x belonging to ε-argmin$F$ i.e. $F(x) \le \inf F + \varepsilon$. From inequality (3.4) it follows that

$$F(x) \le \inf G + d_{\rho_1}(F, G) + \varepsilon \ . \tag{3.8}$$

Moreover since $(x, F(x)) \in (\text{epi}F)_{\rho_1}$, for every $\mu > 0$ there exists some $(y_\mu, \lambda_\mu) \in \text{epi}\, G$ such that

$$\|x - y_\mu\| \le d_{\rho_1}(F, G) + \mu \tag{3.9}$$

$$|F(x) - \lambda_\mu| \le d_{\rho_1}(F, G) + \mu \ .$$

Hence

$$\lambda_\mu \le F(x) + d_{\rho_1}(F, G) + \mu$$

and since $\lambda_\mu \ge G(y_\mu)$

$$G(y_\mu) \le F(x) + d_{\rho_1}(F, G) + \mu \ . \tag{3.10}$$

Adding (3.8) and (3.10) we obtain

$$G(y_\mu) \le \inf G + \varepsilon + 2d_{\rho_1}(F, G) + \mu \ . \tag{3.11}$$

We complete the proof by relying on Lemma 3.3 below (similar to Lemma 2.2). We apply the Lemma to the convex function G with $\theta = \inf G + \varepsilon$ and $h = 2d_{\rho_1}(F, G) + \mu$. So, inequality (3.11) can be reinterpreted as (see notations in Lemma 3.3 below)

$$y_\mu \in S^G_{\theta + h}$$

and by construction, y_μ satisfies

$$\|y_\mu\| \le \rho_0 + d_{\rho_1}(F, G) + \mu \ .$$

From (3.14), Lemma 3.3, we have

$$d(y_\mu, S^G_\theta) \le (2d_{\rho_1}(F, G) + \mu)\frac{\|y_\mu\| + \rho_0}{2d_{\rho_1}(F, G) + \varepsilon}$$

$$\le (d_{\rho_1}(F, G) + \mu)\frac{4\rho_0 + 2d_{\rho_1}(F, G) + 2\mu}{\varepsilon + 2d_{\rho_1}(F, G)} \ . \qquad (3.12)$$

Noticing that $S^G_\theta = \varepsilon\text{-argmin}\, G$, from (3.9), (3.12) by using the Lipschitz contraction property of the distance function $d(\cdot\,, \varepsilon\text{-argmin}\, G)$ we finally derive (let $\mu \to 0$)

$$d(x, \varepsilon\text{-argmin}\, G) \le d_{\rho_1}(F, G)\left[1 + \frac{4\rho_0 + 2d_{\rho_1}(F, G)}{\varepsilon + 2d_{\rho_1}(F, G)}\right] .$$

This being true for any $x \in \varepsilon$-argmin$F$, inequality (3.7) follows. □

LEMMA 3.3 *Let F be a convex function from a normed linear space X into $\mathbf{R} \cup \{+\infty\}$. Let us assume there exists some $\rho_0 > 0$ such that*

$$\forall \varepsilon > 0 \quad \varepsilon\text{-argmin} F \cap B_X(\rho_0) \neq \phi \ . \qquad (3.13)$$

For any $\theta \in \mathbf{R}$ let us denote by $S^F_\theta := \{x \in X;\ F(x) \le \theta\}$. Then, for any θ, $h \in \mathbf{R}$ satisfying

$$\theta + h \ge \theta > \inf F$$

the following inequality holds:

$$\forall y \in S^F_{\theta + h} \qquad d(y, S^F_\theta) \le h \ . \ \frac{\|y\| + \rho_0}{h + \theta - \inf F} \ . \qquad (3.14)$$

PROOF Let $\Theta + h \geq \Theta > \inf F$. Given $y \in S^F_{\Theta+h}$ let us construct some $\xi \in S^F_\Theta$ such that (3.14) is satisfied. Take for any $\mu \in]0, 1[$ some point y_μ such that

$$\begin{cases} F(y_\mu) \leq \inf F + \mu(\Theta - \inf F) = \mu\Theta + (1-\mu)\inf F \\ \|y_\mu\| \leq \rho_0 \end{cases}$$

and introduce for any $\lambda \in [0, 1]$

$$\xi_{\lambda,\mu} = \lambda y + (1-\lambda) y_\mu \ .$$

Let us choose λ and μ in order to have $\xi_{\lambda,\mu} \in S^F_\Theta$, that means $F(\xi_{\lambda,\mu}) \leq \Theta$. By convexity

$$\begin{aligned} F(\xi_{\lambda,\mu}) &\leq \lambda F(y) + (1-\lambda) F(y_\mu) \\ &\leq \lambda(\Theta + h) + (1-\lambda)[\mu\Theta + (1-\mu)\inf F] \\ &\leq [\lambda + (1-\lambda)\mu]\Theta + \lambda h + (1-\lambda)(1-\mu)\inf F \ . \end{aligned} \tag{3.15}$$

In order to have the second member of (3.15) less or equal than Θ take

$$\lambda = \frac{(1-\mu)(\Theta - \inf F)}{h + (1-\mu)(\Theta - \inf F)} = 1 - h \cdot \frac{1}{h + (1-\mu)(\Theta - \inf F)} \ .$$

So $\xi_{\lambda,\mu} \in S^F_\Theta$ and

$$\begin{aligned} \|y - \xi_{\lambda,\mu}\| &= (1-\lambda)\|y - y_\mu\| \\ &= h \frac{1}{h + (1-\mu)(\Theta - \inf F)} \|y - y_\mu\| \\ &\leq h \frac{1}{h + (1-\mu)(\Theta - \inf F)} [\|y\| + \rho_0] \ . \end{aligned}$$

This inequality being true for any $0 < \mu < 1$

$$d(y, S^F_\Theta) \leq h \cdot \frac{\|y\| + \rho_0}{h + (\Theta - \inf F)} \ . \quad \square$$

REMARK 3.4 a) Let us examine what conclusion can be derived when assumption (3.3) is dropped. Clearly

$$\begin{aligned} |\inf F - \inf G| &= \mathrm{haus}(\mathrm{prj}_{\mathbf{R}} \mathrm{epi} F, \mathrm{prj}_{\mathbf{R}} \mathrm{epi} G) \\ &\leq \mathrm{haus}(\mathrm{epi} F, \mathrm{epi} G) \end{aligned}$$

But taking $d(F, G) = \mathrm{haus}(\mathrm{epi} F, \mathrm{epi} G)$ is not convenient since this quantity is $+\infty$

as soon as F and G have distinct recession cones.

b) When assumption (3.6) is dropped then conclusion (3.7) fails to be true as shown by the following example:

Take $X = \mathbf{R}$, $F(x) = x^+$

and for every $n \in \mathbf{N}$

$$G_n(x) = \begin{cases} x & \text{if } x \geq 0 \\ -\frac{x}{n} & \text{if } x \leq 0 \end{cases} .$$

Clearly assumption (3.3) is satisfied and indeed $\inf F = \inf G_n$ is achieved in both cases at $x = 0$. Moreover for every $\rho > 0$, $d_\rho(F, G) = \frac{\rho}{n}$ while $\varepsilon\text{-argmin} F =]-\infty, \varepsilon]$, $\varepsilon\text{-argmin} G_n = [-\varepsilon n, \varepsilon]$ and

$$\text{haus}\,(\varepsilon\text{-argmin} F, \varepsilon\text{-argmin} G_n) \equiv +\infty \;! \square$$

With a similar type of argument as in Theorem 3.2 one can obtain the Lipschitz continuity property of the map

$$F \longrightarrow S_\lambda^F$$

where $S_\lambda^F = \{x \in X; F(x) \leq \lambda\}$:

PROPOSITION 3.5 *Let F, G be two convex functions from X, a normed linear space, into $\mathbf{R} \cup \{+\infty\}$ and let λ be some real number such that* $\lambda > \inf F$, $\lambda > \inf G$. *Let us assume that* $S_\lambda^F \subset B_X(\rho_0)$, $S_\lambda^G \subset B_X(\rho_0)$ *for some* $\rho_0 > 0$. *Then*

$$\text{haus}\,(S_\lambda^F, S_\lambda^G) \leq d_{\rho_1}(F, G)\left[1 + \frac{2\rho_0 + d_{\rho_1}(F, G)}{d_{\rho_1}(F, G) + \lambda - \sup\{\inf F, \inf G\}}\right],$$

where $\rho_1 = \sup\{\rho_0; |\lambda|; |\inf F|; |\inf G|\}$.

PROOF Take $x \in S_\lambda^F$, i.e., $F(x) \leq \lambda$, then $\|x\| \leq \rho_0$ and $(x, \lambda) \in (\text{epi} F)_{\rho_1}$ with ρ_1 given as above. By definition of d_{ρ_1} for every $\varepsilon > 0$ there exists some $(\xi_\varepsilon, \mu_\varepsilon) \in \text{epi}\, G$ such that

$$\|x - \xi_\varepsilon\| \leq d_{\rho_1}(F, G) + \varepsilon \tag{3.16}$$

$$|\lambda - \mu_\varepsilon| \leq d_{\rho_1}(F, G) + \varepsilon$$

Hence,

$$G(\xi_\varepsilon) \le \mu_\varepsilon \le \lambda + d_{\rho_1}(F, G) + \varepsilon \ .$$

Applying Lemma 3.3, we derive

$$d(\xi_\varepsilon, S_\lambda^G) \le (d_{\rho_1}(F, G) + \varepsilon)\frac{\|\xi_\varepsilon\| + \rho_0}{d_{\rho_1}(F, G) + \lambda - \inf G} \ . \tag{3.17}$$

From (3.16) and (3.17) if follows

$$d(x, S_\lambda^G) \le d_{\rho_1}(F, G)\left[1 + \frac{2\rho_0 + d_{\rho_1}(F, G)}{d_{\rho_1}(F, G) + \lambda - \inf G}\right] .$$

This being true for any $x \in S_\lambda^F$ and by exchanging the role of F and G we finally obtain

$$\text{haus}\,(S_\lambda^F, S_\lambda^G) \le d_{\rho_1}(F, G)\left[1 + \frac{2\rho_0 + d_{\rho_1}(F, G)}{d_{\rho_1}(F, G) + \lambda - \sup\{\inf F, \inf G\}}\right] . \quad \square$$

Let us now show how the Lipschitz continuity of the ε-subdifferential of a convex function, proved by Nurminski [17] and Hiriart-Urruty [12], can easily be derived from Theorem 3.2. We recall that given $F: X \longrightarrow \mathbf{R} \cup \{+\infty\}$

$$\partial_\varepsilon F(x) = \{x^* \in X^*;\, F(y) \ge F(x) + \langle x^*, y - x\rangle - \varepsilon \quad \forall y \in X\}$$

Equivalently

$$\partial_\varepsilon F(x) = \{x^* \in X^*;\, F(x) + F^*(x^*) - \langle x, x^*\rangle \le \varepsilon\}$$

where

$$F^*(x^*) = \sup\{\langle x^*, x\rangle;\, x \in X\}$$

is the conjugate (Legendre-Fenchel transform) of F.

COROLLARY 3.6 *Let F be a closed convex function from a Banach space X into $\mathbf{R} \cup \{+\infty\}$. Then for every $\varepsilon > 0$, $\partial_\varepsilon F: X \longrightarrow 2^{X^*}$ has the following locally Lipschitz property:*

For any $x_1, x_2 \in X$ satisfying $\partial_\varepsilon F(x_1) \cup \partial_\varepsilon F(x_2) \subset B_{X^*}(\rho_0)$

$$\text{haus}\,(\partial_\varepsilon F(x_1), \partial_\varepsilon F(x_2)) \le C\left(\rho_0, \frac{1}{\varepsilon}\right) \cdot \|x_1 - x_2\| \tag{3.18}$$

where C depends continuously on ρ_0 *and* $1/\varepsilon$.

PROOF We characterize the ε-subdifferential at x as some ε-approximate solution. Move precisely

$$\partial_\varepsilon F(x) = \{x^* \in X^*;\ F^*(x^*) + (F^*)^*(x) - <x, x^*> \leq \varepsilon\} .$$

Here we used the fact $F = (F^*)^*$. Hence

$$\partial_\varepsilon F(x) = \{x^* \in X^*;\ F^*(x^*) + <x, \xi^*> - F^*(\xi^*) - <x, x^*> \leq \varepsilon \ \forall \xi^* \in X^*\}$$

$$= \{x^* \in X^*;\ F^*(x^*) - <x, x^*> \leq F^*(\xi^*) - <x, \xi^*> + \varepsilon \ \forall \xi^* \in X^*\}$$

$$= \varepsilon\text{-argmin}\,\{F^*(\cdot) - <x, \cdot>\} .$$

Given two points x_1 and x_2 such that $\partial_\varepsilon F(x_1) \cup \partial_\varepsilon F(x_2) \subset B_{X^*}(\rho_0)$ applying Theorem 3.2, formula (3.7), we derive

$$\text{haus}\,(\partial_\varepsilon F(x_1), \partial_\varepsilon F(x_2)) \leq d_{\rho_1}(\Phi_1, \Phi_2) \left[1 + \frac{4\rho_0 + 2d_{\rho_1}(\Phi_1, \Phi_2)}{\varepsilon + 2d_{\rho_1}(\Phi_1, \Phi_2)}\right]$$

where

$$\Phi_1(x^*) = F^*(x^*) - <x_1, x^*>$$

$$\Phi_2(x^*) = F^*(x^*) - <x_2, x^*>$$

$$\rho_1 = \sup\,\{\rho_0;\ \sup\,\{|\inf \Phi_1|;\ |\inf \Phi_2|\} + 1\} . \tag{3.19}$$

An elementary computation yields

$$d_{\rho_1}(\Phi_1, \Phi_2) \leq \rho_1 \cdot \|x_1 - x_2\| ,$$

that is

$$\text{haus}\,(\partial_\varepsilon F(x_1), \partial_\varepsilon F(x_2)) \leq \rho_1 \left[1 + \frac{4\rho_0 + 2\rho_1\|x_1 - x_2\|}{\varepsilon + 2\rho_1\|x_1 - x_2\|}\right] \cdot \|x_1 - x_2\| \tag{3.20}$$

where an upper bound for ρ_1 can be obtained from (3.19) and clearly depends in a continuous way on ρ_0 and $1/\varepsilon$. □

Since $x \mapsto \partial_\varepsilon F(x)$ is a multivalued locally (pseudo) Lipschitz mapping, from classical selection theorem there exists a Lipschitzian selection. Indeed and this will make the link with the other classical approximation of ∂F by Lipschitzian map, more precisely by the Yosida approximation $\nabla F_\lambda = (\partial F)_\lambda$. We prove that ∇F_λ provides a Lipschitz selection of $\partial_\varepsilon F$.

PROPOSITION 3.7 *Let F be a convex Lipschitz function from a Hilbert space H into $\mathbf{R}$ with Lipschitz constant k. Then for any $\lambda > 0$, $\varepsilon > 0$ satisfying*

$$\lambda \leq \frac{4\varepsilon}{k^2} , \tag{3.21}$$

we have

$$\nabla F_\lambda \in \partial_\varepsilon F \tag{3.22}$$

and the bound determined by (3.21) *is sharp.*

PROOF We use the following notation

$$A_\lambda^F \, x := \nabla F_\lambda(x) =: \frac{1}{\lambda}(x - J_\lambda^F x)$$

where $J_\lambda^F x = (I + \lambda\, \partial F)^{-1} x$ and recall that $A_\lambda^F x \in \partial F(J_\lambda^F x)$. Let us prove that under condtion (3.21), $A_\lambda^F x \in \partial_\varepsilon F(x)$, i.e.

$$F(x) + F^*(A_\lambda^F x) - \langle x, A_\lambda x \rangle \leq \varepsilon \ . \tag{3.23}$$

From $A_\lambda x \in \partial F(J_\lambda^F x)$

$$F(J_\lambda^F x) + F^*(A_\lambda^F x) - \langle J_\lambda^F x, A_\lambda^F x \rangle = 0 \ .$$

Hence (3.23) is equivalent to

$$F(x) - F(J_\lambda^F x) - \langle x - J_\lambda^F x, A_\lambda^F x \rangle \leq \varepsilon \ ,$$

that is

$$F(x) - F(J_\lambda^F x) \leq \frac{1}{\lambda} \| x - J_\lambda^F x \|^2 + \varepsilon \ . \tag{3.24}$$

From the k-Lipschitz property of F, (3.24) will be satisfied if

$$k \| x - J_\lambda^F x \| \leq \frac{1}{\lambda} \| x - J_\lambda^F x \|^2 + \varepsilon$$

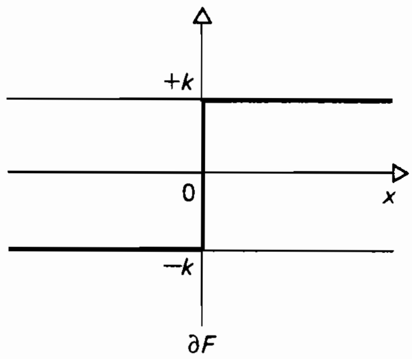
Noticing that for $\lambda \leq 4\varepsilon / k^2$ the function $x \to x^2 - \lambda k x - \varepsilon\lambda$ takes only positive values the relation (3.22) follows.

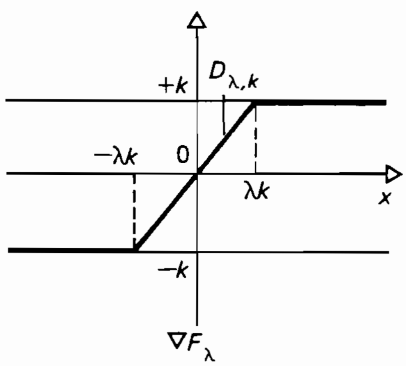
Let us prove that the bound $\lambda \leq 4\varepsilon / k^2$ is sharp. Take $H = \mathbf{R}$ and $F(x) = k|x|$. Let us compute ∇F_λ and $\partial_\varepsilon F$:

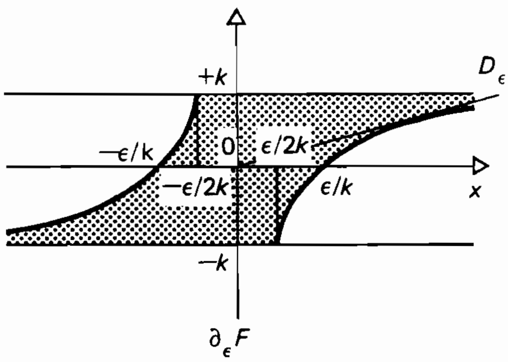
a) Clearly $\partial F(x) = \begin{cases} k & \text{if } x > 0 \\ [-k, +k] & \text{if } x = 0 \\ -k & \text{if } x < 0 \end{cases}$

b) $\nabla F_\lambda(x) = \begin{cases} k & \text{if } x \geq \lambda k \\ \dfrac{x}{\lambda} & \text{if } -\lambda k \leq x \leq \lambda k \\ -k & \text{if } x \leq -\lambda k \end{cases}$

c) $\partial_\varepsilon F(x) = \begin{cases} \left[k - \dfrac{\varepsilon}{x}, +1\right] & \text{if } x \geq \varepsilon / 2k \\ [-1, +1] & \text{if } -\varepsilon/2k \leq x \leq +\varepsilon/2k \\ \left[-k, -k - \dfrac{\varepsilon}{x}\right] & \text{if } x \leq -\varepsilon/2k \end{cases}$







Let us compute the slope m of the ray passing through the origin which is tangent to the curve deliminating $\partial_\varepsilon F$, which equation is $y(x) = +k - \varepsilon / x$. An elementary computation yields $m = k^2 / 4\varepsilon$. In order that the ray $D_{\lambda,k}$, which defines ∇F_λ, to be included in $\partial_\varepsilon F$ one needs

$$\frac{1}{\lambda} \geq \frac{k^2}{4\varepsilon}$$

that is $\lambda \leq 4\varepsilon / k^2$ i.e. condition (3.21) is optimal. □

Thanks to the above result we are able to do the connection between the two approaches described in section 1, that is ε-solution and conditioning.

COROLLARY 3.8 *Let F be a convex function from an Hilbert space H into $\mathbf{R} \cup \{+\infty\}$. For any $\lambda > 0$, let us consider x_λ the solution of the λ-well conditioned problem*

$$\min_{x \in X} \left\{ F(x) + \frac{\lambda}{2} \|x\|^2 \right\}$$

and for any $\varepsilon > 0$, assuming $\inf F > -\infty$, let us consider the set of ε-approximate solutions

$$\varepsilon\text{-}\operatorname{argmin} F \ .$$

Let us suppose that F^* is k-Lipschitz and $\lambda < 4\varepsilon / k^2$. then

$$x_\lambda \in \varepsilon\text{-}\operatorname{argmin} F \ .$$

PROOF Let us write the optimality conditions that characterize respectively x_λ and ε-argmin$F$:

a) $\partial F(x_\lambda) + \lambda x_\lambda \ni 0$ i.e.

$$x_\lambda = (I + \frac{1}{\lambda} \partial F)^{-1}(0)$$

$$= -\frac{1}{\lambda} (\partial F)_\lambda(0)$$

$$= (\partial F^*)_\lambda(0)$$

b) $x \in \varepsilon\text{-}\operatorname{argmin} F$

$$\Updownarrow$$

$$\partial_\varepsilon F(x) \ni 0$$

$$\Updownarrow$$

$$\partial_\varepsilon F^*(0) \ni x$$

i.e. $\varepsilon\text{-argmin} F = \partial_\varepsilon F^*(0)$

Thus the inclusion $x_\lambda \in \varepsilon\text{-argmin} F$ can be translated into $(\partial F^*)_\lambda(0) \subset \partial_\varepsilon F^*(0)$. Applying Proposition 3.7 to F^*, the conclusion follows. □

REMARK 3.9 The condition F^* k-Lipschitz is equivalent to "$F = +\infty$ outside of a ball of radius k". The Lipschitz assumption can be weakened, just assume F^* locally Lipschitz.

THEOREM 3.10 *Let X be a normed linear space and $F, G : X \to \mathbf{R} \cup \{+\infty\}$ two convex functions such that*

- $\inf F$ *and* $\inf G$ *are finite*
- *there exists some* $\rho_0 > 0$ *such that for every* $\varepsilon > 0$

$$\varepsilon\text{-argmin} F \cap B_X(\rho_0) \neq \phi, \ \varepsilon\text{-argmin}\, G \cap B_X(\rho_0) \neq \phi \ . \tag{3.25}$$

Then, the following Lipschitz continuity property of the ε-approximate solutions holds: for any $\rho > \rho_0$, and any $\varepsilon > 0$,

$$\text{haus}\,(\varepsilon\text{-argmin} F \cap B_X(\rho), \ \varepsilon\text{-argmin}\, G \cap B_X(\rho)) \leq C \cdot d_{\alpha(\rho)}(F, G) \tag{3.26}$$

where

$$\alpha(\rho) = \sup\,\{\rho;\ |\inf F| + 1;\ |\inf G| + 1\} \ . \tag{3.27}$$

$$C = 1 + (2\rho + d_{\alpha(\rho)}) \cdot \sup\left\{\frac{1}{\rho - \rho_0 + d_{\alpha(\rho)}} ; \frac{2}{\varepsilon + 2d_{\alpha(\rho)}}\right\} . \tag{3.28}$$

PROOF Let us follow the lines of the proof of Theorem 3.2: Take $x \in \varepsilon\text{-argmin} F \cap B_X(\rho)$ i.e.

$$\begin{cases} \|x\| \leq \rho \\ F(x) \leq \inf F + \varepsilon \ . \end{cases} \tag{3.29}$$

From inequality (3.4)

$$|\inf F - \inf G| \leq d_{\alpha(\rho)}(F, G) \ . \tag{3.30}$$

Since $(x, F(x))$ belongs to $(\text{epi} F)_{\alpha(\rho)}$ for every $\mu > 0$ there exists some (y_μ, λ_μ) belonging to $\text{epi}\, G$ such that

$$\|x - y_\mu\| \leq d_{\alpha(\rho)}(F, G) + \mu \tag{3.31}$$

$$|F(x) - \lambda_\mu| \le d_{\alpha(\rho)}(F, G) + \mu \ . \tag{3.32}$$

From (3.29), (3.30) and (3.32) it follows that

$$\begin{aligned}\lambda_\mu &\le F(x) + d_{\alpha(\rho)}(F, G) + \mu \\ &\le \inf F + \varepsilon + d_{\alpha(\rho)}(F, G) + \mu \\ &\le \inf G + \varepsilon + 2d_{\alpha(\rho)}(F, G) + \mu \ ,\end{aligned}$$

and since $\lambda_\mu \ge G(y_\mu)$,

$$G(y_\mu) \le \inf G + \varepsilon + 2d_{\alpha(\rho)}(F, G) + \mu \ . \tag{3.33}$$

Moreover, from (3.31)

$$\begin{aligned}\|y_\mu\| &\le \|x\| + d_{\alpha(\rho)}(F, G) + \mu \\ &\le \rho + d_{\alpha(\rho)}(F, G) + \mu \ .\end{aligned} \tag{3.34}$$

Let us introduce for any $0 < k < 1$ and thanks to assumption (3.25) some y_k satisfying

$$\begin{cases}\|y_k\| \le \rho_0 \\ G(y_k) \le \inf G + k\varepsilon\end{cases} \tag{3.35}$$

and take for any $\Theta \in]0, 1[$

$$\xi_{k,\vartheta} = \Theta y_\mu + (1 - \Theta) y_k \ . \tag{3.36}$$

By convexity, and from (3.34), (3.35)

$$\begin{aligned}\|\xi_{k,\vartheta}\| &\le \Theta \|y_\mu\| + (1 - \Theta)\|y_k\| \\ &\le \Theta(\rho + d_{\alpha(\rho)} + \mu) + (1 - \Theta)\rho_0 \ .\end{aligned}$$

Thus $\|\xi_{k,\vartheta}\| \le \rho$ as soon as

$$\Theta(\rho + d_{\alpha(\rho)} + \mu) + (1 - \Theta)\rho_0 \le \rho \ ,$$

that is

$$\Theta \in \left[0, \frac{\rho - \rho_0}{\rho - \rho_0 + d_{\alpha(\rho)} + \mu}\right] . \tag{3.37}$$

Similarly

$$G(\xi_{k,\vartheta}) \leq \Theta\, G(y_\mu) + (1-\Theta)\, G(y_k)$$

which from (3.33) and (3.35) implies

$$G(\xi_{k,\vartheta}) \leq \Theta[\inf G + \varepsilon + 2d_{\alpha(\rho)} + \mu] + (1-\Theta)[\inf G + k\varepsilon] \ .$$

Thus $G(\xi_{k,\vartheta})$ is less or equal than $\inf G + \varepsilon$ as soon as

$$\Theta(\varepsilon + 2d_{\alpha(\rho)} + \mu) + (1-\Theta)k\varepsilon \leq \varepsilon \ ,$$

and thus

$$\Theta \in \left[0, \frac{(1-k)\varepsilon}{(1-k)\varepsilon + 2d_{\alpha(\rho)} + \mu}\right] . \tag{3.38}$$

Taking

$$\Theta_k = \inf\left\{\frac{\rho - \rho_0}{\rho - \rho_0 + d_{\alpha(\rho)} + \mu}, \frac{(1-k)\varepsilon}{(1-k)\varepsilon + 2d_{\alpha(\rho)} + \mu}\right\}$$

we obtain

$$\begin{cases} \|\xi_{k,\vartheta_k}\| \leq \rho \ , \\ G(\xi_{k,\vartheta_k}) \leq \inf G + \varepsilon \ , \end{cases}$$

and

$$\|y_\mu - \xi_{k,\vartheta_k}\| = (1-\Theta_k)\|y_\mu - y_k\| \ ,$$

$$= \sup\left\{\frac{d_{\alpha(\rho)} + \mu}{\rho - \rho_0 + d_{\alpha(\rho)} + \mu} ; \frac{2d_{\alpha(\rho)} + \mu}{(1-k)\varepsilon + 2d_{\alpha(\rho)} + \mu}\right\}\|y_\mu - y_k\| \ ,$$

$$\leq (d_{\alpha(\rho)} + \mu) \sup\left\{\frac{1}{\rho - \rho_0 + d_{\alpha(\rho)}} ; \frac{2}{(1-k)\varepsilon + 2d_{\alpha(\rho)}}\right\}(2\rho + d_{\alpha(\rho)} + \mu) \ .$$

Combining the preceding inequality with (3.31)

$$\|x - \xi_{k,\vartheta_k}\| \leq (d_{\alpha(\rho)}(F, G) + \mu)$$

$$\left[1 + (2\rho + d_{\alpha(\rho)} + \mu)\sup\left\{\frac{1}{\rho - \rho_0 + d_{\alpha(\rho)}} ; \frac{2}{(1-k)\varepsilon + 2d_{\alpha(\rho)}}\right\}\right] .$$

The above argument being valid for any $k \in]0, 1[$ and $\mu > 0$ by letting $k \longrightarrow 0$ and $\mu \longrightarrow 0$ we obtain

$$d(x, (\varepsilon\text{-argmin}\, G) \cap B_X(\rho))$$

$$\leq d_{\alpha(\rho)}(F, G)\left[1 + (2\rho + d_{\alpha(\rho)}) \cdot \sup\left\{\frac{1}{\rho - \rho_0 + d_{\alpha(\rho)}} ; \frac{2}{\varepsilon + 2d_{\alpha(\rho)}}\right\}\right] .$$

The formulas (3.26), (3.27), (3.28) clearly follow. □

As a direct consequence of theorem 3.10 we obtain the following result from Hiriart-Urruty [12, Theorem 3.3].

COROLLARY 3.11 *Let us consider* $F : X \to \mathbf{R} \cup \{+\infty\}$ *be a proper, lower semi-continuous convex function from a Banach space* X *into* $\mathbf{R} \cup \{+\infty\}$. *Let* x_1 , *and* x_2 *be in the effective domain of* F *such that for some* $\rho_0 > 0$ *and any* $\varepsilon > 0$

$$\partial_\varepsilon F(x_1) \cap B_x(\rho_0) \neq \phi, \; \partial_\varepsilon F(x_2) \cap B_x(\rho_0) \neq \phi \; . \tag{3.39}$$

Then for all $\rho > \rho_0$ *there exists some constant* $C(1/\rho - \rho_0, 1/\varepsilon)$ *such that*

$$\text{haus}\,(\partial_\varepsilon F(x_1) \cap B_{X^*}(\rho), \partial_\varepsilon F(x_2) \cap B_{X^*}(\rho)) \leq C\left(\frac{1}{\rho - \rho_0}, \frac{1}{\varepsilon}\right)\|x_1 - x_2\| \; . \tag{3.40}$$

PROOF As in Corollary 3.6, we use the following characterization of $\partial_\varepsilon F(x)$:

$$\partial_\varepsilon F(x) = \varepsilon\text{-argmin}\,\{F^*(\cdot) - \langle x, \cdot \rangle\} \; .$$

Then apply Theorem 3.10 to the two functions

$$\Phi_1(x^*) = F^*(x^*) - \langle x_1, x^* \rangle$$

$$\Phi_2(x^*) = F^*(x^*) - \langle x_2, x^* \rangle \; .$$

Clearly $x_1 \in \text{dom} F$, $x_2 \in \text{dom} F$ imply $\inf \Phi_1 > -\infty$, $\inf \Phi_2 > -\infty$ and condition (3.25) is equivalent to (3.39). □

4. PROPERTIES OF THE EPIGRAPHICAL DISTANCE: $\mathbf{d}_\rho$

Let us first describe the following practical result which makes the epigraphical distance easy to handle. Without ambiguity, we use the same notation B_ρ for $B_X(\rho)$ and $B_{B \times \mathbf{R}}(\rho)$.

THEOREM 4.1 *Let X be a normed linear space and F, G be two proper convex functions from X into $\mathbf{R} \cup \{+\infty\}$. Let $\rho_0 > 0$ be such that* $\operatorname{epi} F \cap B_{\rho_0} \neq \phi$, *epi* $G \cap B_{\rho_0} \neq \phi$.

a) *Let us assume that for all $\rho \geq \rho_0$, $d_\rho(F, G) < +\infty$. Then the following condition – called* Kenmochi's *condition – holds: for all $\rho \geq \rho_0$ and $x \in \operatorname{dom} F$ such that $|x| \leq \rho$, $|F(x)| \leq \rho$, for every $\varepsilon > 0$ there exists some $\tilde{x}_\varepsilon \in \operatorname{dom} G$ that satisfies*

$$\begin{cases} \|x - \tilde{x}_\varepsilon\| \leq d_\rho(F, G) + \varepsilon \\ G(\tilde{x}_\varepsilon) \leq F(x) + d_\rho(F, G) + \varepsilon \end{cases} \tag{4.1}$$

and symmetrically, exchanging the role of F and G.

b) *Conversely, assuming that there exists some constant (that depends on ρ) $c(\rho) \in \mathbf{R}^+$ such that for all $x \in \operatorname{dom} F$ such that $|x| \leq \rho$, $|F(x)| \leq \rho$, there exists some $\tilde{x} \in \operatorname{dom} G$ that satisfies*

$$\begin{cases} \|x - \tilde{x}\| \leq c(\rho) \ , \\ G(\tilde{x}) \leq F(x) + c(\rho) \ , \end{cases} \tag{4.2}$$

and the symmetric condition (interchanging F and G), then the following inequality holds:

$$d_\rho(F, G) \leq c(\rho_1) \tag{4.3}$$

where $\rho_1 = \sup\{\rho;\ \alpha(1 + \rho)\}$ and α is such that $-\alpha(1 + \|\cdot\|)$ minorizes F and G. Moreover the following estimation holds:

$$d_\rho(F, G) \leq \operatorname{haus}((\operatorname{epi} F)_\rho, (\operatorname{epi} G)_\rho) \leq \frac{2\rho}{\rho - \rho_0} d_{\rho_1}(F, G) \ . \tag{4.4}$$

PROOF It suffices to observe the following

(i) $h_\rho(\operatorname{epi} F, \operatorname{epi} G) \leq k$, if and only if, for every $\varepsilon > 0$

$$(\operatorname{epi} F)_\rho \subset \operatorname{epi} G + (k + \varepsilon)B$$

$$(\operatorname{epi} G)_\rho \subset \operatorname{epi} F + (k + \varepsilon)B$$

where B is the unit ball of X (Definition 3.1).

(ii) that these inclusions yield exactly the Kenmochi condition (4.1) if one remembers that epi G is an epigraph.

(iii) the estimate (4.3) is obtained by calculating an upper bound on k in terms of the given coefficients ρ and $c(\rho)$. We do that next.

Given $(x, \mu) \in \text{epi} F \cap B_\rho$, i.e., $\|x\| \le \rho$, $|\mu| \le \rho$, $\mu \ge F(x)$, we have, introducing an affine function that minorizes F (it exists because F is proper), for all $y \in X$,

$$F(y) \ge -\alpha(1 + \|y\|) ,$$

$$|F(x)| \le \sup\{\rho; + \alpha(1+\rho)\} =: \rho_1 . \tag{4.5}$$

By (4.2) there exists some $\tilde{x} \in \text{dom}\, G$ such that

$$\|x - \tilde{x}\| \le c(\rho_1)$$

$$G(\tilde{x}) \le F(x) + c(\rho_1)$$

$$\le \mu + c(\rho_1) . \tag{4.6}$$

We distinguish the two separate cases:

(i) If $G(\tilde{x}) \le F(x)$ then $(\tilde{x}, \mu) \in \text{epi}\, G \cap B_{\rho + c(\rho_1)}$ where we have used: $\|\tilde{x}\| \le \|x\| + c(\rho_1) \le \rho + c(\rho_1)$. Hence

$$d((x, \mu), (\text{epi}\, G)_{\rho + c(\rho_1)}) \le \|(x, \mu) - (\tilde{x}, \mu)\|$$

$$\le c(\rho_1) .$$

(ii) If $G(\tilde{x}) \ge F(x)$, take $(\tilde{x}, G(\tilde{x}) + \mu - F(x)) \in \text{epi}\, G$ and observe that by (4.6),

$$|G(\tilde{x}) - F(x) + \mu| \le |\mu| + G(\tilde{x}) - F(x) ,$$

$$\le \rho + c(\rho_1) .$$

Hence $(\tilde{x}, G(\tilde{x}) + \mu - F(x)) \in \text{epi}\, G \cap B_{\rho + c(\rho_1)}$ and

$$d((x, \mu), (\text{epi}\, G)_{\rho + c(\rho_1)}) \le \sup\{c(\rho_1); |\mu - (G(\tilde{x}) + \mu - F(x))|\}$$

$$\le \sup\{c(\rho_1); G(\tilde{x}) - F(x)\}$$

$$\le c(\rho_1) .$$

and that is (4.3). In order to prove (4.4) we next rely on the estimate provided by Proposition 4.2, that yields

$$\text{haus}((\text{epi}\, G)_{\rho + c(\rho_1)}, (\text{epi}\, G)_\rho) \le \left[1 + \frac{2\rho_0}{\rho - \rho_0}\right] c(\rho_1) \ ,$$

which combined with

$$d((x, \mu), (\text{epi}\, G)_{\rho + c(\rho_1)}) \le c(\rho_1) \ ,$$

gives

$$d((x, \mu), (\text{epi}\, G)_\rho) \le 2\left[1 + \frac{\rho_0}{\rho - \rho_0}\right] c(\rho_1)$$

and that is (4.4). □

To prove the next proposition we could use an argument that parallels that of Lemma 3.3, we give here a proof based on duality.

PROPOSITION 4.2 *Let X be a Banach space and C a closed convex set such that $C \cap B_{\rho_0} \ne \phi$. Then for any $\rho > \rho_0$, for any $d\rho \ge 0$*

$$\text{haus}(C \cap B_{\rho + d\rho}, C \cap B_\rho) \le L \cdot d\rho \tag{4.7}$$

where the Lipschitz constant L is given by

$$L = 1 + \frac{2\rho_0}{\rho - \rho_0} \ . \tag{4.8}$$

PROOF From Hormander classical duality formula, see [4, Section 3] for example,

$$\text{haus}(C \cap B_{\rho + d\rho}, C \cap B_\rho) = \sup\{|s(C \cap B_{\rho + d\rho}, x^*) - s(C \cap B_\rho, x^*)|; \|x^*\| \le 1\}$$

where $s(K, x^*) = \sup\{<x^*, x>; x \in K\}$ is the support function of K.

Note that

$$s(C \cap B_\rho) = (\delta_C + \delta_{B_\rho})^*$$

and that δ_{B_ρ} is continuous at a point of the domain of δ_C (because of the assumption $C \cap B_{\rho_0} \ne \phi$ and $\rho > \rho_0$), which means that

$$s(C \cap B_\rho) = \delta_C^* \,\square\, \delta_{B_C}^* \ ,$$

i.e., the inf-condition □ is exact. Hence

$$s(C \cap B_\rho, x^*) = \min\{s(C, y^*) + \rho\|x^* - y^*\|;\ y^* \in X^*\}$$

and the min is achieved at some point y^*_ρ. Similarly

$$s(C \cap B_{\rho+d\rho}, x^*) = \min\{s(C, y^*) + (\rho + d\rho)\|x^* - y^*\|;\ y^* \in X^*\}$$

Thus

$$s(C \cap B_{\rho+d\rho}, x^*) - s(C \cap B_\rho, x^*)$$

$$\leq \{s(C, y^*_\rho) + (\rho + d\rho)\|x^* - y^*_\rho\|\} - \{s(C, y^*_\rho) + \rho\|x^* - y^*_\rho\|\}\ ,$$

$$\leq d\rho \cdot \|x^* - y^*_\rho\|\ .$$

Indeed

$$s(C, y^*_\rho) + \rho\|x^* - y^*_\rho\| \leq s(C \cap B_\rho, x^*)\ .$$

Since $C \cap B_{\rho_0} \neq \phi$, taking some $x_0 \in C \cap B_{\rho_0}$

$$s(C, y^*_\rho) \geq \langle y^*_\rho, x_0\rangle\ .$$

Moreover

$$s(C \cap B_\rho, x^*) \leq \sup\{\langle x^*, u\rangle;\ u \in B_\rho\}$$

$$\leq \rho\|x^*\|\ .$$

Hence

$$\rho\|x^* - y^*_\rho\| \leq \rho_0\|y^*_\rho\| + \rho\|x^*\|\ ,$$

and

$$(\rho - \rho_0)\|y^*_\rho\| \leq 2\rho\|x^*\|\ ,$$

$$\|x^* - y^*_\rho\| \leq \left(1 + \frac{2\rho_0}{\rho - \rho_0}\right)\|x^*\|\ .$$

Finally

$$|s(C \cap B_{\rho+d\rho}, x^*) - s(C \cap B_\rho, x^*)| \leq d\rho\left(1 + \frac{2\rho_0}{\rho - \rho_0}\right)\|x^*\|\ ,$$

and

$$\text{haus}(C \cap B_{\rho + d\rho}, C \cap B_\rho) \le d\rho \cdot \left[1 + \frac{2\rho_0}{\rho - \rho_0}\right] . \quad \square$$

REMARK 4.3 Theorem 4.1 tells us that in order to compute $d_\rho(F, G)$ we have to find the best constant $c(\rho)$ for which the condition (4.2) holds. Indeed this condition that we called Kenmochi's condition had been introduced in Kenmochi [13], see also Attouch and Damlamian [2] in order to study the existence of strong solutions to evolution problems of the following type:

$$0 \in \frac{du}{dt} + \partial F(t, u(t));\ u(0) = u_0 .$$

The time dependence of F with respect to t, with our terminology, can now be expressed as an absolute continuity property of the application $t \to F(t)$. Let us recall this condition: there exists some $b \in C([0, T]; H) \cap W^{1,2}([0, T]; H)$ and a function a, increasing such that:

$$\forall 0 \le s \le t \le T \ \ \forall x \in \text{dom} F(s, \cdot),\ \exists \tilde{x} \in \text{dom} F(t, \cdot) \text{ such that} \tag{4.9}$$

$$\|x - \tilde{x}\| \le |b(t) - b(s)| \cdot [1 + \|x\|] ,$$

$$F(t, \tilde{x}) \le F(s, x) + (a(t) - a(s))[|F(s, x)| + \|x\|^2 + 1] .$$

Thus, $\forall x \in \text{dom} F(s, \cdot)$ with $\|x\| \le \rho$, $|F(s, \cdot)| \le \rho$ we have the existence of some $\tilde{x} \in \text{dom} F(t, \cdot)$ such that

$$\|x - \tilde{x}\| \le (1 + \rho)|b(t) - b(s)| ,$$

$$F(t, \tilde{x}) \le F(s, x) + (1 + \rho + \rho^2)(a(t) - a(s)) .$$

Taking $c(\rho) = \sup\{(1 + \rho)|b(t) - b(s)|;\ (1 + \rho + \rho^2)(a(t) - a(s))\}$, we see that condition (4.2) is satisfied. □

To conclude let us examine the connection between the epigraphic distance and the Moreau-Yosida type distance $d_{\lambda,\rho}$ introduced in [4].

PROPOSITION 4.4 *Let X be a Hilbert space and $F, G : X \to \mathbf{R} \cup \{+\infty\}$ be two proper lower semicontinuous convex functions. For any $\lambda > 0$ and $\rho > 0$ we have the following estimate:*

$$d_{\lambda,\rho}(F, G) \le L \cdot d_{C_\lambda(\rho)}(F, G) \tag{4.10}$$

where $L = 1/\lambda(\rho + C_\lambda(\rho))$ *and* $C_\lambda(\rho)$ *is given by* (4.15)

PROOF Let us fix $x \in X$ and some $\lambda > 0$. By definition of F_λ, G_λ

$$F_\lambda(x) = \min\{F(y) + \frac{1}{2\lambda}\|x - y\|^2;\ y \in X\} ,$$

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

$$G_\lambda(x) = \min\{G(y) + \frac{1}{2\lambda}\|x - y\|^2;\ y \in X\} ,$$

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

Hence

$$F_\lambda(x) - G_\lambda(x) = F(J_\lambda^F x) - G(J_\lambda^G x) + \frac{1}{2\lambda}(\|x - J_\lambda^F x\|^2 - \|x - J_\lambda^G x\|^2) . \tag{4.11}$$

Let us consider the point $J_\lambda^G x \in \operatorname{dom} G$. We first notice that

$$\|J_\lambda^G x\| \leq \|J_\lambda^G 0\| + \|x\| .$$

Introducing $\alpha \geq 0$ such that

$$F(y) + \alpha(1 + \|y\|) \geq 0 \tag{4.12}$$

$$G(y) + \alpha(1 + \|y\|) \geq 0$$

by definition of $J_\lambda^G 0$

$$G(J_\lambda^G 0) + \frac{1}{2\lambda}\|J_\lambda^G 0\|^2 \leq G(x_0) + \frac{1}{2\lambda}\|x_0\|^2$$

where we have picked some point $x_0 \in \operatorname{dom} G$. Hence

$$-\alpha(1 + \|J_\lambda^G 0\|) + \frac{1}{2\lambda}\|J_\lambda^G 0\|^2 \leq G(x_0) + \frac{1}{2\lambda}\|x_0\|^2$$

and

$$\|J_\lambda^G 0\| \leq 2\lambda\,\alpha + \|x_0\| + \sqrt{2\lambda(\alpha + |G(x_0)|)} .$$

Hence

$$\|J_\lambda^G x\| \leq \|x\| + 2\lambda\,\alpha + \|x_0\| + \sqrt{2\lambda(\alpha + |G(x_0)|)} . \tag{4.13}$$

Moreover

$$G(J_\lambda^G x) \le G_\lambda(x)$$

$$\le G(x_0) + \frac{1}{2\lambda}\|x - x_0\|^2 .$$

Thus assuming $\|x\| \le \rho$, we have $\|J_\lambda^G x\| \le c(\rho)$, $|G(J_\lambda^G x)| \le c(\rho)$ where

$$c(\rho) =: \sup\Big\{\rho + 2\lambda\,\alpha + \|x_0\| + \sqrt{2\lambda(\alpha + |G(x_0)|)};\ |G(x_0)| + \frac{1}{\lambda}(\rho^2 + \|x_0\|^2)\ ;$$

$$\alpha\Big[1 + \rho + 2\lambda\,\alpha + \|x_0\| + \sqrt{2\lambda(\alpha + |G(x_0)|)}\Big]\Big\} .$$

We note that $c(\rho)$ depends on a continuous and bounded way of α, ρ, λ, $1/\lambda$, $\|x_0\|$, $G(x_0)$ where x_0 is an arbitrary point in $\operatorname{dom} G$. Assuming $\operatorname{epi} G \cap B_{\rho_0} \ne 0$, we have

$$C_\lambda(\rho) =: C(\rho,\ \lambda,\ \frac{1}{\lambda},\ \rho_0,\ \alpha) = \sup\{\rho + 2\lambda\,\alpha + \rho_0 + \sqrt{2\lambda(\alpha + \rho_0)};\ \rho_0 \tag{4.15}$$

$$+ \frac{1}{\lambda}(\rho^2 + \rho_0^2);\ \alpha\,(1 + \rho + 2\lambda\,\alpha + \rho_0 + \sqrt{2\lambda(\alpha + \rho_0)})\} .$$

By definition of the epigraphical distance d_ρ and Theorem 4.1, (4.1), for every $\varepsilon > 0$ there exists $\tilde{x}_\varepsilon \in \operatorname{dom} F$ such that

$$\begin{cases} \|J_\lambda^G x - \tilde{x}_\varepsilon\| \le d_{C_\lambda(\rho)}(F,\ G) + \varepsilon \\ F(\tilde{x}_\varepsilon) \le G(J_\lambda^G x) + d_{C_\lambda(\rho)}(F,\ G) + \varepsilon \end{cases} \tag{4.16}$$

Returning to (4.11), this yields

$$F_\lambda(x) - G_\lambda(x) \le F(J_\lambda^F x) - F(\tilde{x}_\varepsilon) + d_{C_\lambda(\rho)}(F,\ G) + \varepsilon \tag{4.17}$$

$$+ \frac{1}{2\lambda}[\|x - J_\lambda^F x\|^2 - \|x - J_\lambda^G x\|^2] .$$

From the convex subdifferential inequalities

$$F(\tilde{x}_\varepsilon) \ge F(J_\lambda^F x) + \langle A_\lambda^F x,\ \tilde{x}_\varepsilon - J_\lambda^F x\rangle$$

$$\ge F(J_\lambda^F x) + \langle A_\lambda^F x,\ \tilde{x}_\varepsilon - J_\lambda^G x\rangle + \langle A_\lambda^F x,\ J_\lambda^G x - J_\lambda^F x\rangle\ ,$$

$$\|x - J_\lambda^F x\|^2 - \|x - J_\lambda^G x\|^2 - 2\langle x - J_\lambda^F x,\ J_\lambda^G x - J_\lambda^F x\rangle \le 0$$

and (4.17) we deduce

$$F_\lambda(x) - G_\lambda(x) \le \langle A_\lambda^F x,\ J_\lambda^G x - \tilde{x}_\varepsilon\rangle + \langle A_\lambda^F x,\ J_\lambda^F x - J_\lambda^G x\rangle + d_{C_\lambda(\rho)}(F,\ G) + \varepsilon$$

$$+ \frac{1}{\lambda} < x - J_\lambda^F x, J_\lambda^G x - J_\lambda^F x > .$$

Since $A_\lambda^F x = 1/\lambda (x - J_\lambda^F x)$, we obtain,

$$F_\lambda(x) - G_\lambda(x) \le < A_\lambda^F x, J_\lambda^G x - \tilde{x}_\varepsilon > + d_{C_\lambda(\rho)}(F, G) + \varepsilon$$

which with (4.16) implies,

$$F_\lambda(x) - G_\lambda(x) \le [d_{C_\lambda(\rho)}(F, G) + \varepsilon](1 + \|A_\lambda^F x\|) .$$

Observing that

$$\|A_\lambda^F x\| \le \|A_\lambda^F 0\| + \frac{1}{\lambda}\|x\|$$

$$\le \frac{1}{\lambda}(\|J_\lambda^F 0\| + \|x\|) ,$$

$$\le \frac{1}{\lambda}(\rho + C_\lambda(\rho)) ,$$

we finally obtain (let $\varepsilon \to 0$)

$$F_\lambda(x) - G_\lambda(x) \le d_{C_\lambda(\rho)}(F, G)\left[1 + \frac{1}{\lambda}(\rho + C_\lambda(\rho))\right] .$$

Interchanging the role of F and G, this yields (4.10). □

BIBLIOGRAPHY

[1] Attouch, H.: *Variational Convergence for Functions and Operators*. Applicable Mathematics Series Pitman (London), 1984.

[2] Attouch, H. and A. Damlamian: Problèmes d'évolution dans les Hilbert et applications. J. Math. Pure Appl. 54 (1975), 53–74.

[3] Attouch, H. and R. Wets: Approximation and convergence in nonlinear optimization, in *Nonlinear Programming* 4, eds. O. Mangasarian, R. Meyer and S. Robinson, Academic Press, New York, 367–394, 1981.

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

[5] Attouch, H. and R. Wets: Another isometry for the Legendre-Fenchel transform, Techn. Report, AVAMAC, Univ. Perpignan, 1986.

[6] Attouch, H. and R. Wets (1986): On strictly convex minimization problems. Conditioning and Stability, in preparation.

[7] Aubin, J.-P.: Lipschitz behaviour of solutions to convex minimization problems. *Math. Op. Res.* 9 (1984), 87–111.

[8] Aubin, J.-P. and I. Ekeland (1984): *Applied Nonlinear Analysis.* Wiley Interscience Series.

[9] Aubin, J.-P. and H. Frankowska: On Inverse functions theorems for set valued maps. WP-00-00, IIASA, 1984.

[10] Aze, D.: Convergences variationelles et dualité. Applications en calcul des variations et en programmation mathématique. Thése d'état. Univ. Perpignan, 1986.

[11] Dontchev, A.L.: *Perturbations, Approximations and Sensitivity Analysis of Optimal Control Systems*, Springer Verlag Lect. Notes in Contr. and Inf. Sciences vol. 52, Berlin, 1983.

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

[13] Kenmochi, N.: The semi-discretization method and time dependent parabolic variational inequalities. *Proc. Jap. Acad.* 50 (1974), p. 714–717.

[14] Klatte, D. and B. Kummer: On the (Lipschitz) continuity of solutions of parametric optimization problems, *Proc. of the 16th Conf. "Math. Optimization"* Sellin 1986, Seminar bericht No. 64, Humbold Univ. Sekt. Math. Berlin, 50–61, 1984.

[15] Kummer, B. (1983): Stability of generalized equations and Kuhn-Tucker points of perturbed convex prorams, in: *Proc. of 11th Conference Copenhagen.*

[16] Kutateladze (1979): Convex ε-programming, *Soviet Mathematics Doklady* 20 (1979) 391–393.

[17] Nurminski, E.A. (1978): Continuity of ε-subgradient mappings, *Kibernetika* (Kiev), 5, 148–149, see also IIASA working paper 78–58.

[18] Robinson, S.M.: Stability theory for systems of inequalities. Part I: linear systems *SIAM J. Numerical Analysis* vol. 12, No. 5, p.754–769, 1975.

[19] Robinson, S.M.: Stability theory for systems of inequalities. Part II: Differentiable Nonlinear systems *SIAM J. Numerical Analysis* 13 (1976), No. 4. p. 497–513.

[20] Rockafellar, R.T.: Lipschitzian stability in optimization: the role of nonsmooth analysis, IIASA, WP-86-46, Laxenburg, Austria 1986.

[21] Salinetti, G. and R. Wets: On the convergence of sequence of convex sets in finite dimensions, *SIAM Review* 21 (1979), 18–33.

[22] Schultz, R.: Estimates for Kuhn-Tucher points of perturbed convex programs Tech. Report Humbold Univ. Berlin, Sektion Math., 1986.

[23] Strodiot, J.J., V. Him Nguyen, N. Heukemes: ε-optimal solutions in nondifferentiable convex programming ... *Mathematical Programming* 25 (1983), 307–327.

[24] Mosco, U.: Convergence of convex sets and of solutions of variational inequalities; Advances in Math., 3, 510–585 (1969).