International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Quantitative Stability of Variational Systems: II. A Framework for Nonlinear Conditioning

Attouch, H. and Wets, R.J.-B.

IIASA Working Paper

WP-88-009

February 1988

WORKING PAPER

QUANTITATIVE STABILITY OF VARIATIONAL SYSTEMS: II. A FRAMEWORK FOR NONLINEAR CONDITIONING

Hedy Attouch
Roger J-B Wets

February 1988
WP-88-9

IIASA
International Institute for Applied Systems Analysis

QUANTITATIVE STABILITY OF VARIATIONAL SYSTEMS: II. A FRAMEWORK FOR NONLINEAR CONDITIONING

Hedy Attouch
Roger J-B Wets

February 1988
WP-88-9

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

FOREWORD

This paper follows Attouch and Wets [7] where the authors introduce a new distance function, namely the hausdorff epigraphical distance, which is specially fitted to the study of the quantitative stability of the solutions of optimization problems. They now focuss their attention on well conditioned minimization problems, which amounts to some control of the "curvature' of the function around its local minimum, and obtain with the help of the above distance hölderian and lipschitzian stabilty results, which are in some sense optimal.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

CONTENTS

QUANTITATIVE STABILITY OF VARIATIONAL SYSTEMS: II. A FRAMEWORK FOR NONLINEAR CONDITIONING

Hedy Attouch and Roger J-B Wets***

*Université de Perpignan – **University of California-Davis

Abstract It is shown that for well-conditioned problems (local) optima are holderian with respect to the epi-distance.

1. INTRODUCTION

During the last few years much effort has been devoted to the study of the stability of the solutions of optimization problems under various perturbations of the original problem. Much has been said about the continuity properties of the optimal value and of the set of optimal solutions as a function of these perturbations. This is also the purpose of this paper, as well as its companion [8]. However, we make a break with the standard approach in at least two ways. First, we do not just consider a particular class of perturbations, but allow for perturbations of a global character. The reference to *variational systems* in the title, cf. Rockafellar and Wets [42], is intended to stress this concern in our approach. Second, we are interested in *quantitative* results that could be used in estimating the rate of convergence of an approximation scheme, or to obtain error estimates for the current solution in an algorithmic procedure. To measure the distance between optimization problems, we rely on the *epi-distance*. We deal mainly with the properties of local optima when the problems are appropriately conditioned, and derive hölderian and lipschitzian inequalities in terms of the epi-distance. In [8], we obtain lipschitzian properties for the ϵ-approximate solutions of convex optimization problems.

An overview of the stability results of topological nature could be gather from the work of Evans and Gould [21], Fiacco [22], Bank, Guddat, Klatte, Kummer and Tammer [16], Dolecki [19], Gauvin [21], Hogan [26], Robinson [35], [36], Rockafellar [38], and Zolezzi [51]. Recently, the introduction of the concept of epi-convergence, has allowed us to unify a large number of there results, cf., Mosco [33], Wets [48], Attouch and Wets [4], Attouch [2], and Robinson [37]. And it is also in that framework that we place ourselves in this paper.

Quantitative results, that provide (computable) bounds on the sensitivity of the solution to changes in the values of the parameters, are much more limited. An approach initiated by Aubin [9], Aubin and Frankowska [11], (see also Rockafellar [39], and Aubin and Wets [12]), relies on nonsmooth analysis and the Inverse Function Theorem for multifunctions. Locally, the optimal solution of an optimization problem $\min_x f(x)$ is characterized by the optimality condition

$$0 \in \partial f(x) \quad ,$$

where ∂f is some generalized gradient of f. Then, with a surjectivity assumption on the tangent cone to the graph of ∂f at $(x, 0)$, the solution set is proved to be pseudo-lipschitz with respect to the data. The counterpart of the great generality and flexibility attained here is the need to calcualte second (generalized) derivatives of f and that could be quite involved.

In this article, we present a completely different approach that does not use optimality conditions. Recall that a local minimizer x_f of f is characterized by

$$f(x_f) \leq f(y) \qquad \text{for all } y \text{ in } X \text{ such that } \| y - x_f \| \leq \rho, \rho > 0 \ .$$

Since our aim is a *nonlinear conditioning theory* for optimization problems, we must be able to deal with all kind of perturbations of f and predict the worst possible effect on x_f of some change or slight error in the coefficients of the objective function and/or the constraints. We already know that the good notion of topological deformation that yields stability for minimization problems, is ***epi-continuity***, we are thus naturally led to the study of metrics (for extended real valued functions) that induce epi-convergence. In [5], and also [6], we exhibit such classes of metrics obtained via regularization by epigraphical sum (inf-convolution). Distances are defined in terms of (uniform) bounds on the differences of these so-regularized functions on bounded subsets of X. In [7], we define another metric, called the (hausdorff-) ***epi-distance***, which also induces the topology of epi-convergence. It is this latter notion of distance that we shall use to derive our results. (In view of [7, Theorems 3.4, 3.7 and 3.9], it is always possible to re-express the results in terms of the metrics defined in [5], [6].) In [8], where the attention is restricted to convex functions, the epi-distance is also used to prove that the multifunction $f \mapsto \varepsilon-\text{argmin}\, f$ is pseudo-lipschitzian.

To be able to consider any possible perturbation of f and still obtain quantitative results, we need some geometric assumptions about f. Clearly we have to control the "curvature" of f at x_f. This is done with the help of the radial regularization of f at x_f (see Section 2), i.e., the largest function $\varphi \geq 0$ such that t_n goes to 0 whenever $\varphi(t_n)$ goes to 0, and

$$f(x) \geq f(x_f) + \varphi(\| x - x_f \|) \text{ for all } x \text{ such that } \| x - x_f \| \leq \rho,\ \rho > 0 \ . \tag{1.1}$$

A classical property of f that provides this ***strong*** local minimization property is ***uniform convexity***. Let us however stress the fact that this assumption (existence of such a φ) is of local character, and in general we place ***no*** convexity restrictions on f.

The main result is Theorem 4.1 where x_f is proved to be hölder stable. For example in the normalized case ($x_f = 0$ and $f(x_f) = 0$), when f is quadratically "conditioned" at 0, i.e., $f(x) \geq \| x \|^2 = \varphi(x)$ for $\| x \| \leq 1$, and g is some approximation or perturbation of f, with x_g a corresponding minimizer, Theorem 4.1 asserts that

$$\| x_g - x_f \| \leq [4 \, \mathrm{haus}_\rho(f, g)]^{1/2} \tag{1.2}$$

provided that $\mathrm{haus}_\rho(f, g)$, the epi-distance (of parameter ρ) between f and g, is sufficiently small. We show in Section 5 that this ***hölderian stability result is optimal.*** In fact, this estimate is consistent with related, but more specialized, results that have been obtained in various areas: Moreau [32] the sweeping problem (le problème de rafle), Attouch and Wets [5] isometries for the Legendre-Fenchel transform, Rabier and Thomas [34] approximations for the solutions of elliptic p.d.e., Dontchev [20] approximations and perturbations of optimal control problems, and Daniel [18] and Schultz [44] for specific perturbations in nonlinear mathematical programming.

2. CONDITIONING FOR MINIMIZATION PROBLEMS

Let X be a normed space and let $\|\cdot\|$ denote the norm of X. Given $x \in X$ and $\rho > 0$, we denote by $B(x, \rho)$ the closed ball of radius ρ centered at x. We also write ρB for $B(0, \rho)$. Given $f: x \to \overline{\mathbf{R}}$, a real extended valued function, a point $x \in X$ satisfying

$$f(y) \geq f(x) \quad \text{for all } y \in B(x, \rho) \tag{2.1}$$

for some $\rho > 0$, is called a *local minimizer*

of f.

Our main objective is the study of the stability of the solution of such minimization problems with respect to data perturbation (that is with respect to f). To that end, let us introduce the following class of well behaved minimization problems.

DEFINITION 2.1 *A function* $\varphi: \mathbf{R}_+ \to \overline{\mathbf{R}}_+$ *is called* admissible *if* $\varphi(t_n) \to 0$ *implies* $t_n \to 0$. *Let* $f: X \to \overline{\mathbf{R}}$ *and* $x \in X$ *be such that*

$$f(y) \geq f(x) + \varphi(\|y - x\|) \;\; \forall\, y \in B(x, \rho) \tag{2.2}$$

for some $\rho > 0$ and some admissible function φ. Then x is called a φ -local minimizer *of f.*

This notion will play a key role in our development. Let us mention that under assumption (2.2), x is a unique minimizer of f on $B(x, \rho)$ and that every minimizing sequence does converge strongly to x. That is precisely the notion of ***well posed minimization problem in Tykhonov's sense***, see [46].

Indeed as noticed by T. Zolezzi [50, Corollary 1], to say that the local minimization is well posed in the above sense is equivalent to the existence of an admissible function φ for which (2.2) holds. The choice of the above terminology, "φ-local minimizer", is motivated by the fact that we are interested in quantitative stability, that very much depends on the shape of φ. Figures 1 and 2 illustrate two typical situations.

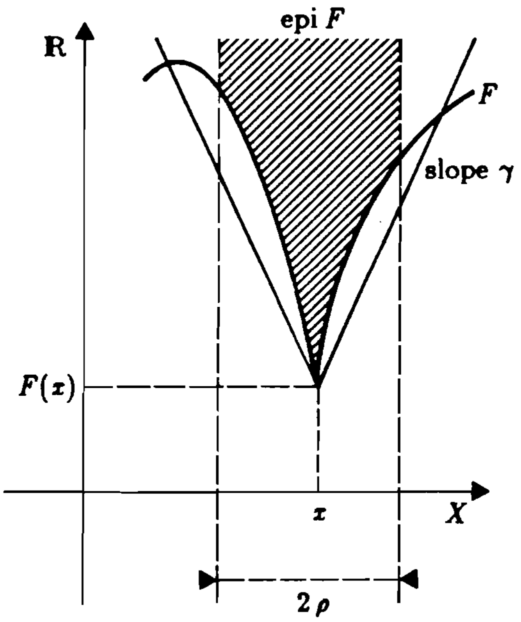

FIGURE 1 $\varphi(r) = \gamma|r|$.

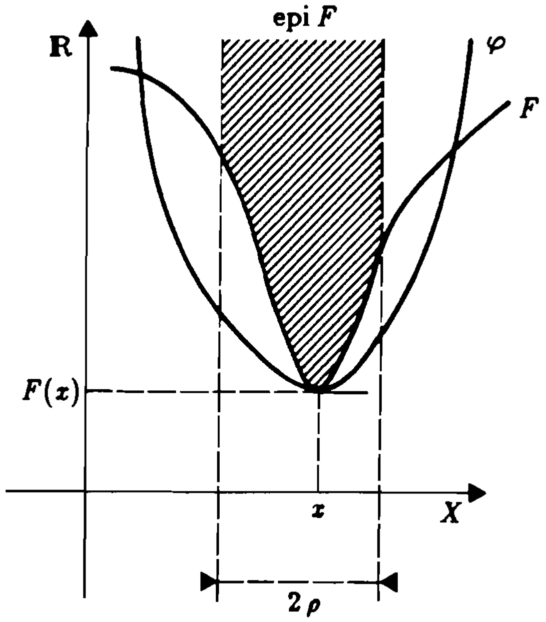

FIGURE 2 $\varphi(r) = cr^2$.

When $\varphi(r) = \gamma|r|$, for some $\gamma > 0$, (see Figure 1), the function f is sharply pointed at x. In that case, we shall be able to derive ***lipschitz stability*** of the local minimizer x with respect of f. When $\varphi(r) = cr^2$, the function f may be smooth at x (see Figure 2).

These examples illustrate the importance of a good understanding of what is the "best" admissible function φ for which (2.2) holds. As we shall see, the sharpest stability results are obtained by taking the largest admissible function φ for which (2.2) holds. We now characterize such functions φ.

PROPOSITION 2.2 *Let us assume that* (2.2) *holds. Then there exists a largest admissible function, which we denote by* φ_f, *such that the inequality* (2.2) *holds. It is given by the following formula*

$$\varphi_f(r) = \inf\{f(y) - f(x);\ \|y - x\| = r\},\ \forall 0 \le r \le \rho\ . \tag{2.3}$$

It is called the radial regularization of f *at* x.

The proof of Proposition 2.2 is quite elementary. Just notice that if $(\varphi_i)_{i\in I}$ is the family of admissible functions such that (2.2) holds, then $\bigvee_{i\in I}\varphi_i$ is still an admissible function and (2.2) still holds. Then take

$$\varphi_f = \bigvee_{i\in I}\varphi_i\ .$$

The function $y \mapsto \varphi_f(\|y - x\|)$ is the largest radial function which minorizes $y \mapsto f(y) - f(x)$. This justifies the terminology of radial regularization which plays an important role in the theory of Orlicz spaces (see A. Fougères [23]). It is an interesting question to characterize the properties of φ_f from the properties of f.

PROPOSITION 2.3 *Assume that* f *is convex and that* (2.2) *holds for some admissible function* φ. *Then the radial regularization* φ_f *of* f *at* x *is such that* $r \mapsto 1/r\,\varphi_f(r)$ *is increasing, and hence* φ_f *is strictly increasing.*

PROOF The proof is patterned after that of Proposition 2 of Zolezzi [50]. Let $0 \leq r_1 \leq r_2 \leq \rho$ and $y_2 \in X$ such that $\|y_2 - x\| = r_2$. Take $y_1 = (1 - r_1/r_2)x + r_1/r_2 y_2$. Then

$$\|y_1 - x\| = \frac{r_1}{r_2} \|y_2 - x\| = r_1 \ .$$

Hence

$$\varphi_f(r_1) \leq f(y_1) - f(x)$$

which by convexity of f yields

$$\varphi_f(r_1) \leq (1 - \frac{r_1}{r_2})\, f(x) + \frac{r_1}{r_2}\, f(y_2) - f(x) \leq \frac{r_1}{r_2}\, (f(y_2) - f(x)) \ .$$

This inequality being true for any $y_2 \in X$ satisfying $\|y_2 - x\| = r_2$, it follows

$$\varphi_f(r_1) \leq \frac{r_1}{r_2}\, \varphi_f(r_2) \ .$$

This means that $r \rightarrow (\varphi_f(r)/r)$ is increasing. Noticing that $\varphi_f(r) \neq 0$ as soon as $r \neq 0$ (this is a consequence of Definition 2.1 of admissible functions), it follows that $r \mapsto \varphi_f(r)$ is strictly increasing on $(0, \rho]$. □

REMARK Even if f is convex, φ_f is not convex in general. Indeed, $\varphi_f(r)/r \geq \varphi'_f(0)$, that is, φ_f is convex "near the origin". To avoid this difficulty one may work with the radial-convex regularization where the admissible functions φ are also required to be convex.

When considering a concrete minimization problem, think for example of a mathematical program with a large number of variables, the construction of an admissible function φ such that the inequality (2.2) is satisfied could be quite involved. The main reason is that the point x which actually minimizes f is a priori unknown. For this reason it is important to know what global properties for f automatically insure that an inequality of the type (2.2) is satisfied around x (which turns to be a global minimizer). This is where the notion *of uniform convexity* turns out to be useful. An abundant literature has been devoted to this subject (see e.g., Zalinescu [49], Vladimirov, Nestorov and Chekanov [47]) and its connection with stability in optimization and control (Sonntag [45], Dontchev [20]). A survey, with some new results, can be found in D. Azé [14]. For the sake of simplicity, we first consider the Hilbert case (see Section 5 for more general results).

PROPOSITION 2.4 *Let H be a real Hilbert space and $f \in \Gamma_0(H)$ the space of convex, lower semicontinuous, proper functions from H into $\mathbf{R} \cup \{+\infty\}$. The following statements are equivalent*

$$(i)\ f(tx+(1-t)y) \le tf(x)+(1-t)f(y)-\gamma t(1-t)\|x-y\|^2, \quad \forall x, y \in \operatorname{dom} f, \forall t \in [0,1]$$

$$(ii)\ f(y) \ge f(x) + <\partial f(x), y-x> + \gamma\|x-y\|^2, \quad \forall (x, \partial f(x)) \in \partial f, \forall y \in \operatorname{dom} f$$

$$(iii)\ <\partial f(y) - \partial f(x), y-x> \ \ge 2\gamma\|y-x\|^2, \quad \forall (x, \partial f(x)) \in \partial f, \forall (y, \partial f(y)) \in \partial f$$

$$(iv)\ f \in \Gamma_\gamma(H) \quad i.e. \quad f - \gamma\|\cdot\|^2 \in \Gamma_0(H)\ .$$

Then f is said to be γ-strongly convex.

As a direct corollary of the equivalence (ii) $\Longleftrightarrow$ (iv) we obtain

COROLLARY 2.5 *Let us assume that $f \in \Gamma_\gamma(H)$ for some $\gamma > 0$. Then f reaches its minimum at a unique point x_f that satisfies*

$$f(x) \ge f(x_f) + \gamma\|x - x_f\|^2 \quad \forall x \in H \tag{2.4}$$

In the terminology of Definition 2.1, x_f is a γ-quadratic global minimizer of f. This leads to introduce the notion of a strong local minimum (see R.T. Rockafellar [39]):

DEFINITION 2.6 *Given $f: X \to \overline{\mathbf{R}}$, $x \in X$ is a local minimum of f in the strong sense if there exists $\gamma > 0$ such that for all y near x,*

$$f(y) \geq f(x) + \gamma \|y - x\|^2 \ . \tag{2.5}$$

A local minimum in the strong sense is a φ-minimum, see Definition 2.1 with $\varphi(r) = \gamma r^2$, $\gamma > 0$. When writing sufficient optimality condition for a local minimum in terms of second derivatives one is naturally led to the notion of strong local minimum. This explains the importance of this notion. Recent results of R.T. Rockafellar [41] allow us to characterize the best γ in (2.5) in terms of a lower bound for second derivatives, for a quite general class of functions f. A function $f: \mathbf{R}^k \to \overline{\mathbf{R}}$, a lower semicontinuous function, is *epi-differentiable* at x (see R.T. Rockafellar [40]) if

$$\Phi_{x,t}(\xi) = \frac{1}{t}\,[f(x + t\xi) - f(x)] \tag{2.6}$$

epi-converges as $t \downarrow 0$. The epi-limit is denoted by f'_x (with $f'_x(0) > -\infty$). f is said to be *twice epi-differentiable at x relatively to v* if it is epi-differentiable and the functions

$$\Psi_{x,v,t}(\xi) = \frac{2}{t^2}\,[f(x + t\xi) - f(x) - t\langle\xi, v\rangle] \tag{2.7}$$

epi-converges as $t \downarrow 0$. The epi-limit is denoted $f''_{x,v}$ (with $f''_{x,v}(0) > -\infty$). When f is twice epi-differentiable at x relative to every pseudo-gradient v then f is said to be *twice epi-differentiable at x.*

PROPOSITION 2.7 [41, Theorem 2.2] *Let $f: \mathbf{R}^k \to \overline{\mathbf{R}}$ be a lower semicontinuous function and x be a point where f is finite and twice epi-differentiable. If 0 is a subgradient of f at x and*

$$f''_{x,0}(\xi) > 0 \text{ for all } \xi \neq 0$$

then f has a local minimum in the strong sense, at x. Moreover, taking $\gamma_0 = \min_{|\xi|=1} f''_{x,0}(\xi)$ *one has*

$$f(y) \geq f(x) + \frac{1}{2}\gamma_0\|y - x\|^2 + o(\|y - x\|^2) \tag{2.8}$$

that is, for all $\gamma < \gamma_0/2$*, there exists* $\rho_\gamma > 0$ *such that*

$$f(y) \geq f(x) + \gamma\|y - x\|^2 \text{ for } all\ y \in B(x, \rho_\gamma) \ . \tag{2.9}$$

Let us end this section and examine how the preceding notions are connected with *conditioning theory*. We need the following definition, see C. Lemaire-Misonne [30] for an introduction to nonlinear conditioning.

DEFINITION 2.8 *Let* $f: X \to \overline{\mathbf{R}}$ *be a real extended valued function and* x_0 *a local minimizer of f, i.e., there exists some* $\rho > 0$ *such that*

$$f(x_0) \leq f(x) \quad \text{for } all\ \ x \in B(x_0, \rho) \ .$$

Let us assume that there is existence and uniqueness of such local minimizer for all linear perturbations of f (with sufficiently small norm): for all $v \in X^*$ *with* $\|v\| \leq \epsilon$ *there exists a unique* $x_v \in X$ *such that*

$$f(x_v) - \langle v, x_v \rangle \leq f(x) - \langle v, x \rangle \ \forall x \in B(x_v, \rho_v) \ . \tag{2.10}$$

The conditioning number of x_0 *relatively to linear perturbations of f is the positive real number defined by*

$$C_l(x_0; f) = \lim_{r \downarrow 0} \sup_{\|v\| \leq r} \frac{\|x_v - x_0\|_X}{\|v\|_{X^*}} \tag{2.11}$$

$$= \lim_{v \to 0} \sup \frac{\|x_v - x_0\|}{\|v\|} \ .$$

When f is uniformly convex one can obtain rather easily a sharp upper bound on $C_l(x_f;\, f)$:

PROPOSITION 2.8 *Let H be a real Hilbert space and $f \in \Gamma_\gamma(H)$ for some $\gamma > 0$ (see* Proposition 2.4*). Let us denote by x_f the unique minimizer of f,*

$$f(x_f) \leq f(x) \quad \text{for } all \;\; x \in X \;\; . \tag{2.12}$$

Then

$$C_l(x_f;\, f) \leq \frac{1}{2\gamma} \quad . \tag{2.13}$$

PROOF Denoting by ∂f the subgradient map of f, the optimality condition (2.12) means that

$$0 \in \partial f(x_f) \quad .$$

The optimality condition for the linearly perturbed problem (2.10) is

$$v \in \partial f(x_v) \quad .$$

From Proposition 2.4, in particular the equivalence (iv) $\Longleftrightarrow$ (iii),

$$<\partial f(x_v) - \partial f(x_f),\, x_v - x_f> \; \geq 2\gamma\|x_v - x_f\|^2 \quad ,$$

that is

$$<v,\, x_v - x_f> \; \geq 2\gamma\|x_v - x_f\|^2 \quad . \tag{2.14}$$

From the Cauchy-Schwarz inequality it follows that for every $v \in H$

$$\frac{\|x_v - x_f\|}{\|v\|} \leq \frac{1}{2\gamma}$$

and hence

$$C_l(x_f; f) \leq \frac{1}{2\gamma} \quad . \qquad \square$$

REMARK 2.9 Inequality (2.13) is sharp. Consider the case where the Cauchy-Schwarz inequality in (2.14) turns to be an equality, that is

$$\lambda(x_v - x_f) \in \partial f(x_v) \quad \text{for some } \lambda \in \mathbf{R} \quad ,$$

$x_v - x_f$ being a nonlinear eigenvector for the operator $\partial f(\cdot + x_f)$. (See Clément [17] for a detailed presentation of this theory).

The preceding (locally lipschitz) stability result takes only into account the linear perturbations of f. In the linear case (we mean by linear the case where $A = \partial f$ is a linear operator) the linear conditioning number C_l does not depend on the point x_f and is the same when considering general perturbations of A. In the nonlinear case, we shall see that the situation is far more delicate. A major obstacle in the development of general results, is to understand the meaning to attach to "(small) perturbations" of f, and "dependence" of x_f on f. This is the purpose of the next section.

3. THE EPI-DISTANCE

This notion was introduced in Attouch and Wets [7]. We made first use of it in the study of the pseudo-lipschitz stability properties of the ϵ-approximate solutions of convex optimization problems [8]. We recall its definition, and mention those results that will be useful in the sequel.

Unless otherwise specifically mentioned, we always denote by $(X, \|\cdot\|)$ a normed linear space, and by d the distance function generated by the norm. For any subset C of X,

$$d(x, C) := inf_{y \in C} \|x - y\|,$$

denotes the *distance from* x *to* C; if $C = \emptyset$ we set $d(x, C) = \infty$. For any $\rho \geq 0$, ρB denotes the ball of radius ρ and for any set C,

$$C_\rho := C \cap \rho B \ .$$

For C, $D \subset X$, the *"excess" function* of C on D is defined as,

$$e(C, D) := \sup_{x \in C} d(x, D),$$

with the (natural) convention that $e = 0$ if $C = \emptyset$. Note that the definition implies $e = \infty$ if C is nonempty and D is empty. For any $\rho \geq 0$, the *ρ-(Hausdorff-)distance* between C and D is given by

$$\text{haus}_\rho(C, D) = \sup \{ e(C_\rho, D), e(D_\rho, C) \}.$$

DEFINITION 3.1 *For $\rho \geq 0$, the ρ-*(Hausdorff-) epi-distance *between two extended real valued functions f, g defined on X, is*

$$\text{haus}_\rho(f, g) := \text{haus}_\rho(\text{epi}\, f, \text{epi}\, g),$$

where the unit ball of $X \times \mathbf{R}$ is the set $B := B_{X \times \mathbf{R}} = \{(x, \alpha) : \| x \| \leq 1, | \alpha | \leq 1\}$.

Convergence with respect to the family of epi-distances $\{\text{haus}_\rho, \rho > 0\}$ is closely related to epi-convergence, which in some sense, is the weakest form of convergence that will guarantee the convergence of the solutions of variational problems, cf. for example, Attouch [2], Rockafellar and Wets [42]. For the connections between this two types of convergences, we refer to [5, Section 4]. At this point, it suffices to know that when X is finite dimensional, the two types of convergences coincide. When X is infinite dimensional, we can show, at least in the convex case, that convergence with respect to the epi-distances implies epi-convergence with respect to both the weak and the strong topologies on X. Please consult [7, Section 4] for further details, and the relationship between convergence with respect to the (Hausdorff-) epi-distances and other pseudo-distances.

A very useful criterion, that allows us to compute or at least estimate, the epi-distance, is provided by the Kenmochi conditions.

THEOREM 3.2 [7, Theorem 2.1] *Suppose f, g are proper extended real valued functions defined on a normed linear space X, both minorized by $-\alpha_0 \|\cdot\|^p - \alpha_1$ for some $\alpha_0 \geq 0$, $\alpha_1 \in \mathbf{R}$ and $p \geq 1$. Let $\rho_0 > 0$ be such that $(\operatorname{epi} f)_{\rho_0}$ and $(\operatorname{epi} g)_{\rho_0}$ are nonempty.*

a) *Then the following conditions – to be called the* Kenmochi *conditions – hold: for all $\rho > \rho_0$ and $x \in \operatorname{dom} f$ such that $\|x\| \leq \rho$, $|f(x)| \leq \rho$, for every $\epsilon > 0$ there exists some $\tilde{x}_\epsilon \in \operatorname{dom} g$ that satisfies*

$$\begin{cases} \|x - \tilde{x}_\epsilon\| \leq \operatorname{haus}_\rho(f, g) + \epsilon \\ g(\tilde{x}_\epsilon) \leq f(x) + \operatorname{haus}_\rho(f, g) + \epsilon \end{cases} \tag{3.1}$$

as well as a symmetric condition with the role of f and g interchanged.

b) *Conversely, assuming that for all $\rho > \rho_0 > 0$ there exists a "constant" $\eta(\rho) \in \mathbf{R}_+$, depending on ρ, such that for all $x \in \operatorname{dom} f$ with $\|x\| \leq \rho$, $|f(x)| \leq \rho$, there exists $\tilde{x} \in \operatorname{dom} g$ that satisfies*

$$\begin{cases} \|x - \tilde{x}\| \leq \eta(\rho) \ , \\ g(\tilde{x}) \leq f(x) + \eta(\rho) \ , \end{cases} \tag{3.2}$$

and the symmetric condition (interchanging f and g), then with $\rho_1 := \rho + \alpha_0 \rho^p + \alpha_1$.

$$\operatorname{haus}_\rho(f, g) \leq \eta(\rho_1) \, . \tag{3.3}$$

Let us conclude by observing that there are many other ways to define metrics on the space of extended real-valued functions that induce epi-convergence. In fact, in view of Theorems 3.4, 3.7 and 3.9 of [7], we know a number of them that are equivalent to that generated by the family of pseudo-distances $\{\text{haus}_\rho,\ \rho > 0\}$. We state our results in terms of the epi-distance, because in many applications it is easier to handle, and possibly easier to "visualize".

4. STABILITY RESULTS

We now turn to the main result that implies lipschitzian – when the problem is (sub)linearly conditioned (see Figure 1, Section 2) –, and more generally hölderian stability of the solutions with respect to perturbations measured in terms of the epi-distance. To state our result, it is convenient to use the following notation: to any pair $(\hat{x}, \hat{\alpha})$ we can associate a ***translation map*** τ such that for any function h

$$\tau h(x) := h(x + \hat{x}) - \hat{\alpha} \ . \tag{4.1}$$

Then epi $\tau h = \operatorname{epi} h + \{-\hat{x}, -\hat{\alpha}\}$. If $\hat{x}$ is a local minimum of h and $h(\hat{x}) = \hat{\alpha}$, then the function τh has a local minimum at 0 with $\tau h(0) = 0$. By $\varphi_{[1]}$ we denote the *epigraphical sum* $\varphi \underset{e}{+} |\cdot|$ of φ with the norm, i.e.,

$$\varphi_{[1]}(t) = \inf\{\varphi(s) + |t - s| : s \in \mathbf{R}\}$$

STABILITY THEOREM 4.1 *Let X be a normed linear space, f and g two proper extended real valued functions defined on X. Suppose that x_f is a φ-local minimizer of f, i.e.,*

$$f(x) \geq f(x_f) + \varphi(\|x - x_f\|) \ \ \forall x \in B(x_f, \rho) \tag{4.2}$$

for some $\rho > 0$ with φ an admissible function, and x_g is a local minimizer of g with respect *to a ball of the same radius,*

$$g(x) \geq g(x_g) \quad \forall x \in B(x_g, \rho) \ . \tag{4.3}$$

Let τ be the translation map associated with the pair $(x_f, f(x_f))$. Suppose also that the function g is close enough to f so that

$$\rho > \frac{3}{2} \text{ haus}_\rho(\tau f, \tau g) \ , \tag{4.4}$$

and

$$\rho \geq 3 \sup \{ \|x_f - x_g\|, |f(x_f) - g(x_g)| \} \ . \tag{4.5}$$

Then

$$|f(x_f) - g(x_g)| \leq \text{haus}_\rho(\tau f, \tau g) \ , \tag{4.6}$$

and

$$\varphi_{[1]}(\|x_f - x_g\|) \leq 4 \text{ haus}_\rho(\tau f, \tau g) \ . \tag{4.7}$$

Moreover, if φ is a finite valued, convex, even function, and if the right-derivative of φ, denoted φ'_+, satisfies

$$\varphi'_+(\|x_f - x_g\|) \leq 1$$

then

$$\|x_f - x_g\| \leq \varphi^{-1}(4\text{haus}_\rho(\tau f, \tau g)) \ . \tag{4.8}$$

PROOF Let us first prove (4.6). The optimality conditions could equivalently be stated as

$$\tau f(x) \geq \varphi(\|x\|), \qquad \forall x \in B(0, \rho) \ ,$$

$$\tau g(x) \geq g(x_g) - f(x_f) = \tau g(x_g - x_f), \qquad \forall x \in B(x_g - x_f, \rho) \ .$$

Observe that assumption (4.5) implies that

$$(y, \beta) := (x_g - x_f, g(x_g) - f(x_f)) \in B_{X \times \mathbf{R}}(0, \rho) =: \rho B \ ,$$

and hence

$$(y, \beta) \in (\operatorname{epi} \tau g)_\rho, (0, 0) \in (\operatorname{epi} \tau f)_\rho \ ;$$

Figure 3 illustrates the situation.

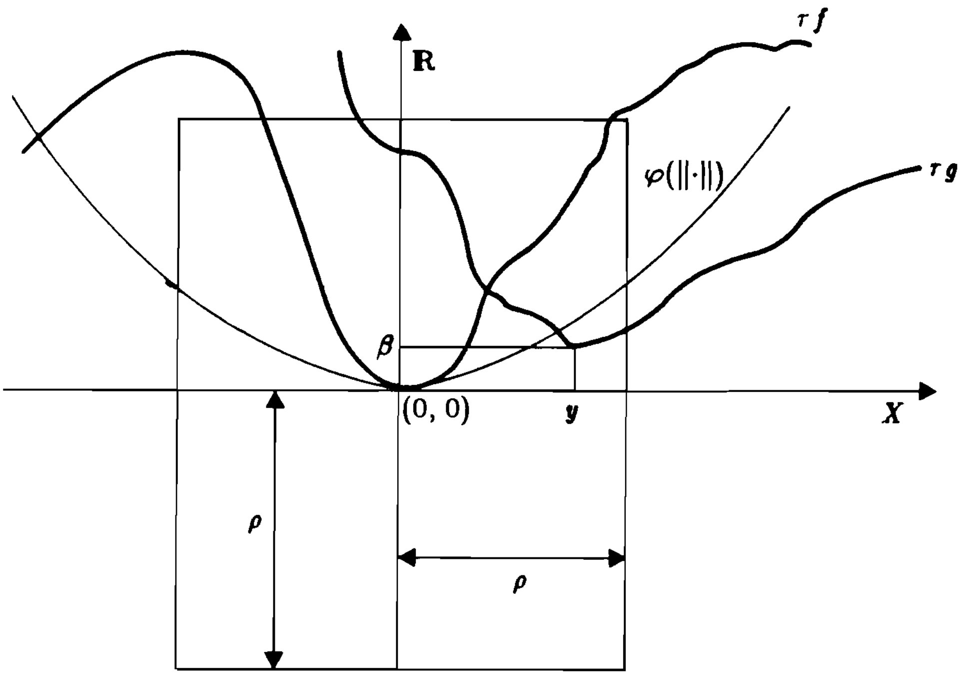

FIGURE 3 τf, τg and $\varphi(\|\cdot\|)$.

Let $\eta := \operatorname{haus}_\rho(\tau f, \tau g)$. By definition of haus_ρ, we have

$$\eta \geq d((0,0), \operatorname{epi} \tau g) \ ,$$

$$\eta \geq d((y, \beta), \operatorname{epi} \tau f) \ .$$

Next, let us observe that

$$d((0,0), \operatorname{epi} \tau g) = d((0,0), (\operatorname{epi} \tau g)_{2\rho/3}) \ , \tag{4.9}$$

$$d((y,\beta), \operatorname{epi} \tau f) = d((y,\beta), (\operatorname{epi} \tau f)\rho) \tag{4.10}$$

These identities are justified by the following argument. For any $(x, \alpha) \notin \rho B$,

$$\begin{aligned} \|(y,\beta) - (x,\alpha)\| &= \max\left[\|(x_g - x_f) - x\|, |g(x_g) - f(x_f) - \alpha|\right] \\ &\geq \max\left[\|x\| - \rho/3, |\alpha| - \rho/3\right] \\ &\geq \|(x,\alpha)\| - \rho/3 \geq 2\rho/3 \ . \end{aligned}$$

Where (4.5) was used to obtain the second inequality. Thus,

$$d((y,\beta), \operatorname{epi} \tau f \cap (X \times \mathbf{R} \backslash \rho B)) \geq 2\rho/3 \ .$$

But, as follows from (4.4)

$$d((y,\beta), \operatorname{epi} \tau f) \leq \eta < 2\rho/3 \ ,$$

and this confirms (4.10). The same argument yields (4.9). In particular this implies that

$$d((0,0), (\operatorname{epi} \tau g)_{2\rho/3}) \leq \eta \ . \tag{4.11}$$

and

$$d((y,\beta), (\operatorname{epi} \tau f)_\rho) \leq \eta \ . \tag{4.12}$$

If $\beta < 0$, noticing that $\tau f \geq 0$ on $B(0,\rho)$, it follows from (4.12) that $|\beta| \leq \eta$. If $\beta \geq 0$ the other inequality (4.11) implies too that $|\beta| \leq \eta$, since $\alpha \geq \beta$ for any $(y,\alpha) \in (\operatorname{epi} \tau g)_{2\rho/3}$. This completes the proof of (4.6).

From (4.10), it follows that $\eta \geq d((y, \beta), (\operatorname{epi} \tau f)_\rho)$. In turn this yields

$$\eta \geq d((y, \beta), \operatorname{epi} \varphi(\|\cdot\|)_\rho)$$

since $\varphi(\|\cdot\|) \leq \tau f$ on ρB. Repeating once more the argument that gave us (4.10), we see that

$$d((y, \beta), \operatorname{epi} \varphi(\|\cdot\|)) = d((y, \beta), \operatorname{epi} \varphi(\|\cdot\|)_\rho) \quad ,$$

and thus $d((y, \beta), \operatorname{epi} \varphi(\|\cdot\|)) \leq \eta$. Now from the triangle inequality and (4.6)

$$d((y, 0), \operatorname{epi} \varphi(\|\cdot\|)) \leq d((y, \beta), \operatorname{epi} \varphi(\|\cdot\|)) + |\beta| \leq 2\eta \quad . \tag{4.13}$$

Next, we calculate an upper bound for $d((y, 0), \operatorname{epi} \varphi(\|\cdot\|))$:

$$\begin{aligned} d((y, 0), \operatorname{epi} \varphi(\|\cdot\|) &= \inf_x \{\sup [\|y - x\|, \alpha] : \alpha \geq \varphi(\|x\|)\} \quad , \\ &\geq \inf_x \{\sup [\|y - x\|, \varphi(\|x\|)]\} \\ &\geq \frac{1}{2} \inf\nolimits_{x \in X} \{\|y - x\| + \varphi(\|x\|)\} \\ &\geq \frac{1}{2} \inf\nolimits_{x \in X} \{| \, \|y\| - \|x\| \, | + \varphi(\|x\|)\} \\ &= \frac{1}{2} \varphi_{[1]}(\|y\|) \quad . \end{aligned}$$

This, with (4.13) yields (4.7), i.e., $\varphi_{[1]}(\|y\|) \leq 4\eta$.

In order to obtain (4 :.8), we use a result of Hiriart-Urruty [25] which shows that $\varphi_{[1]} = \varphi$ whenever $\varphi'_+ \leq 1$. Thus in that case, $\varphi(\|y\|) \leq 4\eta$, from which (4.8) follows. $\square$

REMARK 4.2

a) We stress the fact that in Theorem 4.1, we only *assume that one of the two functions f or g is well-conditioned*, say f. In particular that means that f has a unique local minimizer x_f. When perturbing (or approximating) f, let us call g the perturbed (or approximating) function, it could happen that the local minimization of g has many solutions. Theorem 4.1 tells us that

$$\operatorname{argmin} g \subset B(x_f, \varphi_f^{-1}\eta) \quad ,$$

where φ_f denotes the radial regularization of f at x_f, and $\eta = \text{haus}_\rho(f, g)$ with ρ large enough so that all the significant elements are contained in $B((x_f, f(x_f)), \rho)$.

b) When f is well-conditioned at x_f, and g at x_g, let

$$f(x) \geq f(x_f) + \varphi_f(\|x - x_f\|) \quad \forall x \in B(x_f, \rho) \quad ,$$

$$g(x) \geq g(x_g) + \varphi_g(\|x - x_g\|) \quad \forall x \in B(x_g, \rho) \quad .$$

Let us assume that both φ_f and φ_g are convex, finite-valued. From (4.8) we obtain

$$(\varphi_f \vee \varphi_g)(\|x_f - x_g\|) \leq 4 \, \text{haus}_\rho(f, g) \quad ,$$

provided $(\varphi'_f \vee \varphi'_g)(\|x_f - x_g\|) \leq 1$. Thus the larger of the two functions prevails.

c) The best stability estimates in (4.7) and (4.8) are, of course, obtained by taking φ as large as possible, i.e., by choosing for φ the radial regularization, see Proposition 2.2.

To conclude this section, we state as corollary of Theorem 4.1, a version of that Theorem that is useful in many applications.

COROLLARY 4.2 *Let X be a normed linear space, f and g two proper, extended real valued functions defined on X. Suppose x_f is a local minimizer of f so that*

$$f(x) \geq f(x_f) + \gamma \|x - x_f\|^p \quad \forall x \in B(x_f, \rho) \ ,$$

for some $p \geq 1$, $\gamma > 0$ and $\rho > 0$, and x_g is a local minimizer of g, so that

$$g(x) \geq g(x_g) \quad \forall x \in B(x_g, \rho) \ .$$

Suppose moreover that

$$\rho \geq 3 \sup \{\|x_f - x_g\|, |f(x_f) - g(x_g)|\}$$

and, with τ the translation map associated with $(x_f, f(x_f))$,

$$\rho \geq \frac{3}{2} \operatorname{haus}_\rho(\tau f, \tau g)$$

Then, we have the following estimate:

$$\|x_f - x_g\| \leq \left[\frac{4}{\gamma} \operatorname{haus}_\rho(\tau f, \tau g)\right]^{1/p}$$

provided $\|x_f - x_g\| \leq (\gamma p)^{-1/p-1}$.

Are these estimates optimal? The following examples show that they are sharp, unless additional assumptions enter into play.

5. EXAMPLES

EXAMPLE 5.1 Let us start with the following elementary example. Take $X = \mathbf{R}$, $f = \theta^{p-1}|\cdot|$ and $g = \theta^{p-1}|\cdot - \theta|$, and define

$$f_\theta(x) = \theta^{p-1}|x| + \frac{1}{p}|x|^p$$

$$g_\theta(x) = \theta^{p-1}|x - \theta| + \frac{1}{p}|x|^p$$

where $p \in [1, +\infty)$ and θ is a positive parameter that will go to zero.

Clearly f_θ and g_θ are two convex continuous functions that achieve their minimum respectively at

$$x_f(\theta) = 0$$

and

$$x_g(\theta) = \theta$$

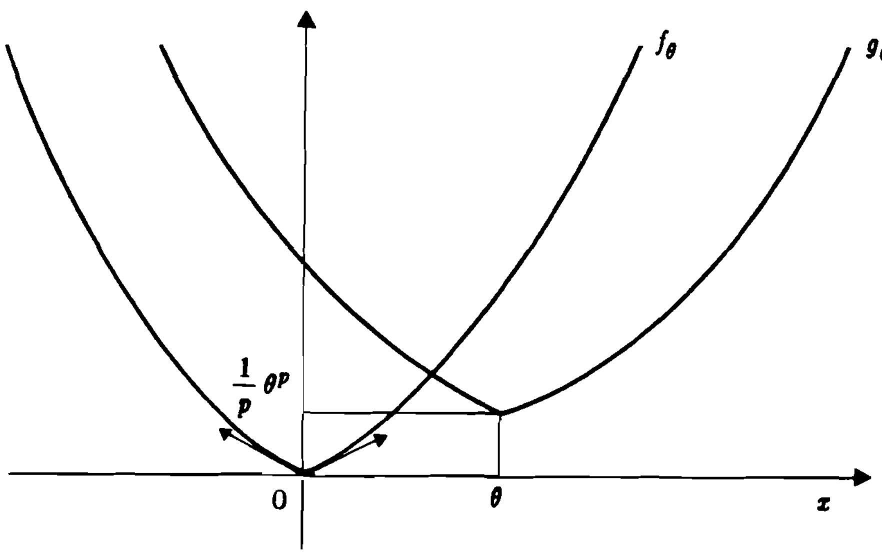

FIGURE 4

Hence $|x_f(\theta) - x_g(\theta)| = \theta$, while $|f_\theta(x) - g_\theta(x)| = \theta^{p-1}|(|x| - |x - \theta|)|$, and thus

$$\sup_x |f_\theta(x) - g_\theta(x)| = \theta^p \quad .$$

We have exactly

$$|x_f(\theta) - x_g(\theta)| = d(f_\theta, g_\theta)^{1/p} \quad , \tag{5.1}$$

the distance being computed with the uniform norm on $\mathbf{R}$. (Notice that $\mathrm{haus}_\rho(f_\theta, g_\theta) \leq d(f, g) \leq (1 + \theta^{p-1} + \rho^{p-1})\,\mathrm{haus}_\rho(f_\theta, g_\theta)$ for all $\theta \geq 0$, and $\rho \geq 0$). Let us interpret this result with the help of Theorem 4.1: $x_f(\theta) = 0$ is a φ-minimum of f_θ with $\varphi(r) = (1/p)r^p$. Indeed the largest admissible function φ (independent of θ) such that

$$f_\theta(x) \geq f_\theta(0) + \varphi(|x|) \text{ is } \varphi(r) = \frac{1}{p}|r|^p \ ! \qquad \square$$

EXAMPLE 5.2 Let us now examine the projection on a convex set. Let X be a reflexive Banach space, C a closed convex non empty subset of X and $x_0 \in X$. The minimization problem

$$\min\{\|x_0 - x\| : x \in C\} \tag{5.2}$$

has a unique solution $p_C(x_0)$ which is the projection of x_0 on C. The minimization problem (5.2) can be rewritten as

$$\min\{f(x) : x \in X\}$$

with δ_C the indicator function of the set C,

$$f(x) = \|x_0 - x\| + \delta_C(x) \ . \tag{5.3}$$

Let us examine the stability of $p_C(x_0)$ with respect to x_0 and C. We first assume X to be a real Hilbert space: it is a well known result that $x_0 \mapsto p_C(x_0)$ is a contraction. Let us now study the mapping $C \mapsto p_C(x_0)$ and prove the following

PROPOSITION 5.3 *Let C and D two closed convex non empty subsets of an Hilbert space H. Given $x_0 \in H$ and*

$$\rho = \|x_0\| + d(x_0, C) + d(x_0, D) \ , \tag{5.4}$$

we have the following estimation:

$$\|p_C(x_0) - p_D(x_0)\| \leq \rho^{1/2} \text{haus}_\rho(C, D)^{1/2} \ . \tag{5.5}$$

PROOF Let us write the classical optimality conditions

$$<x_0 - p_C(x_0),\, z - p_C(x_0)> \leq 0 \quad \forall z \in C \tag{5.6}$$

$$<x_0 - p_D(x_0),\, y - p_D(x_0)> \leq 0 \quad \forall y \in D \tag{5.7}$$

that characterize $p_C(x_0)$ and $p_D(x_0)$. From

$$p_C(x_0) \leq \|x_0\| + d(x_0, C) \leq \rho$$

$$p_D(x_0) \leq \|x_0\| + d(x_0, D) \leq \rho$$

there exists some $\tilde{z} \in C$ such that

$$\|p_D(x_0) - \tilde{z}\| = d(p_D(x_0), C) \leq e(D_\rho, C)$$

and some $\tilde{y} \in D$ such that

$$\|p_C(x_0) - \tilde{y}\| = d(p_C(x_0), D) \leq e(C_\rho, D)$$

which by definition of $\mathrm{haus}\,\rho(C, D)$ yields

$$\sup\{\|p_C(x_0) - \tilde{y}\|;\, \|p_D(x_0) - \tilde{z}\|\} \leq \mathrm{haus}_\rho(C, D) \ . \tag{5.8}$$

Take $z = \tilde{z}$ in (5.6), $y = \tilde{y}$ in (5.7) and add the two inequalities. We obtain

$$<x_0 - p_C(x_0),\, \tilde{z} - p_D(x_0) + p_D(x_0) - p_C(x_0)> +$$

$$+ <x_0 - p_D(x_0),\, \tilde{y} - p_C(x_0) + p_C(x_0) - p_D(x_0)> \leq 0 \ ,$$

that is

$$\|p_C(x_0) - p_D(x_0)\|^2 \leq \max\{\|p_D(x_0) - \tilde{z}\|;\, \|p_C(x_0) - \tilde{y}\|\}(d(x_0, C) + d(x_0, D)) \ .$$

Using inequality (5.8), we finally obtain

$$\|p_C(x_0) - p_D(x_0)\|^2 \leq \rho \cdot h_\rho(C, D) \ . \qquad \square$$

We now turn to the following questions. Is the hölder exponent 1/2 optimal? How is this exponent related to the geometry of the space? And how to interpret this result with the help of Theorem 4.1?

EXAMPLE 5.4 Let us first examine the question of optimality. Take $X = \mathbf{R}^2$ equipped with its euclidian structure and consider the following picture:

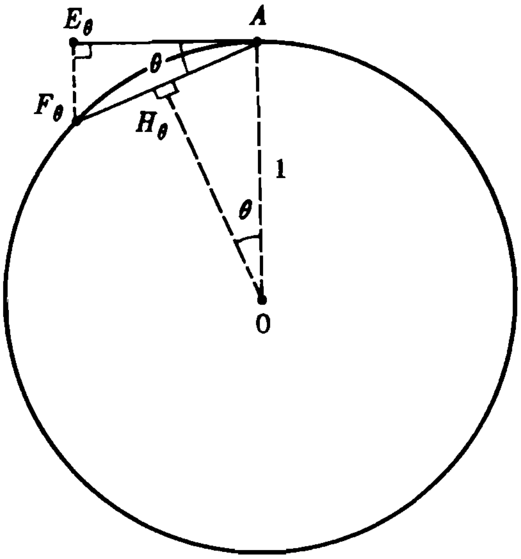

FIGURE 5 Projection on a convex set.

Take $C_\theta = [AE_\theta]$ and $D_\theta = [AF_\theta]$ as convex sets depending on the parameter θ and $x_0 = 0$. Then $p_{C_\theta}(0) = A$, $p_{D_\theta}(0) = H_\theta$, i.e., $\|p_{C_\theta}(0) - p_{D_\theta}(0)\| = \sin\theta$. On the other hand

$$\text{haus}\,(C_\theta,\, D_\theta) = d(E_\theta,\, F_\theta) = 2\,\sin^2\theta$$

that is

$$\|p_{C_\theta}(0) - p_{C_\theta}(0)\| = \frac{1}{\sqrt{2}}\,\text{haus}\,(C_\theta,\, D_\theta)^{1/2} \ .$$

For this example one has no better than the 1/2 hölder continuity provided by Proposition 5.1 for the map $C \mapsto \mathrm{proj}_C x_0$.

Let us now explain how the 1/2-hölder continuity result can be derived from Theorem 4.1 and how it is related to the hilbertian structure. At this point, we need a general version of Proposition 2.4 concerning uniformly convex functions, cf. to the recent survey of Azé [14].

PROPOSITION 5.5 *Let X be a Banach space and $f \in \Gamma_0(X)$, the space of extended real-valued, proper, lower semicontinuous, convex functions. Let us consider the following statements:*

(i) $\forall x_0, x_1 \in \mathrm{dom}\, f \;\forall t \in (0, 1), \qquad f(x_t) \le t f(x_1) + (1-t) f(x_0) - t(1-t) k(\|x_1 - x_0\|)$ *where* $x_t = t x_1 + (1-t) x_0,\ k(s) \ge 0,\ k(0) = 0$.

(ii) $\forall (x_0, y_0) \in \partial f \;\forall x \in X \;\; f(x) \ge f(x_0) + \langle y_0, x - x_0\rangle + w(\|x - x_0\|)$ *where* $w(s) \ge 0$, $w(0) = 0$.

(iii) $\forall (x_0, y_0) \in \partial f,\ \forall (x_1, y_1) \in \partial f \;\; \langle y_1 - y_0, x_1 - x_0\rangle \ge W(\|x_1 - x_0\|)$ *where* $W(s) \ge 0$, $W(0) = 0$.

Then (i) $\Longrightarrow$ (ii) *with* $w = k$,

(ii) $\Longrightarrow$ (iii) *with* $W = 2\omega$,

(iii) $\Longrightarrow$ (i) *with* $w(r) = \int_0^r (W(s)/s)\, ds$ *if X is reflexive,*

(ii) $\Longrightarrow$ (i) *with* $k(r) = 2w(r/2)$ *if $\|\cdot\|^2$ is uniformly convex.*

In order to derive Proposition 5.1 from Theorem 4.1, we notice that $p_C(x_0)$ is solution of the following minimization problem

$$\min \{\delta_C(x) + \|x_0 - x\|^p : x \in H\}$$

for every $p \in [1, +\infty)$, with $\delta_C(\cdot)$ we denote the indicator function of C, i.e., $\delta_C(x) = \infty$ if $x \notin C$, and $\delta_C = 0$ on C. Here,

$$f(x) = \delta_C(x) + \|x_0 - x\|^p \ . \tag{5.9}$$

We then note that in a Hilbert space the function $x \mapsto \|x\|^p$ is uniformly convex as soon as $p \geq 2$. Indeed, with $A := \partial(1/p\|\cdot\|^p)$, we have

$$<Ax - Ay, x - y> \geq \frac{1}{2^{p-2}} \|x - y\|^p \quad \forall x, y \in H \ . \tag{5.10}$$

Hence, from Proposition 5.5, (iii) $\Longrightarrow$ (ii),

$$f(x) \geq f(p_C(x_0)) + \frac{1}{p \cdot 2^{p-2}} \|x - p_C(x_0)\|^p, \quad \forall x \in H \ . \tag{5.11}$$

From Theorem 4.1, it follows that $f \to p_C(x_0)$ and hence $C \to p_C(x_0)$ is $1/p$-hölder continuous for any $p \geq 2$. Clearly the sharpest result is obtained by taking $p = 2$ which provides $1/2$-hölder continuity of $C \to p_C(x_0)$.

We are now able to explain *how this exponent* $1/2$ *is related to the hilbertian structure*. Let us assume we work in a space of type p: like $(\mathbf{R}^m, \|\cdot\|_p)$ where $\|x\|_p = (\sum_{i=1}^m |x_i|^p)^{1/p}$, or $l^p(\mathbf{N})$, $L^p(\Omega)$, $W^{m,p}(\Omega), \cdots$. These spaces are characterized by the fact that the Clarkson's inequality holds (see R.A. Adams [1], Theorem 2.28):

$$\text{if } p \geq 2 \ \left\|\frac{u+v}{2}\right\|_p^p \leq \frac{1}{2}\|u\|_p^p + \frac{1}{2}\|v\|_p^p - \frac{1}{2^p}\|v - u\|_p^p \quad \forall u, v \in X \ ,$$

$$\text{if } p \leq 2 \ \left\|\frac{u+v}{2}\right\|_p^{p'} \leq \frac{1}{2}\|u\|_p^{p'} + \frac{1}{2}\|v\|_p^{p'} - \frac{1}{2^{p'}}\|v - u\|_p^{p'} \quad \forall u, v \in X$$

where p' is the conjugate exponent of p, $p' = (p/p - 1)$.

Thus f given by (5.9) satisfies the same type of inequality. It follows from Proposition 5.5 (i) $\Longrightarrow$ (ii), that $p_C(x_0)$ is a φ-minimizer of f with

$$\varphi(r) = \frac{1}{2^{p-2}} r^p \ \text{ and } \|\cdot\|_p \ \text{ if } p \geq 2$$

and

$$\varphi(r) = \frac{1}{2^{p'-2}} r^{p'} \text{ and } \|\cdot\|_p \text{ if } p \leq 2 \ .$$

Let us summarize this in the following proposition.

PROPOSITION 5.6 *Let X be a Banach space of type p (say $L^p(\Omega)$, $l^p(\mathbb{N}),\cdots$) with $1 < p < +\infty$. Let $x_0 \in X$ and C a closed convex set in X. Then the mapping $C \mapsto p_C(x_0)$ is hölder continuous with exponent $1/p$ if $p \geq 2$, and with exponent $1/p'$, if $p \leq 2$.*

Figure 6 shows the variation of the hölder exponent as a function of p, $(X = L^p,\cdots)$.

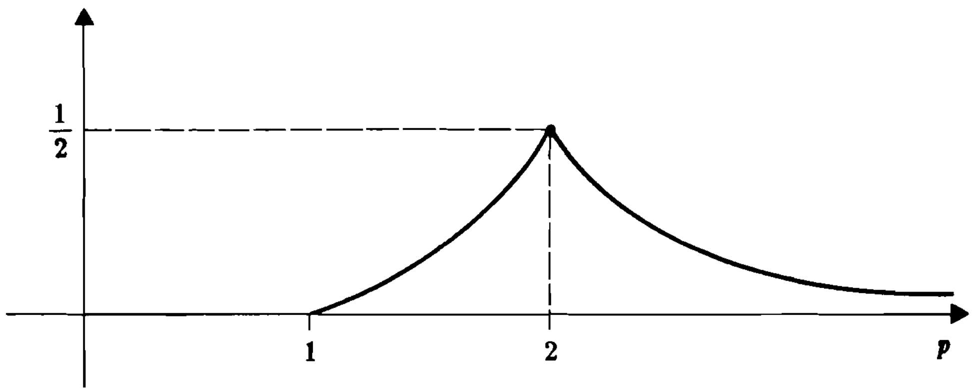

FIGURE 6

It is for the hilbertian structure that we have the best stability result. So the hilbertian metric is well adapted to approximation theory. On the opposite when $p \to 1$ or $p \to +\infty$ the hölder exponent goes to zero. Indeed in a Banach space of type l^1, $L^1\cdots$ the solution of the minimization problem (5.2) may fail to exist or, because of the lack of strict convexity of the norm, it may not be reduced to a singleton.

Let consider the case $X = \mathbb{R}^m$ equipped with the l_1-norm $\|x\|_1 = \sum_{i=1}^m |x_i|$. Then $p_C(x_0)$ is a nonempty convex set. One may conjecture, that when $C_\nu \to C$ for the hausdorff metric then $p_{C_\nu}(x_0) \to p_C(x_0)$ for the hausdorff metric. This is false as shown by the following counter example:

Take $X = \mathbf{R}^2$, $\|x\| = |x_1| + |x_2|$, $x_0 = 0$, $C = \{\lambda(0, 1) + (1 - \lambda)(1, 0);\ 0 \leq \lambda \leq 1\}$, and $C_n = \{\lambda(0, 1) + (1 - \lambda)(1 + (1/n), 0); 0 \leq \lambda \leq 1\}$

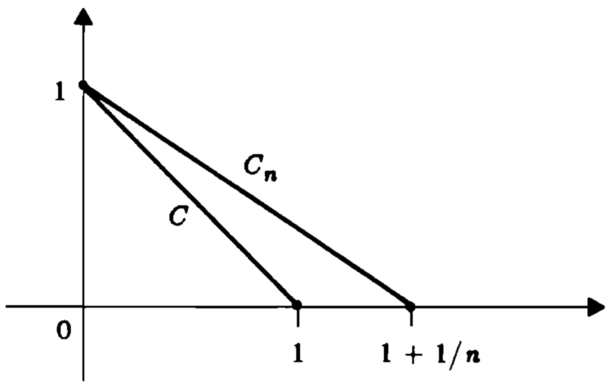

FIGURE 7

Indeed when working in non reflexive Banach spaces, the good notion that still enjoys stability properties is the notion of ϵ-solution. It is proved in Attouch & Wets [8] that the mapping

$$f \to \epsilon\text{-argmin}\, f \quad (\epsilon > 0)$$

is lipschitz, when the space of convex functions is equipped with the epi-distance, and the distance between the ϵ-solutions sets, ϵ-argmin f, is measured in terms of the (hausdorff-) ρ-distance.

6. APPLICATION TO CONVEX PROGRAMMING AND PENALIZATION

The purpose of this section is to suggest the arguments that could be used to exploit the Stability Theorem. Let us begin with the case when we are approximating a convex programming problem $(\mathcal{P}_f)$:

$$\text{minimize } f_0(x)$$

subject to $f_i(x) \leq 0, \quad i = 1, \ldots, m$, (6.1)

by another convex program $(\mathcal{P}_g)$:

minimize $g_0(x)$

subject to $g_i(x) \leq 0, \quad i = 1, \ldots, m$,

where the $\{f_i,\, i = 0, \ldots, m\}$ and $\{g_i,\, i = 0, \ldots, m\}$ are proper, lower semicontinuous, convex functions defined on a Banach space X with values in $\mathbf{R} \cup \{\infty\}$. If we denote by

$$C_f := \{x | f_i(x) \leq 0,\, i = 1, \ldots, m\}, \text{ and } C_g := \{x | g_i(x) \leq 0,\, i = 1, \ldots, m\} \quad ,$$

and

$$f := f_0 + \delta_{C_f}, \text{ and } g := g_0 + \delta_{C_g} \quad ,$$

we can apply the Stability Theorem. Our formulation allows for perturbations that affect f_0, as well as some or all of the constraints. We are interested in their global effect on $\mathcal{P}_f$, in particular in how they affect the optimal solutions. Since $f = f_0 + \delta_{C_f}$, we are led to the study of the stability of a sum, a problem that has received a lot of attention in a topological setting, cf. for example Attouch, Azé and Wets [3, Theorem 4.1]. A standard feature of such results is the need for some constraint qualification. Recently Azé and Penot [15] obtained a version of these results that provide estimates in terms of the epi-distance.

PROPOSITION 6.1 [15, Corollary 2.8] *Let X be a Banach space, $(\varphi_i,\, i = 1, \ldots, n)$ and $(\psi_i,\, i = 1, \ldots, n)$ proper, lower semicontinuous, convex functions defined on X with values in $\mathbf{R} \cup \{\infty\}$. Assume that these functions are minorized by $-\alpha(\|\cdot\| + 1)$ for some $\alpha \geq 0$, and*

$$(\sigma B) \subset \operatorname{diag} X^n \cap (\gamma B)^n - \prod_{i=1}^{n} (\operatorname{lev}_\gamma \varphi_i)_\gamma$$

for some $\gamma \geq 0$, $\sigma > 0$, *where*

$$\operatorname{diag} X^n := \operatorname{diag} \prod_{i=1}^{n} X_{(i)}, \text{ with each } X_{(i)} = X \ ,$$

$$\operatorname{lev}_\gamma \varphi_i := \{x : \varphi_i(x) \leq \gamma\} \ .$$

Then, for all $\rho \geq n\gamma + \sigma$, *assuming that* $\sum_{i=1}^{n} \operatorname{haus}_{\rho_1}(\varphi_i, \psi_i) < \sigma$,

$$\operatorname{haus}_\rho \left[\sum_{i=1}^{n} \varphi_i, \sum_{i=1}^{n} \psi_i \right] \leq \frac{n\gamma + \sigma + \rho}{\sigma} \sum_{i=1}^{n} \operatorname{haus}_{\rho_1}(\varphi_i, \psi_i) \ ,$$

where $\rho_1 = \rho + (n + 1)[\alpha(\rho + \sigma + 1) + \sigma]$.

To apply this to our situation, we take $\varphi_0 := f_0$, $(\varphi_i := \delta_{\{f_i \leq 0\}}, i = 1, \ldots, m)$, and similarly $\psi_0 = g_0$, $(\psi_i = \delta_{\{g_i \leq 0\}}\ i = 1, \ldots, m)$. We use the upper estimate for

$$\operatorname{haus}_\rho(\varphi_i, \psi_i) = \operatorname{haus}_\rho(\delta_{\operatorname{lev}_0 f_i}, \delta_{\operatorname{lev}_0 g_i}) \ .$$

Assuming that $\inf f_i < 0$, and $\inf g_i < 0$, a result in Attouch and Wets [8] allows us to estimate the epi-distance between the level-sets:

$$\operatorname{haus}_\rho(\varphi_i, \Psi_i) \leq \eta \operatorname{haus}_{\rho_2}(f_i, g_i)$$

where

$$\rho_2 = \sup\{\rho, |\inf f_i| + 1, |\inf g_i| + 1\}$$

$$\eta = 1 + \frac{2\rho + \operatorname{haus}_{\rho_2}(f_i, g_i)}{\operatorname{haus}_{\rho_2}(f_i, g_i) - \sup\{\inf f_i, \inf g_i\}} \cdot$$

Now combining this with Proposition 6.1, we obtain the following

PROPOSITION 6.2 *Consider problems (P_f) and (P_g) as defined above, such that the functions are minorized by* $-\alpha(\|\cdot\|^p + 1)$ *where* $\alpha \geq 0$, $p \geq 1$. *Suppose that*

(i) **constraint qualification:** *for some* $\gamma \geq 0$, $\sigma > 0$,

$$(\sigma B) \subset \operatorname{diag} X^{m+1} \cap (\gamma B)^{m+1} - \prod_{i=1}^{m} (\mathrm{lev}_0 f_i)_\gamma \times (\mathrm{lev}_\gamma f_0)_\gamma$$

and $(\inf f_i) < 0$, $(\inf g_i) < 0$, *for* $i = 1, \ldots, m$,

(ii) **strong convexity of f_0:** *for all* x_0, $x_1 \in \operatorname{dom} f_0$, $\lambda \in [0, 1]$

$$f_0(x^\lambda) \leq (1 - \lambda) f_0(x^0) + \lambda f_0(x^1) - \lambda(1 - \lambda) w(\|x^0 - x^1\|)$$

for w, *a convex finite valued function, and* $x^\lambda = (1 - \lambda)x^0 + \lambda x^1$.

Then, there exist $\rho_0 \geq 0$ *and* $\eta > 0$ *such that whenever*

$$\sup_{0 \leq i \leq m} \mathrm{haus}_\rho(f_i, g_i) \leq \eta \ ,$$

for all $\rho > \rho_0$,

$$w(\|x_f - x_g\|) \leq \kappa \sup_{0 \leq i \leq m} \mathrm{haus}_\gamma(f_i, g_i)$$

with γ *and* κ *depending (boundedly) on* ρ.

For our next example, let us consider approximations based on penalization. One of the implications of the next proposition, is that in general one should not expect good convergence rates for numerical methods based on penalization. Let f_0 be a locally lipschitz, finite valued function to be minimized over a nonempty subset $C \subset X$. We approximate the minimization problem

find x that minimizes $f(x) := f_0(x) + \delta_C(x)$ on X, by a problem of the type

find x that minimizes $f_\theta(x) := f_0(x) + \varphi_\theta(x)$ on X

where $\{\varphi_\theta : X \to \mathbf{R}_+, \theta \geq 0\}$ is a cast of functions such that

(i) $\varphi_\theta = 0$ on C

(ii) $\varphi_\theta(x) \geq \alpha\theta[d(x, C)]^p$ for all x in X,

where $p \geq 1$ and $\alpha > 0$. We think of θ as a parameter that tends to ∞.

PROPOSITION 6.3 *Let X be a normed linear space, with f and f_θ as defined above. Let*

$$L(\rho) := \sup\{|f_0(x) - f_0(y)| \cdot \|x - y\|^{-1} : \|x\| \leq \rho, \|y\| \leq \rho\}$$

$$M(\rho) := \sup\{|f_0(x)| : \|x\| \leq \rho\} \; .$$

For any $\rho > 0$, we have that

$$\text{haus}_\rho(f, f_\theta) \leq \gamma\theta^{-1/p} \tag{6.1}$$

where $\gamma = \gamma(\rho)$ is defined in the proof. Moreover, if $x_f \in$ argmin f and

$$f_0(y) \geq f_0(x_f) + w(\|y - x\|), \quad \forall y \in C$$

with w a convex, finite valued, function, then for θ sufficiently large,

$$w(\|x_f - x_\theta\|) \leq 4\gamma(\rho_1)\theta^{-1/p}$$

for all $x_\theta \in$ argmin f_θ, and some $\rho_1 \geq 0$.

PROOF Since $f_\theta \leq f$, with the excess function as defined in Section 3, $e((\text{epi}\, f)_\rho, \text{epi}\, f_\theta) = 0$. To majorize $e((\text{epi}\, f_\theta)_\rho, \tilde{\ } \text{epi}\, f)$, we rely on Kenmochi's conditions, cf. Theorem 3.2. We start with some point with $\|\hat{x}\| \leq \rho$ and $|f_\theta(\hat{x})| \leq \rho$. By definition of f_θ and φ_θ, it follows that

$$f_0(\hat{x}) + \alpha\theta d(\hat{x}, C)^p \leq \rho \ .$$

Because $|f_0| \leq M(\rho)$ on ρB, we have

$$d(\hat{x}, C) \leq [(\alpha\theta)^{-1}(\rho + M(\rho))]^{1/p} \ ,$$

and for every $0 < \epsilon < 1$, there exists $\hat{x}_\epsilon \in C$ such that

$$\|\hat{x} - \hat{x}_\epsilon\| \leq \theta^{-1/p}[\alpha^{-1}(\rho + M(\rho))]^{1/p} + \epsilon \ . \tag{6.2}$$

The following upper bound on $f(\hat{x}_\epsilon)$ is obtained directly from the preceding inequality and the local lipschitz property of f_0: since $\hat{x}_\epsilon \in C$,

$$f(\hat{x}_\epsilon) = f_0(\hat{x}_\epsilon) \leq f_0(\hat{x}) + L(\rho_1)\|\hat{x} - \hat{x}_\epsilon\|$$

where $\rho_1 = \rho + [(\alpha\theta)^{-1}(\rho + M(\rho))]^{1/p} + 1$. If we now take into account the facts that $f_0 \leq f_\theta$ and that $\varphi_\theta \geq 0$, we obtain

$$f(\hat{x}_\epsilon) \leq f_\theta(\hat{x}) + \theta^{-1/p}L(\rho_1)[\alpha^{-1}(\rho + M(\rho))]^{1/p} + \epsilon L(\rho_1) \ .$$

This, with (6.2) shows that the conditions (3.2) are satisfied and thus (6.1) holds with

$$\gamma(\rho) := [\alpha^{-1}(\rho + M(\rho))]^{1/p}\sup\{L(\rho_1), 1\}$$

The second assertion of the theorem is obtained by combining the preceding result with Theorem 4.1. Recall, that in the bounds derived above we need to replace f_0 by $f_0(\cdot + x_f) - f_0(x_f)$, and thus L by $L(\cdot + \|x_f\|)$, and M by $M(\cdot + \|x_f\|) + |f_0(x_f)|$. □

Acknowledgment We are particularly thankful to Dominique Azé (Perpignan), Alain Fougères (Perpignan) and Pierre-Jean Laurent (Grenoble) whose many comments and suggestions have done much to bring this paper to its present form.

REFERENCES

[1] Adams, R.A.: *Sobolev spaces*, Academic Press, 1975.

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

[3] Attouch, H., D. Azé and R. Wets: Convergence of convex-concave saddle functions: continuity properties of the Legendre-Fenchel transform with applications to convex programming and mechanics, Techn. Report, Université de Perpignan, November 1985. (To appear: *Annales de l'Insitut H. Poincaré*).

[4] Attouch, H. and R. Wets: Approximation and convergence in non-linear optimization in *Nonlinear Programming 4*, Mangasarian, R. Meyer and S. Robinson (eds), Academic Press, New York, 1981. 367–394.

[5] Attouch, H. and R. Wets: Isometries for the Legendre-Fenchel transform. *Transactions American Mathematical Society*, 296(1986), 33–60.

[6] Attouch, H. and R. Wets: Another isometry for the Legendre-Fenchel transform, Techn. Report, AVAMAC, Univ. Perpignan, 1986. (forthcoming: *J. Mathematical Analysis and Applications*)

[7] Attouch, H. and R. Wets: Quantitative stability of variational systems: I The epigraphical distance, Techn. Report , University of California-Davis, August 1987.

[8] Attouch, H. and R. Wets: Quantitative stability of variational systems: III ϵ-approximate solutions, WP-87-25 (Title: Lipschitzian stability of ϵ-approximate solutions in convex optimization), IIASA. Laxenburg, 1987.

[9] Aubin, J.-P.: Lipschitz behaviour of solutions to convex minimization problems. *Mathematics of Operations Research*, 9(1984), 87–111.

[10] Aubin, J.-P. and I. Ekeland : *Applied Nonlinear Analysis.* Wiley Interscience Series, New York. 1984.

[11] Aubin, J.-P. and H. Frankowska: On inverse functions theorems for set valued maps, *J. Mathématiques Pures et Appliquées.* (IIASA WP-84-68, Laxenburg).

[12] Aubin, J.-P. and R. Wets: Stable approximations of set-valued maps, IIASA WP-87-74.

[13] Auslender, A. and J.P. Crouzeix: Global regularity theorems. Tech. Report, Université de Clermont-Ferrand II, 1986.

[14] Azé, D.: Characterizations of uniform convexity of functionals. Techn. Report AVAMAC. Université de Perpignan, 1986.

[15] Azé, D. and J.-P. Penot: Operations on convergent families of sets and functions, Manuscript, 1987.

[16] Bank, B., J. Guddat, D. Klatte, B. Kummer and K. Tammer, *Non-linear Parametric Optimization.* Birkhäuser Verlag, Basel. 1983.

[17] Clément, J.–B.: Eigenvalue problem for a class of cyclically maximal monotone operators. *Nonlinear Analysis: Theory, Methods and Applications*, 2(1977), 93–103.

[18] Daniel, J.-W.: Stability of the solution of definite quadratic programs. *Mathematical Programming*, 5 (1973),41–53.

[19] Dolecki, Sz.: Convergence of minima in convergence spaces, *Optimization* 17(1986), 553–572.

[20] 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.

[21] Evans, J.-P. and F.J. Gould: Stability in nonlinear programming, *Operations Research*, 18(1970), 107-118.

[22] Fiacco, A.V.: Convergence properties of local solutions of sequences of nonlinear programming problems in general spaces, *J. Optimization Theory Applications*, 13(1974), 1-12.

[23] Fougères, A.: Convexité et coercivité: structure semi-normée intrinsèque, Travaux du Séminaire Analyse Convexe Montpellier-Perpignan, Exposé no. 6, (1981).

[24] Gauvin, J.: The generalized gradient of a marginal function in mathematical programming, *Mathematics of Operations Research*, 4(1979), 458-463.

[25] Hiriart-Urruty, J.-B.: Lipschitz r-continuity of the approximate subdifferential of a convex function, *Mathematica Scandinavia*, 47(1980), 123-134.

[26] Hogan, W.: Point-to-set maps in mathematical programming, *SIAM Review* 15(1973), 591-603.

[27] 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.

[28] Kummer, B.: Stability of generalized equations and Kuhn-Tucker points of perturbed convex programs, in: *Proceedings of 11th TIMS International Conference, Copenhagen*. 1983.

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

[30] Lemaire-Misonne, C.: Validation des résultats: conditionnement de problèmes. Proceedings "7ème Rencontre franco-belge de statisticiens", Rouen. 1986.

[31] Lucchetti, R., and F. Patrone: Tykhonov well-posedness of a certain class of convex functions, *J. Mathematical Analysis and Applications*, 88(1982), 204-215.

[32] Moreau, J.J.: Rafle par un convexe variable. Séminaire d'Analyse convexe, Montpellier. lère partie, Exposé n°15 (1971), 2éme partie, Exposé n°3 (1972).

[33] Mosco, U.: Convergence of convex sets and of solutions of variational inequalities, *Advances in Mathematics*, 3(1969), 510-585.

[34] Rabier, P. and J.M. Thomas: Exercices d'analyse numérique des équations aux dérivées partielles. Mathématiques appliquées pour la maîtrise. Masson. 1985.

[35] Robinson, S.: Stability theory for systems of inequalities. Part I: linear systems *SIAM J. Numerical Analysis*, 12(1975), 754-769.

[36] Robinson, S.: Stability theory for systems of inequalities. Part II: Differentiable Nonlinear systems, *SIAM J. Numerical Analysis*, 13(1976), 497-513.

[37] Robinson, S.: Local epi-continuity and local optimization, *Mathematical Programming*, 37(1987), 208-222.

[38] Rockafellar, R.T.: Directional differentiability of the optimal value in a nonlinear programming problem, *Mathematical Programming Study*, 21(1984), 87-111.

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

[40] Rockafellar, R.T.: First and second order pseudo differentiability in nonlinear programming, University of Washington, Manuscript, 1987.

[41] Rockafellar, R.T.: Second order optimality conditions in nonlinear programming obtained by way of pseudo-derivatives, University of Washington, Manuscript, 1987.

[42] Rockafellar, R.T. and R. Wets: Variational systems, an introduction, in *Multifunctions and Integrands*, G. Salinetti (ed.), Springer Verlag Lecture Notes in Mathematics No. 1091, 1984. pp 1-54.

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

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

[45] Sonntag, Y.: Approximation et pénalisation en optimisation. Publications des Mathématiques appliquées, Université de Provence (Marseille) n°83-2. (1983).

[46] Tykhonov, A.N.: On the stability of functional optimization problem, *Comput. Mathematics Physics*, 6(1966), 28–33.

[47] Vladimirov, A.B., J.E. Nestorov and J.N. Chekanov: On uniformly convex functionals (in Russian), Manuscript.

[48] Wets, R.: Convergence of convex functions, variational inequalities and convex optimization problems, in *Variational Inequalities and Complementarity Problems*, R. Cottle, F. Gianessi, and J.-L. Lions (eds.), J. Wiley, London, 1980. 405-419.

[49] Zalinescu, C.: On uniformly convex functions. *J. Mathematical Analysis and Applications*, 95(1983), 344–374.

[50] Zolezzi, T.: On equi-wellset minimum problems, *Optimization*, 4(1978), 209–223.

[51] Zolezzi, T.: A characterization of well posed optimal control systems, *SIAM J. Control and Optimization*, 19(1981), 604–616.