ANNALES DE L'I. H. P., SECTION C

A. A. AUSLENDER
J. P. CROUZEIX

Well behaved asymptotical convex functions

Annales de l'I. H. P., section C, tome S6 (1989), p. 101-121
<http://www.numdam.org/item?id=AIHPC_1989__S6__101_0>

WELL BEHAVED ASYMPTOTICAL CONVEX FUNCTIONS

A.A. AUSLENDER & J.P. CROUZEIX
Département de Mathématiques Appliquées, Université Blaise Pascal, B.P. 45, 63170 Aubière, France

ABSTRACT : In this paper, we introduce the class $\mathcal{F}$ of proper closed convex functions for $\mathbb{R}^N$ which satisfy the following property.

"For all sequences $\{x_n\}$, $\{c_n\}$ such that $c_n \in \partial f(x_n)$ and $\lim_{n\to\infty} c_n = 0$ we have $\lim_{n\to\infty} f(x_n) = \inf(f(x) \mid x \in \mathbb{R}^N)$"

Characterizations and applications are then given.

I INTRODUCTION : In this paper all the functions f are proper-closed convex functions and we consider the optimization problem (P)

"Find a minimizing sequence $\{y_n\}$ of (P), i.e., such that :

$\lim_{n \to \infty} f(y_n) = m = \inf(f(x) \mid x \in \mathbb{R}^N)$

In mathematical programming, for proving the convergence of algorithms, one always supposes that the sequences $\{x_n\}$ constructed by the algorithm are bounded, or that the function f is inf-compact (i.e., for all λ the set $\{x : f(x) \leqslant \lambda\}$ is bounded), which in general ensures that the sequence $\{x_n\}$ is bounded. What happens when $\{x_n\}$ is unbounded, when f is not inf-compact? This question has not been considered in literature. For instance, for the proximal method, the Rockafellar convergence theorem [8] claims : "If the optimal set of solutions of P is nonempty then $\{x_n\}$ converges to an optimal solution, else we have $\lim_{n\to\infty} ||x_n|| = +\infty$". In fact, we shall prove that for a large class of convex functions which has a good behaviour at the infinity the previous sequence $\{x_n\}$ is a minimizing sequence.

For defining such a class, let us remark that numerical methods generate decreasing sequences $\{f(x_n)\}$ which are expected to be converging to m. In practice the iterative process is stopped when some condition on the iterate is satisfied. For such a stopping rule in the differentiable case one can think of $||\nabla f(x_n)|| \leqslant \varepsilon$ where ε is positive. Since a necessary and sufficient condition for optimality at $\bar{x}$ is that $\nabla f(\bar{x}) = 0$ it can be expected that if ε is small enough $f(x_n)$ is closed to m. In fact, as soon as the sequence $\{x_n\}$ is bounded, the criteria $\nabla f(x_n) \to 0$ implies that $f(x_n)$ converges to m. This is in particular the case when f is inf-compact. Unfortunately, in the unbounded case, there are convex functions and sequences $\{x_n\}$ such that $\{\nabla f(x_n)\}$ tends to the null vector but the sequence $\{f(x_n)\}$ does not tend to m. Consider for example the function f given by Rockafellar [9]:

$$f(x_1,x_2) = x_2^2/x_1 \text{ if } x_1 > 0, = 0 \text{ if } x_1 = x_2 = 0, = +\infty \text{ elsewhere}$$

$$\text{then } m = 0,\ \nabla f(p_n^2,p_n) = \left(-\frac{1}{p_n^2}, \frac{2}{p_n}\right) \to 0 \text{ if } p_n \to +\infty \text{ and } f(p_n^2,p_n) = 1 \qquad (1.0)$$

In this case the sequence $\{(p_n^2,p_n)\}$ is unbounded so that one can speak of a bad asymptotical behaviour of the function f. By definition, we shall say that a closed-proper convex function f on $\mathbb{R}^N$ has a good asymptotical behaviour if :

"For all sequences $\{x_n\}$, $\{c_n\}$ $c_n \in \partial f(x_n)$ (subdifferential of f at x_n) :

$$\lim_{n\to\infty} c_n = 0 \Rightarrow \lim_{n\to\infty} f(x_n) = m." \qquad (1.1)$$

The class of these functions will be denoted by $\mathcal{F}$. In fact, we shall restrict in general ourselves in the paper to the class of functions $f \in \mathcal{F}$ such that

$$S_\lambda(f) = \{x : f(x) \leqslant \lambda\} \subset \text{ri}(\text{dom } f) \quad \forall\lambda > m \qquad (1.2)$$

and we shall denote this class by $\mathcal{F}_1$.

As an application of functions in $\mathcal{F}_1$ one can consider the concept of entropy which is very important in sciences. The most used measure of entropy is the "x log x" entropy defined on the nonnegative orthant $\mathbb{R}_+^N$ by

$$\text{ent}_1(x) = -\sum_{i=1}^{N} x_i \text{ Log } x_i . \qquad (1.3)$$

Other measures have been considered. In particular, the "Log x" entropy defined on the same set by

$$\mathrm{ent}_2(x) = \sum_{i=1}^{N} \mathrm{Log}\ x_i \ . \tag{1.4}$$

See for example, the recent paper [4] by Censor and Lent. At this time we shall remark only that the first one is inf compact while the second one is only in $\mathcal{F}_1$. In section II we shall study the classes $\mathcal{F}$ and $\mathcal{F}_1$; some equivalent characterizations will be given. In section III we shall prove that these classes of functions allow us to study the convergence of classical optimization methods which generate unbounded sequences. We shall restrict ourselves to two fundamental algorithms based on the proximal idea initiated by Moreau. Through the following $(.,.)$ denotes the usual inner-product, $||.||$ the associated Euclidian norm .

II CHARACTERIZATIONS

Let $\partial_\varepsilon f(x)$ be the ε-subdifferential of f at x with $\varepsilon > 0$:

$$\partial_\varepsilon f(x) = \{x^*: f(y) \geq f(x) + (x^*,y-x) - \varepsilon \quad \forall y \in \mathbb{R}^N\}.$$

Proposition 2.1 f belongs to $\mathcal{F}$ iff for every sequence $\{x_n,c_n,\varepsilon_n\}$ such that

$$\varepsilon_n \geq 0,\ c_n \in \partial_{\varepsilon_n} f(x_n),\ \lim_{n\to 0} \varepsilon_n = 0,\quad \lim_{n\to\infty} c_n = 0 \tag{2.1}$$

we have

$$\lim_{n\to\infty} f(x_n) = m .$$

Proof We have only to prove that if $f \in \mathcal{F}$ and if (2.1) is satisfied then $\{x_n\}$ is a minimizing sequence. From the theorem of Brøndstedt and Rockafellar [3] there exist $\bar{x}_n$ and $\bar{c}_n$ such that $\bar{c}_n \in \partial f(\bar{x}_n)$,

$$||x_n - \bar{x}_n|| \leq \sqrt{\varepsilon_n} \quad \text{and} \quad ||c_n - \bar{c}_n|| \leq \sqrt{\varepsilon_n}. \tag{2.2}$$

From (2.1) and (2.2) we obtain $\lim_{n\to\infty} \bar{c}_n = 0$ and since $f \in \mathcal{F}$ it follows :

$$\lim_{n\to\infty} f(\bar{x}_n) = m. \tag{2.3}$$

But since

$$f(\bar{x}_n) \geq f(x_n) + (c_n, \bar{x}_n - x_n) - \varepsilon_n \geq m + (c_n, \bar{x}_n - x_n) - \varepsilon_n .$$

It follows from (2.1), (2.2) and (2.3)

$$\lim_{n \to \infty} f(x_n) = m. \qquad \square$$

For characterizing convex functions which have a good asymptotical behaviour, that is to say, which satisfy (1.1) it is useful to study the asymptotical behaviour on the boundary of the level set $S_\lambda(f) = \{x : f(x) \leq \lambda\}$ for $\lambda > m$. For this purpose let $d(x|S_\lambda(f))$ be the Euclidean distance from x to $S_\lambda(f)$ and :

$$r(\lambda) = \inf(||c|| \mid c \in \partial f(x),\ f(x) = \lambda), \tag{2.4}$$

$$\ell(\lambda) = \inf\left\{\frac{f(x)-\lambda}{d(x|S_\lambda(f))} \;\middle|\; x \notin S_\lambda(f)\right\}, \tag{2.5}$$

$$k(\lambda) = \inf\left\{f'(x; \frac{d}{||d||}) \mid d \in \partial f(x),\ f(x) = \lambda\right\}, \tag{2.6}$$

where $f'(x;d)$ denotes the directional derivative of f at x in direction d.

Set $\text{dom} f = \{x : f(x) < +\infty\}$ and let $\text{ri}(\text{dom } f)$ be the relative interior of dom f.

<u>Lemma 2.1</u> If $S_\lambda(f) \subset \text{ri}(\text{dom } f)$ and $\lambda > m$ then

$$k(\lambda) = \ell(\lambda).$$

<u>Proof</u> The proof is based on ideas used in Auslender and Crouzeix [2] in a more general context. For the sake of completeness we shall develop it briefly. Let $x \notin S_\lambda(f)$. Let y be the projection of x on $S_\lambda(f)$:

$$y \in S_\lambda(f), \quad d(x|S_\lambda(f)) = ||x-y||.$$

Clearly $f(y) = \lambda$ and there exists $\alpha > 0$ such that $\alpha(x-y) \in \partial f(y)$. Then

$$f(x) - \lambda = f(x)-f(y) \geq (c,x-y) \quad \forall c \in \partial f(y).$$

Let $h = \alpha(x-y)$, then :

$$\frac{f(x)-\lambda}{||x-y||} \geq \sup_{c \in \partial f(y)} (c, \frac{h}{||h||}) = f'(y; \frac{h}{||h||}) \geq k(\lambda).$$

From what we deduce that $\ell(\lambda) \geq k(\lambda)$. Now let y and h be such that $f(y) = \lambda$ and $h \in \partial f(y)$, then since $h \neq 0 (\lambda > m)$ we have

$$\inf_{t>0}\left\{\frac{f(y+\frac{th}{||h||})-f(y)}{t}\right\} = f'(y;\frac{h}{||h||}) > 0.$$

Taking $x = y + \frac{th}{||h||}$, then it follows that $x \notin S_\lambda(f)$ for $t > 0$ and since $t = ||x-y|| = d(x|S_\lambda(f))$ we have $k(\lambda) \leq f'(y;\frac{h}{||h||})$ and finally

$$k(\lambda) \leq \ell(\lambda).$$ □

<u>Theorem 2.2</u> Let f a proper-closed convex function on $\mathbb{R}^N$ satisfying (1.2) then we have the equivalences :

1) $f \in \mathcal{F}$.

2) $r(\lambda) > 0 \quad \forall\lambda > m$.

3) $k(\lambda) > 0 \quad \forall\lambda > m$.

<u>Proof</u> 1) ⇒ 2). Suppose that $f \in \mathcal{F}$ and there exists some $\lambda > m$ such that $r(\lambda) = 0$, then from (2.4) there exists sequences $\{x_n\}$ and $\{c_n\}$ with $c_n \in \partial f(x_n)$ such that

$$f(x_n) = \lambda, \quad \lim_{n\to\infty} c_n = 0$$

which contradicts that $f \in \mathcal{F}$.

2) ⇒ 3). This is an immediate consequence from the inequality

$$r(\lambda) \leq k(\lambda). \tag{2.7}$$

3) ⇒ 1). Suppose that $k(\lambda) > 0$ for all $\lambda > m$ and that $f \notin \mathcal{F}$; then there exists a sequence $\{x_n, c_n\}$ and $\lambda > m$ such that

$$c_n \in \partial f(x_n), \quad \lim_{n\to\infty} c_n = 0, \quad \lambda \leq f(x_n). \tag{2.8}$$

Let $\bar\lambda \in (m,\lambda)$. Then from lemma 2.1, $k(\bar\lambda) = \ell(\bar\lambda)$ and

$$d(x|S_{\bar\lambda}(f)) \leq \frac{f(x)-\bar\lambda}{k(\bar\lambda)} \quad \forall x \notin S_{\bar\lambda}(f). \tag{2.9}$$

Let $\bar x_n$ be such that

$$f(\bar x_n) = \bar\lambda, \; ||x_n - \bar x_n|| = d(x_n|S_{\bar\lambda}(f)) \tag{2.10}$$

so that from (2.9) :

$$k(\bar\lambda)\; ||x_n - \bar x_n|| \leq f(x_n) - \bar\lambda. \tag{2.11}$$

Since f is convex and $c_n \in \partial f(x_n)$

$$\bar{\lambda} = f(\bar{x}_n) \geq f(x_n) + (c_n, \bar{x}_n - x_n) \geq -\|c_n\| \, \|\bar{x}_n - x_n\| + f(x_n) \qquad (2.12)$$

which, combined with (2.11) yields to

$$0 < k(\bar{\lambda}) \leq \|c_n\|$$

which is not possible since $\{c_n\}$ converges to 0. □

What is more surprising is that the functions r(.) and k(.) are non decreasing. For proving this we must investigate the properties of the support function of $S_\lambda(f)$, $\lambda > m$. Let

$$F(x^*,\lambda) = \sup\{(x^*,x) \mid x \in S_\lambda(f)\}$$

Then the following properties obviously hold.

i) $-\infty < F(x^*,\lambda) \leq \delta^*(x^* | \text{dom } f) = \sup\{(x,x^*) | x \in \text{dom } f\} \quad \forall\lambda > m.$

ii) for all $\lambda > m$, $F(.,\lambda)$ is closed proper convex and positively homogeneous.

iii) $F(0,\lambda) = 0 \quad \forall\lambda > m.$

iv) $F(x^*,.)$ is nondecreasing $\forall x^*$.

Furthermore, set $\Psi_{x^*}(x,\lambda) = (x,x^*) - \delta((x,\lambda)|\text{epi } f)$; then $F(x^*,\lambda) = \sup_x \Psi_{x^*}(x,\lambda$

Since f is convex then Ψ_{x^*} is concave and therefore

v) $F(x^*,.)$ is concave $\forall x^*$.

For all $\lambda > m$ we denote by K_λ the barrier cone of $S_\lambda(f)$ i.e., $K_\lambda = \{x^*: F(x^*,\lambda) < +\infty\}$. Then K_λ is convex. Because the function $F(x^*,.)$ is concave and the interior of its domain is $(m,+\infty)$, then $F(x^*,.)$ is continuous on (m,∞) and we have

$$K_\lambda = K_\mu \qquad \forall\lambda,\ \mu > m.$$

Then in the following we denote by K the common barrier cone of level sets $S_\lambda(f)$ $\lambda > m$; furthermore, F is a finite convex-concave function on $K \times (m,+\infty)$. Let us denote by $\partial_{x^*}F(x^*,\lambda)$ the subdifferential of the convex function $F(.,\lambda^*)$ at x^*, $\partial_\lambda F(x^*,\lambda)$ the super-differential of the concave

concave function $F(x^*,.)$ at λ and define

$$S(x^*,\lambda) = \{x \in S_\lambda(f) : (x,x^*) = F(x^*,\lambda)\}.$$

Proposition 2.3

a) $S(x^*,\lambda) = \partial_{x^*} F(x^*,\lambda)$ for all $x^* \in K$, $\lambda > m$.

b) $S(x^*,\lambda)$ is non-empty for all $x^* \in ri(K)$, $\lambda > m$.

c) For $\lambda > m$ if $F(x^*,\lambda) < \delta^*(x^*|\text{dom } f)$ then $f(x) = \lambda$ for all $x \in S(x^*,\lambda)$.

Proof

a) Since $F(.,\lambda)$ is the support function of the closed convex set $S_\lambda(f)$ we have :

$$x \in \partial F_{x^*}(x^*,\lambda) \Leftrightarrow \delta(x|S_\lambda(f)) + F(x^*,\lambda) = (x,x^*) \Leftrightarrow x \in S(x^*,\lambda).$$

b) The subdifferential of a convex function is non-empty on the relative interior of its domain.

c) Suppose that $x \in S(x^*,\lambda)$ and $f(x) < \lambda$. Then $F(x^*,\mu) = F(x^*,\lambda)$ for all $\mu \in [f(x),\lambda]$.

Since $F(x^*,.)$ is concave, $F(x^*,\lambda) = \sup_\mu F(x^*,\mu) = \partial^*(x^*|\text{dom } f)$. □

Proposition 2.4 Assume that $x \in S(x^*,\lambda)$ and $f(x) = \lambda > m$.

a) if $F(x^*,\lambda) < \delta^*(x^*|\text{dom } f)$, then $\partial_\lambda F(x^*,\lambda) = \{\lambda^* / x^* \in \lambda^* \partial f(x)\}$.

b) if $F(x^*,\lambda) = \delta^*(x^*|\text{dom } f)$, then $\partial_\lambda F(x^*,\lambda) = \{\lambda^* / x^* \in \lambda^* \partial f(x)\} \cup \{0\}$.

Proof Set $\Phi(\mu) = -F(x^*,\mu)$, then Φ is a proper convex function which is continuous on $(m,+\infty)$. Consider Φ^*, the conjugate function of Φ :

$$\Phi^*(\mu^*) = \sup_{\mu,y} [\mu^*\mu + \langle x^*,y\rangle : f(y) \leqslant \mu].$$

Then

$$\Phi^*(\mu^*) = \begin{cases} +\infty & \text{if } \mu^* > 0, \\ \delta^*(x^*|\text{dom } f) & \text{if } \mu^* = 0, \\ (-\mu^*) f^* \left[-\dfrac{x^*}{\mu^*} \right] & \text{if } \mu^* < 0. \end{cases}$$

and therefore, since $f(x) = \lambda$ and $\Phi(\lambda) = -(x,x^*)$, λ^* belongs to $\partial\Phi(\lambda)$ iff

$$\Phi^*(\lambda^*) - (x,x^*) = \lambda^* f(x)$$

Then a) and b) follow straightforwardly. □

Conversely, we have

Proposition 2.5 If $f(x) > m$,

$$\partial f(x) = \{x^*/x \in S(x^*,f(x)) = \partial_{x^*}F(x^*,f(x)) \text{ and } 1 \in \partial_\lambda F(x^*,f(x)\}.$$

Proof Observe first that from sufficient optimality conditions we have

$$x^* \in \partial f(x) \Rightarrow x \in S(x^*,f(x)).$$

Now let x,x^* be such that $x \in S(x^*,f(x))$; then as before $-1 \in \partial\Phi(f(x))$ iff $f^*(x^*) - (x,x^*) = -f(x)$, that is iff $x^* \in \partial f(x)$. Furthermore, $-1 \in \partial\Phi(f(x))$ iff $1 \in \partial_\lambda F(x^*,f(x))$. □

In view of the strong connection between $\partial f(x)$ and $\partial_\lambda F(x^*,\lambda)$ we introduce the following function $\tilde{k}$ defined on $(m,+\infty)$

$$\tilde{k}(\lambda) = \inf_{x^*} \sup_{\mu} \{\frac{1}{\mu} : \mu \in \partial_\lambda F(x^*,\lambda),\ x^* \in K,\ ||x^*|| = 1\} \tag{2.13}$$

with by convention $\frac{1}{\mu} = +\infty$ if $\mu = 0$.

Proposition 2.5

a) $\tilde{k}(\lambda) \geq 0 \quad \forall\lambda > m$,

b) $\tilde{k}$ is non decreasing on $(m,+\infty)$,

c) $$\tilde{k}(\lambda) = \inf_{x^*} \sup_{\mu} \{\frac{1}{\mu}: \mu \in \partial_\lambda F(x^*,\lambda),\ x^* \in ri(K),\ ||x^*|| = 1\}. \tag{2.14}$$

Proof a) and b) are direct consequences of the monotonicity and concavity properties of functions $F(x^*,.)$.

c) Let $\bar{x}^* \in K$, $\bar{\mu} = \inf_{\mu}[\mu/\mu \in \partial_\lambda F(\bar{x}^*,\lambda)]$. We shall show that for any ξ such that $\xi < \bar{\mu}$, there is some $x^* \in ri(K)$ such that $\xi < \mu$ for all $\mu \in \partial_\lambda F(x^*,\lambda)$. Hence c) will follow.

Let $x_1^* \in ri(K)$ and $x_n^* = \bar{x}^* + \frac{1}{n}(x_1^* - \bar{x}^*)$. Then from Corollary 7.5.1 [9] we obtain

$$\lim_{n\to\infty} F(x_n^*,t) = F(\bar{x}^*,t) \qquad \text{for all } t \in (m,+\infty).$$

Set $\theta(t) = F(\bar{x}^*, t)$, $\theta_n(t) = F(x_n^*, t)$. Since θ is concave and

$$\xi < \bar{\mu} = \lim_{t \downarrow \lambda} \frac{\theta(t)-\theta(\lambda)}{t-\lambda}$$

there exists some $\bar{t} > \lambda$ such that

$$\xi > \frac{\theta(\bar{t})-\theta(\lambda)}{\bar{t}-\lambda}$$

But then there exists n such that

$$\xi < \frac{\theta_n(\bar{t})-\theta_n(\lambda)}{\bar{t}-\lambda} \leqslant \inf_{\mu} \left[\mu : \mu \in \partial\theta_n(\lambda)\right]$$

Take $x^* = x_n^*$. □

<u>Proposition 2.6</u> Let λ and λ' be such that $m < \lambda < \lambda' < +\infty$ and $S_{\lambda'}(f) \subset \mathrm{ri}(\mathrm{dom}\, f)$ then

$$\tilde{k}(\lambda) \leqslant k(\lambda) \leqslant \tilde{k}(\lambda') \leqslant k(\lambda')$$

so that k is non decreasing on $(m,+\infty)$ when (1.2) is satisfied.

a) Prove that $\tilde{k}(\lambda) \leqslant k(\lambda)$. Let x and c be such that $f(x) = \lambda$, $c \in \partial f(x)$ and $f'(x;c) < +\infty$. Such a couple always exists since $S_\lambda(f) \subset \mathrm{ri}(\mathrm{dom}\, f)$. Take $x^* = ||c||^{-1}c$, then $F(x^*,\lambda) = (x^*,x) < \delta^*(x^* | \mathrm{dom}\, f)$ and $x \in S(x^*,\lambda)$. Hence, by proposition 2.4

$$\sup_{\mu}\left[\frac{1}{\mu} : \mu \in \partial_\lambda F(x^*,\lambda)\right] = \sup_{\mu}\left[\mu : \mu x^* \in \partial f(x)\right] \leqslant \sup\left[(x^*,d) : d \in \partial f(x)\right]$$

It follows from 2.6 and 2.13 that $\tilde{k}(\lambda) \leqslant k(\lambda)$.

b) Prove that $k(\lambda) \leqslant \tilde{k}(\lambda')$. Of course, we assume that $\tilde{k}(\lambda') < +\infty$. Let any $\varepsilon > 0$, by relation (2.14) there is some $x^* \in \mathrm{ri}(K)$ such that $||x^*|| = 1$ and

$$\sup\left[\frac{1}{\mu} : \mu \in \partial_\lambda F(x^*,\lambda')\right] \leqslant \tilde{k}(\lambda') + \varepsilon .$$

Next, by propositions 2.3 and 2.4 there are some $x_o \in S(x^*,\lambda')$ and $\mu_o > 0$ such that

$$\mu_o x^* \in \partial f(x_o) \text{ and } \mu_o \leqslant \tilde{k}(\lambda') + \varepsilon .$$

Now, let x_1 be the Euclidean projection of x_o on the closed convex set $S_\lambda(f)$. Then $f(x_1) = \lambda$ and there exist $c \in \partial f(x_1))$ and $\mu_1 > 0$ such that

$x_o - x_1 = \mu_1 c$. Since f is convex we have

$$f'(x_1;||c||^{-1}c) \leq -f'(x_o; -||c||^{-1}c) \leq (||c||^{-1}c,d) \quad \text{for all } d \in \partial f(x_o)$$

then

$$k(\lambda) \leq f'(x_1; \frac{c}{||c||}) \leq (\mu_o x^*, ||c||^{-1}c) \leq \mu_o \leq \tilde{k}(\lambda') + \varepsilon$$

Letting $\varepsilon \to 0$ we deduce that $k(\lambda) \leq \tilde{k}(\lambda')$.

Proposition 2.7 Let f be a differentiable convex function on an open convex set C of $\mathbb{R}^n$ then $k(\lambda) = \tilde{k}(\lambda)$ for all $\lambda > m$.

Proof In this case $\inf(f'(x; \frac{d}{||d||}) \mid d \in \partial f(x)) = ||\nabla f(x)||$.
On the other hand, for all $x^* \in ri(K)$ such that $||x^*|| = 1$ and $\lambda > m$ $F(x^*,.)$ is differentiable at λ and $\nabla_\lambda F(x^*,\lambda) = ||\nabla f(x)||^{-1}$ where x is any point in $S(x^*,\lambda)$. Then it follows easily that $k(\lambda) = \tilde{k}(\lambda)$.

Proposition 2.8 Let λ and λ' be such that $m < \lambda < \lambda' < +\infty$ and $S_{\lambda'}(f) \subset ri(\text{dom } f)$ then

$$r(\lambda) \leq k(\lambda) \leq r(\lambda') \leq k(\lambda')$$

so that r is non decreasing on $(m,+\infty)$ when (1.2) is satisfied.

Proof Let $\lambda^* \in]\lambda,\lambda'[$, x and d be such that $f(x) = \lambda'$ and $d \in \partial f(x)$. Set $x^* = \frac{d}{||d||}$. Then $||d||^{-1} \in \partial_\lambda F(x^*,\lambda)$ and consequently, by concavity of $F(x^*,.)$

$$||d||^{-1} \leq \mu \quad \text{for all } \mu \in \partial_\lambda F(x^*,\lambda^*) \quad \lambda^* \in]\lambda,\lambda'[$$

Then from (2.13) it follows

$$\tilde{k}(\lambda^*) \leq ||d||$$

so that

$$\tilde{k}(\lambda^*) \leq r(\lambda')$$

Since from (2.7) $r(\lambda') \leq k(\lambda')$ then from proposition 2.6 it follows that

$$r(\lambda) \leq k(\lambda) \leq \tilde{k}(\lambda^*) \leq r(\lambda') \leq k(\lambda'). \qquad \square$$

<u>Proposition 2.9</u> Let f and g be two closed proper convex functions on $\mathbb{R}^N$ and $h = f \nabla g$ be the inf convolution of f and g :

$$h(x) = \inf(f(x-y) + g(y) \,|\, y \in \mathbb{R}^N)$$

Suppose that at each x the infinum is attained, and that f is inf-compact then $g \in \mathcal{F}$ iff $h \in \mathcal{F}$.

<u>Proof</u> 1) Set $m = \inf(h(x) \,|\, x \in \mathbb{R}^N)$, $m_1 = \inf(g(x) \,|\, x \in \mathbb{R}^N)$ $m_2 = \inf(f(x) \,|\, x \in \mathbb{R}^N)$. Obviously, we have :

$$m = m_1 + m_2. \tag{2.15}$$

2) Suppose that $g \in \mathcal{F}$ and let $\{x_n, x_n^*\}$ be a sequence such that

$$x_n^* \in \partial h(x_n) \qquad \lim_{n \to \infty} x_n^* = 0. \tag{2.16}$$

From proposition 6.6.4 [7] if y_n satisfies $h(x_n) = f(x_n - y_n) + g(y_n)$ then

$$\partial h(x_n) = \partial f(x_n - y_n) \cap \partial g(y_n).$$

Since f, $g \in \mathcal{F}$ it follows from (2.16) that

$$\lim_{n \to +\infty} g(y_n) = m_1, \ \lim_{n \to \infty} f(x_n - y_n) = m_2$$

and then from (2.15) we obtain

$$\lim_{n \to \infty} h(x_n) = m_1 + m_2 = m.$$

3) Suppose now that $h \in \mathcal{F}$ and let $\{y_n, y_n^*\}$ be a sequence such that

$$y_n^* \in \partial g(y_n), \qquad \lim_{n \to \infty} y_n^* = 0. \tag{2.17}$$

Since f is inf-compact, f^* is continuous at 0. It follows that there exists a neighbourhood V^* of 0 such that dom $\partial f \supset V^*$ and then for n large enough there exists z_n with $-y_n^* \in \partial f(z_n)$. Set $x_n = z_n + y_n$. Sufficient optimality conditions imply that $h(x_n) = f(x_n - y_n) + g(y_n)$ and from proposition 6.6.4 [7] $y_n^* \in \partial h(x_n)$. Since f and h belong to $\mathfrak{F}$ we obtain from (2.17) :

$$\lim_{n \to \infty} h(x_n) = m, \qquad \lim_{n \to \infty} f(z_n) = m_2$$

and it follows from (2.15) that

$$\lim_{n \to \infty} g(y_n) = m_1$$

Remark There are two particularly important cases where the infimum in h(x) is attained for each x. First when $f = ||.||$ and g is bounded from below, secondly when $f = \frac{1}{2}||.||^2$. We obtain then the corollary

Corollary 2.5 Let g be a closed proper convex function then

1) $g \in \mathfrak{F}$ iff $g \triangledown \frac{1}{2}||.||^2 \in \mathfrak{F}$

2) if g is bounded from below, $g \in \mathfrak{F}$ iff $g \triangledown ||.|| \in \mathfrak{F}$

Remark It is not obvious to find examples of convex functions defined on the whole space $\mathbb{R}^n$, differentiable everywhere which do not belong to the class $\mathfrak{F}_1$. Taking the function f given in the introduction by formula (1.0) and $g = f \triangledown \frac{1}{2}||.||^2$, it follows from the corollary that g is differentiable on $\mathbb{R}^2$ and does not belong to $\mathfrak{F}$.

Given g_1 and g_2 two closed proper convex functions having the same domain D we say that $g_1 \sim g_2$, if there exists a real valued strictly increasing continuous function k on $g_1(D)$ such that

$$g_2(x) = k[g_1(x)] \quad \forall x \in D.$$

Clearly, the relation $\sim$ is reflexive, symmetric and transitive. Denote by G the class of closed proper convex functions which are equivalent to g_1 by relation $\sim$. Then there exists in G a function $\tilde{g}$ which is minimal in the sense that for all $g \in G$, the function k_g, such that $g = k_g \circ \tilde{g}$, is convex (Debreu [5], Kannai [6]).

Proposition 2.10 If $g_1 \in \mathfrak{F}_1$ and $g_2 \sim g_1$ then $g_2 \in \mathfrak{F}_1$.

Proof Let us consider G the class of functions which are equivalent to g_1 and $\tilde{g}$ a minimal function in G then there exists two increasing convex functions k_1 and k_2 such that $g_1 = k_1 \circ \tilde{g}$ and $g_2 = k_2 \circ \tilde{g}$. Note that $\partial g_i(x) = \{\lambda x^* : \lambda \in \partial k_i(\tilde{g}(x)),\ x^* \in \partial\tilde{g}(x)\}$. It follows that $g_i \in \mathfrak{F}$ if and only if $\tilde{g} \in \mathfrak{F}$.

Examples The class $\mathfrak{F}_1$ contains, of course, the class of inf-compact convex functions. But many other functions belong to $\mathfrak{F}_1$. Let us consider, for instance, the positive semi-definite quadratic function

$$f(x) = \frac{1}{2}(x,Ax) - (a,x) + \alpha.$$

This function can be decomposed as

$$f(x_1,x_2) = \frac{1}{2}(x_1,A_1.x_1) - (a_1,x_1) - (a_2,x_2) + \alpha$$

with A_1 symmetric positive definite. If $a_2 \neq 0$ then $||\nabla f(x_1,x_2)|| \geqslant ||a_2||$ and necessarily $f \in \mathfrak{F}_1$. If $a_2 = 0$, then $\nabla f(x_1,x_2) = (A_1x_1-a_1, 0)$. It follows that if $\nabla f(x_1^n,x_2^n) \longrightarrow 0$, then $f(x_1^n,x_2^n) \longrightarrow \inf f(x_1,x_2)$.

An interesting example is the "Log x" entropy defined on the positive orthant by

$$h(x) = \sum_i \text{Log } x_i.$$

Then -h belongs to $\mathfrak{F}$ but is not inf-compact in contrast with the usual "x Log x" entropy. The fact that "x Log x" is inf-compact seems to be one of the reasons why the "x Log x" entropy has been often preferred to the "Log x" entropy, even if the last one has some interest. Entropy arises in various fields of applications including chemistry, image processing, statistics, ... A recent reference is Censor and Lent [4] where the "Log x" entropy is discussed. Closely related to the "Log x" entropy are the Cobb-Douglas functions

$$r(x) = \prod_{i=1} x_i^{\alpha_i}, \quad x_i > 0,\ \alpha_i > 0.$$

then $-r \in \mathfrak{F}$. The Cobb-Douglas functions are very much used in economics.

<u>Remark</u> The class $\mathcal{F}$ appears to be useful in several fields of mathematics dealing with the asymptotical behaviour of convex functions. Let us consider, for instance, the differential inclusion problem : Find x differentiable : $[0, +\infty[\to \mathbb{R}^n$ such that

$$\dot{x}(t) \in -\partial f(x(t)) \text{ for } t \in (0,+\infty), \; x(0) = x_o \in \text{dom } f \qquad (2.18)$$

When f is a proper closed convex function, then there exists a unique solution x to (2.18). If, in addition, the set of optimal solutions of (P) is non-empty, then x(t) converges to an optimal solution of (P) when $t \to +\infty$ (theorem 2, p.160 [1]).

When (P) has no optimal solutions but $f \in \mathcal{F}$, then we have the following result.

$$f(x(t)) \to m \text{ when } t \to +\infty.$$

To see that, we report to the proof given in [1] page 160. Defining $A_\varepsilon = \{t > 0 : \dot{x}(t) < \varepsilon\}$ we have $\text{meas}(A_\varepsilon) = \infty$. Then there exists a sequence $\{x(t_n)\}$ such that $c_n \in \partial f(x(t_n))$ and $c_n \to 0$. Since $f \in \mathcal{F}$, then $f(x(t_n)) \to m$. But the function $t \to f(x(t))$ decreases and therefore $f(x(t)) \to m$ when $t \to +\infty$.

III CONVERGENCE OF CLASSICAL ALGORITHMS

As said in the introduction, the convergence of classical algorithms in unconstrained optimization suppose in general that the generated sequences $\{x_n\}$ are bounded. (This will be the case in particular when f is inf-compact). In this section we shall restrict ourselves on two methods : the approximate proximal method and a gradient method, and we shall prove that convergence can be also obtained for unbounded sequences, provided that $f \in \mathcal{F}_1$.

3.1 Convergence of the approximate proximal method

Let $\{\varepsilon_n\}$ be a sequence of positive reals converging to zero. Then the classical approximate proximal method consists to generate, from a starting point x_o. a sequence $\{x_n\}$ by the rule :

$$\Phi_n(x_{n+1}) \leq \min(\Phi_n(x) \mid x \in \mathbb{R}^N) + \varepsilon_n \tag{3.1}$$

where Φ_n is given by :

$$\Phi_n(x) = f(x) + \frac{1}{2}||x-x_n||^2 \qquad \forall x \in \mathbb{R}^N \tag{3.2}$$

We shall assume in addition that

$$\Phi_n(x_{n+1}) \leq \Phi_n(x_n)$$

which is equivalent to :

$$f(x_{n+1}) + \frac{1}{2}||x_{n+1} - x_n||^2 \leq f(x_n) \tag{3.3}$$

<u>Remark</u> Given x_n, such a point x_{n+1} always exists and is obtained in the following way. Let $\bar{x}_n$ be such that

$$\Phi_n(\bar{x}_n) \leq \min(\Phi_n(x) \mid x \quad \mathbb{R}^N) + \varepsilon_n$$

If $\Phi_n(\bar{x}_n) \leq \Phi_n(x_n)$ set $x_{n+1} = \bar{x}_n$ otherwise set $x_{n+1} = x_n$.

<u>Theorem 3.1</u> Suppose that $f \in \mathcal{F}_1$, then $\{x_n\}$ is a minimizing sequence.

<u>Proof</u> By construction, the sequence $\{f(x_n)\}$ is non increasing. Denote by ℓ its limit and assume for contradiction that $\ell > m$. From (3.3), it follows that $||x_{n+1} - x_n||$ converges to 0.

Denote by y_n the point where Φ_n reaches its minimum. Since Φ_n is strongly convex, we have :

$$\frac{1}{2}||x_{n+1} - y_n||^2 + \Phi_n(y_n) \leqslant \Phi_n(x_{n+1}) \leqslant \Phi_n(y_n) + \varepsilon_n$$

from what we deduce

$$||x_{n+1} - y_n|| \leqslant \sqrt{2\varepsilon_n} \tag{3.4}$$

and $||x_n - y_n||$ converges to 0. On the other hand,

$$f(x_{n+1}) + \frac{1}{2}||x_{n+1}-x_n||^2 \leqslant f(y_n) + \frac{1}{2}||y_n-x_n||^2 + \varepsilon_n \leqslant f(x_n) + \varepsilon_n$$

from what we deduce that $\{f(y_n)\}$ converges to ℓ.

Let $\lambda \in (m,\ell)$. There exists n_o such that $f(y_n) \geqslant \lambda$ for all $n \geqslant n_o$. From optimality conditions there exists $d_n \in \partial f(y_n)$ such that

$$d_n + (y_n - x_n) = 0.$$

But then $\{d_n\}$ converges to 0, in contradiction with $f \in \mathcal{F}_1$ and $f(y_n) \geqslant \lambda > m$.

3.2 Convergence of a proximal-gradient method

In this section we suppose that f is differentiable on $\mathbb{R}^N$. We shall modify slightly the stepsize rules of the gradient method. Instead of minimizing f along the descent half-line $\{x_n - t\nabla f(x_n) \mid t \geqslant 0\}$, we shall minimize the regularized proximal function $\Phi_n = f + \frac{1}{2}||. - x_n||^2$.

This is necessary for defining correctly the algorithm. Indeed, if f is not inf-compact there does not exist necessarily a real t_n which minimizes f on this half line.

Furthermore, in order to obtain convergence results we suppose in addition that $\nabla f(.)$ is uniformly continuous on each level set of f, that is to say :

$$\lim_{\delta \to 0} w_\lambda(\delta) = 0 \qquad \forall \lambda > m$$

where

$$w_\lambda(\delta) = \sup(||\nabla f(x)-\nabla f(x')|| \mid ||x-x'|| \leqslant \delta,\ f(x) \leqslant \lambda,\ f(x') \leqslant \lambda).$$

This assumption is satisfied trivially for example for quadratic functions, $f(x) = e^{-x}$, ...

3.2.1 Proximal-gradient method with exact minimization

Starting from an arbitrary point x_o we construct the sequence $\{x_n\}$ by the following rule : suppose x_n computed. If $\nabla f(x_n) = 0$ stop, else

$$x_{n+1} = x_n - t_n \nabla f(x_n) \tag{3.5}$$

where t_n minimizes on $\mathbb{R}_+$ the function $t \to \Phi_n(x_n - t\nabla f(x_n))$.

<u>Theorem 3.2</u> Suppose that $f \in \mathcal{F}_1$ then for each n $f(x_{n+1}) \leqslant f(x_n)$ and $\{x_n\}$ is a minimizing sequence.

<u>Proof</u> Without loss of generality we can suppose that $\nabla f(x_n) \neq 0$ for all n. Then x_{n+1} satisfies the equation

$$(\nabla f(x_{n+1}) + x_{n+1} - x_n, \nabla f(x_n)) = 0 \tag{3.6}$$

and we have

$$f(x_{n+1}) + \frac{1}{2}||x_{n+1} - x_n||^2 \leqslant f(x_n) \tag{3.7}$$

From (3.7) it follows that

$$f(x_{n+1}) \leqslant f(x_n) \quad \forall n \tag{3.8}$$

Set $\ell = \lim_{n \to \infty} f(x_n)$. If $\ell = -\infty$, the theorem is proved, if not ℓ is finite and from (3.7) we obtain then :

$$\lim_{n \to +\infty} ||x_{n+1} - x_n|| = 0 \tag{3.9}$$

Dividing both numbers of (3.6) by $||\nabla f(x_n)||$ we obtain :

$$-||\nabla f(x_n)|| = (\nabla f(x_{n+1}) - \nabla f(x_n), \frac{\nabla f(x_n)}{||\nabla f(x_n)||}) + (x_{n+1} - x_n, \frac{\nabla f(x_n)}{||\nabla f(x_n)||})$$

so that

$$||\nabla f(x_n)|| \leqslant w_{f(x_o)}(||x_{n+1} - x_n||) + ||x_{n+1} - x_n||$$

and from (3.9) we obtain then

$$\lim_{n \to \infty} \nabla f(x_n) = 0$$

Since $f \in \mathcal{F}_1$ it follows that $\{x_n\}$ is a minimizing sequence.

Remark This algorithm can be also interpreted as being the proximal method in which, instead of minimizing at each step the function Φ_n on $\mathbb{R}^N$, one does only the first step of a gradient method.

Except for quadratic functions, minimizing Φ_n on the half line $\{x_n - t\nabla f(x_n) \mid t \geqslant 0\}$ is not in general an implementable rule. We introduce now, an implementable step-size rule (Goldstein-Armijo) for which we shall see that convergence can also be obtained.

3.2.2 Gradient proximal method with Goldstein-Armijo stepsize rule

Starting from an arbitrary point x_o we construct the sequence $\{x_n\}$ by the following rule :

Suppose x_n computed. If $\nabla f(x_n) = 0$ stop. Else :

$$x_{n+1} = x_n - t_n \frac{\nabla f(x_n)}{||\nabla f(x_n)||^2} \tag{3.10}$$

where t_n is given by the implementable rule :

$$t_n = \sup\{t : t = \frac{1}{2^i}, \ i \in \mathbb{N} : \Phi_n(x_n - t \frac{\nabla f(x_n)}{||\nabla f(x_n)||^2}) - f(x_n) \leqslant -t^2\} \tag{3.11}$$

Proposition 3.3 If $\nabla f(x_n) \neq 0$ then there exists i_n such that

$$t_n = \frac{1}{2^{i_n}}$$

Proof Let us remark first that

$$\Phi_n(x_n) = f(x_n), \ \nabla\Phi_n(x_n) = \nabla f(x_n)$$

so that (3.6) is equivalent to :

$$t_n = \sup\{t : t = \frac{1}{2^i} \ i \in \mathbb{N} : \Phi_n(x_n - t \frac{\nabla\Phi_n(x_n)}{||\nabla\Phi_n(x_n)||^2}) - \Phi_n(x_n) \leqslant -t^2\} \tag{3.12}$$

If the supremum in (3.11) is not reached then for each $i > 1$, we have :

$$\frac{\Phi_n(x_n - \frac{1}{2^i} \frac{\nabla\Phi_n(x_n)}{||\nabla\Phi_n(x_n)||^2}) - \Phi_n(x_n)}{\frac{1}{2^i}} > -\frac{1}{2^i} > -\frac{1}{2}$$

Passing to the limit we obtain then

$$-1 \geqslant -\frac{1}{2}$$

which yields to a contradiction. □

<u>Theorem 3.4</u> The sequence $\{f(x_n)\}$ converges to m.

<u>Proof</u> From (3.10) and (3.11) we obtain

$$\frac{1}{2}||x_{n+1}-x_n||^2 + f(x_{n+1}) \leqslant f(x_n) - t_n^2 \tag{3.13}$$

Then the sequence $\{f(x_n)\}$ is non increasing. Let ℓ be its limit. Assume for contradiction that $\ell > m \geqslant -\infty$. Since $f(x_n) - f(x_{n+1})$ tends to 0, then from (3.13)

$$t_n \to 0 \text{ and } ||x_{n+1}-x_n|| \to 0 \text{ when } n \to \infty \tag{3.14}$$

By definition of t_n we have :

$$\phi_n(x_n - 2t_n \frac{\nabla f(x_n)}{||\nabla f(x_n)||^2}) - \phi_n(x_n) > -4t_n^2$$

and

$$\phi_n(x_n - t_n \frac{\nabla f(x_n)}{||\nabla f(x_n)||^2}) - \phi_n(x_n) \leqslant -t_n^2$$

By continuity of ϕ_n, there exists $\theta_n \in [1,2)$ such that

$$\phi_n(x_n - \theta_n t_n \frac{\nabla f(x_n)}{||\nabla f(x_n)||^2}) - \phi_n(x_n) = -\theta_n^2 t_n^2 < 0 \tag{3.15}$$

and by the mean value theorem, there exists $\xi_n \in (0,1)$ such that

$$\theta_n t_n (\frac{\nabla f(x_n)}{||\nabla f(x_n)||^2}, \nabla f(y_n) + y_n - x_n) = \theta_n^2 t_n^2 \tag{3.16}$$

where $y_n = x_n - \theta_n \xi_n t_n \dfrac{\nabla f(x_n)}{||\nabla f(x_n)||^2}$, (3.17)

then

$$(\frac{\nabla f(x_n)}{||\nabla f(x_n)||^2}, \nabla f(y_n)) = \theta_n t_n (1 - \frac{\xi_n}{||\nabla f(x_n)||^2}). \tag{3.18}$$

Since $f \in \mathfrak{F}$ and $f(x_n) \geqslant \ell > m$, there exists $\alpha > 0$ such that

$$||\nabla f(x_n)|| \geqslant \alpha \text{ for all } n$$

and the second member of (3.18) converges to 0.

On the other hand, since ϕ_n is convex and $\xi_n \in (0,1)$, we deduce from (3. 15) that

$$f(y_n) \leqslant \phi_n(y_n) \leqslant \phi_n(x_n) = f(x_n) \leqslant f(x_o).$$

By (3.17), we have also

$$||y_n - x_n|| = \theta_n \xi_n ||x_{n+1} - x_n|| \leqslant 2||x_{n+1} - x_n||$$

and

$$\left| \left(\frac{\nabla f(x_n)}{||\nabla f(x_n)||^2}, \nabla f(y_n) \right) - 1 \right| = \left| \left(\frac{\nabla f(x_n)}{||\nabla f(x_n)||^2}, \nabla f(x_n) - \nabla f(y_n) \right) \right| \leqslant$$

$$\frac{||\nabla f(x_n) - \nabla f(y_n)||}{||\nabla f(x_n)||} \leqslant \frac{1}{\alpha} \quad w_{f(x_o)} \quad (2||x_{n+1} - x_n||).$$

Since $||x_{n+1} - x_n|| \longrightarrow 0$, the first member of (3.18) tends to 1 and we have here a contradiction. □

REFERENCES

[1] H.P. Aubin and A. Cellina, Differential Inclusions, Springer Verlag,Berlin-Heidelberg (1981).

[2] A. Auslender and J.-P. Crouzeix, "Global regularity theorems", to appear in Math. of Oper. Research.

[3] A. Brøndstedt and R.T. Rockafellar, "On the subdifferentiability of convex functions",Proceedings of the American Mathematical Society, 16 (1965) pp. 605-611.

[4] Y. Censor and A. Lent, "Optimization of "Log x" entropy over linear equality constraints", S.I.A.M. Journal Control & Optimization, Vol. 25 (1987) pp. 921-933.

[5] Debreu "Least concave utility functions", Journal of Math. Economics, Vol.3 (1976 pp. 121-129.

[6] Kannai "Concavifiability of convex preferences", Journal of Math. Economics, Vol.4 (1977) pp. 1-56.

[7] P.J. Laurent, Approximation et optimisation, Hermann (1972).

[8] R.T. Rockafellar "Monotone operators and the proximal point algorithm", S.I.A.M. Journal on Control and Optimization 14 (1976) pp. 877-898.

[9] R.T. Rockafellar, Convex Analysis, Princeton University Press - Princeton, N.J. 1970 .