J-B HIRIART-URRUTY

ε- subdifferential calculus

J.-B. HIRIART-URRUTY

Université de Clermont
Complexe Scientifique des Cezeaux
Département de Mathematiques Appliquées
Boite Postale 45
63170 Aubière
France

PURPOSE AND SCOPE

The appropriate subdifferential of a convex function has been proved to be a useful tool in convex analysis, from the theoretical viewpoint as well as for practical purposes. Given a convex function f defined on X, the approximate subdifferential $\partial_{\bullet} f(.)$ which assigns to $(x,\varepsilon) \in X \times \mathbb{R}_+$ the so-called ε-subdifferential of f at x turns out to have remarkable properties. The properties of the subdifferential set-valued mapping

$x \rightrightarrows \partial f(x)$ (i.e. for $\varepsilon = 0$) as well as calculus rules on

subdifferentials are widely known in convex analysis [3, 8, 12, 22, 28, 33, 40]. As for the ε-subdifferential $\partial_{\varepsilon} f$, whose definition is just a "perturbation by ε" of that of ∂f, it enjoys, for $\varepsilon > 0$, some noteworthy properties different as for their nature from those of the "exact" subdifferential. Moreover, *exact* rules on ε-subdifferentials, generalizing those established for $\varepsilon = 0$, do not seem to be well-known. Our aim in this paper is two-fold : firstly to get a better insight into the local behaviour of $\partial_{\varepsilon} f(x)$ around a given $x_o \in X$, secondly, to survey the main chain rules on ε-subdifferentials.

One of the reasons why the properties of ∂f and $\partial_{\varepsilon} f$ may be different is that ∂f is a local notion while $\partial_{\varepsilon} f$ is a global one. To be more explicit, the ε-subdifferential of f at x_o is defined as the set of $x^* \in X^*$ satisfying

$$f(x) \geq f(x_o) + \langle x^*, x-x_o \rangle - \varepsilon$$

for all x. $\partial_{\varepsilon} f(x_o)$ may be very sensitive to the variations of f, even when those variations do not hold in a neighborhood of x_o. From the geometrical viewpoint, $\partial_{\varepsilon} f(x_o)$ is closely related to the closed (convex) cone with apex $(x_o, f(x_o) - \varepsilon)$ generated by the epigraph of f (epif). When $\varepsilon > 0$, this cone is known to be composed of two convex cones: the cone with apex $(x_o, f(x_o) - \varepsilon)$ generated by epif and the asymptotic cone (or recession cone) of epif [9]. The latter cone takes into account the behaviour of f when x is "far" from x_o. The necessity, a priori, of knowing the behaviour of f on all of X for the calculation of $\partial_\varepsilon f(x_o)$ is corroborated by the fundamental approximation result due to Brøndsted and Rockafellar[7]. Roughly speaking, the result states that the more precisely you desire to know $\partial_\varepsilon f(x_o)$, the farther from x_o you need to know $\partial f(x)$. However, this disadvantage is weighted against the good effects wrought by the perturbation by ε. For example, the use of the ε-directional derivative $f'_\varepsilon(x_o;.)$(i.e. the support function of $\partial_\varepsilon f(x_o)$) as a substitute for the usual directional derivative has been proved advantageous in many algorithms of convex optimization; see [2,29,30]and references therein. As for the continuity property of $\partial_. f(.)$ as a set-valued mapping of both x and ε, the main result goes back to the comprehensive study by Asplund and Rockafellar [O]. In particular, they proved that for a lower-semicontinuous convex function defined on a Banach space, the approximate subdifferential

$\partial_. f(.) : (x,\varepsilon) \rightrightarrows \partial_\varepsilon f(x)$ was continuous in the Hausdorff sense on int(domf) $\times \mathbb{R}^*_+$. More recently, Nurminskii [35] showed the locally Lipschitzian behaviour of the ε-subdifferential of a finite convex function defined on $\mathbb{R}^n$; this result was improved and generalized in various ways in the author's companion paper [19]. Section I and VII of the present study are along the lines of the above described local properties of the approximate subdifferential.

Like for the subdifferentials, calculus rules on ε-subdifferentials are of importance; in many situations it is of interest to have expressions of ε-subdifferentials of a convex function g which has been constructed from other convex functions whose properties are better known. For example, g might be a sum or supremum of convex functions f_i. Various formulas yielding *some* elements of $\partial_\varepsilon g(x_o)$ in terms of the $\partial_\eta f_i(x_o)$ have been established in the literature. However, very little seems to be known concerning the *exact* expression of $\partial_{\varepsilon} g(x_o)$ in terms of the $\partial_{\eta} f_i(x_o)$, except maybe when g is a sum of functions ($g = f_1+f_2$) or results from the composition of f with a linear mapping $A(g = f \circ A)$. In this paper we drew up a panorama of the main calculus rules on ε-subdifferentials and we join to them various conditions sufficient for this or that formula to be valid. The reader will have noted that all the chain rules presented here generalize what is known for subdifferentials and the sufficient conditions for their applicability are exactly the same as for $\varepsilon = 0$. Concerning these calculus rules, one should mention that general results dealing with convex operators (i.e. convex vector-valued functions) have been announced recently in a note by Kutateladze [27]. To a certain extent, Kutateladze's formulas cover the greatest part of chain rules displayed here. Nevertheless, we seize the opportunity of treating real-valued functions to present conditions of applicability peculiar to that context and to provide the proofs of formulas.

We assume that the reader is familiar with basic definitions and properties from convex analysis.

1. PRELIMINARY DEFINITIONS AND PROPERTIES

As it is customary in the context of convex analysis, we work in the setting of two locally convex (real) topological vector spaces X and X^* *paired in duality* by a bilinear form $(x,x^*) \mapsto \langle x,x^*\rangle$ (see, e.g., [33,§6], [28, §6.3] or [41,§3]). The most usual example of paired spaces is obtained by considering as X a locally convex Hausdorff topological vector space, as X^* the topological dual space of X, and $\langle x,x^*\rangle = x^*(x)$ as the pairing on $X \times X^*$.

Throughout we shall deal with *proper convex* functions (a function f is said to be *proper* if f is not identically equal to $+\infty$ and if $f(x) > -\infty$ for all x), and we shall denote by $\Gamma^0(X)$ the set of proper convex functions which are *lower-semicontinuous* (l.s.c.).

Given a proper function f, the ε-*subdifferential* of f at $x_0 \in \operatorname{dom} f$ (dom f is the set where f is finite) is defined *for each* $\varepsilon \geq 0$ as the set of vectors $x^* \in X^*$ satisfying

$$f(x) \geq f(x_0) + \langle x^*, x-x_0\rangle - \varepsilon \tag{1.1}$$

for all $x \in X$.

The set of such vectors, denoted by $\partial_\varepsilon f(x_0)$, is closed convex set in X^* which reduces to the subdifferential $\partial f(x_0)$ when $\varepsilon = 0$. Moreover, if $f \in \Gamma_0(X)$, $\partial_\varepsilon f(x_0)$ is nonempty whenever ($x_0 \in \operatorname{dom} f$ and) $\varepsilon > 0$. Geometrically (1.1) says that the epigraph of the affine function passing through $(x_0, f(x_0)-\varepsilon)$ and of slope x^* contains the epigraph of f. This definition, as it is, shows that the behaviour of f on the whole space X may be relevant to the construction of $\partial_\varepsilon f(x)$ for $\varepsilon > 0$. There are two fundamental ways of characterizing $\partial_\varepsilon f(x)$: through the conjugate function f^* and with its support function. Since they both will be used in the sequel, we recall these characterizations as well as some easy consequences which can be derived from them.

Proposition 1.1

$$x^\star \in \partial_\varepsilon f(x_o) \text{ \underline{if and only if}} :$$
$$f(x_o) + f^\star(x^\star) - \langle x_o, x^\star\rangle \leq \varepsilon \ . \qquad \square \qquad (1.2)$$

When $f \in \Gamma_o(X)$, f and $f^\star (\in \Gamma_o(X^\star))$ play a symmetric role; then (1.2) is equivalent to : $x_o \in \partial_\varepsilon f^\star(x^\star)$. As the above formula illustrates , the knowledge of $f^\star$ suffices for the calculation of $\partial_\varepsilon f(x_o)$. Actually, all the chain rules concerning the ε-differentials which will be displayed later hinge on formula (1.2).

Example 1.1. Let $f \in \Gamma_o(X)$ be positively homogeneous (i.e.) $f(\lambda x) = \lambda f(x)$ for all $x \in X$ and $\lambda > 0$). For such an f, $f^\star$ is the indicator function of $\partial f(0)$. Hence, for all $x_o \in \text{dom } f$,

$$\partial_\varepsilon f(x_o) = \{x^\star \in \partial f(0) \mid \langle x_o, x^\star\rangle \geq f(x_o) - \varepsilon\}.$$

In particular, let X = E be a normal vector space and $X^\star = E'$ be its topological dual space endowed, for example, with the $\sigma(E',E)$- topology; let ϕ designate the normal function on E and ψ the dual norm on E'. Then

$$\partial_\varepsilon \phi(x_o) = \{x^\star \mid \psi(x^\star) \leq 1, \langle x_o, x^\star\rangle \geq \phi(x_o) - \varepsilon\}. \qquad \square$$

Example 1.2. Let C be a nonempty closed convex set in X. The set $N_\varepsilon(C;x_o)$ of ε-*normals* to C at $x_o \in C$ is defined as the ε-subdifferential of the indicator function $\delta(.|C)$ at x_o, i.e.

$$N_\varepsilon(C;x_o) = \{x^\star \in X^\star \mid \langle x^\star, x-x_o\rangle \leq \varepsilon \text{ for all } x \in C\}.$$

In a dual formulation, $x^\star \in N_\varepsilon(C;x_o)$ if and only if

$$\delta^*(x^*|C) - \langle x_o, x^* \rangle \leq \varepsilon. \qquad \square$$

As a general rule, $N_\varepsilon(C;x_o)$ is no more a cone so that duality results on cones cannot be invoked when dealing with it. Incidentially, note that the notion of ε-normality is unhelpful for defining a concept of "approximate tangent cone" to C at x_o. What can only be said in general is that $N_\varepsilon(C;x_o)$ is a closed convex set contained in the barrier cone of C, and whose recession cone is $N(C;x_o)$.

Example 1.3. Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a polyhedral convex function [40, Section 19]. For such a function, the level sets are polyhedral convex sets and the conjugate function is polyhedral. Consequently, for all $x_o \in \text{dom } f$, $\partial_\varepsilon f(x_o)$ is a polyhedral convex set. $\square$

The characterization of $\partial_\varepsilon f(x_o)$ in terms of its support function is given by the following result [40, pp. 219-220],[33, p. 671].

Proposition 1.2. Let $f \in \Gamma_o(X)$; then the support function of $\partial_\varepsilon f(x_o)$ is given by

$$d \longmapsto f'_\varepsilon(x_o;d) = \inf_{\lambda > o} \frac{f(x_o+\lambda d) - f(x_o) + \varepsilon}{\lambda}. \qquad \square \quad (1.3)$$

Observe that, as it is, the formula giving $f'_\varepsilon(x_o;d)$ again emphasizes that, for $\varepsilon > 0$, the infimum of the approximate differential quotient $[f(x_o+\lambda d) - f(x_o) + \varepsilon]\lambda^{-1}$ (or an infimum within $\alpha > 0$) may be achieved "very far" from the concerned point x_o. The example of $x \longmapsto |x|$ at $x_o = 0$ is illustrative of that.

Example 1.4. Let f be a *quadratic* function defined on $\mathbb{R}^n$ as

$$f(x) = \tfrac{1}{2}\langle Ax, x \rangle + \langle b, x \rangle + c,$$

where A is a symmetric positive-definite $n \times n$ matrix, b a vector in $\mathbb{R}^n$ and c a real number. This is an example where the ε-directional derivative $f'_\varepsilon(x_o;d)$ is easy to calculate. Given

$x_o \in \mathbb{R}^n$ and a d direction, the definition (1.3) yields for the particular f involved

$$f'_\varepsilon(x_o;d) = \langle Ax_o,d\rangle + \langle b,d\rangle + (2\varepsilon \langle Ad,d\rangle)^{\frac{1}{2}}.$$

The function $d \mapsto (2\varepsilon\langle Ad,d\rangle)^{\frac{1}{2}}$ is known to be the support function of the set $\{x^* | \langle x^*, A^{-1}x^*\rangle \leq 2\varepsilon\}$ (see [40, p.119]). Hence we have that

$$\partial_\varepsilon f(x_o) = Ax_o + b + \{x^* | \langle x^*, A^{-1}x^*\rangle \leq 2\varepsilon\}$$

$$= \nabla f(x_o) + \{Ay^* | \langle Ay^*, y^*\rangle \leq 2\varepsilon\}. \quad (*)$$

We note in this example that even if $f'(x_o;d) = \lim_{\varepsilon \to 0^+} \downarrow f'_\varepsilon(x_o;d)$, the difference $f'_\varepsilon(x_o;d) - f'(x_o;d)$ may decrease slowly, as slowly as $\varepsilon^{\frac{1}{2}}$.

For fixed $\varepsilon > 0$ $x_o \in \text{dom } f$ $(f \in \Gamma_o(X))$ and $d \neq 0$, the behaviour of the function

$$q_f : \lambda \longmapsto \frac{f(x_o+\lambda d) - f(x_o) + \varepsilon}{\lambda}$$

on $\mathbb{R}^*_+$ is of particular importance in the way of approximating $f'_\varepsilon(x_o;d)$. Of course, d is chosen among those which are of interest , i.e. those for which there exists $\bar{\lambda} \in]0,+\infty]$ such that $\{x_o + \lambda d | 0 \leq \lambda < \bar{\lambda}\} \subset \text{dom } f$. The behaviour of q_f near 0^+ and $+\infty$ is known since

$$\lim_{\lambda \to 0^+} q_f(\lambda) = +\infty \quad ,$$

(*) Of course, there are various ways of obtaining this formula; see for example [30, p. 38].

and

$$\lim_{\lambda\to+\infty} q_f(\lambda) = \sup_{\lambda>0} \frac{f(x_0+\lambda d) - f(x_0)}{\lambda} = f_\infty(d).$$

Here, f_∞ is what is known as the *recession function* of f (or the *asymptotic function* of f). The function $f_d : \lambda \longrightarrow f(x_0+\lambda d)$ occurring in the numerator of q_f is a function of $\Gamma_0(\mathbb{R})$ which is finite at least on $[0,\bar{\lambda}[$. Obviously q_f is a lower- semi-continuous quasi-convex function on $\mathbb{R}_+^*$. Even more, q_f enjoys a *pseudo-convexity* property in the sense that the stationary points of q_f in $\mathbb{R}_+^*$ are also the global minima of q_f on $\mathbb{R}_+^*$. Let us make this more precise.

Proposition 1.3. $\lambda_0 \in \mathbb{R}^*$ is a global minimum of q_f on $\mathbb{R}^*$ if and only if

$$q_f(\lambda_0) \in \partial f_d(\lambda_0). \qquad \square \quad (1.4)$$

Proof. Immediate from the definitions. $\square$

Under mild assumptions on f, we have that

$$\partial f_d(\lambda_0) = \langle \partial f(x_0+\lambda_0 d), d\rangle \quad ,$$

so that (1.4) is rewritten as

$$q_f(\lambda_0) \in \langle \partial f(x_0+\lambda_0 d), d\rangle. \qquad (1.5)$$

This relation deserves some more explanation. For the sake of simplicity, let X = E be a Banach space and $f : E \longrightarrow \mathbb{R}$ be a continuous convex function. Such a function is locally Lipschitz on E and q_f is now locally Lipschitz from $\mathbb{R}_+^*$ into $\mathbb{R}$. A necessary condition for $\lambda_0 \in \mathbb{R}_+^*$ to be a minimum of q_f on $\mathbb{R}_+^*$ is that

$$0 \in \partial\, q_f(\lambda_0) \quad , \qquad (1.6)$$

where ∂q_f stands for Pshenichnyi's quasi-differential [38, Ch. III]or, equivalently here, Clarke's generalized gradient of q_f [10]. It is now just a matter of applying existing chain rules [4], [17, 18] to obtain that

$$\partial q_f(\lambda_o) = \frac{1}{\lambda_o} [<\partial f(x_o+\lambda_o d), d> - q_f(\lambda_o)].$$

Plugging this expression in the condition (1.6) just yields (1.5).

The reason why the necessary condition for optimality (1.6) is also sufficient can be explained by the following:

<u>Proposition 1.4</u>. <u>The function</u> r_f <u>defined on</u> $\mathbb{R}_+^*$ <u>by</u>

$$r_f(\mu) = q_f(\frac{1}{\mu})$$

<u>is convex</u>.

While dealing with finite coercive functions, Lemaréchal and Nurminskii [31] noticed the above property as a by-product of a duality result. However, the result can be derived in the general case from noticing that the function $\mu \longmapsto \mu\, f(x_o + \frac{d}{\mu})$ is convex on $\mathbb{R}_+^*$. The latter merely comes from the following : given an interval $I \subset \mathbb{R}_+^*$, $\theta(\mu)$ is convex on I if and only if $\mu\,\theta(\frac{1}{\mu})$ is convex on I.

Obviously, $\lambda_o \in \mathbb{R}_+^*$ is a minimum of q_f on $\mathbb{R}_+^*$ if and only if $\mu_o = 1/\lambda_o$ is a minimum of r_f on $\mathbb{R}_+^*$. Now the necessary and sufficient condition for optimality of $\mu_o > 0$, $0 \in \partial r_f(\mu_o)$, is made equivalent to (1.6) by a chain rule (on generalized gradients of arbitrary functions) which states that

$$\partial r_f(\mu_o) = -\partial q_f(\lambda_o)\lambda_o^2 \qquad (\lambda_o\mu_o = 1).$$

The next example illustrates the foregoing.

<u>Example 1.5</u>. Let $f : \mathbb{R} \longrightarrow (-\infty, +\infty]$ be defined by

$$f(x) = \begin{cases} x & \text{if } 0 \leq x \leq 1, \\ 2x-1 & \text{if } x \geq 1, \\ +\infty & \text{if } x < 0. \end{cases}$$

Let $x_o = 0, d = +1$ and $\varepsilon = 1/2$. Then

$$q_f(\lambda) = \begin{cases} 1+1/2\lambda & \text{if } 0 < \lambda \leq 1, \\ 2-1/2\lambda & \text{if } \lambda \geq 1, \end{cases}$$

and the only λ_o for which $q_f(\lambda_o) \in \partial f(\lambda_o)$ is $\lambda_o = 1$.
It might be more advantageous to work with r_f rather than q_f. From the computational viewpoint, it is fairly easier to minimize the *convex* function r_f on $\mathbb{R}_+$ especially as

$$\lim_{\mu \to +\infty} r_f(\mu) = +\infty ,$$

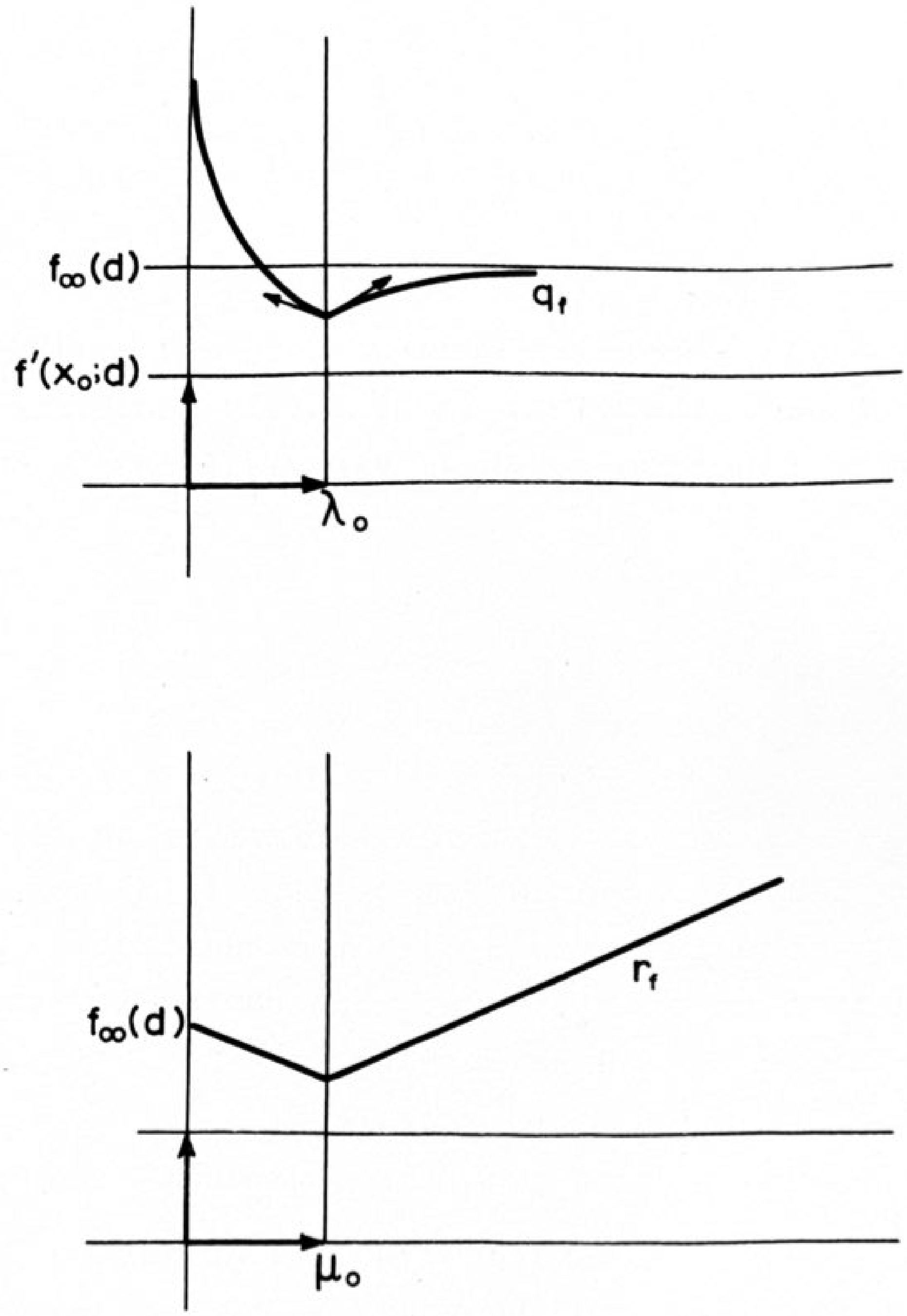
$$\lim_{\mu \to 0^+} r_f(\mu) = f_\infty(d) > -\infty$$

The case where q_f does not achieve the infimum value $f'_\varepsilon(x_o;d)$ on $\mathbb{R}^*_+$ corresponds to the situation where $q_f(\lambda) > f_\infty(d)$ for all $\lambda > 0$; in terms of the function r_f, that means that $\mu_o = 0$ is the unique minimum of r_f on $\mathbb{R}_+$. One then can be led to the consideration of those $\lambda_\alpha > 0$ satisfying

$$q_f(\lambda_\alpha) \leq f'_\varepsilon(x_o;d) + \alpha \qquad (\alpha > 0). \tag{1.7}$$

When f is a convex Lipschitz function on a Banach space (with Lipschitz constant $r > 0$), one easily checks that all the λ_α satisfying (1.7) belong to the interval $\left[\frac{\varepsilon}{2r\|d\| + \alpha} , +\infty \right[$.

As noticed above, a necessary and sufficient condition for q_f to achieve its minimum on $\mathbb{R}^*_+$ is to suppose that there exists $\lambda_* > 0$ for which $q_f(\lambda_*) \leq f_\infty(d)$. That is certainly true for the d directions satisfying $f_\infty(d) = +\infty$.



The functions in Example 1·5

2. THE ε-SUBDIFFERENTIALS OF $f_1 + f_2$ AND $f \circ A$.

2.1. Let f_1 and f_2 be two proper convex functions on X. As indicated in the previous paragraph, the description of the ε-subdifferentials of f_1+f_2 can be derived from the expression of $(f_1+f_2)^*$ in terms of f_1^* and f_2^*. The basic assumption for that is the following :

$$(H^+) \qquad (f_1+f_2)^*(x^*) = \min\{f_1^*(x_1^*) + f_2^*(x_2^*) \mid x^* = x_1^* + x_2^*\} \quad \forall x^*.$$

Theorem 2.1. Let f_1 and f_2 be two proper convex functions for which (H^+) holds. Then

$$\partial_\varepsilon (f_1+f_2)(x_0) = \bigcup_{\substack{\varepsilon_1 \geq 0, \varepsilon_2 \geq 0 \\ \varepsilon_1+\varepsilon_2 = \varepsilon}} \{\partial_{\varepsilon_1} f_1(x_0) + \partial_{\varepsilon_2} f_2(x_0)\} \quad (2.1)$$

for all $x_0 \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$. □

Proof. As for the proofs of all chain rules in the sequel, use the characterization given in Proposition 1.1. □

There are various assumptions guaranteeing (H^+) which are displayed in the literature devoted to convex analysis. Let us list some of them, beginning with the finite-dimensional case $X = \mathbb{R}^n$.

(H_1^+) ri dom $f_1 \cap$ ri dom $f_2 \neq \emptyset$, where "ri" stands for the relative interior; the ri requirement may be deleted for either index i for which f_i may happen to be polyhedral;

(H_2^+) $0 \in \operatorname{int} \Delta$ where

$$\Delta = \{(x_1,x_2) \mid x_i \in \mathbb{R}^n,\ x_2-x_1 \in \operatorname{dom} f_1 - \operatorname{dom} f_2\};$$

(H_3^+) for some x^* and some real α the set

$$\{(x_1^*, x_2^*) \mid x_i^* \in \text{dom } f_i,\ x^* = x_1^* + x_2^*,\ f_1^*(x_1^*) + f_2^*(x_2^*) \leq \alpha\}$$

is nonempty and bounded;

(H_4^+) for all x^* the condition $[f_1^*]_\infty(x^*) + [f_2^*]_\infty(-x^*) \leq 0$ implies $[f_1]_\infty(-x^*) + [f_2^*]_\infty(x^*) \leq 0$;

(H_5^+) f_1 is polyhedral and whenever x^* satisfies $[f_1^*]_\infty(x^*) +$ $[f_2^*]_\infty(-x^*) \leq 0$ for $[f_1^*]_\infty(-x^*) + [f_2^*]_\infty(x^*) > 0$ it follows that $[f_1^*]_\infty(x^*) = [f_2^*]_\infty(x^*)$;

(H_6^+) f_1 and f_2 are polyhedral and $[f_1^*]_\infty(x^*) + [f_2^*]_\infty(-x^*) \leq$ for all x^*.

In the infinite-dimensional setting, we have the following :

(H_7^+) there exists $\bar{x} \in \text{dom } f_1$ at which f_2 is finite and continuous;

(H_8^+) X is a Banach space (in the designated topology compatible with the pairing), f_1 and f_2 are in $\Gamma_o(X)$, and 0 lies in the algebraic interior of Δ for Δ as in (H_2^+);

(H_9^+) f_1 and f_2 are in $\Gamma_o(X)$, and for some open set Ω in X^* the set

$$\{x_1^*, x_2^*) \mid x_i^* \in \text{dom } f_i^*, x_1^* + x_2^* \in \Omega,\ f_1^*(x_1^*) + f_2^*(x_2^*) < \alpha\}$$

is nonempty and equicontinuous;

(H_{10}^+) X is a Fréchet space, X^* its dual space, f_1 and f_2 are in $\Gamma_o(X)$. For any continuous semi-norm p on X there exists a continuous semi-norm q on X such that

$$p^* + (f_1^* \nabla f_2^*) \geq f_1^* \nabla (f_2^* + q^*).$$

$$[\text{or } p \nabla (f_1 + f_2) \leq f_1 + (f_2 \nabla q)].$$

The sufficiency of conditions (H_1^+) through (H_6^+) are proved in Rockafellar's book [40]; conditions (H_4^+)- (H_6^+) stated using the recession functions can be seen to be the dualized versions of the conditions contained in (H_1^+), as shown by McLinden [32, p. 165]. In the infinite-dimensional context, the sufficiency of conditions (H_7^+) through (H_9^+) is Theorem 20 of Rockafellar's monograph [41] where the reader will find the appropriate references. Note that (H_7^+) is the most widely used condition for securing a property like (H^+).

Conditions like (H_2^+) and (H_8^+) are appealing because firstly they are symmetric in f_1, f_2 and secondly they do not require the interior of dom f_1 (or dom f_2) to be nonempty. (H_8^+) is a requirement akin to the following

$$0 \in \text{int} (\text{dom } f_1 - \text{dom } f_2) , \tag{2.2}$$

which can be found in some textbooks on convex analysis, while (H_7^+) reduces to (compare with (2.2)) :

$$0 \in \text{dom } f_1 - \text{int} (\text{dom } f_2). \tag{2.3}$$

Conditions like (2.2) are considered in the recent literature for stability questions and regularity conditions in mathematical programming problems. In particular, when f_1 and f_2 are positively homogeneous, refinements of condition (2.2) are stated by imposing that dom f_1 - dom f_2 is a subspace of X; they then are "in general position" (Kutateladze's terminology [26]) or "transversal" (Penot's terminology [37]).

Conditions (H_{10}^+) is due to Joly [24] and has a particular flavour when f_1 and f_2 are indicator functions. Let X be a Banach space and A_1, A_2 be two closed convex sets in X,; the condition displayed in (H_{10}^+) can be translated by saying that the "codistance" between A_1 and A_2 [24, p. 437] must be strictly positive. This condition on the codistance is appealing by its geometrical nature (see [24] for the properties); it allows, as well as the other displayed conditions, to decompose the set of ε-normals to $A_1 \cap A_2$ in terms of ε_i-normals to A_i.

2.2. Let f be a proper convex function on a certain locally convex space Y (paired with a space $Y^\star$) and let $A : X \longrightarrow Y$ be a *linear* transformation. Then the function g defined on X by

$$g(x) = \begin{cases} (f \circ A)(x) & \text{if } x \in \text{dom } A \\ +\infty & \text{elsewhere} \end{cases} \tag{2.4}$$

is a convex function and the question is how to express $\partial_\varepsilon g$ in terms of ∂f and the adjoint transformation $A^\star : Y^\star \longrightarrow X^\star$ (where as earlier $X^\star$ is paired with X). As usual in such a situation it is assumed that A (and $A^\star$) is *densely defined* with *closed graph*. The basic assumption in our context is the following:

$$(H^a) \qquad g^\star(x^\star) = \min\{f^\star(y^\star) \mid Y^\star \in \text{dom } A^\star,\ A^\star y^\star = x^\star\} \quad \forall x^\star.$$

Theorem 2.2. Let f be a proper convex function on Y, let $A : X \longrightarrow Y$ be a densely defined linear operator with closed graph. Assume that (H^a) holds for $g = f \circ A$ defined as in (2.4). Then

$$\partial_\varepsilon (f \circ A)(x_o) = A^\star \partial_\varepsilon f(Ax_o) \tag{2.5}$$

for all $x_o \in$ dom A that $Ax_o \in$ domf. □

Actually, the statements of Theorem 2.2 and Theorem 2.1 are equivalent in consideration of their contents. Indeed one can pass from the framework "f_1+f_2" to the framework "$F \circ A$" and vice versa by simple transformations. For example, in [41] the framework "$F \circ A$" is considered first and the results on "f_1+f_2" are derived afterwards. Therefore it is not surprising to find as conditions ensuring (H^a) a list of conditions (H_1^a) merely corresponding to the (H_1^+). We do not consider to list all the

(H_i^a) since they can be picked from [40] and [41, Theorem 19]. We shall be content with mentioning the most significative ones. In the finite-dimensional case :

(H_1^a) there exists $\bar{x}$ such that $\bar{A}x \in \mathrm{ri}(\mathrm{dom}\, f)$;

(H_2^a) f is polyhedral and there exists $\bar{x}$ and such that $A\bar{x} \in \mathrm{dom} f$. As for the infinite-dimensional setting :

(H_3^a) there exists $\bar{x} \in \mathrm{dom}\, A$ such that f is finite and continuous at $A\bar{x}$;

(H_4^a) X and Y are Banach spaces, $f \in \Gamma_o(Y)$ and 0 lies in the algebraic interior of (dom f - range A).

A further assumption, suitable for applications to abstract control problems, may be found in [46].

Remark 1. Along the same lines, a formula generalizing (2.1) can be derived for the "continuous" case, i.e. for the continuous sum $\int_T f_t d\mu : x \in X \longmapsto \int_T f_t(x) d\mu(t)$ [14] [15] [16, §4], as for integral functionals $I_f : x \in L \longmapsto \int_T f_t(x(t)) d\mu(t)$ (see pages 58 to 64 in [41]). □

Remark 2. We do not claim any novelty in producing formulas for the ε-subdifferential of $f_1 + f_2$ and $f \circ A$. The expression of $\partial_\varepsilon(f_1+f_2)$ under the assumptions (H_1^+) or (H_7^+) was used by the author in [14], [15] and [16]; for finite functions, it was rediscovered in [11]. Anyway, the two quoted formulas were likely to be known by those who are familiar with convex analysis. □

2.3. Some Examples of Applications. The next examples illustrate the utilization of chain rules (2.1) and (2.5).

Example 2.1. Let $f \in \Gamma_o(X)$, let $x_o \in \mathrm{dom}\, f$ and let d be a non null direction. Suppose, for example, that either

. $X = \mathbb{R}^n$, f is polyhedral and $(x_o+\mathbb{R}d) \cap \mathrm{dom}\, f \neq \emptyset$,

or

. $X = \mathbb{R}^n$, $(x_o+\mathbb{R}d) \cap \mathrm{ri}\, (\mathrm{dom}\, f) \neq \emptyset$,

or

. f is finite and continuous at some point of $(x_o+\mathbb{R}d)$.

Then the function $f_{x_o,d} : \lambda \longmapsto f(x_o+\lambda d)$ has its ε-subdifferential at λ_o (where it is finite) given as

$$\partial_\varepsilon f_{x_o,d}(\lambda_o) = \langle \partial_\varepsilon f(x_o+\lambda_o d),\ d\rangle \tag{2.6}$$

This is obtained by just applying Theorem 2.2. to f and $A : \lambda \longmapsto \lambda d$.

Example 2.2. Let f be a proper convex function on X, and consider the problem of finding an approximate minimum of f over a nonempty convex subset C. More precisely, given ε, we are looking for $x_o \in C \cap \mathrm{dom}\, f$ satisfying

$$f(x_o) \leq \inf_{x\in C} f(x) + \varepsilon. \tag{2.7}$$

The problem is the same as finding the ε-minima of $x \longmapsto f(x) + \delta(x|C)$ over X, and its solutions are the points x_o such that $0 \in \partial_\varepsilon[f+\delta(.|C)]\ (x_o)$. Theorem 2.1 gives sufficient conditions for the latter relation to be expressible as

$$0 \in \bigcup_{\substack{\varepsilon_1,\varepsilon_2\geq 0\\ \varepsilon_1+\varepsilon_2=\varepsilon}} \{\partial_{\varepsilon_1} f(x_o) + N_{\varepsilon_2}(C;x_o)\}. \tag{2.8}$$

This condition means that, for some positive ε_1, ε_2 with $\varepsilon_1+\varepsilon_2 = \varepsilon$, there is an element $x^* \in \partial_{\varepsilon_1} f(x_o)$ such that $-x^*$ is an ε_2-normal to C at x_o. The condition can be simplified if x_o is the apex of a convex cone, so that $N_{\varepsilon_2}(C;x_o) = N(C;x_o)$ for all ε_2 (see Example 1.1). When C is represented as an inequality constaint, i.e.

$$C = \{x \in X \mid g(x) \leq 0\}\ ,$$

it is natural to try to express $N_\varepsilon(C;x_o)$ in terms of $\partial_\varepsilon g(x_o)$. Unlike the case $\varepsilon = 0$, one cannot restrict to those x_o lying on the boundary of C and one cannot invoke cone properties of $N_\varepsilon(C;x_o)$. The comparison result, generalizing what is known for $\varepsilon = 0$, is displayed in [43]. We now illustrate relation (2.8) in a situation particularly relevant for problems of best approximation [20,28]. Let $f \in \Gamma_o(X)$ be continuous at least at one point of its domain and let V be a linear subspace of V, of dimension n. We assume that

$$\text{int}(\text{dom } f) \cap V \neq \emptyset,$$

and we consider the $x \in V$ minimizing f over V within ε (relation (2.7)). Then, a generalization of what is derived when $\varepsilon = 0$ [28, §8] is the following : a *necessary and sufficient condition for* $x_o \in V$ *to be an* ε*-minimum of* f *over* V *is that there exists* r (≥ 1) *extreme points* $x_1^*, \ldots, x_r^*$ of $\partial_\varepsilon f(x_o)$, s *extremal directions* $d_1, \ldots, d_s$ *of* $\partial_\varepsilon f(x_o)$, *with* $r+s \leq n+1$, *and positive* $\rho_1, \ldots, \rho_r, \sigma_1, \ldots, \sigma_s$, $\sum_{i=1}^{r} \rho_i = 1$ *such that*

$$\sum_{i=1}^{r} \rho_i x_i^* + \sum_{i=1}^{s} \sigma_i d_i \in V^\perp . \qquad \square$$

3. THE ε-SUBDIFFERENTIAL OF $f_1 \nabla f_2$

3.1. Given two proper functions f_1, f_2 one defines a new function in the following way:

$$g(x) = \inf_{\substack{x_1, x_2 \in X \\ x_1 + x_2 = x}} \{f_1(x_1) + f_2(x_2)\}.$$

g is said to be the *infimal convolution* of f_1 and f_2, and we shall use the notation $g = f_1 \nabla f_2$. The infimal convolution is said to be exact at $x_o = x_o^1 + x_o^2$ if one has

$$f_1(x_o^1) + f_2(x_o^2) = \min_{\substack{u, v \in X \\ u+v=x_o}} \{f_1(u) + f_2(v)\}.$$

Assumption (H^+) was merely requiring that $(f_1+f_2)^* = f_1^* \nabla f_2^*$ with the infimal convolution exact on X^*. As a general rule, $(f_1 \nabla f_2)^* = f_1^* + f_2^*$ so that the operations "+" and "∇" are dual to each other with respect to the conjugacy operation. The description of $\partial_\varepsilon(f_1 \nabla f_2)$ at a point where the infimal convolution is finite and exact does not require any condition.

Theorem 3.1. Let f_1, f_2 be proper functions, let $x_o = x_o^1 + x_o^2$ be a point where the infimal convolution is finite and exact. Then

$$\partial_\varepsilon(f_1 \nabla f_2)(x_o) = \bigcup_{\substack{\varepsilon_1 \geq 0, \varepsilon_2 \geq 0 \\ \varepsilon_1 + \varepsilon_2 = \varepsilon}} \{\partial_{\varepsilon_1} f_1(x_o^1) \cap \partial_{\varepsilon_2} f_2(x_o^2)\}. \quad (3.1)$$

Conditions ensuring that the infimal convolution is exact at all $x_o \in \operatorname{dom}(f_1 \nabla f_2)$ are "dual" to those ensuring (H^+). As for example, the following hypotheses secure the exactness and the resulting $f_1 \nabla f_2$ in $\Gamma_o(X)$:

(H_1^{∇}) $f_1, f_2 \in \Gamma_o(\mathbb{R}^n)$ and $\operatorname{ri}(\operatorname{dom} f_1) \cap \operatorname{ri}(\operatorname{dom} f_2) \neq \emptyset$; the ri requirement may be deleted for either index i for which f_i may happen to be polyhedral;

(H_2^{∇}) $f_1, f_2 \in \Gamma_o(X)$ and there is $\bar{x}^* \in \operatorname{dom} f_1^*$ at which f_2^* is finite and continuous.

Remark. A formula generalizing (3.1) to some extent can be produced for the continuous infimal convolution

$$\oint_T f_t \, d\mu : x \longrightarrow \inf\{\int_T f_t(x(t))d\mu(t) \,\Big|\, \int_T x(t)d\mu(t) = x\};$$

for such purposes, see [21], [47], [41, p. 63], [16 Chapter 4].

3.2. Applications. Let here X be a real Banach space E and let f be a proper convex function on E. Performing the infimal convolution of f with another "regular" function yields a "smoothing" or "regularizing" effect , as is usual with operations of the "convolution" type. If $\|\cdot\|$ denotes the norm (function) and r a positive constant, the function $f_r = f \nabla r\|\cdot\|$ does have some interesting properties which are described in detail in[19, §3]. In particular, it is noteworthy that f_r is either identically equal to $-\infty$ or is Lipschitz (on all of E) with Lipschitz constant r. The coincidence set of f and f_r, i.e. the set of $x \in E$ for which $f(x) = f_r(x)$, can be fully described and for all x_o where f and f_r coincide, the following holds:

$$\partial_\varepsilon f_r(x_o) = \{x^* \in \partial_\varepsilon f(x_o) \mid \|x^*\|_* \leq r\}, \tag{3.2}$$

where $\| \cdot \|_*$ denotes the dual norm [19, Proposition 2.3]. We now turn our attention to a particular case of the above regularization and to another regularization process.

Example 3.1. Let S be a nonempty closed set of E, different from E. Since the distance function d_S is unable to make a distinction between int S and bd S (the boundary of S), we introduced in [18] the following function :

$$\Delta_S(x) = d_S(x) - d_{S^c}(x),$$

where S^c stands for the complementary set of S in E. Actually, Δ_S is nothing more than a regularized version of the following convex function μ_S ($\Delta_S = \mu_S \nabla \| \cdot \|$):

$$\mu_S(x) = +\infty \text{ if } x \in S^c, -d_{S^c}(x) \text{ if } x \in S.$$

Properties of μ_S and Δ_S from the convex analysis viewpoint are displayed in [18]. Concerning the ε-subdifferential, we observe that the infimal convolution of μ_S and $\| \cdot \|$ is exact at $x_o = x_o + 0$ for all $x_o \in S$. Thus, according to (3.2), one has that

$$\forall x_o \in S, \; \partial_\varepsilon \Delta_S(x_o) = \{x^* \in \partial_\varepsilon \mu_S(x_o) \mid \| x^* \|_* \leq 1\}. \qquad \square$$

Example 3.2. Let H be a Hilbert space and let $f \in \Gamma_o(H)$. Another regularizing process which is widely used in nonlinear analysis consists in taking for any $r > 0$,

$$\phi_r = f \nabla \frac{r}{2} \| \cdot \|^2 .$$

ϕ_r is everywhere finite on H and the unique point x_r where the function $u \longmapsto f(u) + \frac{r}{2} \| u-x \|^2$ achieves its minimum is

$$x_r = \mathrm{prox}_{f/r}(x),$$

the "proximal point of x relatively to $\frac{f}{r}$" [34]. The optimality condition yields that $r(x-x_r) \in \partial f(x_r)$, which can be rewritten as :

$$x_r = (I + \frac{1}{r}\,\partial f)^{-1}(x).$$

The mapping $(I + \frac{1}{r}\,\partial f)^{-1}$ is the so-called "resolvant mapping" of ∂f [5]. ϕ_r is known to be $C^{1,1}$ (i.e. ϕ_r is differentiable and its gradient mapping is Lipschitz) with $\nabla\phi_r(x) = r(x-x_r)$ ($\in \partial f(x_r)$). As for the ε-subdifferential of ϕ_r, we easily deduce from the rule (3.1) and from the ε-subdifferential of $\frac{r}{2}\,\|\cdot\|^2$ that

$$\partial_\varepsilon \phi_r(x) = \bigcup_{\substack{\varepsilon_1,\varepsilon_2 \geq 0 \\ \varepsilon_1+\varepsilon_2=\varepsilon}} \left\{\partial_{\varepsilon_1} f(x_r) \cap [r(x-x_r) + \sqrt{2r\,\varepsilon_2}\,B]\right\},$$

where B is the closed unit ball in H.

4. THE ε-SUBDIFFERENTIAL OF $\max_{i \in I} f_i$

4.1. Let $\{f_i \mid i \in I\}$ be a collection of proper convex functions and let $f = \max_{i \in I} f_i$. It is known that, under suitable assumptions, any subgradient of f at x_o can be expressed as a convex combination of subgradients at x_o of those f_i which satisfy $f_i(x_o) = f(x_o)$. The situation is different for the ε-differential; due to its non-local nature, the knowledge of $\partial_\varepsilon f(x_o)$ requires a priori the knowledge of $\partial_\varepsilon f_i(x_o)$ for all $i \in I$. The calculation of the conjugate function f^* in terms of f_i^* is not without trouble; for more convenience we shall deal with the case where I is a finite index set $\{1,\ldots,m\}$.

Theorem 4.1. Let $f_1,\ldots,f_m$ be convex functions finite on the entire space X, and suppose that all of them except, possibly, one are continuous. Then $x^* \in \partial_\varepsilon f(x_o)$ if and only if there exist vectors x_i^*, $i = 1,\ldots,m$, non-negative α_i, $i = 1,\ldots,m$, adding up to 1, and non-negative ε_i, $i=1,\ldots,m$, such that

(a) $x^* = \sum_{i=1}^{m} \alpha_i x_i^*$;

(b) $\sum_{i=1}^{m} \varepsilon_i + f(x_o) - \sum_{i=1}^{m} \alpha_i f_i(x_o) = \varepsilon$,

(c) $x_i^* \in \partial_{\frac{\varepsilon_i}{\alpha_i}} f_i(x_o)$ for all i such that $\alpha_i > 0$.

Proof. Under the above assumptions, for every $x^* \in \operatorname{dom} f^*$, there exist vectors $x_i^* \in \operatorname{dom} f_i^*$, $i=1,\ldots,m$, and positive α_i, $i=1,\ldots,m$, adding up to 1 such that

$$f^*(x^*) = \sum_{i=1}^m \alpha_i f_i^*(x_i^*),$$

$$x^* = \sum_{i=1}^m \alpha_i x_i^* \quad \text{(see [21,p.66] or [23,p.178]).}$$

$x^* \in \partial_\varepsilon f(x_o)$ is known to satisfy $f^*(x^*) + f(x_o) - \langle x_o, x^* \rangle \leq \varepsilon$.
This relation can be rewritten through the above-mentioned relation as

$$\sum_{i=1}^m \alpha_i [f_i^*(x_i^*) + f_i(x_o) - \langle x_o, x_i^* \rangle] \leq \varepsilon + \sum_{i=1}^m \alpha_i f_i(x_o) - f(x_o).$$

This inequality is equivalent to the existence of non-negative ε_i, $i = 1,\ldots,m$, satisfying :

$$\sum_{i=1}^m \varepsilon_i = \varepsilon + \sum_{i=1}^m \alpha_i f_i(x_o) - f(x_o),$$

$$f_i^*(x_i^*) + f_i(x_o) - \langle x_o, x_i^* \rangle \leq \frac{\varepsilon_i}{\alpha_i} \text{ for all } i \text{ such that } \alpha_i > 0.$$

Whence we derive the desired result. □

In a set-formulation, the result of Theorem 4.1 can be rephrased as

$$\partial_\varepsilon f(x_o) = \{\sum_{i=1}^m \partial_\varepsilon(\alpha_i f_i)(x_o) \,|\, \alpha_i \geq 0,\ \sum_{i=1}^m \alpha_i = 1;\ \varepsilon_i \geq 0,$$
$$\sum_{i=1}^m \varepsilon_i + f(x_o) - \sum_{i=1}^m \alpha_i f_i(x_o) = \varepsilon\}, \tag{4.1}$$

a formula announced by Kutateladze [27].

Remark. When I is an arbitrary index set and $f_i \in \Gamma_o(\mathbb{R}^n)$ for all $i \in I$, a formula giving $(\sup_{i \in I} f_i)^*$ does exist [40, Theorem 16.5].

The problem however is to give sufficient conditions ensuring that

$$(H^m) \qquad (\sup_{i \in I} f_i)^*(x^*) = \min\{ \sum_{i \in I} \alpha_i f_i^*(x_i^*) \}$$

where for each x^* the (attained) infimum is taken over all representations of x^* as a convex combination $\sum_{i \in I} \alpha_i x_i^*$. Moreover, it is known that the infimum can be taken over all expressions of x^* as a convex combination in which at most n+1 of the coefficients α_i are positive and the corresponding x_i^* are affinely independent [40, Corollary 17.1.3]. Usual conditions ensuring (H^m) require some additional assumptions on I (I compact space) and on the mappings $i \longrightarrow f_i(x)$ (upper-semicontinuity); for a generalization of formula (4.1) in such a case, see [43].

4.2. The result of Theorem 4.1 is of importance for theoretical considerations as well as from the computational viewpoint. We illustrate its wide range of applicability by two examples.

Example 4.1. Let $g : X \longrightarrow \mathbb{R}$ be a convex function and consider the function $g^+ (g^+ = \max(0,g))$. What is the exact expression of $\partial_\varepsilon g^+$ in terms of $\partial_\eta g$? According to formula (4.1), we have that

$$\partial_\varepsilon g^+(x_o) = \{\partial_\eta(\alpha g)(x_o) \mid 0 \leq \alpha \leq 1, \eta \geq 0, \eta + g^+(x_o) - \alpha g(x_o) = \varepsilon\}.$$

For each $\alpha \in [0,1]$, let $\eta(\alpha) = \varepsilon - g^+(x_o) + \alpha\, g(x_o)$; then

$$\partial_\varepsilon g^+(x_o) = \bigcup_{0 \leq \alpha \leq 1} \partial_{\eta(\alpha)}(\alpha g)(x_o), \tag{4.2}$$

with the convention that $\partial_\eta(\alpha g)(x_o)$ is empty whenever $\eta < 0$. □

Example 4.2. Let $a_1^*, \ldots, a_m^*$ be in X^*, let $c_1, \ldots, c_m$ be real numbers. We set $f(x) = \max_{i=1,\dots,m} (\langle a_i^*, x\rangle + c_i)$. Then one easily checks that

$$\partial_\varepsilon f(x_o) = \{ \sum_{i=1}^{m} \alpha_i a_i^* \mid \alpha_i \geq 0, \sum_{i=1}^{m} \alpha_i = 1,$$

$$f(x_o) - \sum_{i=1}^{m} \alpha_i (\langle a_i^*, x_o\rangle + c_i) \leq \varepsilon \}. \tag{4.3}$$

When f is the maximum of a finite number of quadratic functions, the formula giving exactly $\partial_\varepsilon f(x_o)$ is derived from (4.1) and results in Example 1.4. □

5. THE ε-SUBDIFFERENENTIAL OF $\sigma \circ f$

Let $f : X \longrightarrow (-\infty, +\infty]$ be a proper convex function and let $\sigma : \mathbb{R} \longrightarrow (-\infty, +\infty]$ be an increasing convex function. By posing $\sigma(+\infty) = +\infty$, we get a new convex function $\sigma \circ f : X \longrightarrow (-\infty, +\infty]$. The question in this section is to give the exact formulation of $\partial_\varepsilon(\sigma \circ f)$ in terms of $\partial_\varepsilon \sigma[f(x_o)]$ and $\partial_\varepsilon f(x_o)$.

We first note that $\partial_\varepsilon \sigma(t_o) \subset \mathbb{R}_+$ for all $t_o \in \operatorname{dom} \sigma$. A general (usable) result is hopeless without any assumption on the overlapping of $f(\operatorname{dom} f)$ and $\operatorname{dom} \sigma$. The following (mild) assumption will be made (Kutateladze [25], [26, §3.7]) :

(H^o) $\quad f(\operatorname{dom} f) \cap \operatorname{int} (\operatorname{dom} \sigma) \neq \emptyset$.

Theorem 5.1. Let x_o be such that $f(x_o) \in \operatorname{dom} \sigma$. Then $x^* \in \partial_\varepsilon(\sigma \circ f)(x_o)$ if and only if there exist non-negative ε_1, ε_2 and t^* such that

(a) $\quad \varepsilon_1 + \varepsilon_2 = \varepsilon$,

(b) $\quad t^* \in \partial_{\varepsilon_1} \sigma[f(x_o)]$, $x^* \in \partial_{\varepsilon_2}(t^* f)(x_o)$. □

Proof. Under the assumption (H^o), the following holds (Kutateladze [25], [26, §3.7]) :

$$(\sigma \circ f)^*(x^*) = \min (t^* f)^*(x^*) + \sigma^*(t^*) \mid t^* \geq 0\}.$$

Let $x^* \in \partial_\varepsilon(\sigma \circ f)(x_o)$; then there exists $t^* \geq 0$ such that

$$(t^* f)^*(x^*) + \sigma^*(t^*) + \sigma[f(x_o)] - \langle x_o, x^* \rangle \leq \varepsilon$$

which can be rewritten as

$$(t^* f)^*(x^*) + t^* f(x_o) - \langle x_o, x^* \rangle$$
$$+ \quad \sigma^*(t^*) + \sigma[f(x_o)] - t^* f(x_o) \leq \varepsilon.$$

Whence the announced result is easily derived. □

The result of Theorem 5.1 can be simplified when σ is positively homogeneous (see Example 1.1). As for example, if $\sigma : t \longmapsto t^+$, one extends formula (4.2) to arbitrary convex functions f.

Another application worthwhile mentioning is now displayed.

Let f be a proper convex function on X and let C be defined as

$$C = \{x \in X \mid g(x) \leq 0\},$$

where g is a finite convex function. We consider the problem of characterizing the ε-minima of f over C (see Example 2.2). By setting $\sigma(t) = 0$ if $t \leq 0$, $+\infty$ elsewhere, the above problem is equivalent to finding the ε-minima of $x \longmapsto f(x) + (\sigma \circ g)(x)$ over X, and its solutions are thus the points x_o satisfying

$$0 \in \partial_\varepsilon (f + \sigma \circ g)(x_o).$$

We suppose that

$$(H) \quad \begin{cases} \cdot \text{ there exists a point of C where f is finite and} \\ \quad \text{continuous,} \\ \cdot \text{ there is } \bar{x} \in C \text{ such that } g(\bar{x}) < 0. \end{cases}$$

Observe that the latter assumption is nothing more than (H^o) for the particular invoked σ. For such a σ, we clearly have that

$$\partial_\eta \, \sigma[g(x_o)] = \{t^* \geq 0 \mid \eta + t^* g(x_o) \geq 0\} \quad \forall \; x_o \in C.$$

Hence a straightforward application of Theorem 2.1 and Theorem

5.1 yields :

Theorem 5. 2. Under the assumption (H), a necessary and sufficient condition for $x_o \in C$ to be an ε-minimum of f over C is there exist non-negative $\varepsilon_1, \varepsilon_2, \varepsilon_3$ adding up to ε and a non-negative t^* satisfying

(a) $\quad 0 \in \partial_{\varepsilon_1} f(x_o) + \partial_{\varepsilon_2}(t^* g)(x_o),$

(b) $\quad \varepsilon_3 + t^* g(x_o) \geq 0.$

The above result was announced by Kutateladze [27] in a setting dealing with convex operators, and proved by Strodiot *et al* [43] through a different approach.

6. THE ε-SUBDIFFERENTIAL OF A MAGINAL FUNCTION

Let $f : X \times Y \longrightarrow (-\infty, +\infty]$ be a proper convex function, where as usual X is paired with X^* and Y with Y^*. The *marginal function* ϕ is defined on X by

$$\phi(x) = \inf_{y \in Y} f(x,y). \tag{6.1}$$

For all x_o such that $\phi(x_o) \in \mathbb{R}$, let $M(x_o)$ denote the set of elements (if any) for which the infimum in (6.1) is attained. If $M(x_o)$ is non-empty, it comes from the ε-subgradient inequality that

$$\partial_\varepsilon \phi(x_o) = \{x^* \in X^* \mid (x^*, 0) \in \partial_\varepsilon f(x_o, y_o)\} \tag{6.2}$$

for all $y_o \in M(x_o)$.
Let us particularize the result (6.2) in the case where the marginal function is defined at x through y constrained to a set $F(x)$. Let thus $F : X \rightrightarrows Y$ be a set-valued mapping whose graph $\{(x,y) \in X \times Y \mid y \in F(x)\}$ is denoted by $\mathcal{F}$, and let ϕ_F defined as

$$\phi_F(x) = \inf_{y \in F(x)} f(x,y).$$

We denote $M_F(x_o)$ the set of $y_o \in F(x_o)$ for which $\phi_F(x_o) = f(x_o, y_o)$. The next result gives the description of the ε-subdifferential of ϕ_F.

Theorem 6.1. *Let f be a proper convex function on $X \times Y$, let $F : X \rightrightarrows Y$ be a set-valued mapping with a nonempty convex graph $\mathcal{F}$. Assume moreover that there is a point in $\mathcal{F}$ at which f is finite and continuous. At a point x_o where ϕ_F is finite and $M_F(x_o)$*

non-empty, let us choose any y_o in $M_F(x_o)$; then we have the following : $x^* \in \partial_\varepsilon \phi_F(x_o)$ if and only if there exist ε_1, ε_2, adding up to ε such that

$$(x^*,0) \in \partial_{\varepsilon_1} f(x_o,y_o) + N_{\varepsilon_2}(F;(x_o,y_o)). \quad (6.$$

Proof. Since $\delta(y|F(x)) = \delta((x,y)|F)$, ϕ_F can be rewritten as

$$\phi_F(x) = \inf_{y \in Y}[f+\delta(.|F)](x,y).$$

According to (6.2)), we have that

$$\partial_\varepsilon \phi_F(x_o) = \{x^* | (x^*,0) \in \partial_\varepsilon [f+\delta(.|F)] \quad (x_o,y_o)\}$$

whatever $y_o \in M_F(x_o)$. Theorem 2.1 applied under the assumption (H_7^+) thus yields the announced result.

As an illustration, let f be dependent on the only variable y and let $F : x \rightrightarrows F(x) = \{y \in Y | Ay = x\}$, where $A : Y \to X$ is a continuous linear mapping. The marginal function associated with these data is then the so-called *image of f under A*,

$$(Af)(x) = \inf \{f(y) | Ay = x\}.$$

Here, the ε_1-subdifferential at (x_o,y_o) of f considered as a function on $X \times Y$ is $\{0\} \times \partial_{\varepsilon_1} f(y_o)$, while $N_{\varepsilon_2}(F;(x_o,y_o))$ does not depend on ε_2 and is reduced to

$$\{(x^*, - A^* x^*) \mid x^* \in X^*\}.$$

Consequently, the result of Theorem 6.1 is rewritten as

$$\partial_\varepsilon (Af)(x_o) = \{x^* | A^* x^* \in \partial_\varepsilon f(y_o)\}, \quad (6.5)$$

where y_o satisfies $Ay_o = x_o$ and $(Af)(x_o) = f(y_o)$

Remark. The calculus rule (6.4) is a general result which can be proved in different ways. Actually, results (6.2) and (6.4) are of an equivalent nature since one can pass from one framework to the other one by simple transformations.

7. LOCAL BEHAVIOUR OF THE ε-SUBDIFFERENTIAL

Throughout this section, $X = E$ is a Banach space (paired with $X^* = E^1$, topological dual space of E) and $f \in \Gamma_o(E)$. We denote by B (resp. B^*) the closed unit ball in E (resp. in E'). In this section we are interested in the local properties of

$$\partial_{\bullet} f(.) : \text{dom } f \times \mathbb{R}_+ \longrightarrow \mathcal{C}_\sigma(E')$$

$$(x,\varepsilon) \rightleftarrows \partial_\varepsilon f(x),$$

where $\mathcal{C}_\sigma(E')$ denotes the collection of all $\sigma(E',E)$-closed subsets of E'. When $x \in \text{int(dom } f)$, $\partial_\varepsilon f(x)$ is, for any $\varepsilon \in \mathbb{R}_+$, a non-empty $\sigma(E',E)$-closed bounded (convex) subset of E'. We recall that the Hausdorff-topology on the collection $\mathcal{C}_{\sigma,b}(E')$ of all non-empty $\sigma(E',E)$-closed bounded convex subsets of E' is the topology in which, for each $C^* \in \mathcal{C}_{\sigma,b}(E')$, the sets of the form

$$\{D^* \in \mathcal{C}_{\sigma,b}(E') \,|\, D^* \subset C^* + \alpha B^* \text{ and } C^* \subset D^* + \alpha B^*\}$$

constitute a fundamental system of neighborhoods of C^* as α ranges over $\mathbb{R}_+^*$. The Hausdorff-topology on $\mathcal{C}_{\sigma,b}(E')$ can be defined by a metric (the so-called Hausdorff-distance h) whose definition in a dual way is as follows :

$$\forall\, C^*, D^* \in \mathcal{C}_{\sigma,b}(E') \quad h(C^*,D^*) = \sup_{d \in B} |\delta^*(d|C^*) - \delta^*(d|D^*)|$$

7.1 <u>Behaviour of $\partial_\varepsilon f(x_o)$ as a Function of ε.</u> Evidently, $\partial_\varepsilon f(x_o)$ decreases as ε decreases to 0, and the intersection of the nest $\partial_\varepsilon f(x_o)$ is just $\partial f(x_o)$. The rate of convergence of $\partial_\varepsilon f$

towards $\partial f(x_o)$ when ε goes to 0^+ may be very bad; see for example the case of quadratic functions in finite dimensions (cf. Example 1.4) where $h(\partial f(x_o), \partial_\varepsilon f(x_o))$ behaves as $\varepsilon^{\frac{1}{2}}$. Even more, when E is not finite-dimensional, it is not certain that for all $\delta > 0$, there exists $\alpha > 0$ such that

$$\forall \varepsilon \in [0,\alpha[\quad , \partial_\varepsilon f(x_o) \subset \partial f(x_o) + \delta B^*. \tag{7.1}$$

The next statement, due to Robert [39, Part I] gives conditions for an approximation result like (7.1) to hold. For the rest of this section, we shall assume that *there is a non-empty open set on which f is bounded above*.

Theorem 7.1. Let x_o lie in the interior of dom f. The following assertions are equivalent

(a) $f(x_o + h) = f(x_o) + f'(x_o;h) + \varepsilon(h)\|h\|$, with $\lim_{h\to 0} \varepsilon(h) = 0$;

(b) $[f(x_o+\lambda d) - f(x_o)]\lambda^{-1}$ converges to $f'(x_o;d)$ uniformly in $d \in B$ when $\lambda \to 0^+$;

(c) for all $\delta > 0$, there exists $\alpha > 0$ such that $\forall \varepsilon \in [0,\alpha[\ , \partial_\varepsilon f(x_o) \subset \partial f(x_o) + \delta B^*$;

(d) for all $\delta > 0$, there is a neighborhood V of x_o such that $\forall x \in V$, $\partial f(x) \subset \partial f(x_o) + \delta B^*$. □

f is said to be *Fréchet-subdifferentiable* at those points where one of the above equivalent conditions is satisfied. The discrepancy which may occur between the convergence of $[f(x_o+\lambda d) - f(x_o)]\lambda^{-1}$ to $f'(x_o;d)$ for each d and uniformly in $d \in B$ is of the same nature as the difference between Hadamard-differentiability and Fréchet-differentiability. According to Robert's result, the rate of convergence of $\partial_\varepsilon f(x_o)$ towards $\partial f(x_o)$ is closely related to the rate of approximation of $f'(x_o;d)$ by $[f(x_o+\lambda d) - f(x_o)]\lambda^{-1}$. Actually, in proving the equivalence of (b) and (c), Robert showed the following relationship. Let $\delta > 0$; if (b) is assumed, there exists $\lambda_0 > 0$ such that

$$\forall\lambda \in]0,\lambda_0],\ \forall d \in B,\ [f(x_0+\lambda d)-f(x_0)]\lambda^{-1} \leq f'(x_0;d) + \delta/2.$$

Then, $\alpha = (\delta\lambda_0)/2$ satisfies the requirement of (c). When ε and ε' lie in $\mathbb{R}_+^*$, the Hausdorff-distance between $\partial_\varepsilon f(x_0)$ and $\partial_{\varepsilon'} f(x_0)$ is estimated in the following manner [19, Theorem 3.3].

<u>Theorem 7.2. Let</u> $x_0 \in \operatorname{int}(\operatorname{dom} f)$. <u>Then for all</u> $\bar{\varepsilon}>0$ <u>there exists</u> k <u>such that</u>

$$h(\partial_\varepsilon f(x_0),\ \partial_{\varepsilon'} f(x_0)) \leq \frac{k}{\min(\varepsilon,\varepsilon')}\ |\varepsilon-\varepsilon'| \tag{7.2}$$

<u>for all</u> ε, ε' <u>in</u> $]0,\bar{\varepsilon}]$.

<u>7.2. Behaviour of $\partial_\varepsilon f(x)$ as a function of x.</u> Let us begin by noting two properties. Firstly, let $rg(\partial_\varepsilon f)$ denote (for $\varepsilon \geq 0$) the range of $\partial_\varepsilon f$, i.e.

$$rg(\partial_\varepsilon f) = \bigcup_{x \in \operatorname{dom} f} \partial_\varepsilon f(x).$$

It is a mere consequence of the approximation result of Brøndsted and Rockafellar [7] that the closure of $rg(\partial_\varepsilon f)$ in E' is independent of ε, namely

$$cl\ rg(\partial_\varepsilon f) = cl\ rg(\partial f)$$

for all $\varepsilon \geq 0$. Secondly, due to the fact that f is locally Lipschitz on int(dom f), for all $x_0 \in \operatorname{int}(\operatorname{dom} f)$ there is a neighborhood V of x_0 such that $\bigcup_{x \in V} \partial_\varepsilon f(x)$ is a bounded set of E'. The main result concerning the behaviour of $\partial_\varepsilon f(.)$ is the following one [19, Corollary 3.4].

Theorem 7.3. Let $x_o \in \text{int}(\text{dom } f)$, let $\varepsilon>0$. Then there exists a neighborhood V of x_o and a positive k such that

$$h(\partial_\varepsilon f(x),\ \partial_\varepsilon f(x')) \leq \frac{k}{\varepsilon}\| x-x' \| \tag{7.3}$$

for all x, x' in V. □

Since

$$h(\partial_\varepsilon f(x),\ \partial_\varepsilon f(x')) = \sup_{d\in B}|f'_\varepsilon(x;d) - f'_\varepsilon(x';d)|,$$

$f'_\varepsilon(x;d)$ is, for each $d\in E$, a locally Lipschitz function as a function of x. A natural question which arises now is : *does the (usual) directional derivative of* $f'_\varepsilon(.;d)$ *exist?* This question was answered recently by Lemaréchal and Nurminskii [31] in a particular setting. We shall recall the statement of their result and provide an interpretation of it as well as some corollaries. Let $f : \mathbb{E}^n \longrightarrow \mathbb{E}$ be a convex function and *suppose that* f^* *is finite everywhere*. This assumption is known to be equivalent to

$$f_\infty(d) = +\infty \quad \text{for all } d \neq 0.$$

In Rockafellar's terminology, such a function is called *co-finite* [40, p.116]. If f is co-finite, the same obviously holds true for f^*. As already seen, the support function $f'_\varepsilon(x_o;d)$ in the d direction of $\partial_\varepsilon f(x_o)$ is the optimal value of

$$(P)\quad \begin{cases} \max\langle x^*, d\rangle \\ \\ f^*(x^*) + f(x_o) - \langle x_o, x^*\rangle - \varepsilon \leq 0. \end{cases}$$

The assumption that f^* is finite everywhere is a technical one to ensure that the function defining the constraint set in (P) is finite everywhere. We note incidently that this assumption implies that

$$\partial f(x_o) \subset \mathrm{int}(\partial_\varepsilon f(x_o))$$

for all $x_o \in \mathbb{R}^n$. Let now $x_o \in \mathbb{R}^n$, $\varepsilon>0$ and $d \neq 0$ be fixed. A further consequence of having f co-finite is that

$$M_d(x_o) = \{\mu_o \in \mathbb{R}_+ \mid r_f(\mu_o) = \min_{\mu \in \mathbb{R}_+} r_f(\mu)\} \quad (7.4)$$

is a (nonempty) compact interval of $\mathbb{R}^*_+$, which turns out to equal

$$\{1/\lambda_o \mid q_f(\lambda_o) = \min_{\lambda \in \mathbb{R}^*} q_f(\lambda)\} \quad (7.5)$$

(see Propositions 1.3 and 1.4 in §1). The Kuhn-Tucker coefficients μ_o for the program (P) are those satisfying

$$\begin{cases} \langle x^*, d\rangle = f'_\varepsilon(x_o;d) \\ -d \in \mu_o(\partial f^*(x^*) - x_o). \end{cases}$$

Since $\mu_o>0$, the latter relation can be rewritten as $x^* \in \partial f(x_o + \frac{d}{\mu_o})$.

Lemaréchal and Nurminskii [31] showed that the above coefficients μ_o were just the μ_o defined in (7.4). Now, their main result comes as follows :

Theorem 7.4. For a given direction δ, the differential quotient $[f'_\varepsilon(x_o+s\delta;d) - f'_\varepsilon(x_o;d)]\, s^{-1}$ has a limit $f''_\varepsilon(x_o;d,\delta)$, and

$$f''_\varepsilon(x_o;d,\delta) = \min_{\mu \in M_d(x_o)} \ \max_{x^* \in \partial_\varepsilon f(x_o)_d} \{\mu[\langle x^*,\delta\rangle - f'(x_o;\delta)]\}, \quad \ldots \quad (7.6)$$

where $\partial_\varepsilon f(x_o)_d = \{x^* \in \partial_\varepsilon f(x_o) \mid \langle x^*, d\rangle = f'_\varepsilon(x_o;d)\}$. Moreover, the operations "max" and "min" can commute in (7.6). □

We *interpret* formula (7.6) as a *sensitivity result* for the mathematical program (P). Indeed, $f'_\varepsilon(x_o;d)$ is the optimal value of program (P) while $f'_\varepsilon(x_o+s\delta;d)$ is the optimal value of

$$(P_s)\quad \begin{cases} \max\langle x^*,d\rangle \\ f^*(x^*)+f(x_o+s\delta) - \langle x_o+s\delta,x^*\rangle - \varepsilon\leq 0. \end{cases}$$

The difference between (P) and (P_s) lies in the definition of the constraint set. Actually, (P_s) can be viewed as a perturbed version of (P). Let x^* be a solution of (P_s), i.e. $x^*\in\partial_\varepsilon f(x_o+s\delta)_d$; we clearly have

$$f^*(x^*)+f(x_o)-\langle x_o,x^*\rangle-\varepsilon\leq s\left[\max_{x^*\in\partial_\varepsilon f(x_o+s\delta)_d}\langle x^*,\delta\rangle-f'(x_o;\delta)\right] \qquad \ldots\quad (7.7)$$

so that

$$f'_\varepsilon(x_o+s\delta;d) \leq \max_{x^*\in C_s^+}\langle x^*,d\rangle \qquad (7.8)$$

where C_s^+ is (for all $s\geq 0$) the constraint set defined in (7.7.)
Due to the upper-semicontinuity of the set-valued mapping

$s \rightleftarrows \partial_\varepsilon f(x_o+s\delta)_d$, the expression occuring in the right-hand side of inequality (7.7) is " approximately " s.u. with

$$u = \max_{x^*\in\partial_\varepsilon f(x_o)_d}[\langle x^*,\delta\rangle-f'(x_o;\delta)]. \qquad (7.9)$$

Now, since $f'_\varepsilon(x_o;d) = \max_{x^*\in C_o^+}\langle x^*,d\rangle$, we have that

$$\frac{f'_\varepsilon(x_o+s\delta;d)-f'_\varepsilon(x_o;d)}{s} \leq \frac{1}{s}\left(\max_{x^*\in C_s^+} \langle x^*,d\rangle - \max_{x^*\in C_o^+} \langle x^*,d\rangle \right) \quad \ldots (7.10)^+$$

One could operate in a similar way to obtain a lower bound on $f'_\varepsilon(x_o+s\delta;d)$ of the form $\max_{x^*\in C_s^-} \langle x^*,d\rangle$, where C_s^- is a constraint set akin to (C_s^+) in its definition. Now, due to the interpretation of the set of Kuhn-Tucker coefficients in terms of a marginal function associated with perturbed versions of (P), one can interpret $f''_\varepsilon(x_o;d,\delta)$ (through $(7.10)^+$ and a companion inequality $(7.10)^-$) as the support function (except for the sign) of $M_d(x_o)$ in the direction u defined in (7.9).

Comment. As indicated earlier, Theorem 7.4 was proved by Lemaréchal and Nurminskii under the assumption that f (is finite and) has an everywhere finite conjugate function f^*. Under the only assumption that f is a finite convex function, the same formula (7.6) has been proved very recently by Auslender [1].

We now turn our attention to particular situations where the formulation of $f''_\varepsilon(x_o;d,\delta)$ can be reduced to simpler expressions. The next corollaries can be viewed as complements to works [31] and [1].

Let $v_{\varepsilon,d}$ denote the function $x \longmapsto v_{\varepsilon,d}(x) = f'_\varepsilon(x;d)$. According to Theorem 7.4, $v_{\varepsilon,d}$ admits a directional derivative $\delta \longmapsto v'_{\varepsilon,d}(x_o;\delta)$ at those points x_o where it is differentiable. Firstly, suppose that δ satisfies the following assumption:

(A_d) the linear form $x^* \longmapsto \langle x^*,\delta\rangle$ is constant on $\partial_\varepsilon f(x_o)_d$.

This assumption merely says that the width of $\partial_\varepsilon f(x_o)_d$ in the direction δ is null, or equivalently that δ belongs to the orthogonal subspace to the affine hull of $\partial_\varepsilon f(x_o)_d$. Clearly,

$\delta = \rho d$ ($\rho \in \mathbb{R}$) is among the δ satisfying (A_d). For those directions δ such that (A_d) holds, we have that

$$v'_{\varepsilon,d}(x_o;\delta) = \min_{\mu \in M_d(x_o)} \{\mu[\alpha - f'(x_o;\delta)]\} ,$$

where α is the constant $<\partial_\varepsilon f(x_o)_d, \delta>$. In particular, (A_d) is satisfied by *all* directions δ whenever $\partial_\varepsilon f(x_o)_d$ is reduced to *one* element x^*_d . In such a case,

$$\delta \longmapsto v'_{\varepsilon,d}(x_o;\delta) = \min_{\mu \in M_d(x_o)} \{\mu[<x^*_d,\delta> - f'(x_o;\delta)]\}$$

is a *concave* function, so that $-v_{\varepsilon,d}$ is quasi-differentiable at x_o in Pshenichnyi's sense [38, Chapter 3]. We therefore have:

Corollary 7.5. Let x_o and d be such that $\partial_\varepsilon f(x_o)_d$ is reduced to a single element x^*_d. Then $-v_{\varepsilon,d}$ is quasi-differentiable at x_o and the quasi-differential ∂^* $(-v_{\varepsilon,d}(x_o)$ of $-v_{\varepsilon,d}$ at x_o is given as

$$\partial^*(-v_{\varepsilon,d})(x_o) = M_d(x_o) \quad [\partial f(x_o) - x^*_d] \qquad (7.11)$$

Proof. $\partial^*(-v_{\varepsilon,d})(x_o)$ is defined as the set of x^* satisfying

$$<x^*,\delta> \leq - v'_{\varepsilon,d}(x_o;\delta) = \max_{\mu \in M_d(x_o)} \{\mu[f'(x_o;\delta) - <x^*_d,\delta>]\}$$

for all δ. Consequently, due to the expression for the subdifferential of a maximum of convex functions (see for example [28, p. 355]), we have that

$$\partial^*(-v_{\varepsilon,d})(x_o) = \mathrm{co}\{ \bigcup_{\mu \in M_d(x_o)} \mu(\partial f(x_o) - x^*_d)\}$$

But since $M_d(x_o)$ does not contain negative elements, the above set is nothing else than $M_d(x_o)\ [\partial f(x_o)-x_d^*]$.

Since $v_{\varepsilon,d}$ is a locally Lipschitz function, it has a generalized gradient $\partial v_{\varepsilon,d}$ in Clarke's sense [10] at all $x_o \in \mathbb{R}^n$. What is the *exact* evaluation of $\partial v_{\varepsilon,d}(x_o)$? The expression for $\partial v_{\varepsilon,d}$ as well as consequences of it are given in [1] *in the case where* $M_d(x)$ *is single-valued in a neighborhood of* x_o. The result described in Corollary 7.5 suggests instead to look at the counterpart situation, namely *when* $\partial_\varepsilon f(x)_d$ *is single-valued in a neighborhood of* x_o. We than have the following result :

Corollary 7.6. Suppose $\partial_\varepsilon f(x)_d$ is single-valued in a neighborhood of x_o. Then

$$\partial v_{\varepsilon,d}(x_o) = -\partial^*(-v_{\varepsilon,d})(x_o) = M_d(x_o)[x_d^* - f(x_o)], \qquad (7.12)$$

where x_d^* stands for $\partial_\varepsilon f(x_o)_d$.

Proof. At a point x around x_o where $v_{\varepsilon,d}$ is differentiable, both $v_{\varepsilon,d}$ and $-v_{\varepsilon,d}$ are quasi-differentiable with

$$\partial^* v_{\varepsilon,d}(x) = -\partial^*(-v_{\varepsilon,d})(x) = \{\nabla v_{\varepsilon,d}(x)\}.$$

Therefore, a mere consequence of (7.11) is that

$$v_{\varepsilon,d} \text{ is differentiable at } x \iff \begin{cases} \text{both } M_d(x) \text{ and } \partial f(x) \text{ are} \\ \text{single-valued at } x. \end{cases}$$

Thus, again from (7.11), at all x_i (in a neighborhood of x_o) where $v_{\varepsilon,d}$ is differentiable, we have that

$$\nabla v_{\varepsilon,d}(x_i) = \mu_d(x_i)[x_d^*(x_i) - \nabla f(x_i)].$$

Now, the mapping which assigns to x the unique element $x_d^*(x)$ of of $\partial_\varepsilon f(x)_d$ is continuous (since, as a general rule, the set-valued mapping $x \rightrightarrows \partial_\varepsilon f(x)_d$ is upper-semicontinuous). Similarly,

Similarly, the upper-semicontinuity of the set-valued mapping $x \rightrightarrows M_d(x)$ (see [31] or [1]) makes that $\lim_{i\to\infty} \sup\{\mu_d(x_i)\} \subset M_d(x_o)$. Hence the result (7.12) is easily derived. □

There is a situation where the assumption of the above corollary is automatically satisfied around *any* x_o in $\mathbb{R}^n$, that is when f is differentiable on all of $\mathbb{R}^n$. The precise statement is as follows :

Corollary 7.7. Let f be differentiable on $\mathbb{R}^n$. Then, for all $x \in \mathbb{R}^n$,

$$\partial v_{\varepsilon,d}(x) = M_d(x) \, [x_d^*(x) - \nabla f(x)], \tag{7.13}$$

where $x_d(x)$ is the unique element of $\partial_\varepsilon f(x)_d$. □

Proof. Since f is differentiable, f* is strictly convex on $\mathbb{R}^n$[40, p. 253]. Therefore, the program (P) whose constraint set is defined through f^* has only one solution x_d^*. □

If f is strictly convex, $M_d(x)$ is single-valued for all x (see for example [31] or [1]). Thus, as a by-product of (7.13), we obtain the following "global" statement already observed in [1].

Corollary 7.8. Suppose that f is differentiable and strictly convex on $\mathbb{R}^n$. Then $v_{\varepsilon,d}$ is continuously differentiable and

$$\nabla v_{\varepsilon,d}(x) = \mu_d(x)[x_d(x) - f(x)] \tag{7.14}$$

for all x. □

Since the "strict convexity" and the "differentiability" are dual properties [40, Section 26], if f is differentiable and strictly convex, the same holds for f^*. The correspondence between f and f* is precisely the Legendre transform [40, Theorem 26.6]; it would be worth getting a better insight into the relationship of $v_{\varepsilon,d}$ (associated with f) with the corresponding $v^*_{\varepsilon,d}$ associated with f^*. A(rather trival) illustration of (7.14) is in considering the example of quadratic functions (see Example 1.4). In such a case, tne unique $x_d^*(x_o)$ of $\partial_\varepsilon f(x_o)_d$ is

$$x_d^*(x_o) = Ax_o+b+\left(\frac{2\varepsilon}{<Ad,d>}\right)^{1/2} Ad,$$

while $M_d\ (x_o)$ is reduced to $\mu_d(x_o) = \left(\frac{<Ad,d>}{2\varepsilon}\right)^{1/2}$.

Consequently, we have that

$$f''_\varepsilon(x_o;d,\delta) = <Ad,\delta>$$

whatever $\varepsilon \geq 0$. Of course, the result could have been obtained at once from the expression for $f'_\varepsilon(x;d)$. Nevertheless, this example shows a noteworthy feature (refer to (7.14)): when $\varepsilon \to 0^+$, then $\mu_d(x) \to +\infty$ while $[x_d(x) - \nabla f(x)] \longrightarrow 0$. So, even in the situation of Corollary 7.8, the behaviour of $\nabla v_{\varepsilon,d}(x)$ could be wild when $\varepsilon \to 0^+$.

8. CONCLUSION

In this paper, we reviewed the main properties and calculus rules of the ε-subdifferential of a convex function. The definition of $\partial_\varepsilon f$ was peculiar to convex functions and all the material related to it heavily rested on tools from convex analysis. Actually, another definition was proposed by Taylor [44, p, 745] but for different purposes. For a class of functions close to that of convex functions, the so-called weakly convex functions, Nurminskii and Zhelikhovski [36] proposed a concept of ε-quasigradient and gave an iterative procedure for the minimization of weakly convex functions, formulated in terms of ε-quasi-gradients. In the locally Lipschitz case, the only concept of ε-generalized gradient we are aware of is the one given by Goldstein [13]. However his definition is a local one and cannot reduce for convex functions to the one used in this study.

As for the function $(d,\delta) \longmapsto f''_\varepsilon(x_o;d,\delta)$, it is not clear whether it could be of some help for defining a generalized Hessian matrix for convex functions. Introducing such an object, tractable from the computational viewpoint, is of main concern in the current research in convex analysis.

REFERENCES

[0] E. Asplund and R.T. Rockafellar, Gradients of convex functions, Trans. Amer. Math. Soc. 139 (1969), 443-467.

[1] A. Auslender, Differential properties of the ε-directional derivative of a convex function, to appear.

[2] M.L. Balinski and P. Wolfe, (eds.) Nondifferentiable optimization, Math. Programming Study 3, North-Holland (1975).

[3] V. Barbu and Th. Precupanu, Convexity and optimization in Banach spaces, Sijthoff & Noordhoff, Int. Publishers (1978).

[4] J.M. Borwein, Fractional programming without differentiability, Math. Programming 11 (1976), 283-290.

[5] H. Brezis, Opérateurs maximaux monotones, North-Holland, Amsterdam (1973).

[6] A. Brøndsted, On the subdifferential of the supremum of two convex functions, Math. Scand. 31 (1972), 225-230.

[7] A. Brøndsted and R.T. Rockafellar, On the subdifferentiability of convex functions, Proc. Amer. Math. Soc. 16(1965), 605-611.

[8] C. Castaing and M. Valadier, Convex analysis and measurable multifunctions, Lecture notes in Mathematics 580, Springer-Verlag (1977).

[9] G. Choquet, Ensembles et cônes convexes faiblement complets, Note aux Comptes Rendus Acad. Sc. Paris, t. 254 (1962), 1908-1910.

[10] F.H. Clarke, A new approach to Lagrange multipliers, Math. of Operations Res., 1(1976), 165-174.

[11] V.F. Dem'janov and V.K. Šomesova, Conditional subdifferentials of convex functions, Soviet Math. Dokl. 19(1978), 1181-1185.

[12] I. Ekeland and R. Temam, Analyse convexe et problèmes variationnels, Bordas, Paris(1974).

[13] A.A. Goldstein, Optimization of Lipschitz continuous functions, Math. Programming 13 (1977), 14-22.

[14] J.-B. Hiriart-Urruty, Etude de quelques propriétés de la fonctionnelle moyenne et de l'inf-convolution continue en analyse convexe stochastique, Note aux Comptes rendus Acad. Sc. Paris, t.280(1975), 129-132.

[15] J.-B. Hiriart-Urruty, About properties of the mean value functional and of the continuous infimal convolution in stochastic convex analysis, Proceedings of the 7th I.F.I.P. Conference (Nice, 1975); Lecture Notes in Computer Science (J. Cea. Editor) Springer-Verlag (1976), 763-789.

[16] J.-B. Hiriart-Urruty, Contributions à la programmation Mathématique: cas deterministe et stochastique. Thèse de Doctorat ès -Sciences Mathématiques, Université de Clermont-Ferrand II (1977).

[17] J.-B. Hiriart-Urruty, Gradients généralisés de fonctions composées. Applications. Note aux Comptes Rendus Acad. Sc. Paris, t.285(1977), 781-784.

[39] R. Robert, Contributions à l'analyse non linéaire, Thèse de Doctorat ès-Sciences Mathématiques, Université de Grenoble (1976).

[40] R.T. Rockafellar, Convex analysis, Princeton University Press (1970).

[41] R.T. Rockafellar, Conjugate duality and optimization, Reg. Conf. Ser. in Appl. Math., 16, SIAM Publications (1974).

[42] J. Stoer and C. Witzgall, Convexity and optimization in finite dimensions, Springer-Verlag (1970).

[43] J.-J. Strodiot, Nguyen van Hien and N. Heukemes, Caractérizations des solutions à ε-près en programmation convexe, preprint 1980.

[44] P.D. Taylor, Subgradients of a convex function obtained from a directional derivative, Pac. J. Math. 44(1973), 739-747

[45] M. Thera, Calcul ε-sous-differentiel des applications convexes, Note aux Comptes Rendus Acad. Sc. Paris, t.290(1980), 549-551.

[46] D. Tiba, Subdifferentials of composed functions and applications to optimal control, An Stiin. ale Univ. din Iasi, t. XXIII (1977), 381-386.

[47] M. Valadier, Intégration d'ensembles convexes fermés, notamment d'épigraphes, Rev. Francaise Inf. Rech. Op. 4(1970), 57-73.