Revista de la
Unión Matemática Argentina
Volumen 31, 1984.

ON THE ε-SUBDIFFERENTIAL OF A CONVEX FUNCTION

Telma Caputti

1. INTRODUCTION.

The ε-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.

Throughout this paper, f is a lower-semicontinuous convex function from R^n (the usual vector space of real n-tuples) into $(-\infty,+\infty]$. Given such a function and $\varepsilon > 0$ the ε-subdifferential of f at $x_0 \in \mathrm{dom} f$ (domf is the set where f is finite) is denoted by $\partial_\varepsilon f(x_0)$ and defined by

$$\partial_\varepsilon f(x_0) = \{x \in R^n : f(x_0) + f^*(x) - \langle x,x_0\rangle \leqslant \varepsilon\}$$

where f* designates the Frenchel conjugate of f defined by

$$f^*(x) = \sup_{x_0} \{\langle x_0,x\rangle - f(x_0)\} \qquad [1]$$

and $\langle x_0,x\rangle$ is the usual inner product of two vectors x_0,x.

Let p be a non null vector in R^n; throughout the sequel we shall assume that $x_0 \in \mathrm{int(dom}f)$ (int(domf) is the interior of domf). Then it is well known that $\partial_\varepsilon f(x_0)$ is a nonempty compact convex set so that we can denote

$$v(x_0) = \sup_{x\in\partial_\varepsilon f(x_0)} \langle p,x\rangle = \inf_{\lambda>0} \frac{f(x_0+\lambda p) - f(x_0) + \varepsilon}{\lambda}$$

Moreover a major aim of research is to define a concept of second derivative for a nondifferentiable function. In this respect Nurminski [2] proved that the set-valued mapping $\partial_\varepsilon f(.)$ is locally Lipschitz when f is real-valued. More recently, Hiriart-Urruty [3, Corollary 3.4] proved that this last assumption could be omitted and that $\partial_\varepsilon f(.)$ is locally Lipschitz on int(domf).

Hence v is locally Lipschitz on int(domf) and following Clarke [4] the generalized directional derivative of v at x_0 in direction d, denoted $v^\circ(x_0;d)$ is given by

$$v^\circ(x_0;d) = \lim_{\substack{h\to 0\\ \lambda\to 0^+}} \sup \frac{v(x_0+h+\lambda d) - v(x_0+h)}{\lambda}$$

It follows from a fundamental theorem of Clarke [4, Proposition 1.4] that

$$v^\circ(x_0;d) = \sup_{z\in\partial v(x_0)} \langle z,d\rangle$$

where (since v has at almost all points a derivative) $\partial v(x_0)$ is the convex hull of the set of limits of the form $\nabla v(x_0 + h_n)$ when $h_n \to 0$ as $n \to +\infty$; $\partial v(x_0)$ is called the generalized gradient of v at x_0. We always have $v'(.,.) \leqslant v^\circ(.,.)$.

In the first part (Section 2) some properties of $v(x_0)$ and $v'(x_0;d)$ are proved and in the second part (Section 3) p will be considered as a variable. We shall denote

$$f'_\varepsilon(x_0;d) = v(x_0) \quad ; \quad f''_\varepsilon(x_0;p;d) = v'(x_0,;d)$$

and we shall study the properties of the functions

$$p \longrightarrow f''_\varepsilon(x_0;p;p)$$
$$p \longrightarrow f'_\varepsilon(x_0;p) + \frac{1}{2} f''_\varepsilon(x_0;p;p)$$

since one of the possible applications of the formula giving $f''_\varepsilon(x_0;p;p)$ would be to define a Newton type method for minimizing a nondifferentiable convex function. Following this idea we propose a convergent algorithm similar to defined by Bertsekas-Mitter [5]. In this section we shall describe a descent algorithm for the minimiza tion of a convex function subject to convex constraints. Rather than considering explicity the constraints, however, we shall allow the function to be minimized to take the value $+\infty$.

Thus the problem of finding the minimum of a function g over a set X is equivalent to finding the minimum of the extended real-valued function $f(x) = g(x) + \delta(x/X)$ where $\delta(./X)$ is the indicator function of X, i.e., $\delta(x/X) = 0$ for x in X; $\delta(x/X) = \infty$ for $x \notin X$.

Stating the problem formally: Find $\inf_x f(x)$ where $f\colon R^n \to (-\infty,+\infty]$ is a convex function which is lower semicontinuous with $\inf_x f(x) > -\infty$ and $f(x) < +\infty$ for at least one x in R^n. With this assumption, the function f is a closed proper convex function as defined in [1].

A basic concept for the algorithm that we shall present is the notion of ε-subgradient. This notion was introduced in [6], [7] in connection with investigations related to the existence and characterization of subgradients of convex functions.

PRELIMINARIES AND NOTATIONS.

If we consider the optimization problem

$$\text{(P)} \qquad v(x_0) = \sup_{x \in \partial_\varepsilon f(x_0)} \langle p,x\rangle$$

we can associate the usual dual problem

$$\text{(D)} \qquad \alpha(x_0) = \inf_{u \geq 0} \theta(x_0;u)$$

where
$$\theta(x_0;u) = \sup_{x \in R^n} L(x;x_0;u) \tag{1.1}$$

with

$$L(x;x_0;u) = \begin{cases} \langle p,x\rangle - u(f(x_0) + f^*(x) - \langle x_0,x\rangle - \varepsilon) & \text{if } x \in \text{dom}f^* \quad (1.2) \\ -\infty & \text{otherwise} \quad (1.3) \end{cases}$$

Denote by $U(x_0)$ the set of optimal solutions of (D), that is, $U(x_0) = \{u \geqslant 0: \alpha(x_0) = \theta(x_0;u)\}$ and let $M(x_0)$ be the set of optimal solutions of (P)

$$M(x_0) = \{x \in \partial_\varepsilon f(x_0): v(x_0) = \langle p,x\rangle\}.$$

Since $\partial_\varepsilon f(x_0)$ is compact convex and nonempty, $M(x_0)$ is a nonempty convex compact set. Furthermore, since $\partial_\varepsilon f(.)$ is locally Lipschitz on int(domf) M(.) is closed and locally bounded on int(domf) (the set-valued mapping M(.) is said to be locally bounded at x_0 if there exists a neighborhood V of x_0 such that $\bigcup_{z \in V} M(z)$ is bounded). Also, $U(x_0)$ is a nonempty convex and compact set and since $f = f^{**}$ it follows that

$$\theta(x_0;u) = \begin{cases} u(f(x_0 + \frac{p}{u}) - f(x_0) + \varepsilon) & \text{if} \quad u > 0 \quad (1.4) \\ \sup_{x \in \text{dom}f^*} \langle p,x\rangle & \text{if} \quad u = 0 \; . \quad (1.5) \end{cases}$$

Now, using the methodology of Hogan [8, Theorem 2] we use the following theorem, the Lemarechal-Nurminski theorem [9], deleting the coercivity assumption.

THEOREM 1.1. [9]. *The directional derivative of* v *at* x_0 *in the direction* d *is given as*

$$v'(x_0;d) = \max_{x \in M(x_0)} \; \min_{u \in U(x_0)} - u(f'(x_0;d) - \langle x,d\rangle) \tag{1.6}$$

and the operators max-min commute.

2. PROPERTIES OF THE FUNCTIONS $v(x_0)$ AND $v'(x_0;d)$.

According to the expression of $v'(x_0;d)$ in the Lemarechal-Nurminski theorem and considering p as a variable, we set

$$f'_\varepsilon(x_0;d) = v(x_0) \quad ; \quad f''_\varepsilon(x_0;p;d) = v'(x_0;d).$$

We can study very interesting properties of the following functions

$$p \longrightarrow f''_\varepsilon(x_0;p;p) \tag{2.1}$$

$$p \longrightarrow f'_\varepsilon(x_0;p) + \frac{1}{2} f''_\varepsilon(x_0;p;p). \tag{2.2}$$

We set $U_\varepsilon(x_0;p) = U(x_0)$. Then for all $\lambda > 0$ the relation

$$U_\varepsilon(x_0;\lambda p) = \lambda U_\varepsilon(x_0;p) \tag{2.3}$$

is valid.

According to (1.4) the following statements are equivalent for $u > 0$:

i) $u \in U_\varepsilon(x_0;p)$; ii) $f'_\varepsilon(x_0;p) = u(f(x_0 + \frac{p}{u}) - f(x_0) + \varepsilon)$;

iii) $\lambda u(f(x_0 + \frac{\lambda p}{\lambda u}) - f(x_0) + \varepsilon) = f'_\varepsilon(x_0;p)$; iv) $\lambda u \in U_\varepsilon(x_0;p)$.

PROPOSITION 2.1. a) $f''_\varepsilon(x_0;p;p) \geqslant 0$ *for all* p. (2.4)

b) $f''_\varepsilon(x_0;\lambda p;\lambda p) = \lambda^2 f''_\varepsilon(x_0;p;p)$ *for all* $\lambda > 0$. (2.5)

Proof. From (1.6) in Theorem 1.1 we have

$$f''_\varepsilon(x_0;p;p) = \min_{u \in U_\varepsilon(x_0;p)} - u(f'(x_0;p) - f'_\varepsilon(x_0;p)) \tag{2.6}$$

from which we obtain inequality (2.4) since $u \geqslant 0$ and since $f'_\varepsilon \geqslant f'$. The relation (b) is an immediate consequence of the above proposition and formula (2.6). (q.e.d.)

Throughout the sequel we shall assume henceforth that f is real-valued.

Suppose now that f is strongly convex, that is, there exists $\delta > 0$ such that for each x, y and $\lambda \in [0,1]$ we have

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y) - \lambda(1-\lambda)\delta\|x-y\|^2$$

where $\|.\|$ denotes the usual Euclidean norm in R^n.

It is very easy to establish the following property: If the function $\lambda \longrightarrow \varphi(\lambda) = f(x_0 + \lambda p)$ is strictly convex on R^+, then $U(x_0)$ is reduced to a single point $u(x_0)$.

This property is an immediate consequence of the convexity of f and the properties of the subgradient of $\theta(u)$ with x_0 fixed. Then f is strictly convex and $U_\varepsilon(x_0;p)$ is reduced to a single point $u_\varepsilon(x_0;p)$. Moreover $u_\varepsilon(.,.)$ is strictly positive. So if we define $\bar{u}_\varepsilon(x_0) =$ $= \min\{u_\varepsilon(x_0;p) : \|p\| = 1\}$ we have $\bar{u}_\varepsilon(.) > 0$.

The set $\partial_\varepsilon f(x_0)$ has some interesting properties from the algorithmic point of view as shown by the following two propositions:

PROPOSITION 2.2. *Let* x_0 *be a vector such that* $f(x_0) < \infty$. *Then*

$$0 \leqslant f(x_0) - \inf_z f(z) \leqslant \varepsilon \quad \textit{if and only if} \quad 0 \in \partial_\varepsilon f(x_0).$$

Proof. By definition of ε-subdifferential of f at x_0, that is, $x \in R^n$ is said to be an ε-subgradient of f at x_0 if $f(z) \geqslant f(x_0) - \varepsilon + \langle z - x_0, x\rangle$ for all z in R^n.

In consequence, $0 \in \partial_\varepsilon f(x_0)$ if and only if $f(z) \geqslant f(x) - \varepsilon$ for all z in R^n which is equivalent to the desired relation. (q.e.d.)

PROPOSITION 2.3. *Let* x_0 *be a point such that* $f(x_0) < \infty$ *and* $0 \notin \partial_\varepsilon f(x_0)$. *Let* p *be any vector such that*

$$v(x_0) = f'_\varepsilon(x_0;p) = \sup_{x\in\partial_\varepsilon f(x_0)} \langle p,x\rangle < 0. \tag{2.7}$$

Then we have

$$f(x_0) - \inf_{\lambda>0} f(x_0 + \lambda p) > \varepsilon. \tag{2.8}$$

Proof. Assume the contrary, i.e., $\inf_{\lambda\geq 0} f(x_0 + \lambda p) - f(x_0) + \varepsilon \geqslant 0$, then we have

$$\frac{f(x_0 + \lambda p) - f(x_0) + \varepsilon}{\lambda} \geqslant 0 \quad , \quad \text{for all } \lambda > 0.$$

Using the definition of $v(x_0)$ this implies that

$$\sup_{x\in\partial_\varepsilon f(x_0)} \langle p,x\rangle = \inf_{\lambda>0} \frac{f(x_0 + \lambda p) - f(x_0) + \varepsilon}{\lambda} \geqslant 0 .$$

Since $\partial_\varepsilon f(x_0)$ is closed this implies that $0 \in \partial_\varepsilon f(x_0)$ which contradicts the hypotesis. (q.e.d.)

In the case $0 \notin \partial_\varepsilon f(x_0)$, a possible method for finding a vector $\bar{y}(x_0)$ in R^n such that $\sup_{x\in\partial_\varepsilon f(x_0)} \langle \bar{y}(x_0),x\rangle < 0$ is the following:

Let $x^*(x_0)$ be the unique vector of minimum norm in $\partial_\varepsilon f(x_0)$. Then the vector $\quad \bar{y}(x_0) = -x^*_\varepsilon(x_0)/\|x^*_\varepsilon(x_0)\|$ (2.9)

satisfies $\quad \sup_{x\in\partial_\varepsilon f(x_0)} <\bar{y}(x_0),x> = -\|x^*_\varepsilon(x_0)\| < 0.$

Propositions 2.2 and 2.3 form the basis for the algorithm that we shall present later.

PROPOSITION 2.4. *If* f *is strongly convex, then the functions*

$$p \longrightarrow f''_\varepsilon(x_0;p;p) \quad \textit{and} \quad p \longrightarrow f'_\varepsilon(x_0;p) + \frac{1}{2} f''_\varepsilon(x_0;p;p)$$

satisfy the following relations

$$f''_\varepsilon(x_0;p;p) = k_\varepsilon(x_0) \|p\|^2 \quad \textit{for all } p \tag{2.10}$$

$$(f'_\varepsilon(x_0;p) + \frac{1}{2} f''_\varepsilon(x_0;p;p)) \geqslant \|p\| \ (-\|x^*_\varepsilon(x_0)\| + \frac{1}{2} k_\varepsilon(x_0)\|p\|) \tag{2.11}$$

Proof. We remark that

$$\frac{1}{\|p\|} v(x_0) \geqslant \min_{\|d\|\leq 1} \max_{z\in\partial_\varepsilon f(x_0)} <z,d> = -\|x^*(x_0)\| .$$

Moreover, if f is strictly convex, we have

$$\frac{f(x_0 + \lambda p) - f(x_0) + \varepsilon}{\lambda} \geqslant f'(x_0;p) + \lambda\|p\|^2\delta + \frac{\varepsilon}{\lambda} , \text{ for all } \lambda > 0.$$

This inequality implies

$$\inf_{\lambda>0} \frac{f(x_0 + \lambda p) - f(x_0) + \varepsilon}{\lambda} \geqslant f'(x_0;p) + \min_{\lambda>0} \{\lambda\|p\|^2 + \frac{\varepsilon}{\lambda}\}$$

which is equivalent to

$$f'_\varepsilon(x_0;p) - f'(x_0;p) \geqslant 2\sqrt{\varepsilon\delta} \ \|p\|$$

and since $U_\varepsilon(x_0;p)$ is homogenous in p and is reduced to a single point $\bar{u}_\varepsilon(x_0)$ we obtain the relations (2.10) and (2.11) respectively. (q.e.d.)

REMARK 2.1. If $0 \in \partial_\varepsilon f(x_0)$ then $f'_\varepsilon(x_0;p) \geqslant 0$ for each p and from (2.4) we have $f'_\varepsilon(x_0;p) + \frac{1}{2} f''_\varepsilon(x_0;p;p) \geqslant 0$ for all p.
If $0 \notin \partial_\varepsilon f(x_0)$, then there exists p such that $f'_\varepsilon(x_0;p) < 0$. Consequently, there exists p satisfying:

$$\|p\| \leqslant 1 \quad ; \quad f'_\varepsilon(x_0;p) + \frac{1}{2} f''_\varepsilon(x_0;p;p) < 0. \tag{2.12}$$

Therefore,

$0 \notin \partial_\varepsilon f(x_0)$ if and only if $\min\limits_{\|p\| \leq 1} \{f'_\varepsilon(x_0;p) + \frac{1}{2} f''_\varepsilon(x_0;p;p)\} < 0.$

If f is strongly convex from Proposition 2.4 we obtain the following equivalence

$0 \notin \partial_\varepsilon f(x_0)$ if and only if $\min\limits_{p \in R^n} \{f'_\varepsilon(x_0;p) + \frac{1}{2} f''_\varepsilon(x_0;p;p)\} < 0.$

REMARK 2.2. One can prove that $U_\varepsilon(x_0;.)$ is locally bounded and closed at each $p \neq 0$.

Then from (2.6) it follows that the function

$$p \rightarrow f'_\varepsilon(x_0;p) + \frac{1}{2} f''_\varepsilon(x_0;p;p)$$

is lower semicontinuous.

3. APPLICATIONS IN ALGORITHMS.

In connection with Propositions 2.2 and 2.3 we can state that whenever the value f(x) exceeds the optimal value by more than ε, then by a descent along a vector x satisfying (2.7) in Proposition 2.3 we can decrease the value of the cost by at least ε.

Consider the following descent algorithm for the minimization of a convex function subject to convex constraints which is a descent numerical method for optimization problems with nondifferentiable cost functionals:

STEP 1. Select a vector x_0 such that $f(x_0) < \infty$, a scalar $\varepsilon_0 > 0$ and a scalar a, $0 < a < 1$.

STEP 2. Given x_n and $\varepsilon_n > 0$, set $\varepsilon_{n+1} = a^k \varepsilon_n$ where k is the smallest non-negative integer such that $0 \notin \partial_{\varepsilon_{n+1}} f(x_n)$.

STEP 3. Choose a vector y_n that satisfies

$$f'_{\varepsilon_{n+1}}(x_n;y_n) + \frac{1}{2} f''_{\varepsilon_{n+1}}(x_n;y_n;y_n) < 0 .$$

From Remark 2.1, such a vector exists if $0 \notin \partial_{\varepsilon_{n+1}} f(x_n)$, and (2.7) is valid.

STEP 4. Set $x_{n+1} = x_n + \lambda_n y_n$ where $\lambda_n > 0$ is such that $f(x_n) - f(x_{n+1}) > \varepsilon_{n+1}$. Return to Step 2.

REMARK 3.1. If x_n is not a minimizing point of f there always exists a non-negative integer k such that $0 \notin \partial_{a^k\varepsilon_n} f(x_n)$ since by Proposition 2.2 we have

$0 \notin \partial_{\varepsilon_{n+1}} f(x_n)$ if and only if $f(x_n) - \inf_x f(x) > \varepsilon_{n+1} = a^k \varepsilon_n$

and by Proposition 2.3 there exists a scalar ε_n such that

$$f(x_n) - f(x_n + \lambda_n y_n) > \varepsilon_{n+1} \tag{3.1}$$

thus showing that Step 4 can always be carried out. One way of finding a scalar λ_n satisfying (3.1) is by means of the one-dimensional minimization

$$f(x_n + \lambda_n y_n) = \min_{\lambda>0} f(x_n + \lambda y_n)$$

assuming the minimum is attained. This in turn can be guaranteed whenever the set of minimizing points of f is nonempty and compact, since in this case all the level sets are compact [1].

REMARK 3.2. We note that Steps 2 and 3 of the algorithm can be carried out by means of the auxiliary minimization problem:

$$\min_{x \in \partial_{a^k \varepsilon_n} f(x_n)} \|x\| . \tag{3.2}$$

Now clearly we have $0 \in \partial_{a^k \varepsilon_n} f(x_n)$ if and only if (3.2) has a zero optimal value and therefore Step 2 of the algorithm can be carried out by solving problem (3.2) successively for $k = 0,1,\ldots$. There exists an integer k for which the problem (3.2) has a nonzero optimal value. Let x^* be the optimal solution of problem (3.2) for the first such integer k. Then a suitable direction of descent y_n satisfying (2.7) in Step 3 of the algorithm is given by $y_n = -x^*/\|x^*\|$.

REMARK 3.3. This algorithm is the same as defined by Bertsekas and Mitter in their paper but the kind of choice for y_n is different. However, the proof of convergence given in [5] is always valid with this kind of choice. Certainly, a good choice of y_n would be a vector that minimizes the function

$$p \longrightarrow f'_{\varepsilon_{n+1}}(x_0;p) + \frac{1}{2} f''_{\varepsilon_{n+1}}(x_0;p;p)$$

on the unit ball.

We are now attempting to implement such a choice.

After the release of the preprint of this article, the author has been informed about the fact that a recent work along similar lines has been published by J.B.Hiriart-Urruty. Unfortunally she has not been able to read it and verify the overlap between both papers.

REFERENCES

[1] R.T.ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton New Jersey, 1970.

[2] E.A.NURMINSKI, *Nondifferentiable optimization with ε-subdifferential methods*, Working Paper 78-55, I.I.A.S.A. Luxemburg, 1978.

[3] J-B.HIRIART-URRUTY, *Lipschitz r-continuity of the approximate subdifferential of a convex function*, Mathematica Scandinavica 47 (1980) p.p. 123-134.

[4] F.CLARKE, *Necessary conditions for a nonsmooth problem in optimal control and the calculus of variations*, Dissertation, University of Washington, Seattle, W.A., 1973.

[5] D.P.BERTSEKAS-S.K.MITTER, *A descent numerical method for optimization problems with nondifferentiable cost functionals*, SIAM Journal of Control and Optimization 11 (1973) p.p. 637-652.

[6] R.T.ROCKAFELLAR, *Characterization of the subdifferentials of convex functions*, Pacific Journal of Mathematics 17 (1966), p.p. 497-510.

[7] A.BRONSTED-R.T.ROCKAFELLAR, *On the subdifferentiability of convex functions*, Proc. Math. Soc. 16 (1965), p.p. 605-611.

[8] W.W.HOGAN, *Directional derivatives for extremal value functions with applications to the completely convex case*, Operations Research 21 (1973) p.p. 188-206.

[9] C.LEMARECHAL-E.A.NURMINSKI, *Sur la différentiabilité de la fonction d'appui du sous différentiel approché*, Comptes Rendus de l'Académie des Sciences 290 A(1980), p.p. 855-858.

Departamento de Matemática
Facultad de Ciencias Exactas
Universidad de Buenos Aires.

Recibido en marzo de 1984.