Journal of Convex Analysis
Volume 1 (1994), No.1, 75–86

Bounded Diagonally Stationary Sequences in Convex Optimization

B. Lemaire
Université Montpellier II, Place E. Bataillon,
F-34095 Montpellier Cedex 5

Received 2 December 1993
Revised manuscript received 13 June 1994

Let X be a real normed linear space, $f, f^n,\ n \in \mathbb{N}$, be extended real-valued proper closed convex functions on X. A sequence $\{x_n\}$ in X is called diagonally stationary for $\{f^n\}$ if for all n there exists $x_n^\star \in \partial f^n(x_n)$ such that $\|x_n^\star\|_\star \to 0$. Such sequences arise in approximation methods for the problem of minimizing f. We present some general quantitative convergence results based upon metric variational convergence theory, appropriate equi-well-posedness and conditioning concepts for the limit function f, and Fejér monotonicity.

Keywords : Asymptotical behaviour, conditioning, convex optimization, diagonal stationarity, error estimate, Fejér monotonicity, variational convergence, well-posedness.

1991 Mathematics Subject Classification: 65K10, 90C25, 49D07.

1. Introduction

Let X be a real normed linear space, $f, f^n,\ n \in \mathbb{N}$, be extended real-valued proper closed convex functions on X, and $\{x_n\}$ be a diagonally stationary sequence for $\{f^n\}$, in other words, for all n there exists $x_n^\star \in \partial f^n(x_n)$ such that $\|x_n^\star\|_\star \to 0$. Such sequences are generated by perturbed (or diagonal) constructive processes like regularization or gradient method combined with approximation like discretization or penalization (see the bibliography of [9]).

Some convergence results for such diagonally stationary sequences based upon variational convergence theory and appropriate equi-well-posedness concepts have been presented in [9]. These previous results were of qualitative kind: the distance $d(x_n, S)$ from x_n to the optimal set $S := \text{Argmin} f$ tends to 0, or, in a reflexive Banach space setting, any weak accumulation point of $\{x_n\}$ is in S. The present work is devoted to more precise results for bounded ($\|x_n\| \le \rho$) diagonally stationary sequences based upon metric variational convergence theory, suitable well-posedness and conditioning concepts for the limit function f, and Fejér monotonicity. Two main results are established. The first one is an error estimate: if f is "sufficiently well" conditioned and if the ρ-Hausdorff excess $e_\rho(\partial f^n, \partial f)$ tends to 0, then $d(x_n, S)$ can be estimated in terms of $\|x_n^\star\|_\star$ and $e_\rho(\partial f^n, \partial f)$; these two quantities are under control during the constructive process that

ISSN 0944-6532 / \$ 2.50

generates $\{x_n\}$. Therefore, such an estimate would be of practical interest for deriving a stopping rule in such a process, at least when it is possible to estimate the appropriate constants arising in the error estimating formula. Nevertheless, this gives information on the rate of convergence of the error. In particular, finite termination holds ($x_n \in S$ for n large enough) if f is linearly conditioned and if (what we call exact approximation) $e_\rho(\partial f^n, \partial f) = 0$ for n large enough. The second main result is **norm** convergence to **some** minimizer (which is stronger than $d(x_n, S) \to 0$). For that, two key concepts are used: well posedness and Fejér monotonicity, a property currently encountered in iterative processes but not so much exploited on its own. As a consequence, we show that well posedness gives a new infinite dimensional case of **strong** convergence for the Martinet-Rockafellar's proximal method.

The paper is organized as follows. Section 2 recalls some definitions about metric variational convergence theory and diagonal stationarity. Section 3 is devoted to the error estimate and finite termination. In section 4 we prove that in case of quadratic penalization in convex programming, $e_\rho(\partial f^n, \partial f)$ is of order $1/\sqrt{r_n}$ where r_n is the penalty parameter ($r_n \to +\infty$), and that exact approximation holds true in case of linear penalty. Finally, in section 5 we show that (quasi) Fejér monotonicity jointly with well posedness leads to **norm** convergence of $\{x_n\}$ to some minimizer of f.

2. Notations and definitions

First let us recall some definitions about metric variational convergence theory. Let $(X, \|.\|)$ be a normed linear space and $(X^\star, \|.\|_\star)$ its dual. For all subset C in X (resp. in $X^\star$) we denote the distance from some point x in X (resp. $x^\star$ in $X^\star$) to C by

$$d(x, C) := \inf_{y \in C} \|x - y\| \;\; (\text{resp. } d_\star(x^\star, C))$$

For all $\rho \geq 0$, we denote by ρB the closed ball of radius ρ. For C and D subsets of X, the Hausdorff excess of C over D is defined by $e(C, D) := \sup_{x \in C} d(x, D)$ and the ρ-Hausdorff excess of C over D by $e_\rho(C, D) := e(C_\rho, D)$ where $C_\rho := C \cap \rho B$. The ρ-Hausdorff distance between C and D is defined by

$$\text{haus}_\rho(C, D) := \max\{e_\rho(C, D), e_\rho(D, C)\}.$$

A sequence of extended real valued functions f^n epi-distance (or Attouch-Wets) converges to some function f iff for all $\rho \geq 0, \text{haus}_\rho(f^n, f) \to 0$, the involved functions being identified with their epigraphs in the product space $X \times \mathbb{R}$ endowed with the "box" norm $\max\{\|.\|, |.|\}$ [6].

A sequence of set valued operators A^n from X into $X^\star$ graph-distance converges to some operator A iff for all $\rho \geq 0$, $\text{haus}_\rho(A^n, A) \to 0$, the involved operators being identified with their graphs in the product space $X \times X^\star$ endowed with the "box" norm $\max\{\|.\|, \|.\|_\star\}$.

Definition 2.1. Let $\{f^n\}$ be a sequence of extended real-valued proper closed convex functions on X. A sequence $\{x_n\}$ in X is diagonally stationary for $\{f^n\}$ (for short : $\{f^n\} - DS$) iff

$$\lim_{n \to +\infty} d_\star(0, \partial f^n(x_n)) = 0$$

where ∂f^n denotes the subdifferential of Convex Analysis. In other words, for each $n \in \mathbb{N}$, there exists a subgradient $x_n^\star \in \partial f^n(x_n)$ such that $x_n^\star \to 0$ strongly in $X^\star$.

It is an easy consequence of the Brøndsted-Rockafellar's theorem [12] that a $\{f^n\} - DS$ sequence does exist if X is a Banach space and if, for all n in $\mathbb{N}$, f^n is bounded from below [17], [9].

3. Error estimate and finite termination

In all this section, $f, f^n,\ n = 1, 2, ...$ are extended real-valued proper closed convex functions on X, $S :=$ Argmin$f$ and $\{x_n\}$ is any bounded $\{f^n\} - DS$ sequence ($\|x_n\| \leq \rho$ for some $\rho > 0$), $\{x_n^\star\}$ is a sequence of subgradients associated with $\{x_n\}$ according to definition 2.1.

Let us recall the definition of conditioning.

Definition 3.1. A real extended valued function g on X is said to be conditioned if $S :=$ Argmin$g \neq \emptyset$ and if $\exists \psi : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{+\infty\},\ \ \psi(0) = 0$ such that

$$\forall x,\ \ g(x) \geq \min g + \psi(d(x, S))$$

Remark 3.2. It is known [24], [11], [17] that firm conditioning (this means that, in definition 3.1, ψ has the firmness property: $\forall \{t_n\} \subset \mathbb{R}_+,\ \ \psi(t_n) \to 0 \Rightarrow t_n \to 0$) is equivalent to well posedness (see definition in proposition 5.1 below). In the following we need more than firmness.

Proposition 3.3. *Let us assume that $e_\rho(\partial f^n, \partial f)$ tends to 0 as n tends to $+\infty$ and that f is conditioned with a conditioning function ψ such that the function φ defined by*

$$\varphi(t) := \psi(t)/t \ \text{ if } t > 0,\ \ \varphi(0) := 0,$$

is finite valued, strictly increasing and continuous on $[0, \bar{t}]$ for some $\bar{t} > 0$. Then ($S :=$ Argmin$f$), for n large enough,

$$d(x_n, S) \leq e_\rho(\partial f^n, \partial f) + \varphi^{-1}(\|x_n^\star\|_\star + e_\rho(\partial f^n, \partial f)).$$

Proof. First it must be noted that φ^{-1} the inverse of φ exists, is finite valued, strictly increasing and continuous on $[0, \varphi(\bar{t})]$.

As $\|x_n^\star\|_\star \to 0$, for n large enough, $\|x_n^\star\|_\star \leq \rho$ and so, from the definition of the ρ-Hausdorff excess, $d((x_n, x_n^\star), \partial f) \leq e_\rho(\partial f^n, \partial f)$. Then, from the definitions of the box norm and the distance to some subset, for all $\epsilon > 0$, there exists x_n^ϵ and $x_n^{\epsilon\star}$ such that

$$\|x_n^\epsilon - x_n\| \leq e_\rho(\partial f^n, \partial f) + \epsilon, \quad \|x_n^{\epsilon\star} - x_n^\star\|_\star \leq e_\rho(\partial f^n, \partial f) + \epsilon$$

and $x_n^{\epsilon\star} \in \partial f(x_n^\epsilon)$. So

$$\forall x \in S,\ \ f(x) \geq f(x_n^\epsilon) + \langle x_n^{\epsilon\star}, x - x_n^\epsilon \rangle$$

Therefore

$$f(x_n^\epsilon) - \min f \le \|x_n^{\epsilon\star}\|_\star d(x_n^\epsilon, S) \le \alpha_n^\epsilon d(x_n^\epsilon, S)$$

where

$$\alpha_n^\epsilon := \|x_n^\star\|_\star + e_\rho(\partial f^n, \partial f) + \epsilon$$

Furthermore, from conditioning, $\psi(d(x_n^\epsilon, S)) \le f(x_n^\epsilon) - \min f$. Then,

$$\varphi(d(x_n^\epsilon, S) \le \alpha_n^\epsilon$$

For ϵ small enough and n large enough, $\alpha_n^\epsilon \le \varphi(\bar{t})$. Therefore, as φ^{-1} is increasing on $[0, \varphi(\bar{t}]$, we get

$$d(x_n^\epsilon, S) \le \varphi^{-1}(\alpha_n^\epsilon).$$

We conclude from

$$d(x_n, S) \le \|x_n - x_n^\epsilon\| + d(x_n^\epsilon, S) \le e_\rho(\partial f^n, \partial f) + \epsilon + \varphi^{-1}(\alpha_n^\epsilon)$$

and passing to the limit as $\epsilon \to 0$. □

Corollary 3.4. *If $\psi(t) = \gamma t^p$, $\gamma > 0$, $p > 1$, then*

$$d(x_n, S) \le (1 + 1/\gamma')(\|x_n^\star\|_\star + e_\rho(\partial f^n, \partial f))^{\frac{1}{p-1}}, \quad \forall n$$

where $\gamma' := \gamma^{\frac{1}{p-1}}$.

Proof. Take $\bar{t}$ as large as needed and

$$\varphi^{-1}(s) = \frac{s^{\frac{1}{p-1}}}{\gamma}$$

□

Corollary 3.5. *If for all $\rho > 0$, for all n large enough, $e_\rho(\partial f^n, \partial f) = 0$ (we then say that exact approximation holds true), then, for n large enough,*

$$d(x_n, S) \le \varphi^{-1}(\|x_n^\star\|_\star).$$

Remark 3.6. The assumption of exact approximation implies that $\{x_n\}_{n \ge n_0}$, for some n_0, is f-stationary because $e_\rho(\partial f^n, \partial f) = 0$ is equivalent to $(\partial f^n)_\rho \subset \partial f$ and therefore $x_n^\star \in \partial f(x_n)$ for n large enough.

Remark 3.7. By [6], [7], [21], for a given $\rho > 0$ large enough the ρ-Hausdorff distance between ∂f^n and ∂f can be controled by the γ-Hausdorff distance between f^n and f for some $\gamma > 0$ (that depends on ρ). But the first distance can be nul for n large enough whereas the second one can be different from zero for all n, as it is shown by the following example: $X := \mathbb{R}$,

$$f(x) := e^x \text{ if } x \ge 0, \ +\infty \text{ otherwise}$$

$$f^n(x) := e^x \text{ if } x \geq 0, \quad e^x - r_n x \text{ otherwise, with } r_n \uparrow +\infty.$$

It is proved in [16] that if f has a linear conditioning then any f-stationary sequence **finitely** converges to $S := \mathrm{Argmin} f$ (for instance this is the case for the sequence generated by the Martinet-Rockafellar's proximal method [15]), the converse being true if X is a Banach space. The purpose of the following is to show that, under exact approximation, finite termination is preserved for any $\{f^n\} - DS$ sequence.

Proposition 3.8. *If f is linearly condioned: $\psi(t) := \gamma t$ for some $\gamma > 0$, then, for n large enough,*

$$d(x_n, S) \leq e_\rho(\partial f^n, \partial f).$$

Proof. Returning to the proof of proposition 3.3 we get

$$\gamma d(x_n^\epsilon, S) \leq \alpha_n^\epsilon d(x_n^\epsilon, S)$$

Taking $\epsilon < \gamma$, for n large enough we have

$$\|x_n^\star\|_\star + e_\rho(\partial f^n, \partial f) < \gamma - \epsilon$$

and then $d(x_n^\epsilon, S) = 0$. Therefore

$$d(x_n, S) \leq e_\rho(\partial f^n, \partial f) + \epsilon \qquad \forall\, \epsilon > 0$$

$\square$

The finite termination result is then immediate:

Corollary 3.9. *If f is linearly conditioned and if exact approximation holds true, then $\{x_n\}$ is finitely convergent: for all n large enough x_n is in S.*

4. Penalization in Convex Programming

In this section we give some estimate of $e_\rho(\partial f^n, \partial f)$ in case of exterior penalization in convex programming. In particular, we prove that exact approximation holds true in case of linear penalty.

$X := \mathbb{R}^d$, $f := f_0 + \psi_C$ with

$$C := \{x \in X; \;\; f_i(x) \leq 0, \;\; i = 1, ..., m\}$$

where the $f_i's$ are convex functions from X into $\mathbb{R}$.

Let us recall the definitions of linear and quadratic penalty approximations.

Linear penalty : $f_L^n := f_0 + r_n \sum_{i=1}^m f_i^+, \quad 0 \leq r_n \uparrow +\infty.$

Quadratic penalty : $f_Q^n := f_0 + \frac{r_n}{2} \sum_{i=1}^m (f_i^+)^2, \quad 0 \leq r_n \uparrow +\infty.$

Let us assume the Slater constraint qualification:

$$\exists \tilde{x} \in X, \quad f_i(\tilde{x}) < 0 \quad i = 1, ..., m$$

Lemma 4.1. *For all $\rho > 0$ there exists $\gamma_\rho > 0$ such that for all $x \notin C$, $\|x\| \leq \rho$, there exists i such that*

$$f_i(x) \geq \gamma_\rho \, d(x, C)$$

Proof. Let $-\epsilon := \max_i f_i(\tilde{x}) < 0$ and let $\overline{x} := \lambda \tilde{x} + (1 - \lambda)x$ where $0 < \lambda < 1$, be the boundary point in C that belongs to the open segment $]\tilde{x}, x[$. We have $f_i(\overline{x}) = 0$ for some i and, by convexity,

$$f_i(x) \geq \frac{\lambda \epsilon}{1 - \lambda} \geq \lambda \epsilon$$

But $\lambda = \|\overline{x} - x\| / \|\tilde{x} - x\|$. Therefore, as $\|\overline{x} - x\| \geq d(x, C)$, we are done with $\gamma_\rho := \epsilon / (\|\tilde{x}\| + \rho)$. $\square$

Remark 4.2. If C is bounded, γ_ρ doesn't depend on ρ (in the previous proof take $\gamma_\rho := \epsilon / 2c$ where c is some norm bound of C), a result already stated in [20].

Then conditioning results on the penalty functions follow immediately.

Proposition 4.3.

(i) *The linear penalty function has a local linear conditioning:*

$$\|x\| \leq \rho \quad \Rightarrow \quad \sum_{i=1}^{m} f_i^+(x) \geq \gamma_\rho \, d(x, C).$$

(ii) *The quadratic penalty function has a local quadratic conditioning:*

$$\|x\| \leq \rho \quad \Rightarrow \quad \sum_{i=1}^{m} (f_i^+(x))^2 \geq \gamma_\rho^2 \, d(x, C)^2.$$

From proposition 4.3 we get the following metric estimates.

Theorem 4.4.

(i) *For all $\rho > 0$, $haus_\rho(\partial f_Q^n, \partial f) = O_\rho(r_n^{-\frac{1}{2}})$.*

(ii) *For all $\rho > 0$, for n large enough, $haus_\rho(\partial f_L^n, \partial f) = 0$ $(\Leftrightarrow (\partial f_L^n)_\rho = (\partial f)_\rho)$.*

Proof. (i) 1. Let us prove $e_\rho(\partial f, \partial f_Q^n) = O_\rho(r_n^{-\frac{1}{2}})$.

Let $(x, x^\star) \in (\partial f)_\rho$. It is easily proved that x minimizes $f_0 - \langle x^\star, . \rangle$ on C. So, thanks to the Slater condition, there exists a Kuhn-Tucker vector $\lambda \in \mathbb{R}_+^m$ ([22], theorem 28.2):

$$\forall z \in X, \quad f_0(x) - \langle x^\star, x\rangle \leq f_0(z) - \langle x^\star, z\rangle + \sum_{i=1}^m \lambda_i f_i(z)$$

that implies λ bounded (dependently on ρ):

$$\sum_{i=1}^m \lambda_i \leq [f_0(\tilde{x}) - \min_{\|y\|\leq\rho} f_0(y) + \rho(\rho + \|\tilde{x}\|)]/\epsilon$$

(where $\tilde{x}$ is a Slater point and $-\epsilon := \max_i f_i(\tilde{x}) < 0$) and (using the trick: $2\langle a, b\rangle = \|a\|^2/\alpha + \alpha\|b\|^2, \;\; \alpha > 0$),

$$\forall z \in X, \quad f_Q^n(z) \geq f_0(x) + \langle x^\star, z - x\rangle - \alpha_n$$

viz. $x^\star \in \partial_{\alpha_n} f_Q^n(x)$ where $\alpha_n := (\sum_{i=1}^m \lambda_i)/2r_n = O_\rho(r_n^{-1})$.

Finally, thanks to Brøndsted-Rockafellar's theorem [12], there exists $(x_n, x_n^\star) \in \partial f_Q^n$ such that $\|(x_n, x_n^\star) - (x, x^\star)\| = O_\rho(r_n^{-\frac{1}{2}})$.

2. Let us prove $e_\rho(\partial f_Q^n, \partial f) = O_\rho(r_n^{-\frac{1}{2}})$.

Let $(x_n, x_n^\star) \in (\partial f_Q^n)_\rho$, $\overline{x}_n :=$ proj $_C x_n$ and $\overline{x}_n^\star \in \partial f_0(\overline{x}_n)$. As $\{x_n\}$ is bounded (by ρ), $\{\overline{x}_n\}$ and $\{\overline{x}_n^\star\}$ are bounded (by some constant depending on ρ). Moreover we have

$$\forall z \in C, \quad f_0(z) \geq f_0(\overline{x}_n) + \langle x_n^\star, z - \overline{x}_n\rangle + \langle x_n^\star, \overline{x}_n - x_n\rangle + \langle \overline{x}_n^\star, x_n - \overline{x}_n\rangle + \frac{1}{2} r_n \gamma_\rho^2 \|x_n - \overline{x}_n\|^2$$

from which we get (using again the above trick) $x_n^\star \in \partial_{\alpha_n} f(\overline{x}_n)$ where $\alpha_n := \dfrac{\|x_n^\star - \overline{x}_n^\star\|^2}{2\gamma_\rho^2 r_n} = O_\rho(r_n^{-1})$. Again apply the Brøndsted-Rockafellar's theorem to get $(\tilde{x}_n, \tilde{x}_n^\star) \in \partial f$ such that $\|(x_n, x_n^\star) - (\tilde{x}_n, \tilde{x}_n^\star)\| = O_\rho(r_n^{-\frac{1}{2}})$.

(ii) 1. Let us prove $(\partial f)_\rho \subset \partial f_L^n$ for n large enough.

Let $(x, x^\star) \in (\partial f)_\rho$. Returning to previous (i) 1. and using

$$\sum_{i=1}^m \lambda_i f_i(z) \leq \max_i \lambda_i \sum_{i=1}^m f_i^+(z) \leq r_n \sum_{i=1}^m f_i^+(z) \;\; \forall n \text{ large enough}$$

we get $(x, x^\star) \in \partial f_L^n$.

2. Let us prove $(\partial f_L^n)_\rho \subset \partial f$ for n large enough.

Let $(x_n, x_n^\star) \in (\partial f_L^n)_\rho$. The main point is to show that, for n large enough, $x_n \in C$.

Let $\overline{x}_n :=$ proj $_C x_n$ and $\overline{x}_n^\star \in \partial f_0(\overline{x}_n)$. As $\{x_n\}$ is bounded, so are $\{\overline{x}_n\}$ and $\{\overline{x}_n^\star\}$, we get

$$r_n \gamma_\rho d(x_n, C) \leq r_n \sum_{i=1}^m f_i^+(x_n) \leq c_\rho d(x_n, C), \;\; \text{for some } c_\rho > 0$$

Now if $r_n > c_\rho/\gamma_\rho$ then $d(x_n, C) = 0$. □

From corollary 3.4, corollary 3.9 and theorem 4.4, we get the following error estimates.

Corollary 4.5.

(i) *Let $\{x_n\}$ be a bounded $\{f_Q^n\} - DS$ sequence. If f has a quadratic conditioning (cf. corollary 3.4 with $p = 2$) then*

$$d(x_n, S) = O(\|x_n^\star\|_\star + r_n^{-\frac{1}{2}}).$$

(ii) *Let $\{x_n\}$ be a bounded $\{f_L^n\} - DS$ sequence. If f has a linear conditioning then $\{x_n\}$ is finitely convergent to S.*

Remark 4.6. As the extended real valued function associated with the linear programming problem is linearly conditioned ([20], [16]), from corollary 4.5 (ii), we recover the finite termination property of the linear penalty-prox method for linear programming [8]. Moreover, in this case the Slater condition is no more needed because the linear penalty function being polyhedral condition *i)* in proposition 4.3 is (even globally) fulfilled.

5. Fejér monotonicity and norm convergence

5.1. Quasi Fejér monotonicity

As noted in [10], the notion of Fejér monotonicity goes back at least to [18]. Here we relax this notion in the following definition (cf. [13] for a slightly more general extension).

Definition 5.1. Let (X, d) be a metric space and S be a non empty subset in X. A sequence $\{x_n\}$ in X is quasi Fejér monotone with respect to S (in short S-QFM) iff there exists a sequence of non negative reals $\{\epsilon_n\}$ such that $\sum_{n=1}^{+\infty} \epsilon_n < +\infty$ and

$$\forall x \in S, \quad d(x_n, x) \leq d(x_{n-1}, x) + \epsilon_n, \quad n = 1, 2, ...$$

Remark 5.2. Fejér monotonicity refers to the case where $\epsilon_n = 0$ for all n. A typical example is then given by the iterations of a non expansive mapping with non empty fixed points set S.

The notion of quasi Fejér monotonicity is motivated by the following result.

Theorem 5.3. *Let us assume X to be complete and S to be closed. For all S-QFM sequence $\{x_n\}$, if $d(x_n, S) \to 0$ then there exists $x_\infty \in S$ such that $x_n \to x_\infty$ (of course the converse being always true) with the estimate*

$$d(x_n, x_\infty) \leq 2\ d(x_n, S) + \sum_{k=n+1}^{+\infty} \epsilon_k, \quad \forall n \in \mathbb{N}.$$

Proof. First let us show that $\{x_n\}$ is a Cauchy sequence. Let $m > n$.

$$\forall x \in S, \quad d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) \leq 2d(x_n, x) + \sum_{k=n+1}^{m} \epsilon_k$$

or $d(x_m, x_n) \leq 2d(x_n, S) + \sum_{k=n+1}^{m} \epsilon_k$ that tends to 0 as m and n tend to $+\infty$.

Therefore x_n tends to some x_∞ and, thanks to the continuity of the function $d(., S)$ we have $d(x_\infty, S) = 0$ and, as S is closed, $x_\infty \in S$.

Finally let m tend to $+\infty$ in the above estimate. □

Remark 5.4. Generally, if S is a non empty closed subset in X and if some sequence $\{x_n\}$ is compact then $d(x_n, S) \to 0$ is equivalent to: any accumulation point of $\{x_n\}$ is in S.

If X is a reflexive Banach space, f, f^n, $n = 1, 2, ...$ are extended real-valued proper closed convex functions on X such that f^n Mosco-converges to f and if $\{x_n\}$ is a bounded $\{f^n\} - DS$ sequence then $f^n(x_n) \to \inf f$ and any weak accumulation point of $\{x_n\}$ is a minimizer of f ([9], Proposition 4.3). Furthermore if X is a Hilbert space and $\{x_n\}$ is Argmin$f - QFM$ a classical argument (for instance used in the convergence proof of the proximal method) shows that the whole sequence $\{x_n\}$ weakly converges to some minimizer of f. In the finite dimensional setting theorem 5.3 furnishes an alternative argument of this fact since, in this setting, saying that any accumulation point of a bounded sequence is in S is equivalent to saying that $d(x_n, S) \to 0$.

5.2. Connection with well posedness and diagonal stationarity

As an immediate consequence of theorem 5.3 we get the following proposition.

Proposition 5.5. *Let f be an extended real-valued proper closed function on the complete metric space X. Let us assume f to be well-posed: $S :=$ Argmin$f$ is non empty and*

$$f(x_n) \to \inf f \quad \Rightarrow \quad d(x_n, S) \to 0.$$

Then any S-QFM f-minimizing sequence converges to some minimizer of f.

Example 5.6. Approximate prox.

Let X be a real Hilbert space and f an extended real-valued proper closed convex function on X such that $S :=$ Argmin$f \neq \emptyset$. Let $\{\lambda_n\}$ and $\{\epsilon_n\}$ be sequences of positive reals such that

$$\lambda_n \geq \underline{\lambda} > 0, \quad \sum_{n=1}^{+\infty} \sqrt{\lambda_n \epsilon_n} < +\infty$$

x_0 be given in X, the Auslender's approximate prox method [2] generates a sequence $\{x_n\}$ such that

$$x_n \in \epsilon_n - \text{Argmin}\{\frac{1}{2\lambda_n}\|. - x_{n-1}\|^2 + f\}, \quad n = 1, 2, ...$$

It is known [2] that such a sequence satisfies

$$\|x_n - \operatorname{prox}_{\lambda_n f} x_{n-1}\| \leq \sqrt{2\lambda_n \epsilon_n}, \quad n = 1, 2, ...$$

viz. $\{x_n\}$ is generated by the Rockafellar's approximate prox method [23]. Therefore, it is a direct consequence of the non expansiveness of the prox mapping and because the set of fixed points of this prox mapping is S, that $\{x_n\}$ is $S - QFM$. Moreover [15] $\{x_n\}$ is minimizing (and this does not depend neither of the non vacuity of S nor of the finiteness of $\inf_X f$).

It is well known that the sequence generated by the Rockafellar's approximate prox method is weakly convergent to some minimizer of f. Actually, if f is well posed, proposition 5.5 shows that, in the Auslender's approximate (and *a fortiori* in the exact) prox method, the convergence holds true for the **norm** topology.

Proposition 5.7. *Let X be a real Banach space and f, f^n, $n = 1, 2, ...$ be extended real-valued proper closed convex functions on X. We assume that f is well posed and that ∂f^n graph-distance converges to ∂f. Then any S-QFM (where $S :=$ Argminf) $\{f^n\} - DS$ sequence $\{x_n\}$ is norm convergent to some minimizer of f.*

Proof. As a $S - QFM$ sequence, $\{x_n\}$ is bounded. Then, as ∂f^n graph-distance converges to ∂f, $\exists\{\overline{x}_n\}$ stationary for f such that $\|x_n - \overline{x}_n\| \to 0$. Moreover, f well posed is equivalent to f has a well asymptotical behaviour in solution (for any f-stationary sequence $\{z_n\}$, $d(z_n, S)$ tends to 0) [17]. Therefore $d(\overline{x}_n, S) \to 0$. Finally apply theorem 5.3. □

Remark 5.8. We recall that in a Banach space the graph-distance convergence of ∂f^n to ∂f is implied by the epi-distance convergence of f^n to f [4], [21], [7].

Example 5.9. Diagonal prox.

Let X and f be as in the previous example and $0 < \underline{\lambda} \leq \lambda_n \leq \overline{\lambda} < +\infty$. Let $\{f^n\}$ be a sequence of extended real-valued proper closed convex functions on X such that

$$\forall \rho > 0, \ \sum_{n=1}^{+\infty} \operatorname{haus}_\rho(\partial f^n, \partial f) < +\infty$$

x_0 be given in X we consider the sequence $\{x_n\}$ generated by the diagonal prox method [14], [1]:

$$x_n := \operatorname{prox}_{\lambda_n f^n} x_{n-1} := argmin\{\frac{1}{2\lambda_n}\|. - x_{n-1}\|^2 + f^n\}, \ \ n = 1, 2, ...$$

Lemma 5.10.

$$\exists \rho > 0, \ \|x_n - \ prox_{\lambda_n f} \ x_{n-1}\| \leq (2 + \overline{\lambda}) \ haus_\rho(\partial f^n, \partial f)$$

Proof. Following [19], let us consider the Yosida semi distance between maximal monotone operators [5]

$$\forall \rho \geq 0,\ \forall \lambda > 0,\ \ d_{\lambda,\rho}(A,B) := \sup_{\|x\|\leq\rho} \|J_\lambda^A - J_\lambda^B\|$$

where J_λ^A denotes the resolvant of A with parameter λ (recall that $\operatorname{prox}_{\lambda f}$ is nothing but $J_\lambda^{\partial f}$). A direct application of [6] (theorem 5.1) or [3] (proposition 1.2) gives the following estimate.

$$\forall \rho \geq 0\ \ \exists \rho' > 0\quad d_{\lambda_n,\rho}(\partial f^n, \partial f) \leq (2+\overline{\lambda})\ \operatorname{haus}_{\rho'}(\partial f^n, \partial f)$$

From the non expansiveness of the prox mapping we get easily

$$\forall x \in S,\ \ \|x_n - x\| \leq \|x_{n-1} - x\| + d_{\lambda_n,\|x\|}(\partial f^n, \partial f)$$

that combined with the above estimate shows that $\{x_n\}$ is bounded ($\|x_n\| \leq \rho$). Then the result comes from the immediate estimate

$$\|x_n - \operatorname{prox}_{\lambda_n f} x_{n-1}\| \leq d_{\lambda_n,\rho}(\partial f^n, \partial f)$$

$\square$

As in the previous example $\{x_n\}$ is generated by the Rockafellar's approximate prox method and so is $S-QFM$ and, [23], $\|x_n - x_{n-1}\| \to 0$. Moreover, as $\dfrac{x_n - x_{n-1}}{\lambda_n} \in \partial f^n(x_n)$, $\{x_n\}$ is $\{f^n\}-DS$. If f is well posed, proposition 5.7 shows that, still here, the convergence is **strong**. Taking $f^n = f$ for all n, we again recover (cf. sub-section 6.1) the strong convergence for the (exact) prox method.

Acknowledgment. The author would like to thank an anonymous referee for his valuable remarks and suggestions.

References

[1] P. Alart, B. Lemaire: Penalization in non-classical convex programming via variational convergence, Math. Progr. 51, 1991, 307–331.

[2] A. Auslender: Numerical methods for non-differentiable convex optimization, in Math. Progr. Studies, 30, 1985

[3] H. Attouch, A. Moudafi, H. Riahi: Quantitative stability analysis for maximal monotone operators and semi-groups of contractions, Séminaire d'Analyse Convexe de Montpellier, Vol. 21, 1991.

[4] H. Attouch, J.L. Ndoutoumé, M. Théra: Epigraphical convergence of functions and convergence of their derivatives in Banach spaces, Séminaire d'Analyse Convexe de Montpellier, Vol. 20, exposé n° 9, 1990.

[5] H. Attouch, R. Wets: Isometries for the Legendre-Fenchel transform, Trans. A.M.S., Vol. 296, 1986, 33–60.

[6] H. Attouch, R. Wets: Quantitative stability of variational systems: I. The epigraphical distance, Trans. A.M.S., Vol. 328, 1991, 695-729.

[7] D. Azé, J.P. Penot: Recent quantitative results about the convergence of set and functions, Functional Analysis and Approximation, P.L. Papini ed., Pitagora Editrice, Bologna, 1989, 90-110.

[8] A. Auslender, J.P. Crouzeix, P. Fedit: Penalty proximal methods in convex programming, Journal of Optimization Theory and Applications, Vol. 55, 1, 1987, 1-21.

[9] M.A. Bahraoui, B. Lemaire: Convergence of diagonally stationary sequences in Convex Optimization, Set-Valued Analysis 2, 1994, 1–13

[10] H.H. Bauschke, J.M. Borwein: On the convergence of Von Neumann's Alternating Projection Algorithm for Two Sets, CORR 92–37, Faculty of Mathematics University of Waterloo, nov. 1992.

[11] E. Bednarczuk, J.P. Penot: Metrically well-set minimization problems, Applied Mathematics and Optimization, 26: 273–285, Springer-Verlag, 1992.

[12] A. Brøndsted, R.T. Rockafellar: On the subdifferentiability of convex functions, Proc. Amer. Math. Soc. 16, 1965, 605–611.

[13] A.N. Iusemi, B.F. Svaiter, M. Teboulle: Entropy-like proximal methods in convex programming, Research Report 92–09, Mathematics University of Maryland, May 1992.

[14] B. Lemaire: Coupling optimization methods and variational convergence, in Trends in Mathematical Optimization, International Series of Numerical Mathematics, Vol 84(c), Birkhäuser Verlag, 1988, 163–179.

[15] B. Lemaire: About the convergence of the proximal method, Advances in Optimization, Proceedings Lambrecht 1991, Lecture Notes in Economics and Mathematical Systems 382, Springer-Verlag, 1992, 39–51.

[16] B. Lemaire: Bons comportements et conditionnement linéaire, Journées d'Optimisation et Analyse Non-Linéaire, Avignon, nov. 1991.

[17] B. Lemaire: Bonne position, conditionnement, et bon comportement asymptotique, Séminaire d'Analyse Convexe de Montpellier, exposé $n^0 5$, 1992.

[18] T.S. Motzkin, I.J. Schoenberg: The relaxation method for linear inequalities, Canadian Journal of Mathematics, 6: 393–404, 1954.

[19] A. Moudafi: Coupling proximal methods and variational convergence, to appear in Zeit. Operation Research.

[20] B.T. Polyak: Sharp minimum, International Workshop on Augmented Lagragians, Vienna, 1979.

[21] H. Riahi: Quelques résultats de stabilité en analyse épigraphique : approche topologique et quantitative, Thèse Université Montpellier 2, 1989.

[22] R.T. Rockafellar: Convex Analysis, Princeton University Press, 1970.

[23] R.T. Rockafellar: Monotone Operators and the proximal point algorithm, SIAM J. Control and Optimisation, 1976, 877–898.

[24] T. Zolezzi: On Equiwellset Minimum Problems, Appl. Math. Optim. 4, 1978, 209–223.