MÉMOIRES DE LA S. M. F.

J-B. HIRIART-URRUTY

New concepts in nondifferentiable programming

Mémoires de la S. M. F., tome 60 (1979), p. 57-85
<http://www.numdam.org/item?id=MSMF_1979__60__57_0>

Analyse non convexe [*1977. Pau*]
Bull. Soc. math. France,
Mémoire 60, 1979, p. 57-85.

NEW CONCEPTS IN NONDIFFERENTIABLE PROGRAMMING

J-B. Hiriart-Urruty

Introduction. Numerous studies have been devoted to the determination of the first order necessary optimality conditions for an optimization problem. The study of such conditions and the applications to different problems have first been made in a geometrical form, using for that conical approximations of a subset (Dubovitskii and Milyutin's formalism and its extensions) and separation theorems for convex cones. In differentiable programming, the objective function and the functions defining the constraint set (equalities, inequalities, mixed data) are supposed to be *differentiable* at the considered optimal point. The best known necessary optimality criterion for such a mathematical programming problem is the Kuhn-Tucker criterion. In order for the Kuhn-Tucker criterion to hold, one must impose a constraint qualification on the constraints of the problem; various constraint qualifications have been considered: conditions on the geometry of the constraint, on the representation form of the constraint, conditions combining the objective function and the functions defining the constraints . . . . The introduction of the notion of subdifferential in *convex* analysis has allowed the extension of optimality conditions (and their applications) to nondifferentiable convex problems by replacing the notion of gradient by that of subdifferential. This concept has appeared very fruitful to handle nondifferentiable convex problems.

During the last years, different attempts in considering nondifferentiable nonconvex problems have been made: in the absence of both differentiability and convexity assumptions on the functions involved in the problem, the first step was in defining a new concept coinciding with the notion of gradient in the differentiable case and coinciding with the notion of subdifferential in the convex case. In Part I, we mention some "disconvexifying" processes which have been recently developed. This enumeration, although nonexhaustive, may appear to the reader like a catalogue of definitions. We thought it was not fruitless to recall these different approaches and to show the evolution of ideas and the successive generalizations. In fact, each of these concepts has its own interest and for each introduced notion, there exists a class of functions (including the differentiable or convex ones) which is well adapted. Among all these concepts, we shall emphasize in the sequel the concept of generalized gradient for locally Lipschitz functions.

The Part II of this paper deals with the different conical approximations of a subset S at $x_0 \in S$. Beside the cones of feasible displacements which are classical in mathematical programming, we give some further details about the concept of tangent cone such as introduced by F.H. Clarke in [13] and we studied in [35] in a Banach space setting. The functions μ_S and Δ_S connected with S and introduced in [35] play a rôle similar to those of the indicator function δ_S and the distance function; wespecify some of their properties in the context of convex analysis as well as their influences in the comparison results between tangent cones.

The Part III is concerned with the "functional" part of the "geometrical" notions of Part II. Recalling the definitions of different generalized subdifferentials and generalized gradients, comparison results and examples are given.

The Part IV of this study is devoted to some examples of necessarv (and in certain cases of sufficient) optimality conditions for a nondifferentiable nonconvex optimization problem. The references to developments on the subject are also indicated.

In many mathematical programming problems, the objective function as well as the functions defining the constraints occur to be composite functions. The Part V is exclusively concerned with chain rules on generalized gradients of locally Lipschitz functions. Beside those already existing, we establish a new chain rule for generalized gradients: the first result in this sense is a general inclusion between generalized gradients of the composing functions; after we give sufficient conditions for this inclusion to be an equality.

In the list of references, we just quoted the papers related to the introduction of new concepts in nondifferentiable programming and to their applications to necessary optimality conditions. In particular, we hold apart the papers specifically dealing with the study of algorithms for nondifferentiable problems.

Part I: SOME "DISCONVEXIFYING" PROCESSES.

A.1. B. N. Pschenichnyi's work [55] was probably one of the first attempts in considering nondifferentiable nonconvex functions. Let f be a function defined on a topological vector space E and taking values in $\mathbb{R}$. f is said to be *quasi-differentiable* at $x_0 \in E$ in the sense of B. N. Pschenichnyi if

(1.1) $$f'(x_0;d) = \lim_{\lambda \to 0^+} [f(x_0 + \lambda d) - f(x_0)]\,\lambda^{-1} \text{ exists for all } d$$

and if

(1.2) there exists a nonempty weak* closed subset $M_f(x_0)$ of E^* such that

$$f'(x_0;d) = \underset{x^* \in M_f(x_0)}{\mathrm{Max}} \langle x^*, d\rangle$$

If f is Gâteaux-differentiable at x_0, it is quasi-differentiable at x_0 and $M_f(x_0) = \{\nabla f(x_0)\}$; likewise, if f is a convex function, $\partial f(x_0) = M_f(x_0)$. The properties of quasi-differentiable functions are studied in B. N. Pschenichnyi's book [55] ; the notion of quasi-differentiability is examined and related to fractional programming in [9] ; Lagrangean conditions for a quasi-differentiable optimization problem are considered in [17].

R. Janin generalized the given definition to functions taking values in $\overline{\mathbb{R}}$ by saying that f is *sub-linearizable* at x_0 when the limit exists in (1.1) (possibly $+$ or $-\infty$) and when the function $d \mapsto f'(x_0; d)$ is convex (as a function taking values in $\overline{\mathbb{R}}$). The set $M_f(x_0)$ is defined as in (1.2) by setting:

(1.3) $$x^* \in M_f(x_0) \Leftrightarrow \forall\, d \in E,\ \langle x^*, d\rangle \leqslant f'(x_0; d)$$

Particular sub-classes of the class of sub-linearizable functions (*almost convex functions of the 1st order, of the 2nd order . . .*) are also exhibited by R. Janin; properties of such functions are detailed in the Chapter I of [38].

In these two neighboring definitions ((1.1), (1.2) and (1.3)), it is supposed on the one hand that the limit *exists* in (1.1) and on the other hand that the function $d \mapsto f'(x_0; d)$ is *convex*. This last (stringent) condition permits considering the function $f'(x_0; .)$ as a support function and introducing the convex set $M_f(x_0)$.

We shall consider again the quasi-differentiable functions in the Part V.

A.2. Definitions which come near to the definitions of convex analysis are given by E. A. Nurminskii [48] in the following manner: given a function f from a finite-dimensional euclidean vector space E_n into R; f is said to be *weakly convex* if for every $x_0 \in E_n$ there exists a nonempty set $M_f(x_0)$ of elements x^* such that for all $x \in E_n$

(1.4) $$f(x) \geqslant f(x_0) + \langle x^*, x-x_0 \rangle + r(x_0, x) \qquad \text{where}$$

(1.5) $r(x_0, x)\|x-x_0\|^{-1} \to 0$ for $x \to x_0$, uniformly with respect to x_0 in each compact subset of E_n.

The set $M_f(x_0)$ of elements x^* satisfying (1.4) is a convex compact subset, supposed to be nonempty by definition: that is the set of *quasi-gradients* of f at x_0. Of course, by changing the sense of the inequality (1.4), one has the definition of weakly concave functions. If f is convex, $M_f(x_0)$ is the subdifferential $\partial f(x_0)$ for all points $x_0 \in E_n$.

In the definition of weakly convex functions, the relation (1.5) has to be verified *uniformly* in each compact subset of E_n; so, a continuously differentiable function is weakly convex (and concave) with as unique quasi-gradient at x_0, $\nabla f(x_0)$.

By adding $-\epsilon$ in the right-hand side of the inequality (1.4), the concept of *ϵ-quasi-gradient* is introduced by E. A. Nurminskii and A. A. Zhelikhovskii [49] who also give an iterative procedure for the minimization of weakly convex functions, formulated in terms of ϵ-quasi-gradients.

The class of quasi-differentiable (resp. sub-linearizable, weakly convex) functions is stable for certain usual operations such as addition, maximum of a family of functions. . . .

A.3. If one sets the inequality (1.4) with only the following condition on the residual term r:

(1.6) $$\lim_{x \to x_0} r(x_0, x)\|x-x_0\|^{-1} = 0$$

one finds the notion of *$\geqslant$-gradient* (and of *$\leqslant$-gradient* with the reversed inequality for (1.4)) studied by M. S. Bazaraa, J. J. Goode and Z. Nashed [4]. If f is a convex function, for each point x_0, the set of $\geqslant$-gradients of f at x_0 is the subdifferential $\partial f(x_0)$ and if f is concave, the $\leqslant$-gradients are the supgradients. In relation with the differentiability properties of a function, we have the following results:

(1.7) If x_1^* and x_2^* are $\geqslant$ and $\leqslant$-gradients of f at x_0, then f is Fréchet-differentiable at x_0 and $x_1^* = x_2^* = \nabla f(x_0)$ [4, Theorem 4.1]

(1.8) If f is Fréchet-differentiable at x_0, $\nabla f(x_0)$ is the unique $\geqslant$-gradient and the unique $\leqslant$-gradient. Concerning the support function of the set of $\geqslant$-gradients of f at x_0 (denoted by $\partial \geqslant f(x_0)$), one has necessarily [4, Theorem 3.1]:

(1.9) $$\text{For } x^* \in \partial \geqslant f(x_0),\ \forall d,\ \langle x^*, d\rangle \leqslant \liminf_{\lambda \to 0^+} [f(x_0 + \lambda d) - f(x_0)]\lambda^{-1}$$

that is

(1.10) $$\forall d,\ \delta^*_{\partial \geqslant f(x_0)}(d) \leqslant \liminf_{\lambda \to 0^+} [f(x_0 + \lambda d) - f(x_0)]\lambda^{-1} .$$

As noticed by M. S. Bazaraa *et al* [4, p. 399], the relation (1.9) gives a necessary but not sufficient condition for a vector x^* to be a $\geqslant$-gradient. In fact, there is very little to change in (1.9) to have a necessary and sufficient condition (see A.4 below).

The notions of $\geqslant$ and $\leqslant$-gradients are related to the cones of feasible displacements for the epigraph and for the hypograph of f from $(x_0, f(x_0))$ [4, Theorem 3.2]; we shall come again to these geometrical characterizations later on.

A.4. The condition (1.9) which is necessary for $x^* \in \partial \geqslant f(x_0)$ leads to the consideration at x_0 of different convex subsets the definitions of which are analogous to that of (1.9) but with different lower and upper limits in the right-hand side. This approach (in a general context) is due to J.-P. Penot [52, 53] who, using the way in which the derivatives of Denjoy, Young, Saks generalize the notion of a derivative of a function defined on $\mathbb{R}$, introduced different concepts of *generalized subdifferentials.*

Let E be a real Banach space, let $f : E \to \overline{\mathbb{R}}$ be finite at x_0. In fact, J.-P. Penot's definitions are given in a more general context, by considering functions defined on a topological vector space and taking values in an ordered topological vector space [53]. For our particular case, J.-P. Penot defines successively:

(1.11) $$\forall d \in E, \quad \underline{f}(x_0; d) = \liminf_{\substack{\lambda \to 0^+ \\ v \to d}} [f(x_0 + \lambda v) - f(x_0)]\lambda^{-1}$$
$$\overline{f}(x_0; d) = \limsup_{\substack{\lambda \to 0^+ \\ v \to d}} [f(x_0 + \lambda v) - f(x_0)]\lambda^{-1}$$

and the analogous "radial" definitions:

(1.12) $$\underline{f}_r(x_0; d) = \liminf_{\lambda \to 0^+} [f(x_0 + \lambda d) - f(x_0)]\lambda^{-1}$$
$$\overline{f}^r(x_0; d) = \limsup_{\lambda \to 0^+} [f(x_0 + \lambda d) - f(x_0)]\lambda^{-1}$$

These definitions lead to the definitions of the *lower subdifferential* and of the *upper subdifferential* of f at x_0 by setting:

(1.13) $$x^* \in \underline{\partial} f(x_0) \Leftrightarrow \forall d,\ \langle x^*, d\rangle \leqslant \underline{f}(x_0; d)$$
$$x^* \in \overline{\partial} f(x_0) \Leftrightarrow \forall d,\ \langle x^*, d\rangle \leqslant \overline{f}(x_0; d)$$

The *radial lower subdifferential* $\underline{\partial}_r f(x_0)$ and *the radial upper subdifferential* $\bar{\partial}^r f(x_0)$ are defined in the same way starting from the definitions given in (1.12).

To make connections with what is seen above in Section A.3, one easily shows that $\partial \geqslant f(x_0)$ is exactly $\underline{\partial} f(x_0)$. As for the preceding approaches, one can relate these different notions to that of subdifferentiability for a convex function and to those of differentiabilities for Gâteaux (Hadamard, Fréchet . . .)-differentiable functions.

The different directional derivatives introduced in (1.11) and (1.12) are not generally convex functions (as functions of $d \in E$). The support functions of the different generalized subdifferentials are the biconjugate functions of the corresponding directional derivatives and thereby may be identically equal to $-\infty$ (for instance, $\delta^*_{\underline{\partial} f(x_0)} = [\underline{f}(x_0;.)]^{**}$). When one considers a nondifferentiable function, one cannot say if, for all points, f has $\geqslant$-gradients or $\leqslant$-gradients or if $\bar{\partial} f(x)$, $\bar{\partial}^r f(x)$ . . . are nonempty. That is a difficulty in the utilization of these concepts and in the study of necessary optimality conditions [53, § 5].

To assure that the generalized subdifferentials $\underline{\partial} f(x_0)$ and $\bar{\partial} f(x_0)$ are nonempty, it is necessary to make assumptions which express a *boundedness* property (in a neighborhood of x_0) and a *convexity* property of the functions $\underline{f}(x_0;.)$ and $\bar{f}(x_0;.)$. For that, the class of *unscarped, tangentially convex* functions and the class of *smooth, inwardly tangentially convex* functions are considered by J.-P. Penot [53, 3.6–3.12].

A.5. In mathematical programming, an important class of functions is the class of *quasi-convex* functions. A function $f : E \to \bar{R}$ defined on a topological vector space is said to be quasi-convex if $S_\lambda(f)$ defined by $\{x \in E \mid f(x) \leqslant \lambda\}$ is convex for all λ. For nondifferentiable quasi-convex functions, two neighboring concepts have been recently introduced and studied.

Let f be a quasi-convex function, finite at x_0. H. J. Greenberg and W. P. Pierskalla [28], Y. I. Zabotin, A. I. Korablev and R. F. Khabibullin [63] defined the following set:

$$\partial^* f(x_0) = \{x^* \in E^* \mid f(x) < f(x_0) \Rightarrow \langle x^*, x - x_0\rangle \leqslant 0\} \tag{1.14}$$

Contrary to previous definitions, this concept is a *cone* which is related to the normal cones to the level sets $S_\lambda(f)$. H. J. Greenberg *et al* called $\partial^* f(x_0)$ the *quasi-subdifferential* of f at x_0, whereas Y. I. Zabotin *et al* called it the *generalized support* of f at x_0.

In a slightly different manner, J.-P. Crouzeix [18] introduced the *tangential* of f at x_0 as following:

$$x^* \in Tf(x_0) \Leftrightarrow \forall \lambda < f(x_0),\ \mathrm{Sup}\,[\langle x^*, x - x_0\rangle \mid x \in S_\lambda(f)] < 0 \quad . \tag{1.15}$$

Each of these two concepts has its own interest; the properties of the tangential have been studied in connection with conjugacy and duality theory in quasi-convex analysis [18, 19].

A.6. A notion which is related in a certain sense to $\bar{\partial}_r f$ (or $\bar{\partial} f$) is the notion of *generalized gradient* of a function in the sense of F. H. Clarke [12] (see also N. Z. Shor [59]). Let us recall this definition: given a locally Lipschitz function $f : R^n \to R$, that is to say a function satisfying a Lipschitz condition on all bounded subsets of R^n, the generalized gradient of f at x_0 is the convex compact subset denoted $\partial f(x_0)$ (like the subdifferential) and the support function of which is:

(1.16)
$$d \mapsto f^\circ(x_o; d) = \limsup_{\substack{x \to x_o \\ \lambda \to 0^+}} \, [f(x + \lambda d) - f(x)]\,\lambda^{-1}$$

in other words

(1.17)
$$x^* \in \partial f(x_o) \Leftrightarrow \forall\, d,\ \langle x^*, d\rangle \leqslant f^\circ(x_o; d)$$

Owing to the Lipschitz property of f in a neighborhood of x_o, let us remark that we also have:

$$f^\circ(x_o; d) = \limsup_{\substack{x \to x_o,\, v \to d \\ \lambda \to 0^+}} \, [f(x + \lambda v) - f(x)]\,\lambda^{-1} \ .$$

Moreover, if we denote

$$\forall\, d, \qquad f_\circ(x_o; d) = \liminf_{\substack{x \to x_o \\ \lambda \to 0^+}} \, [f(x + \lambda d) - f(x)]\,\lambda^{-1}$$

an equivalent definition of $\partial f(x_o)$ is:

$$x^* \in \partial f(x_o) \Leftrightarrow \forall\, d,\ \langle x^*, d\rangle \geqslant f_\circ(x_o; d)$$

If we consider any function $f : R^n \to R$, the function $f^\circ(x_o; .)$ is necessarily a *convex* function; if, moreover, f is a locally Lipschitz function, the generalized gradient is *nonempty* for all x. The definitions (1.15) and (1.17) have been considered again in a Banach space context [14, 15].
An equivalent definition of $\partial f(x_o)$ in a finite-dimensional context is the following one: a locally Lipschitz function is, according to a Rademacher's theorem (see [60] or [39]), differentiable almost everywhere; if we denote by $\mathcal{D}$ the set of points where f is differentiable, we have:

(1.18)
$$\partial f(x_o) = \text{co}\,\{\lim_{i \to \infty} \nabla f(x_i) \mid x_i \to x_o,\ x_i \in \mathcal{D}\} \ .$$

This kind of characterization of the generalized gradient of a locally Lipschitz function f has been studied when f is defined on a separable Banach space by L. Thibault [61]. Among all the properties of the generalized gradient, let us recall those ones which will be continually used in the sequel. Let E be a real Banach space, let $f, f_1, f_2, \ldots$ be locally Lipschitz functions defined on E, let the generalized gradient of such functions be defined as in (1.17); then we have [14]:

(1.19) $\partial f(x_o)$ is a nonempty weak*-compact subset of E^*

(1.20) if f is Fréchet-differentiable at x_o, the derivative being strong at x_o ([50, p. 71], [47]), $\partial f(x_o)$ is reduced to the Fréchet-derivative of f at x_o. If f is convex, $\partial f(x_o)$ coincides with the subdifferential of f at x_o in the sense of convex analysis.

(1.21) $\partial(f_1 + f_2)(x_o) \subset \partial f_1(x_o) + \partial f_2(x_o)$

(1.22) $\partial(-f)(x_o) = -\partial f(x_o)$.

General chain rules on generalized gradients will be considered in the last section. The first part of the relation (1.20) shows that the generalized gradient is a generalization of the concept of strong derivative.
A sub-class of locally Lipschitz functions, called *well-behaved* functions, has been considered for numerical

purposes by A. Feuer [23]. When $E = R^n$, a concept of *ε-generalized gradient* has been introduced by A. A. Goldstein [25] in order to define a method of descent for locally Lipschitz functions.
We now consider a nonempty subset S of E and we let d_S be its distance function. In a first approach, F. H. Clarke considered $E = R^n$, S closed, [13], and defined the *normality* to S at x_o as following: the *normal cone* to S at $x_o \in S$ is the closed convex cone denoted by $N(S; x_o)$ (or $N_S(x_o)$) such that

(1.23) $$N(S; x_o) = \overline{cc}\,\partial d_S(x_o) \text{ (closed conical hull of } \partial d_S(x_o))$$

or equivalently

(1.24) $$N(S; x_o) = \overline{co}\,\{\lim_{i\to\infty} \lambda_i(x_i - \overline{x}_i)\} \text{ with } \lambda_i > 0,\ x_i \to x_o \text{ and } \overline{x}_i \text{ a point of } S \text{ closest to } x_i.$$

A slight different definition is considered when E is a Banach space [14, Definition 2]. The generalization of the notion of generalized gradient to a class of functions broader than the class of locally Lipschitz functions has been attempted by the way of the normal cone to the epigraph of f at $(x_o, f(x_o))$ [13, Proposition 3.18]. Let $f : R^n \to R^{\cdot}(= R \cup \{+\infty\})$ be a lower semi-continuous (l.s.c.) function, let x_o be a point where f is finite; then the generalized gradient is defined by

(1.25) $$\partial f(x_o) = \{x^* \in R^n \mid (x^*, -1) \in N_{\text{epif}}(x_o, f(x_o))\} .$$

This is a *geometrical* definition and thereby it is more difficult to work with it. We adopted the same definition in a Banach space context for a function taking possibly values in $\overline{R}$; for details see [35]. For a function taking values in R, $f^{\cdot}(x_o; .)$ defined in (1.16) is convex (as a function from E to $\overline{R}$) and one would have defined a sort of generalized gradient by adopting the same definition as in (1.17). Let us note at once that this notion does not coincide with that one geometrically defined in (1.25); example: $f(x) = -|x|^{1/2}$, $\partial f(0) = \emptyset$ and adopting the definition (1.17), $\partial f(0)$ would be R.
For some comparison results between the generalized gradient and some other concepts recalled in this Part I, see Parts III and V.

Part II: CONES CONNECTED WITH A SUBSET OF A BANACH SPACE.

II.A. Let E be a real Banach space, let S be a subset of E. By *intS, clS* and *bdS*, we denote respectively the interior, the closure and the boundary of S. By S^c we mean the complementary set of S in E and the interior of S^c will be denoted *extS*. In the sequel, by E^*_σ we shall mean the topological dual space E^* endowed with the weak* topology. B^* is the unit ball in E^* and the norm of an element $x^* \in E^*$ is denoted by $\|x^*\|_*$.
If L is a linear topological space and L^* its topological dual space, the *polar cone* of $A \subset L$ is given by:

$$A^\circ = \{x^* \in L^* \mid \forall\, a \in A,\ \langle x^*, a\rangle \leqslant 0\} .$$

We recall the definitions of several kinds of feasible displacements which are classical in mathematical programming. Let S be a nonempty subset of E and $x_o \in clS$.
Let $S_t(x_o) = (S - x_o)t^{-1}$ for every $t > 0$; $V(x_o)$ denoting the filter of neighborhoods of x_o in E, the

family $\{S_t(x_o) \mid t>0;\ V(x_o)\}$ is a filtered family [6, p. 125–126].

Definition 1. $d \in E$ *is said to be an adherent displacement for* S *from* x_o *if and only if*

$$d \in \limsup_{t \to 0^+} S_t(x_o) = \{d \in E \mid \exists\, \mu_n \downarrow 0,\ d_n \to d \text{ with } x_o + \mu_n d_n \in S\} \quad .$$

An equivalent definition of an adherent displacement d is given by saying that there exists a sequence $\{x_n\} \subset S$ converging to x_o and a nonnegative real sequence $\{\lambda_n\}$ such that $d = \lim_{n\to\infty} \lambda_n(x_n - x_o)$. A slightly different definition consists in considering the *weak adherent displacements* d, that is to say by supposing in the definition above that $\lambda_n(x_n - x_o)$ converges weakly to d; see [5, 27].

Definition 2. $d \in E$ *is said to be an interior displacement for* S *from* x_o *iff for every sequence* $\{d_n\}$ *converging to* d *and for every sequence* $\{\lambda_n\} \subset R_+^*$ *converging to* 0, *one has* $x_o + \lambda_n d_n \in S$ *for* n *sufficiently large.*

The set of interior displacements will be denoted by $I(S;x_o)$ (or $I_S(x_o)$) and the set of adherent displacements by $T(S;x_o)$ (or $T_S(x_o)$). $I(S;x_o)$ is an open cone and $T(S;x_o)$ a closed one. For various properties of these cones, we refer to the Chapter I of P.-J. Laurent's book [40].

The *radial* cones of feasible displacements which correspond to $T(S;x_o)$ and $I(S;x_o)$ are defined as following:

Definition 3. $d \in E$ *is said to be a radial adherent displacement for* S *from* x_o *if there exists a sequence* $\{\lambda_n\} \subset R_+^*$ *converging to* 0 *such that* $x_o + \lambda_n d \in S$ *for all* n.

Definition 4. $d \in E$ *is said to be a radial interior displacement for* S *from* x_o *if there exists* $\epsilon > 0$ *such that* $x_o + \lambda d \in S$ *for all* $\lambda \in\,]0, \epsilon[$.

The set of radial adherent displacements will be denoted by $T^r(S;x_o)$ (or $T_S^r(x_o)$) and the set of radial interior displacements by $D(S;x_o)$ (or $D_S(x_o)$). If S is locally star-shaped at x_o, $D(S;x_o) = T^r(S;x_o)$. Generally, $T^r(S;x_o)$ and $D(S;x_o)$ have no topological property.
All these cones have been extensively used (under different names) in mathematical programming since their definitions for geometrical purposes in the thirties. The cone of interior displacements is essentially used when the constraint set S is defined by inequalities and the cone of adherent displacements when S is defined by equalities. The cone $D(S;x_o)$ (sometimes called cone of feasible directions) has a particular interest for the algorithmic point of view (see for instance [64, §2.4]).

Beside these notions, F. H. Clarke introduced in the case where E is finite-dimensional and S a closed subset of E the notion of tangent cone to S at $x_o \in S$ [13, Definition 3.6]. We adopted the same definition in the context of a Banach space for S being an arbitrary nonempty subset of E and $x_o \in c\ell S$ [35, Definition 3]. We recall that the *tangent cone* $\mathcal{T}(S;x_o)$ is the *closed convex cone* defined as following:

$$\mathcal{T}(S;x_o) = [\partial d_S(x_o)]^\circ = \{d \in E \mid d_S^\bullet(x_o;d) = 0\} \quad . \tag{2.1}$$

We have given a *sequential characterization* of $\mathcal{T}(S;x_o)$ when $x_o \in c\ell S$ [35, Theorem 1]. We make here the case $x_o \in bdS$ a little more precise.

Theorem 1. *Let* $S \subset E$ *and* $x_o \in bdS$; $d \in \mathcal{T}(S;x_o)$ *iff for every sequence* $\{x_n\} \subset bdS$ *converging to* x_o *and for every sequence* $\{\lambda_n\} \subset R_+^*$ *converging to* 0, *there exists a sequence* $\{d_n\}$ *converging to* d *such that* $x_n + \lambda_n d_n \in S$ *for all* n.

Proof. The same characterization has been proved by replacing "$\{x_n\} \subset bdS$" by "$\{x_n\} \subset c\ell S$" [35, Theorem 1]; so the announced property holds if $d \in \mathcal{T}(S;x_o)$. Conversely, let us consider a sequence $\{x_n\} \subset c\ell S$ converging to x_o and a sequence $\{\lambda_n\} \subset R_+^*$ converging to 0. We examine the quantity $d_S(x_n + \lambda_n d)$.

If $d_S(x_n + \lambda_n d) > 0$, $x_n + \lambda_n d \in extS$; so, there exists $\bar{x}_n \in [x_n, x_n + \lambda_n d[$ which is on bdS. Let $\sigma_n \in]0,1]$ be such that $\bar{x}_n = x_n + (1-\sigma_n)(x_n + \lambda_n d)$ and we set $\mu_n = \sigma_n \lambda_n$. The sequence $\{\mu_n\} \subset R_+^*$ converges to 0 and $\bar{x}_n + \mu_n d = x_n + \lambda_n d$. $\{\bar{x}_n\} \subset bdS$ converges to x_o and according to the announced property, there exists a sequence $\{d_n\}$ converging to d such that $\bar{x}_n + \mu_n d_n \in S$ for all n. Therefore, $d_S(\bar{x}_n + \mu_n d) \leqslant \mu_n \|d_n - d\| \leqslant \lambda_n \|d_n - d\|$. Briefly, in any case $[d_S(x_n + \lambda_n d)]\lambda_n^{-1} \leqslant \|d_n - d\|$; consequently, $\lim_{n\to\infty} [d_S(x_n + \lambda_n d)]\lambda_n^{-1} = 0$ for the considered sequences $\{x_n\}$ and $\{\lambda_n\}$. So, $d \in \mathcal{T}(S;x_o)$ [35, characterization (1.2)].

Observations. 1. As for $T(S;x_o)$, $\mathcal{T}(S;x_o)$ depends only on adherent points of S ($\mathcal{T}(S;x_o)$ is $\mathcal{T}(c\ell S;x_o)$). Let us bear in mind that $\mathcal{T}(S;x_o)$ is a *convex* cone which *always* is included in the (non necessarily convex) cone $T(S;x_o)$ [35, Theorem 2]. A more precise comparison result between these two types of cones is given for E finite-dimensional in [32].

2. The elements $d \in \mathcal{T}(S;x_o)$ express a certain "tangency" property in a neighborhood of x_o but contrary to the other cones $I(S;x_o)$, $T(S;x_o)$, ..., one cannot say that $\mathcal{T}(S;x_o)$ is really a conical approximation of S at x_o. For example, if $S_1 = \{(x_1,x_2) \mid x_2 + x_1^2 + |x_1| = 0\}$ and $S_2 = \{(x_1,x_2) \mid x_2 - \mathrm{Log}(|x_1|+1) \leqslant 0\}$, at $x_o = (0,0)$ one has: $\mathcal{T}(S_1;x_o) = \{0\}$ and $\mathcal{T}(S_2;x_o)$ is $\{(x_1,x_2) \mid x_2 + |x_1| \leqslant 0\}$.

Concerning the cones of adherent displacements, we only have:

$$bd(T(S;x_o)) \subset T(bdS;x_o) \subset T(S;x_o) \cap T(S^c;x_o) \qquad (x_o \in bdS)$$

whereas for the tangent cones, we always have: $\mathcal{T}(bdS;x_o) = \mathcal{T}(S;x_o) \cap \mathcal{T}(S^c;x_o)$ [35].

3. It is worthwhile observing that if $S_1 \subset S_2$ and if $x_o \in (bdS_1) \cap (bdS_2)$ we generally cannot assert that $\mathcal{T}(S_1;x_o) \subset \mathcal{T}(S_1;x_o)$. Example:

$$S_1 = \{(x_1,x_2) \in R^2 \mid x_2 - x_1^2 \geqslant 0\},\ S_2 = \{(x_1,x_2) \in R^* X R \mid x_2 - x_1^2 \sin 1/x_1 \geqslant 0\} \cup \{(0,0)\}$$

$$S_1 \subset S_2 \text{ and at } x_o = (0,0),\ \mathcal{T}(S_1;x_o) = RXR_+,\ \mathcal{T}(S_2;x_o) = \{(x_1,x_2) \in R^2 \mid x_2 - |x_1| \geqslant 0\}.$$

The radial counterpart of the characterization given in Theorem 1 is the definition of the radial tangent directions.

Definition 5. $d \in E$ *is said to be a radial tangent direction to* S *at* x_o *if for every sequence* $\{x_n\} \subset S$ *converging to* x_o *and for every sequence* $\{\lambda_n\} \subset R_+^*$ *converging to* 0*, one has* $x_n + \lambda_n d \in S$ *for* n *sufficiently large.*

The set of radial tangent directions to S at x_o is called the *radial tangent cone to* S *at* x_o and denoted by $\mathcal{T}^r(S;x_o)$ (or $\mathcal{T}_S^r(x_o)$).
It clearly follows from the definitions that $\mathcal{T}^r(S;x_o) \subset \mathcal{T}(S;x_o)$ and that $\mathcal{T}^r(S;x_o) \subset T^r(S;x_o)$.
Example: $S = \{(x_1,x_2) \in R^2 \mid x_2 + |x_1|^{1/2} \geqslant 0\}$; at $x_o = (0,0)$, one has $\mathcal{T}^r(S;x_o) = \{0\} \times R_+$, $\mathcal{T}(S;x_o) = \{0\} \times R$ and $T^r(S;x_o) = \{(x_1,x_2) \mid x_1 \neq 0$ or $x_2 \geqslant 0\}$.

II.B. Let S be a subset of E; in [35], we have introduced the function μ_S defined as following:

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

When S is nonempty and different from the whole space, μ_S takes its values in $R^\cdot$ and is not identically equal to $+\infty$. From the definition (2.2), it is obvious that $\{x \in E \mid \mu_S(x) < 0\} = \text{int}S$ and that $S_1 \subset S_2 \Leftrightarrow \mu_{S_2} \leqslant \mu_{S_1}$. Moreover, μ_S is a l.s.c. function iff S is a closed subset. Let us also note that $\mu_{c\ell S} \leqslant c\ell \mu_S$.
First of all, we shall study the properties of μ_S in view of convexity.

Proposition 1. μ_S *is a convex function iff* S *is convex.*

Proof. If μ_S is convex, $S = \{x \in E \mid \mu_S(x) \leqslant 0\}$ is convex. Conversely, if S is a nonempty convex set, different from E, it is sufficient to show that for all $\lambda \in [0,1]$ and for all $x,y \in S$

$$d_{S^c}(\lambda x + (1-\lambda)y) \geqslant \lambda d_{S^c}(x) + (1-\lambda) d_{S^c}(y)$$

This is easily done using the convexity property of $c\ell S$.

The convexity of μ_S is suggested in an equivalent form as an exercise by N. Bourbaki [10, p. 150 Exercise 18]. At this stage, we assumed no closedness property of S. If S is a nonempty convex subset of E such that $\text{ext}S \neq \emptyset$, $\mu_{c\ell S}$ is a proper l.s.c. convex function and we easily deduce that $\mu_{c\ell S} = c\ell \mu_S = (\mu_S)^{**}$.
For the properties and characteristics related to μ_S in view of convex analysis, we may suppose without loss of generality that S is closed. $I(S;x_o)$ and $T(S;x_o)$ are referred to as convex approximations to the set S at x_o; some connections between these cones and the subdifferential of μ_S are expounded in the following proposition:

Proposition 2. *Let* S *be a nonempty closed convex subset of* E*, different from* E.

(a) $\partial\mu_S(x)$ *is nonempty for all* $x \in S$.

(b) *if* $x_o \in \text{bd}S$, *we have:*

(b_1) $\partial\mu_S(x_o) \subset N(S;x_o)$ *and* $\lambda\partial\mu_S(x_o) \subset \partial\mu_S(x_o)$ *for all* $\lambda \geqslant 1$.

(b_2) $[\partial\mu_S(x_o)]^\circ = T(S;x_o)$ *and* $\{d \in E \mid \mu_S'(x_o;d) < 0\} = I(S;x_o)$.

Proof. Under the assumptions made, μ_S is a proper l.s.c. convex function with $\mu_S^{-1}(R) = S$. If $\text{int}S = \emptyset$,

μ_S is the indicator function δ_S. So, for all $x \in S$, $\partial\mu_S(x) = N(S;x)$ and the announced results are verified.

(a) Let us consider now the case where *int*$S \neq \emptyset$. If $x_o \in \mathrm{int}S$, $\partial\mu_S(x_o)$ is obviously a nonempty compact subset of E^*_σ. Let $x_o \in \mathrm{bd}S$; we consider the directional derivative h of μ_S at x_o, that is

$$\forall\, d \in E,\ \ h(d) = \mu'_S(x_o; d) = \operatorname*{Inf}_{\lambda>0}\ [\mu_S(x_o+\lambda d)]\,\lambda^{-1} = \lim_{\lambda\to 0^+}\ [\mu_S(x_o+\lambda d)]\,\lambda^{-1}\ .$$

h is a positively homogeneous convex function and we first prove that

(2.3) $$c\ell(\mathrm{dom}\,h) = T(S;x_o)\ .$$

Let $d \in I(S^c;x_o)$; according to the definition of $I(S^c;x_o)$, for every sequence $\{\lambda_n\} \subset R^*_+$ converging to 0, one has $x_o + \lambda_n d \in S^c$ for n sufficiently large. Consequently, $h(d) = +\infty$ and by the relation $I(S^c;x_o) = [T(S;x_o)]^c$, we have: $\mathrm{dom}\,h \subset T(S;x_o)$.

Conversely, let $d \in I(S;x_o)$. For a sequence $\{\lambda_n\} \subset R^*_+$ converging to 0, one has $x_o + \lambda_n d \in S$ if n is sufficiently large. Thus, $\mu_S(x_o+\lambda_n d) = -d_{S^c}(x_o+\lambda_n d)$ and since $-\|d\| \leqslant \overline{\lim_{n\to\infty}}\ [d_{S^c}(x_o+\lambda_n d)]\,\lambda_n^{-1} \leqslant 0$, we have $h(d) \in R$. So, $I(S;x_o) \subset \mathrm{dom}\,h$. But S being a convex set with nonempty interior, $c\ell I(S;x_o)$ is $T(S;x_o)$ [40, Ch. I]. Hence the result (2.3).

Let $d_o \in T(S;x_o)$ and we consider a sequence $\{d_n\}$ converging to d_o. We have

$$\forall\,\lambda>0,\ \ [\mu_S(x_o+\lambda d_n)]\,\lambda^{-1} \geqslant [-d_{S^c}(x_o+\lambda d_n)]\,\lambda^{-1} \geqslant -\|d_n\|\ .$$

So, $h(d_n) \geqslant -\|d_n\|$ and consequently $c\ell h(d_o) \geqslant -\|d_o\|$. $c\ell h$ takes its values in $R^\bullet$; then $(c\ell h)^{**} = c\ell h$ is the support function of a nonempty set, namely of $\partial\mu_S(x_o)$.

(b) $\mu_S \leqslant \delta_S$ and if $x_o \in \mathrm{bd}S$, $\mu_S(x_o) = \delta_S(x_o) = 0$. This implies that $\partial\mu_S(x_o) \subset \partial\delta_S(x_o) = N(S;x_o)$. By definition of $\partial\mu_S(x_o)$,

$$x^* \in \partial\mu_S(x_o) \Leftrightarrow \forall\, x \in S,\ \langle x^*, x-x_o\rangle \leqslant -d_{S^c}(x)\ .$$

Obviously, $\lambda\partial\mu_S(x_o) \subset \partial\mu_S(x_o)$ for all $\lambda \geqslant 1$.

The support function of $\partial\mu_S(x_o)$ is $c\ell h$; so, $[\partial\mu_S(x_o)]^\circ = \{d \in E \mid c\ell h(d) \leqslant 0\}$. Following the construction of $c\ell h$, $c\ell h(d) \leqslant 0$ iff $c\ell h(d)$ is finite. We also have: $\mathrm{dom}\,h = \{d \in E \mid h(d) \leqslant 0\}$. Consequently, $\mathrm{dom}(c\ell h)$ is $c\ell(\mathrm{dom}\,h)$, that is to say $T(S;x_o)$ (by the equality (2.3)).

Since we supposed $\mathrm{int}S \neq \emptyset$ and $x_o \in \mathrm{bd}S$, $0 \notin \partial\mu_S(x_o)$. So, there exists d such that $\mu'_S(x_o;d) < 0$; let us take such a d.

For every sequence $\{d_n\}$ converging to d and every sequence $\{\lambda_n\} \subset R^*_+$ converging to 0, we have for n sufficiently large

$$[-d_{S^c}(x_o+\lambda_n d_n)]\,\lambda_n^{-1} \leqslant [\mu_S(x_o+\lambda_n d)]\,\lambda_n^{-1} + \|d_n - d\|$$

Since $\mu'_S(x_o;d) < 0$, for $n \geqslant n_o$, $-d_{S^c}(x_o+\lambda_n d) < 0$, that is: $x_o + \lambda_n d_n \in S$. Therefore,

$$\{d \in E \mid \mu'_S(x_o;d) < 0\} \subset I(S;x_o)\ .$$

Conversely, let $d \in \mathrm{int}S - x_o$. We consider $u \in \mathrm{int}S$ such that $d = u - x_o$; there exists $\epsilon > 0$ such that

$d_{S^c}(u) > \epsilon$. By the concavity property of the function d_{S^c} on S (Proposition 1), we have:

$$d_{S^c}(x_0 + \lambda_n d) \geqslant \lambda_n \epsilon \quad .$$

Consequently $\mu_S'(x_0; d) < 0$ and $\operatorname{int}S - x_0 \subset \{d \in E \mid \mu_S'(x_0; d) < 0\}$. The result then follows from the relation $I(S; x_0) = R_+^*(\operatorname{int}S - x_0)$ and from the homogeneity property of $\mu_S'(x_0; .)$.

As we shall see now, there are analogies between the conjugate function μ_S^* and the support function δ_S^*.

Proposition 3. S *being a nonempty closed convex subset of* E, *we have*

(a) $\forall\, x^* \in \bigcup_{x \in \operatorname{bd}S} \partial\mu_S(x)$, $\mu_S^*(x^*) = \delta_S^*(x^*)$.

(b) *if* S *is compact*, μ_S^* *is a finite function such that* $\delta_S^* \leqslant \mu_S^* \leqslant \delta_S^* + \alpha$ *where* $\alpha = \mu_S^*(0)$.

Proof. (a) Since $\mu_S \leqslant \delta_S$, $\delta_S^* \leqslant \mu_S^*$. Let $x^* \in \bigcup_{x \in \operatorname{bd}S} \partial\mu_S(x)$; let us consider $x_0 \in \operatorname{bd}S$ such that $x^* \in \partial\mu_S(x_0)$. According to the characterization property of a subgradient [40, Theorem 6.4.2], we have

$$x^* \in \partial\mu_S(x_0) \Leftrightarrow \mu_S^*(x^*) + \mu_S(x_0) = \langle x_0, x^* \rangle \quad .$$

But $\partial\mu_S(x_0) \subset N(S; x_0) = \partial\delta_S(x_0)$; consequently

$$\delta_S^*(x^*) + \delta_S(x_0) = \langle x_0, x^* \rangle \quad .$$

For $x_0 \in \operatorname{bd}S$, $\mu_S(x_0) = \delta_S(x_0) = 0$; hence the equality $\mu_S(x^*) = \delta_S^*(x^*)$.

(b) $\alpha = \mu_S^*(0)$ is by definition $\operatorname{Sup}_{x \in S} d_{S^c}(x)$. If S is compact, μ_S^* is a finite function and

$$\forall\, x^*, \qquad \delta_S^*(x^*) \leqslant \mu_S^*(x^*) \leqslant \delta_S^*(x^*) + \alpha \quad .$$

Example: Let us take the unit ball B in E; then $\mu_B^* = \operatorname{Max}(1, \|\cdot\|_*)$.

Remarks. 1. Under the assumptions of Proposition 2, let $x_0 \in \operatorname{bd}S$. From the definition of μ_S, we have the following equivalences:

$$(0 \notin \partial\mu_S(x_0)) \Leftrightarrow (\forall\, x \in \operatorname{bd}S, 0 \notin \partial\mu_S(x)) \Leftrightarrow (\operatorname{int}S \neq \emptyset) \Leftrightarrow (I(S; x_0) \neq \emptyset) \quad .$$

Otherwise, we recall that the *center of S* [11] is defined as following:

$$C(S) = \{\bar{x} \in S \mid d_{S^c}(\bar{x}) \geqslant d_{S^c}(x) \quad \text{for every} \quad x \in S\} \quad .$$

In this definition, $d_{S^c}(\bar{x})$ is supposed to be finite. If S is compact, the center C(S) is precisely $\partial\mu_S^*(0)$.

The function Δ_S defined by: $\Delta_S(x) = d_S(x) - d_{S^c}(x)$ is obtained by infimal convolution of μ_S and of the norm function $\|\cdot\|$ [35, Proposition 1]. If S is nonempty and different from E, Δ_S is a Lipschitz function with constant 1.

It $S_1 \subset S_2$, we know that $\mu_{S_2} \leqslant \mu_{S_1}$ and consequently $\mu_{S_2} \nabla \|\cdot\| \leqslant \mu_{S_1} \nabla \|\cdot\|$. So, if S_1 and S_2 are closed subsets of E, then

$$S_1 \subset S_2 \Leftrightarrow \Delta_{S_2} \leqslant \Delta_{S_1} \quad .$$

Proposition 4. *If* S *is convex*, Δ_S *is convex. If* Δ_S *is convex, then* $c\ell S$ *is convex.*

Proof. If S is convex, μ_S is convex (Proposition 1) and Δ_S which is $\mu_S \nabla \|\cdot\|$ is also convex [40, Corollary 6.5.3]. Conversely, if Δ_S is convex, $c\ell S = \{x \in E \mid \Delta_S(x) \leqslant 0\}$ is convex.

Proposition 5. *Let* S *be a nonempty closed convex subset of* E. *If* $x_o \in \mathrm{bd}S$,

$$\partial\Delta_S(x_o) = \partial\mu_S(x_o) \cap B^* \quad .$$

Proof. When $x_o \in \mathrm{bd}S$, $\mu_S(x_o) = \Delta_S(x_o) = 0$; so

$$x^* \in \partial\Delta_S(x_o) \Leftrightarrow \Delta_S^*(x^*) = \langle x_o, x^* \rangle \quad .$$

But $\Delta_S^* = (\mu_S \nabla \|\cdot\|)^* = \mu_S^* + \delta_{B^*}$ [40, Corollary 5.5.4]. Consequently,

$$x^* \in \partial\Delta_S^*(x_o) \Leftrightarrow (\mu_S^*(x^*) = \langle x_o, x^* \rangle \text{ and } \delta_{B^*}(x^*) = 0) \quad .$$ Hence the result.

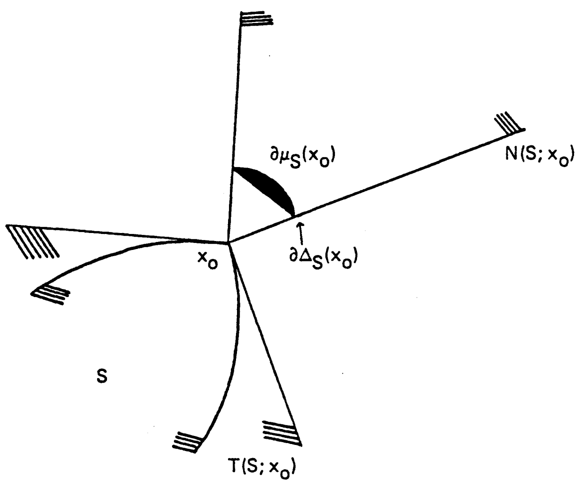
Remarks. 1. Let us remark that for $x_o \in \mathrm{bd}S$, the infimal convolution of the functions μ_S and $\|\cdot\|$ is *exact* at $(x_o, 0)$. The expression of $\partial\Delta_S(x_o)$ at $x_o \in \mathrm{bd}S$ is similar to that of $\partial d_S(x_o)$ which is $N(S; x_o) \cap B^*$. This yields

$$\partial d_S(x_o) \cap \partial\mu_S(x_o) = \partial\Delta_S(x_o) \quad .$$

The result of the preceding proposition completes the properties of Proposition 2 on the structure of $\partial\mu_S(x_o)$ since

(2.4) $$\partial\mu_S(x_o) = \bigcup_{\lambda \geqslant 1} \{\lambda \partial\Delta_S(x_o)\} \quad .$$

The following figure illustrates this special structure.



2. If S is convex, $d_{S^c} = d_S - \Delta_S$ (or $\delta_S - d_{S^c}$) appears as a difference of two convex functions. The conjugate of the difference of two convex functions may be calculated with the $*$-difference introduced by B. N. Pshenichnyi [56]; this gives additional relations between $\delta^*_{S^c}$, μ^*_S and Δ^*_S.

3. If $S \subset E$ and $x_0 \in \text{bd}S$, the *index of angularity* of bdS at x_0, denoted by $\alpha_{\text{bd}S}(x_0)$, is defined by: $\alpha_{\text{bd}S}(x_0) = d(0, \partial\Delta_S(x_0))$ [35, Definition 5]. Moreover, x_0 is said to be a *regular* point of bdS if $\alpha_{\text{bd}S}(x_0) > 0$. If S is a nonempty closed convex set of E and if $x_0 \in \text{bd}S$, we remark that $\alpha_{\text{bd}S}(x_0) = d(0, \partial\mu_S(x_0))$; therefore:

x_0 *is a regular point of* bdS *iff every* x *on the boundary of* S *is a regular point.*

Starting from the structure of $\partial\Delta_S(x_0)$ for an arbitrary subset $S \subset E$, we defined in [35, Definition 4] the concept of *symmetric tangent cone* to S at $x_0 \in \text{bd}S$ by taking the polar cone of $\partial\Delta_S(x_0)$; let us recall this definition:

Definition 6. *The symmetric tangent cone to* S *at* $x_0 \in \text{bd}S$ *is the closed convex cone of* E^*_σ denoted by $\mathcal{U}(S; x_0)$ *and defined by:*

$$\mathcal{U}(S; x_0) = [\partial\Delta_S(x_0)]^\circ = \mathcal{T}(S; x_0) \cap -\mathcal{T}(S^c; x_0) \quad .$$

For comparison results between $\mathcal{U}(S; x_0)$ and $\mathcal{T}(S; x_0)$, between $\text{int}\,\mathcal{U}(S; x_0)$ and $I(S; x_0)$, see [35, §I].

Part III: GENERALIZED SUBDIFFERENTIALS AND GRADIENTS OF A FUNCTION.

Let us denote by $F(E,F)$ the set of functions from E to $F (F = R^\cdot, \bar{R}, \ldots)$; $f \in F(E,\bar{R})$ is said to be Lipschitz in a neighborhood of x_0 if f is finite in a neighborhood V_0 of x_0 and if there exists k such that $|f(x) - f(y)| \leqslant k\|x - y\|$ for all $x,y \in V_0$.

The different generalized subdifferentials described in Part I (A.4, A.6) may be connected with the different kinds of feasible displacements for epif from $(x_0, f(x_0))$ in the following manner:

$$(3.1)\quad \begin{aligned} \underline{f}(x_0; d) &= \text{Inf}\,\{\mu \in R \mid (d, \mu) \in T_{\text{epi}f}(x_0, f(x_0))\} \quad (\inf \emptyset = +\infty) \\ \bar{f}(x_0; d) &= \text{Inf}\,\{\mu \in R \mid (d, \mu) \in I_{\text{epi}f}(x_0, f(x_0))\} \\ \underline{f}_r(x_0; d) &= \text{Inf}\,\{\mu \in R \mid (d, \mu) \in T^r_{\text{epi}f}(x_0, f(x_0))\} \\ \bar{f}^r(x_0; d) &= \text{Inf}\,\{\mu \in R \mid (d, \mu) \in D_{\text{epi}f}(x_0, f(x_0))\} \end{aligned}$$

and

$$(3.2)\quad \begin{aligned} \bar{\partial}f(x_0) &= \{x^* \in E^* \mid (x^*, -1) \in [I_{\text{epi}f}(x_0, f(x_0))]^\circ\} \\ \underline{\partial}f(x_0) &= \{x^* \in E^* \mid (x^*, -1) \in [T_{\text{epi}f}(x_0, f(x_0))]^\circ\} \\ \underline{\partial}_r f(x_0) &= \ldots, \qquad \bar{\partial}^r f(x_0) = \ldots \end{aligned}$$

The radial directional derivatives $\underline{f}_r(x_0; .)$ and $\bar{f}^r(x_0; .)$ have also been used by M. S. Bazaraa and

J. J. Goode [3] ; in their work, they noticed the connection with the radial cones of feasible displacements as in (3.1).

In a similar way, the generalized gradient of f at x_o (Definition (1.25)) may be described as following:

$$\partial f(x_o) = \{x^* \in E^* \mid \langle x^*, d\rangle \leq f^{\square}(x_o; d) \text{ for all } d\}$$

where

$$f^{\square}(x_o; d) = \text{Inf } \{\mu \in R \mid (d, \mu) \in \mathscr{T}_{\text{epi}f}(x_o, f(x_o))\} \quad [35, \text{Proposition } 2].$$

The support function of $\partial f(x_o)$ is the biconjugate function $[f^{\square}(x_o; .)]^{**}$. By the same construction, the concept of radial tangent cone induces the concept of *radial generalized gradient* of f at x_o:

$$\partial^r f(x_o) = \{x^* \in E^* \mid \langle x^*, d\rangle \leq f^{\square r}(x_o; d) \text{ for all } d\}$$

where

$$f^{\square r}(x_o; d) = \text{Inf } \{\mu \in R \mid (d, \mu) \in \mathscr{T}^r_{\text{epi}f}(x_o, f(x_o))\} .$$

According to the inclusions between the cones used in these definitions, we have

$$\underline{\partial} f(x_o) \subset \partial f(x_o) \subset \partial^r f(x_o)$$
$$\underline{\partial}_r f(x_o) \subset \partial^r f(x_o)$$

Examples. 1. $f : x \mapsto -|x|$; $\underline{\partial} f(0) = \underline{\partial}_r f(0) = \emptyset$ and $\partial f(x_o) = \partial^r f(x_o) = [-1, +1]$

2. $f : x \to -|x|^{1/2}$; $\partial f(0) = \emptyset$ and $\partial^r f(x_o) = R$.

As application, we shall examin properties of the generalized gradient of the function μ_S. Some of them are slight generalizations of properties established in the convex case (Proposition 2).

Proposition 6. *Let* S *be a nonempty subset of* E, *different from* E. *Then*

(a) *if* $x_o \in \text{int}S$, $\partial\mu_S(x_o) = -\partial d_{S^c}(x_o)$; *if* $x_o \in S \cap \text{bd}S$, $\partial\mu_S(x_o)$ *is nonempty.*

(b) *for every* $x_o \in S$, $\Pi_E(\mathscr{T}_{\text{epi}\,\mu_S}(x_o, \mu_S(x_o))) = \mathscr{T}(S; x_o)$ *where* Π_E *designates the projection on* E *in a parallel direction to* R.

(c) *if* $x_o \in S \cap \text{bd}S$, $\partial\Delta_S(x_o) \subset \partial\mu_S(x_o) \cap B^*$

(d) *for every* $x_o \in S$, $[\partial\mu_S(x_o)]^\circ = \{d \mid (d,0) \in \mathscr{T}_{\text{epi}\,\mu_S}(x_o, \mu_S(x_o))\}$.

Proof. (a) If $x_o \in \text{int}S$, $\mu_S = -d_{S^c}$ is Lipschitz in a neighborhood of x_o and $\partial\mu_S(x_o)$ is a nonempty convex compact subset of E^*_σ. If $x_o \in S \cap \text{bd}S$, $\mu_S(x_o) = 0$ and $\partial\mu_S(x_o) \neq \emptyset$ iff $(0,-1) \in \mathscr{T}_{\text{epi}\,\mu_S}(x_o,0)$ [35, Theorem 5]. Let us suppose that $(0,-1) \in \mathscr{T}_{\text{epi}\,\mu_S}(x_o,0)$. We consider a sequence $\{x_n\} \subset S$ converging to x_o; the sequence $(x_n, -d_{S^c}(x_n))$ is a sequence of epi μ_S converging to $(x_o,0)$. According to the sequential characterization of the tangent cone, for every sequence $\{\lambda_n\} \subset R^*_+$ converging to 0, there exists a sequence (d_n, σ_n) converging to $(0,-1)$ such that $(x_n + \lambda_n d_n, -d_{S^c}(x_n) + \lambda_n\sigma_n)$ is in epi μ_S for all n. So,

$$\mu_S(x_n + \lambda_n d_n) = -d_{S^c}(x_n + \lambda_n d_n) \leq -d_{S^c}(x_n) + \lambda_n\sigma_n .$$

Consequently

$$-\|d_n\| \leq [d_{S^c}(x_n) - d_{S^c}(x_n + \lambda_n d_n)]\lambda_n^{-1} \leq \sigma_n \ .$$

Now, (d_n, σ_n) converges to $(0,-1)$. This yields the contradiction.

(b) We have $\Pi_E(\text{epi}\,\mu_S) = \text{dom}\,\mu_S = S$. Let $(d,\sigma) \in \mathcal{T}_{\text{epi}\,\mu_S}(x_o, 0)$ and let us consider a sequence $\{x_n\} \subset S$ converging to x_o and a sequence $\{\lambda_n\} \subset R_+^*$ converging to 0. There exists a sequence $\{(d_n, \sigma_n)\}$ converging to (d,σ) such that $(x_n + \lambda_n d_n, -d_{S^c}(x_n) + \lambda_n \sigma_n) \in \text{epi}\,\mu_S$ for all n, that is $x_n + \lambda_n d_n \in S$ for all n. So,

$$\Pi_E(\mathcal{T}_{\text{epi}\,\mu_S}(x_o, 0)) \subset \mathcal{T}(S; x_o) \ .$$

Conversely, let $d \in \mathcal{T}(S; x_o)$. We have to show that there exists $\sigma \in R$ such that $(d,\sigma) \in \mathcal{T}_{\text{epi}\,\mu_S}(x_o, 0)$. Let us take $\sigma > \|d\|$. We consider a sequence $\{(x_n, y_n)\} \subset \text{epi}\,\mu_S$ converging to $(x_o, 0)$ and a sequence $\{\lambda_n\} \subset R_+^*$ converging to 0. Since $d \in \mathcal{T}(S; x_o)$, there exists a sequence $\{d_n\}$ converging to d such that $x_n + \lambda_n d_n \in S$ for all n. We set $\sigma_n = \sigma$;

$$\mu_S(x_n + \lambda_n d_n) = -d_{S^c}(x_n + \lambda_n d_n) \leq -d_{S^c}(x_n) + \lambda_n \|d_n\|$$
$$\leq y_n + \lambda_n \sigma \quad \text{for } n \geq n_o \ .$$

Thus $(d,\sigma) \in \mathcal{T}_{\text{epi}\,\mu_S}(x_o, 0)$.

(c) We know that: $x^* \in \partial\Delta_S(x_o) \Leftrightarrow \forall\, d \in E, \langle x^*, d\rangle \leq \Delta_S^{\cdot}(x_o; d)$.

Δ_S is a Lipschitz function with constant 1; therefore $\Delta_S^{\cdot}(x_o; d) \leq \|d\|$ and $\partial\Delta_S(x_o) \subset B^*$. The inclusion $\partial\Delta_S(x_o) \subset \partial\mu_S(x_o)$ is equivalent to this one:

(3.3)
$$\mathcal{T}_{\text{epi}\,\mu_S}(x_o, 0) \subset \mathcal{T}_{\text{epi}\,\Delta_S}(x_o, 0) \ .$$

Let $(d,\sigma) \in \mathcal{T}_{\text{epi}\,\mu_S}(x_o, 0)$. We consider a sequence $\{(x_n, y_n)\} \subset \text{epi}\,\Delta_S$ converging to $(x_o, 0)$ and a sequence $\{\lambda_n\} \subset R_+^*$ converging to 0. We construct the sequence $\{\bar{x}_n\}$ as following:

$$\begin{cases} \bar{x}_n = x_n \ \text{ if } x_n \in S \\ \bar{x}_n \in S \ \text{ such that } \|x_n - \bar{x}_n\| \leq d_S(x_n) + \lambda_n^2 \ \text{ if } x_n \in S^c \ . \end{cases}$$

Similarly, we set $\bar{y}_n = y_n$ if $x_n \in S$; $\bar{y}_n = 0$ if $x_n \notin S$.

The sequence $\{\bar{x}_n, \bar{y}_n\}$ converges to $(x_o, 0)$ and by the construction itself $(\bar{x}_n, \bar{y}_n) \in \text{epi}\,\mu_S$. Since $(d,\sigma) \in \mathcal{T}_{\text{epi}\,\mu_S}(x_o, 0)$, there exists a sequence $\{d_n, \sigma_n\}$ converging to (d,σ) such that

$$\Delta_S(\bar{x}_n + \lambda_n d_n) = -d_{S^c}(\bar{x}_n + \lambda_n d_n) \leq \bar{y}_n + \lambda_n \sigma_n$$

for all n. If $x_n \in S$, this reduces to: $\Delta_S(x_n + \lambda_n d_n) \leq y_n + \lambda_n \sigma_n$.

If $x_n \in S^c$, by the Lipschitz property (with coefficient 1) of Δ_S, we have

$$\Delta_S(x_n + \lambda_n d_n) \leq \|x_n - \bar{x}_n\| + \lambda_n \sigma_n$$
$$\leq d_S(x_n) + \lambda_n(\sigma_n + \lambda_n) \leq y_n + \lambda_n(\sigma_n + \lambda_n) \ .$$

Finally, $(x_n + \lambda_n d_n, y_n + \lambda_n(\sigma_n + \lambda_n)) \in \text{epi}\,\Delta_S$. Thereby, (d,σ) belongs to $\mathcal{T}_{\text{epi}\,\Delta_S}(x_o, 0)$

and the inclusion (3.3) is verified.

(d) Let $x_o \in S$ and let us consider u such that $(u,0) \in \mathcal{T}_{\text{epi}\,\mu_S}(x_o, \mu_S(x_o))$. This means that $\langle u, n\rangle \leqslant 0$ for all $(n, \theta) \in N_{\text{epi}\,\mu_S}(x_o, \mu_S(x_o))$; consequently $u \in [\partial\mu_S(x_o)]^\circ$.

Conversely, $N_{\text{epi}\,\mu_S}(x_o, \mu_S(x_o))$ being the normal cone to an epigraph, $\theta \leqslant 0$ for all (n, θ) in this cone. We distinguish two cases: $\theta < 0$ and $\theta = 0$. If $\theta < 0$, $-n\theta^{-1} \in \partial\mu_S(x_o)$; for all $u \in [\partial\mu_S(x_o)]^\circ$, $\langle u, -n\theta^{-1}\rangle \leqslant 0$ and $\langle u, n\rangle \leqslant 0$. Let $\theta = 0$. $\partial\mu_S(x_o)$ is nonempty; so, there exists a sequence $\{n_k, \sigma_k\}$ of $N_{\text{epi}\,\mu_S}(x_o, \mu_S(x_o))$ converging to $(n, 0)$ and such that $\sigma_k < 0$ for all k. Consequently, $\langle u, n\rangle = \lim_{k\to\infty} \langle u, n_k\rangle \leqslant 0$, that is to say $(u, 0) \in \mathcal{T}_{\text{epi}\,\mu_S}(x_o, \mu_S(x_o))$.

Remarks. 1. If $x_o \in S \cap \text{bd}S$, the results (c) and (d) of the proposition above bring the discrepancy between the tangent cone and the symmetric tangent cone. We have

$$\mathcal{T}(S; x_o) = \Pi_E(\mathcal{T}_{\text{epi}\,\mu_S}(x_o, 0))$$

$$\mathcal{U}(S; x_o) \supset \{d \in E \mid (d,0) \in \mathcal{T}_{\text{epi}\,\mu_S}(x_o, 0)\} \quad .$$

So, if the domain of the support function $d \mapsto \dot{\mu}_S(x_o; d)$ is equal to the level set $\{d \in E \mid \dot{\mu}_S(x_o; d) \leqslant 0\}$, then $\mathcal{U}(S; x_o) = \mathcal{T}(S; x_o)$.

2. The tangent cone to the epigraph of the indicator function δ_S is $\mathcal{T}(S; x_o) \times R_+$. If $x_o \in S \cap \text{bd}S$, in general we have not the inclusion: $\mathcal{T}(S; x_o) \times R_+ \subset \mathcal{T}_{\text{epi}\,\mu_S}(x_o, 0)$; consequently, the inclusion $\partial\mu_S(x_o) \subset N(S; x_o)$ is not generally true. Likewise, other properties of $\partial\mu_S(x_o)$ established in the convex case (Proposition 2) cannot be extended to the nonconvex case.

Example. Let $S = \{(x_1, x_2) \in R^2 \mid x_2 - |x_1|^{1/2} \leqslant 0\}$; then at $x_o = 0$, $\mathcal{T}(S; x_o)$ is the set $\{(0,x_2) \mid x_2 \in R\}$ whereas $\mathcal{T}_{\text{epi}\,\mu_S}(x_o, 0)$ is equal to $\{(x_1, x_2, x_3) \mid x_1 = 0, \ \text{Max}\,(0, x_2) \leqslant x_3\}$.

In [35] we introduced the notion of *symmetric generalized* gradient of a function as following: let $x_o \in E$ where $f \in F(E, \overline{R})$ is finite; the symmetric generalized gradient of f at x_o, denote by $Df(x_o)$ is defined by

$$Df(x_o) = \{x^* \in E^* \mid \forall\, d \in E, \langle x^*, d\rangle \leqslant f^\vee(x_o; d)\}$$

where

$$f^\vee(x_o; d) = \text{Inf}\,\{\mu \in R \mid (d, \mu) \in \mathcal{U}_{\text{epi}\,f}(x_o, f(x_o))\} \quad .$$

Some elementary properties of $f^\vee(x_o; .)$ are collected in the following Proposition.

Proposition 7. *Let* f *be finite at* x_o. *Then,*

(a) $f^\vee(x_o; .)$ *is a* l.s.c., *convex, positively homogeneous function from* E *to* $\overline{R}$ *and* $f^\square(x_o; .) \leqslant f^\vee(x_o; .)$.

(b) *if there exists* d *such that* $f^\vee(x_o; d) = -\infty$, $\text{dom} f^\vee(x_o; .)$ *which is equal to*

$$\{d \in E \mid f^\vee(x_o; d) = -\infty\}$$

is a nonempty closed convex cone with apex 0.

(c) *the support function of* $Df(x_o)$ *is* $[f^\vee(x_o; .)]^{**}$ *and if* $Df(x_o) \neq \emptyset$, $\delta^*_{Df(x_o)} = f^\vee(x_o; .)$.

(d) $Df(x_0) \neq \emptyset \Leftrightarrow f^{\vee}(x_0;.)$ *takes its values in* $R^{\cdot}$

$\Leftrightarrow (0,-1) \notin \mathcal{U}_{\text{epi} f}(x_0, f(x_0))$

(e) *if the function* $d \mapsto f^{\vee}(x_0; d)$ *is bounded in a neighborhood of* 0, $Df(x_0)$ *is a compact subset of* E^*_{σ}.

Proof. The inclusion $\mathcal{U}_{\text{epi} f}(x_0, f(x_0)) \subset \mathcal{T}_{\text{epi} f}(x_0, f(x_0))$ implies that $f^{\square}(x_0;.) \leqslant f^{\vee}(x_0;.)$. The proofs of different properties are similar to those of [35, Theorem 5].

Remarks. For a nonempty subset S of $E \times R$, let $\check{S}$ be defined by: $(x,\theta) \in S \Leftrightarrow (x,-\theta) \in \check{S}$; we remark that if $(x_0, \theta_0) \in c\ell S$, $\mathcal{T}_{\check{S}}(x_0, -\theta_0) = \check{\mathcal{T}}_S(x_0, \theta_0)$. This remark combined with the general result

$$\mathcal{U}(S; u_0) = -\mathcal{U}(S^c; u_0)$$

yields that $D(-f)(x_0) = -Df(x_0)$. [34, Chapter VII]. This equality (or the equivalent relation $(-f)^{\vee}(x_0; d) = f^{\vee}(x_0; -d)$) allow us to derive optimality conditions for maximization problems from corresponding conditions for minimization problems.

Comparison results. 1. Generally, we have: $\partial f(x_0) \subset Df(x_0)$. From the theorems comparing the tangent cone and the symmetric tangent cone (see [35, §I]), it follows that $\partial f(x_0) = Df(x_0)$ at least in two cases: first, when f is convex (in a neighborhood of x_0) and secondly when f is Lipschitz in a neighborhood of x_0. In fact, when f is a convex function, finite and continuous in a neighborhood of x_0, *all* the notions we have spoken about coincide.

2. If $(x_0, f(x_0))$ is a regular point of the graph of f, $\text{int}\,\mathcal{U}_{\text{epi} f}(x_0, f(x_0))$ is included in $I_{\text{epi} f}(x_0, f(x_0))$ [35, Theorem 4]. So, $c\ell(\bar{f}(x_0;.)) \leqslant f^{\vee}(x_0;.)$ and $\bar{\partial} f(x_0) \subset Df(x_0)$. More particularly, this last inclusion may be directly proved if f is Lipschitz in a neighborhood of x_0.

3. If f is Lipschitz in a neighborhood of x_0, we also have: $\partial^r f(x_0) = \partial f(x_0)$. The question whether the different subdifferentials and the generalized gradient are generically equal has been recently investigated [41]*, [54].

Part IV: NECESSARY OPTIMALITY CONDITIONS, SUFFICIENT OPTIMALITY CONDITIONS.

We consider the optimization problem of mathematical programming in general form. For S a nonempty subset of E and $f \in F(E, \bar{R})$, we consider

(P) *Minimize* f *(locally) on the subset* S.

IV.A. $x_0 \in S$ is said to be a *local minimum* of f on S if f is finite at x_0 and if there exists a neighborhood V_0 of x_0 such that $f(x) \geqslant f(x_0)$ for all $x \in S \cap V_0$. For sake of simplicity, we shall give examples of necessary optimality conditions in particular cases.

Theorem 2. *If* x_0 *is a local minimum of* f *on* S *and if* f *is Lipschitz in a neighborhood of* x_0, *then*

$$f^{\cdot}(x_0; d) \geqslant 0,\ \forall\, d \in T(S; x_0) \quad . \tag{4.1}$$

Theorem 3. *Let* $x_0 \in S$ *and let us consider a convex cone* M *with apex* 0 *included in* $T(S; x_0)$. *If* f *is a function Lipschitz in a neighborhood of* x_0 *and if* x_0 *is a local minimum of* f *on* S, *then*

$$0 \in \partial f(x_0) + M^\circ \tag{4.2}$$

Proofs. See [32] or [35, §III.1].

For developments on necessary optimality conditions in a general context (non necessarily locally Lipschitz functions), we refer the reader to the papers [53, 35]; for the locally Lipschitz case, see more particularly [14, 32].

IV.B. In this section, we give some indications about sufficient optimality conditions. In [31], we introduced the following definition of pseudo-convexity in the nondifferentiable case:

Definition 7. *A locally Lipschitz function* $f : E \to R$ *is said to be locally pseudo-convex at* x_0 *if there exists a neighborhood* V *of* x_0 *such that*

$$\forall x \in V, \quad f^\circ(x_0; x - x_0) \geqslant 0 \Rightarrow f(x) \geqslant f(x_0) \tag{4.3}$$

If the property holds for all $x_0 \in E$, we shall simply say that f is locally pseudo-convex. If the relation (4.3) holds globally (i.e. V = E), f will be simply called pseudo-convex. Let us remark that, as in the differentiable case [44], we have:

$$f \text{ pseudo-convex} \Rightarrow f \text{ "strictly" quasi-convex.}$$

A similar kind of concept was introduced by H. Tuy [37] (*semiconvex* functions) but Tuy's definition requires the existence of the directional derivative $d \mapsto f'(x_0; d)$. We give now some examples of sufficient optimality conditions extracted from [34, Chapter V].

Theorem 4. *Under the following hypotheses,*

(a) f *is locally pseudo-convex at* x_0

(b) *the constraint set* S *verifies the condition* (L) *below*

(L) $\exists V$, *neighborhood of* x_0, *such that for all* $x \in V \cap S$, $x - x_0 \in T(S; x_0)$. *If the necessary optimality condition* (4.1) *is verified at* x_0, *then* x_0 *is a local minimum of* f *on* S.

Theorem 5. *Let* $x_0 \in S$ *be such that* $T(S; x_0)$ *is convex; a necessary condition for* x_0 *to be a local minimum of* f *on* S *is that*

$$\partial f(x_0) \cap -[T(S; x_0)]^\circ \neq \emptyset \tag{4.4}$$

Moreover, if f *is locally pseudo-convex at* x_0 *and if* S *verifies the assumption* (L) *at* x_0, *then* (4.4) *is a sufficient condition for* x_0 *being a local minimum of* f *on* S.

Part V: GENERALIZED GRADIENTS OF COMPOSITE FUNCTIONS. APPLICATIONS.

The principal results of this section have been announced as a Compte Rendu Acad. Sciences de Paris [36].

V.A. In numerous problems (optimal control, fractional programming, best approximation, estimation . . .), the objective function and the functions defining the constraint set occur to be composite functions.

An example: The response of a physical system is a quantity $y(t)$ satisfying a relation of the type $y(t) = \theta(t; a)$ where θ is a known function of t and a, t an auxiliary parameter (time for example) and a an unknown parameter of R^n. For several values of t, $t_1, \ldots, t_m$, one only has access to noisy observations of $y(t_i)$, namely $y_i = y(t_i) + \xi_i$. By making m observations ($m >> n$), one chooses $a^* \in A$ (constraint set) minimizing $f(a) = \Phi(y_1 - \theta(t_1; a), \ldots, y_m - \theta(t_m; a))$ on A. In particular, the choice of $\Phi(u_1, \ldots, u_m) = \sum_{i=1}^{m} u_i^2$ corresponds to the least squares estimation.

A second example: Generalization of the problem of Fermat-Weber.
Let $\{K_i\}_{i=0}^{p}$ be a family of nonempty subsets of E, let $\{\sigma_i\}_{i=1}^{p}$ be a family of functions from R_+ to R. The optimization problem consists on minimizing $\sum_{i=1}^{p} \sigma_i[d_{K_i}(x)]$ on K_0. In the problem of Fermat-Weber, the subsets K_i are reduced to points and $K_0 = E$. The consideration of distances to subsets and the introduction of a constraint set come naturally, specially for localization problems. In particular, the constraint K_0 may have the following structure:

$$d_{F_j}(x) \leqslant d_j \quad \forall j = 1, \ldots, q. \quad d_{F_k}(x) \geqslant d_k \quad \forall k = q+1, \ldots, r \quad \text{where } \{F_\ell\} \text{ is a family of subsets of } E.$$

If the functions σ_i are locally Lipschitz, the criterion $f = \sum_{i=1}^{p} \sigma_i \circ d_{K_i}$ is locally Lipschitz. More generally, one may consider criterions of the following type: $x \to \varphi(d_{K_1}(x), \ldots, d_{K_p}(x))$ with φ locally Lipschitz.

Beside those already existing, we shall establish a new chain rule for the generalized gradient of composite functions.
First of all, let us recall a definition and a property of a class of locally Lipschitz functions.

Definition 8: [15]. A function $f \in F(E, \overline{R})$, Lipschitz in a neighborhood of x_0 is said to be *regular* at x_0 if the directional derivative $f'(x_0; d) = \lim_{\lambda \to 0^+} [f(x_0 + \lambda d) - f(x_0)] \lambda^{-1}$ exists and is equal to $f^\circ(x_0; d)$ for all $d \in E$.
In this definition and in the sequel, we refer to regularity at x_0 only for functions which are Lipschitz in a neighborhood of x_0. Convex or continuously differentiable functions $f : E \to R$ are examples of functions which are regular at each point of E. As an immediate consequence of the definition, we have:

(5.1) *if* f *is regular at* x_0 *and if* $\sigma : R \to \overline{R}$ *is continuously differentiable at* $f(x_0)$ *with* $\sigma'(f(x_0)) \geqslant 0$, *then* $\sigma \circ f$ *is regular at* x_0. In particular,

(5.2) *if* $-f$ *is regular at* x_0 *with* $f(x_0) \neq 0$, $1/f$ *is regular at* x_0.

In (5.1), $\sigma \circ f$ is not regular at x_0 in general when $\sigma'(f(x_0)) < 0$ (take for example $\sigma(u) = -u$). We emphasize that in the context of nondifferentiable problems, the regularity condition is a rather stringent hypothesis; geometrically, the regularity of f at x_0 means that

$$c\ell\, \{I_{\mathrm{epi}f}(x_0, f(x_0))\} = T_{\mathrm{epi}f}(x_0, f(x_0)) = \mathcal{T}_{\mathrm{epi}f}(x_0, f(x_0)) \quad .$$

If $f : E \to R$ is regular at each point, f is clearly quasi-differentiable on E and the quasi-differential $M_f(x_o)$ is equal to $\partial f(x_o)$ (see definitions in Part I). Conversely, F. H. Clarke [15, § 16] proved the following interesting result:

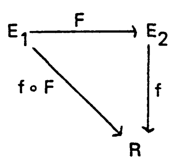
(5.3) *if* $f : E \to \bar{R}$ *is quasi-differentiable, if* M_f *is compact-valued in* E^*_σ *and if the set-valued mapping* $M_f : E \rightrightarrows E^*$ *is upper semicontinuous, then* f *is regular and* $M_f = \partial f$.

A counter-example typical in this respect is the following one: $E = R$, $f_1(x) = x^2 \sin 1/x$ if $x \neq 0$, $f_1(0) = 0$. f_1 is locally Lipschitz, quasi-differentiable at each point (since differentiable); f_1 is not regular at 0 and $M_{f_1}(0) = \{0\}$, $\partial f_1(0) = [-1, +1]$.

V.B. Chain rules for generalized gradients.

$E, E_1, E_2, \ldots$ are real Banach spaces; we shall not distinguish the duality products $\langle .\,,. \rangle$ between the different spaces and their topological dual spaces.
The first two chain rules (Rules No 1, No 2) are from F. H. Clarke [15, § 13, 14]; we recall them here for the convenience of the reader.

Rule No 1

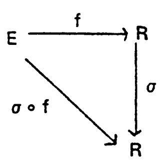


(i) *Let* F *be continuously differentiable, let* f *be locally Lipschitz. Then, by denoting* $DF(x_o) \in \mathcal{L}(E_1, E_2)$ *the differential operator of* F *at* x_o, *one has:*

$$\partial(f \circ F)(x_o) \subset \partial f(F(x_o)) \circ DF(x_o) \quad .$$

(ii) *Equality holds if either* f *(or* $-f$*) is regular at* $F(x_o)$ *or* $DF(x_o)$ *is surjective.*

Rule No 2



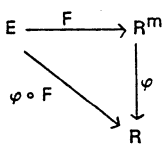
(i) *Let* f *and* σ *be locally Lipschitz; then:*

$$\partial(\sigma \circ f)(x_o) \subset \text{co}\ \{\partial\sigma(f(x_o))\ .\ \partial f(x_o)\} \ .$$

(ii) *Furthermore, if* σ *is continuously differentiable, or if* σ *(or* $-\sigma$*) is regular at* $f(x_o)$ *and* f *continuously differentiable, one has:*

$$\partial(\sigma \circ f)(x_o) = \partial\sigma(f(x_o))\ .\ \partial f(x_o) \quad .$$

Rule No 3



Theorem 6. *Let* $F = (f_1, \ldots, f_m)^T$ *and* φ *be locally Lipschitz. Then,*

$$\partial(\varphi \circ F)(x_o) \subset \text{co}\ \{\ \sum_{i=1}^{m} u_i x^*_i \mid (u_1, \ldots, u_m) \in \partial\varphi(F(x_o))$$

$$(x^*_1, \ldots, x^*_m) \in \prod_{i=1}^{m} \partial f_i(x_o)\} \ .$$

Moreover, if the functions f_i *are regular at* x_o, *if* φ *is regular at* $F(x_o)$ *and if* $\partial\varphi(F(x_o)) \subset R^m_+$, *the equality holds.*

Proof. (i) Let $d \in E$. Let us consider a sequence $\{x_n\}$ converging to x_o and a sequence $\{\lambda_n\} \subset R_+^*$ converging to 0; let us set:

$$E_n = \{\varphi[F(x_n + \lambda_n d)] - \varphi[F(x_n)]\} \cdot \lambda_n^{-1}$$

$\varphi : R^m \to R$ is locally Lipschitz and by the mean value theorem [41], there exist $F_n \in]F(x_n), F(x_n + \lambda_n d)[$ in R^m and $\bar{u}_n \in \partial\varphi(F_n)$ such that

(5.4) $$E_n = \langle F(x_n + \lambda_n d) - F(x_n), \bar{u}_n\rangle \cdot \lambda_n^{-1} \qquad (\bar{u}_n = (\bar{u}_n^1, \ldots, \bar{u}_n^m))$$

According to the same theorem, there exist $\bar{x}_n^i \in]x_n, x_n + \lambda_n d[$ and $\bar{y}_n^i \in \partial f_i(\bar{x}_n^i)$ for all $i = 1, \ldots, m$ such that

$$f_i(x_n + \lambda_n d) - f_i(x_n) = \lambda_n \langle \bar{y}_n^i, d\rangle \ .$$

Briefly, from (5.4):

(5.5) $$E_n = \sum_{i=1}^m \bar{u}_n^i \langle \bar{y}_n^i, d\rangle = \langle \sum_{i=1}^m \bar{u}_n^i \bar{y}_n^i, d\rangle \ .$$

The set-valued mappings $\partial f_i : E \rightrightarrows E_\sigma^*$ and $\partial\varphi : R^m \rightrightarrows R^m$ are upper semicontinuous [14, Proposition 7]. When $n \to \infty$, $F_n \to F(x_o)$ and $\bar{x}_n^i \to x_o$. So, by taking appropriate subsequences,

(5.6) $$\forall\, i = 1, \ldots, m \qquad \bar{y}_n^i \xrightarrow{E_\sigma^*} \bar{y}_i \in \partial f_i(x_o)$$
$$\bar{u}_n^i \longrightarrow \bar{u}^i \in \partial\varphi(F(x_o)) \ .$$

Let us denote by D the set $\{x^* \in E^* \mid x^* = \sum_{i=1}^m u_i x^*_i,\ (u_1, \ldots, u_m) \in \partial\varphi(F(x_o)),\ (x^*_1, \ldots, x^*_m)$ in $\prod_{i=1}^m \partial f_i(x_o)\}$. D is a compact subset of E_σ^* and we derive from (5.5) and (5.6):

$$\limsup_{n\to\infty} E_n \leqslant \text{Max}\ \{\langle x^*, d\rangle \mid x^* \in D\} \ .$$

So, $(\varphi \circ F)^\circ(x_o; d) \leqslant \delta_D^*(d)$ for all $d \in E$ and consequently, $\partial(\varphi \circ F)(x_o) \subset \text{co}\, D$.

(ii) Let us consider the expression $E_n^\circ = \{\varphi[F(x_o + \lambda_n d)] - \varphi[F(x_o)] \cdot \lambda_n^{-1}$. f_i being supposed to be regular at x_o, one has for each i:

$$f_i(x_o + \lambda_n d) - f_i(x_o) = \lambda_n f_i'(x_o; d) + \lambda_n \epsilon_n^i \qquad \text{with} \qquad \lim_{n\to\infty} \epsilon_n^i = 0 \ .$$

Let ϵ_n be the vector $(\epsilon_n^1, \ldots, \epsilon_n^m)^T$, let $V(x_o; d)$ be the vector $(f_1^\circ(x_o; d), \ldots, f_m^\circ(x_o; d))^T$. Since φ is Lipschitz, E_n° may be written as

$$E_n^\circ = \{\varphi[F(x_o) + \lambda_n V(x_o; d)] - \varphi[F(x_o)]\} \cdot \lambda_n^{-1} + O(1)\|\epsilon_n\| \ .$$

φ is regular at $F(x_o)$; so, $\lim_{n\to\infty} E_n^\circ = \varphi^\circ(F(x_o); V(x_o; d))$. Let $(u_1, \ldots, u_m) \in \partial\varphi(F(x_o))$ and $(x^*_1, \ldots, x^*_m) \in \prod_{i=1}^m \partial f_i(x_o)$; since $\partial\varphi(F(x_o)) \subset R_+^m$

$$\sum_{i=1}^m u_i \langle x_i^*, d\rangle \leqslant \sum_{i=1}^m u_i f_i^\circ(x_o; d) \leqslant \varphi^\circ(F(x_o); V(x_o; d)) \ .$$

Consequently, $\delta_D^*(d) \leqslant (\varphi \circ F)^\circ(x_o; d)$ for all d; hence the announced result.

Comments. These different calculus rules have many consequences and applications. Let us begin by the Rule No 3 we have just proved.

■ Rule No 3. As a first consequence, we have

C_1. *If φ is continuously differentiable,*

(5.7) $$\partial(\varphi \circ F)(x_o) \subset \sum_{i=1}^{m} \frac{\partial\varphi}{\partial x_i}(F(x_o))\,\partial f_i(x_o)$$

with equality if the functions f_i are regular at x_o and if $\nabla\varphi(F(x_o)) \in R_+^m$.

One cannot expect the equality in (5.7) with only the hypothesis that φ is continuously differentiable (at least with $m > 1$). That is this estimate (5.7) (with equality if $\frac{\partial\varphi}{\partial x_i}(F(x_o)) \geqslant 0$ for all i) which was proved by A. Auslender [2] for functions of the type $\varphi(x, f_1(x), \ldots, f_n(x))$ where the (regular) functions f_i have a particular structure. We shall give later the exact evaluation of $\partial(\varphi \circ F)(x_o)$ when E is finite-dimensional. By taking $\varphi(x) = \prod_{i=1}^{m} x_i$, one has as a particular application:

(5.8) $$\partial(\prod_{i=1}^{m} f_i)(x_o) \subset \sum_{i=1}^{m} \prod_{j\neq i} f_j(x_o)\,\partial f_i(x_o)$$

with equality if the functions f_i are regular at x_o and if $f_i(x_o) \geqslant 0$ for all i.

C_2. Let us consider φ defined by $\varphi(x) = \operatorname{Max}_{i=1,\ldots,m} x_i$. φ is obviously regular and if e_i is the i-th row vector of the identity (m,m) matrix, one has: $\partial\varphi(F(x_o)) = \text{co}\,\{e_i \mid i, f_i(x_o) = \operatorname{Max}_i f_i(x_o)\}$. Applying the formula of Rule No 3, one finds again the estimate (and the exact evaluation when the functions f_i are regular) of the generalized gradient of $\operatorname{Max}_i f_i$ [14, 34 Chapter V].

C_3. Let us point out that the Rule No 1 (ii) with $E_2 = R^m$ also gives another case implying the equality in the estimate of the Rule No 3 without any regularity assumption on φ, when F is continuously differentiable.

■ Rule No 1

C_4. Let x_o, $d \in E$, let $f_{x_o,d}$ be defined on R by $f_{x_o,d}(\lambda) = f(x_o + \lambda d)$. By applying the Rule No 1 with $E_1 = R$, $E_2 = E$ and $F = F_{x_o,d} : \lambda \to x_o + \lambda d$, one obtains:

(5.9) $$\partial f_{x_o,d}(0) \subset \langle \partial f(x_o), d\rangle$$

with equality if f *(or* $-$f*) is regular at* x_o.

If $E = R^n$, as pointed out by F. H. Clarke, one can state:

(5.10) $$\textit{for almost all } x_o,\ \partial f_{x_o,d}(\lambda) = \langle \partial f(x_o + \lambda d), d\rangle \ \textit{for all } \lambda .$$

The inclusion (5.9) can be directly obtained by proving that for all $v \in R$, $f^{\bullet}_{x_o,d}(0; v) \leqslant f^{\bullet}(x_o; vd)$; with this relation, the mean value theorem can be proved in a straightforward manner. When E_1 and E_2 are finite-dimensional, the estimate of Rule No 1 also appears as a consequence of results of [33].

■ Rule No 2. The estimate (i) given in this rule appears as a particular case of the inclusion given in Theorem 6 but the conditions (ii) ensuring the equality work only when f is real-valued.

Let us give some examples of applications.

C_5. By taking $\sigma(x) = 1/x$, we have:

(5.11) $$\text{if } f(x_0) \neq 0,\ \partial(\tfrac{1}{f})(x_0) = -\frac{\partial f(x_0)}{[f(x_0)]^2}\ .$$

This relation, combined with (5.8) and the remark (5.1) yields:

(5.12) $$\text{if } g(x_0) \neq 0,\ \partial(\tfrac{f}{g})(x_0) \subset \frac{g(x_0)\partial f(x_0) - \partial g(x_0) f(x_0)}{[g(x_0)]^2}$$

with equality if f *and* $-g$ *are regular at* x_0, $f(x_0) \geqslant 0$ *and* $g(x_0) > 0$.

These different relations allow us to apply the necessary (resp. the sufficient) optimality conditions (see Part IV) to *fractional programming,* that is to say to optimization problems where the objective function f has the form $x \mapsto f(x) = [\prod_{i=1}^{m} f_i(x)]\,[\prod_{j=1}^{p} g_j(x)]^{-1}$. In particular, the context of our approach is more general than that of J. M. Borwein [9] who treats the quasi-differentiable case.

C_6. By applying the Rule No 2 (ii) to $\sigma(x) = |x|$, one derives that *if* f *is continuously differentiable and if* $f(x_0) = 0$, $\partial|f|(x_0) = [-\nabla f(x_0), \nabla f(x_0)]$. Generally speaking, we only have

$$\partial|f|(x_0) \subset \text{co}\ \{\partial f(x_0) \cup -\partial f(x_0)\}$$

if $f(x_0) = 0$; the equality does not hold even for functions regular at x_0.

C_7. There is another case where the convexifying operation "co" is unnecessary in the formula of Rule No 2, that is when σ is monotone in a neighborhood of $f(x_0)$. Indeed, if $\sigma : R \to R$ is an increasing (resp. decreasing) function in a neighborhood of u_0, one has $\partial\sigma(u_0) \subset R_+$ (resp. R_-) (this result remains true if $\sigma : R \to \bar{R}$ is not necessarily locally Lipschitz).

The Rule No 3 (ii) applied with $m = 1$ yields another equality case in the frame of Rule No 2, that is:

(5.13) *if* f *is regular at* x_0, *if* σ *is regular at* $f(x_0)$ *and if* $\partial\sigma(f(x_0)) \subset R_+$, *then*

$$\partial(\sigma \circ f)(x_0) = \partial\sigma(f(x_0))\,.\,\partial f(x_0)\ .$$

Example: if f is regular at x_0 and if $f(x_0) = 0$, one has: $\partial f^+(x_0) = [0,1]\,\partial f(x_0)$.

Remarks. Due to the local nature of the notion of generalized gradient, the required properties on the functions (Lipschitz property, continuous differentiability . . .) need to be assumed only in a neighborhood of the considered points.

The results (5.13) and (5.7) are extensions of analogous results established in the convex case by C. Lescarret [42, Propositions 3 and 6].

V.C. An exact chain rule in the finite-dimensional case.

In this section, we shall suppose that $E = R^n$.

For $g : R^r \to R^s$ satisfying a Lipschitz condition in a neighborhood of $x_0 \in R^r$, F. H. Clarke [16] defined the *generalized Jacobian matrix* of g at x_0 as following:

Definition 9. *The generalized Jacobian matrix of* g *at* x_0, *denoted by* $\mathcal{J}(g; x_0)$, *is the set of matrices defined by*

$$\mathcal{J}(g; x_0) = \mathrm{co}\ \{ \lim_{i \to \infty} J(g; x_i) \}$$

in this definition, x_i *converges to* x_0, g *is differentiable at* x_i *for each* i *and* $J(g; x_i)$ *is the usual Jacobian matrix.*

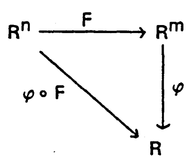
$\mathcal{J}(g; x_0)$ is a set of (s,r) matrices and we denote by $\mathcal{J}^T(g; x_0)$ the set of transposed matrices A^T with $A \in \mathcal{J}(g; x_0)$. If g is continuously differentiable at x_0, $\mathcal{J}(g; x_0)$ is reduced to $\{J(g; x_0)\}$. In the case where $s = 1$, $g'(x)$ is represented by a row vector $(\frac{\partial g}{\partial x_i}(x))$ whereas the column vector is the gradient $\nabla g(x) \in R^r$. In order to preserve the given definition of $\partial g(x_0)$ (see (1.18)) and the coherence of notations, we shall consider the elements of $\partial g(x_0)$ as elements of R^r (and not of $(R^r)^*$). So, we generally have:

$$\mathcal{J}^T(g; x_0) \subset (\partial g_1(x_0), \ldots, \partial g_s(x_0)) \quad . \tag{5.14}$$

This inclusion is strict even for very regular functions g_i; for example, if $g : R \to R^2$ is defined by $g(x) = (|x|, |x|)^T$, we have $\mathcal{J}^T(g; 0) = \{(u, u) \mid u \in [-1, +1]\}$.

Let us consider again the frame of Rule No 3 with a continuously differentiable function φ; we shall give an *exact evaluation* of $\partial(\varphi \circ F)(x_0)$ using the generalized Jacobian matrix of F at x_0.

Rule No 3 bis



Theorem 7. *Let* F *be a locally Lipschitz function, let* φ *be continuously differentiable. Then,*

$$\partial(\varphi \circ F)(x_0) = \mathcal{J}^T(F; x_0) \nabla \varphi(F(x_0))$$

Proof. According to Definition 9, it is easy to show that $\partial(\varphi \circ F)(x_0)$ contains the set $\mathcal{J}^T(F; x_0)\nabla\varphi(F(x_0))$. For the converse inclusion, we follow the proof given by F. H. Clarke [16, Lemma 2] in the case where $m = n$. This proof is based on the following property: if $g : R^q \to R$ is a locally Lipschitz function, if Λ is a subset of R^n of measure 0, one does not modify $\partial g(x_0)$ if the points x_i are constrained to the complement of Λ in the definition (1.18). We then apply this result to $g = \varphi \circ F$ by taking $\Lambda = \{x \in R^n \mid F$ is nondifferentiable at $x\}$.

As examples, *if* $F = (f_1, \ldots, f_m)^T : R^n \to R^m$ *is locally Lipschitz, one has:*

$$\partial(\sum_{i=1}^m f_i)(x_0) = \{ \sum_{i=1}^m A_i \mid (A_1, \ldots, A_m) \in \mathcal{J}^T(F; x_0) \}$$

$$\partial(\prod_{i=1}^m f_i)(x_0) = \{ \sum_{i=1}^m \prod_{j \neq i} f_j(x_0) A_i \mid (A_1, \ldots, A_m) \in \mathcal{J}^T(F; x_0) \}$$

By taking $\varphi(x) = x_i$, we find again the following "projection" property:

$$\partial f_i(x_0) = \{ A_i \mid (A_1, \ldots, A_i, \ldots, A_m) \in \mathcal{J}^T(F; x_0) \} \ .$$

Remarks. 1. An exact chain rule such as in Theorem 7 with a nondifferentiable function φ cannot be obtained; the function $\varphi(x) = \underset{i}{\mathrm{Max}}\, x_i$ allow to construct simple counterexamples.

2. Other generalizations of the notion of derivative have been successively introduced for vector-valued functions: the notion of "derivate containers" by J. Warga [62] and the concept of "screens" by H. Halkin [29]; although conceptually related to the notion of generalized Jacobian matrix, their definitions seem more difficult to handle.

Addendum. After having finished this work, we became aware of three papers which also deal with the generalized gradient and related definitions. The first paper is that of B. Pourciau [J. of Optim. Theory and Applications, Vol. 22, No 3 (1977), 311–351]; B. Pourciau gives a definition of the *generalized derivative* of a function $f : R^n \to R^k$ by taking $\partial^P f(x_0) = \bigcap_{\epsilon>0} \overline{co}\, \{f'(x) \mid x \in B_\epsilon(x_0) \cap L(f')\}$; in this definition, $L(f')$ is the Lebesgue set of f'. When f is real-valued, $\partial^P f(x_0)$ is exactly the generalized gradient but in the vectorial case $(k > 1)$, we do not know whether $\partial^P f(x_0)$ and $\mathscr{J}(f; x_0)$ really coincide. The two other papers are those of R. Mifflin and have reference to *semismooth* and *semiconvex* functions [SIAM J. Control and Optimization, Vol. 15, No 6 (1977), 959–972, Math. of Operations Research, Vol 2, No 2 (1977), 191–207]. Semismooth functions are particular locally Lipschitz functions possessing a semicontinuous relationship between their generalized gradients and directional derivatives. In the first paper quoted above, R. Mifflin gives a chain rule (inclusion) as in Theorem 6 (with $E = R^n$) and shows that a semismooth composition of semismooth functions is semismooth. Note that the functions called quasi-differentiable by R. Mifflin are in fact regular following the terminology of F. H. Clarke. In the second R. Mifflin's paper, the class of *weakly upper semismooth* functions is introduced; such functions also have a semicontinuous relationship between their generalized gradients and directional derivatives but this relationship is weaker than the corresponding one for semismooth functions.

References

[1] A. Auslender. *Optimisation: Méthodes numériques.* Masson et Cie, Paris (1976).

[2] A. Auslender. *Minimisation sans contraintes de fonctions localement Lipschitziennes: application à la programmation mi-convexe, mi-différentiable.* C. R. Acad. Sc. Paris, Série A, t. 284 (1977), 959–961, and Proceedings of the 3rd Symposium on Nonlinear Programming, University of Wisconsin, Madison (to appear).

[3] M. S. Bazaraa and J. J. Goode. *Extension of optimality conditions via supporting functions.* Math. Programming 5 (1973), 267–285.

[4] M. S. Bazaraa, J. J. Goode and Z. Nashed. *On the cones of tangents with applications to mathematical programming.* J. Optimization Theory and Appl. Vol 13, No 4 (1974), 389–426.

[5] M. S. Bazaraa, C. M. Shetty, J. J. Goode and M. Z. Nashed. *Nonlinear programming without differentiability in Banach spaces: necessary and sufficient constraint qualification.* Applicable Anal. Vol 5 (1976), 165–173.

[6] C. Berge. *Espaces topologigues, fonctions multivoques.* Dunod, Paris (1966).

[7] H. Berliocchi and J. M. Lasry. *Principe de Pontryagin pour des systemes régis par une équation différentielle multivoque.* Publications Mathématiques de l'université de Bordeaux I 1973–1974 (2), 25–41.

[8] V. G. Bolt'yanskii. *The method of tents in the theory of extremal problems.* Russian Math. Surveys 30, 3, (1975), 1–54.

[9] J. M. Borwein. *Fractional programming without differentiability.* Math. Programming 11 (1976), 283–290.

[10] N. Bourbaki. *Espaces vectoriels topologiques.* Fascicules XV, XVII, Hermann, Paris (1966–1967).

[11] Bui-Trong-Liêu and P. Huard. *La méthode des centres dans un espace topologique.* Numer. Math. 8 (1966), 58–67.

[12] F. H. Clarke. *Necessary conditions for nonsmooth problems in optimal control and the calculus of variations.* Ph. D. Dissertation, University of Washington (1973).

[13] F. H. Clarke. *Generalized gradients and applications.* Transactions of the American Math. Society, Vol 205 (1975), 247–262.

[14] F. H. Clarke. *A new approach to Lagrange multipliers.* Math. of Operations Res. 2 (1976), 165–174.

[15] F. H. Clarke. *Generalized gradients of Lipschitz functionals.* Madison M. R. C. Technical Summary Report, University of Wisconsin (1976).

[16] F. H. Clarke, *On the inverse function theorem.* Pacific J. Math., Vol 64, No 1 (1976), 97–102.

[17] B. D. Craven and B. Mond. *Lagrangean conditions for quasi-differentiable optimization.* Research Report No 7 (1976), School of Mathematical Sciences, University of Melbourne.

[18] J.-P. Crouzeix. *Conjugacy in quasi-convex analysis.* Convex Analysis and its applications, Lecture Notes in Economics and Mathematical Systems No 144 (A. Auslender, Ed.) Springer-Verlag (1977) pp. 66–99.

[19] J.-P. Crouzeix. *Contributions à l'étude des fonctions quasi-convexes.* Thèse de Doctorat ès-Sciences Mathématiques, Université de Clermont II (1977).

[20] V. F. Dem'yanov and V. N. Malozemov. *Introduction to minimax.* John Wiley and Sons, New York (1974).

[21] J. Durdil. *On the geometrical characterization of differentiability* I. Commentationes Mathematicae Universitatis Carolinae 15, 3 (1974), 521–540.

[22] J. Durdil and J. Daneš. *A note on geometric characterization of Fréchet-differentiability.* Commentationes Mathematicae Universitatis Carolinae 17, 1(1976), 195–204.

[23] A. Feuer. *Minimizing well-behaved functions.* Proceedings of Twelfth Annual Allerton Conference on Circuit and System Theory. Illinois (1974), 25–34.

[24] I. V. Girsanov. *Lectures on mathematical theory of extremum problems.* Lecture Notes in Economics and Mathematical Systems, Springer-Verlag (1972).

[25] A. A. Goldstein. *Optimization of Lipschitz continuous functions.* Math. Programming, Vol 13 (1977) No 1, 14–22.

[26] F. J. Gould and J. W. Tolle. *Geometry of optimality conditions and constraint qualifications.* Math. Programming 2 (1972), 1–18.

[27] F. J. Gould and J. W. Tolle. *Optimality conditions and constraint qualifications in Banach space.* J. Optimization Theory Appl. Vol 15 (1975), No 6, 667–684.

[28] H.J. Greenberg and W. P. Pierskalla. *Quasiconjugate function and surrogate duality.* Cahiers du Centre d'Etudes de Rech. Oper. Vol 15, No 4 (1973), 437–448.

[29] H. Halkin. *Interior mapping theorem with set-valued derivatives.* Journal d'Analyse Mathématique, Vol 30 (1976), 200–207.

[30] M. R. Hestenes. *Optimization Theory: the finite dimensional case.* J. Wiley and Sons (1975).

[31] J-B. Hiriart-Urruty. *Conditions nécessaires d'optimalité en programmation non différentiable.* Note aux Comptes Rendus Acad. Sc. Paris t. 283, Série A (1976), 843–845.

[32] J-B. Hiriart-Urruty. *On necessary optimality conditions in nondifferentiable programming.* Math. Programming, Vol 14, (1978) (to appear).

[33] J-B. Hiriart-Urruty. *Gradients généralisés de fonctions marginales.* SIAM J. on Control and Optimization (1978) (to appear).

[34] J-B. Hiriart-Urruty. *Contributions a la programmation mathématique: cas détérministe et stochastique.* Thèse de Doctorat ès-Sciences Mathématiques, Université de Clermont II (1977).

[35] J-B. Hiriart-Urruty. *Tangent cones, generalized gradients and mathematical programming in Banach spaces.* Math. of Operations Res. (1978) (to appear).

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

[37] Hoang Tuy. *Sur les inégalités linéaires.* Colloq. Math. 13 (1964), 107–123.

[38] R. Janin. *Sur la dualité et la sensibilité dans les problèmes de programme mathématique.* Thèse de Doctorat ès-Sciences Mathématiques, Université de Paris VI (1974).

[39] S. N. Kruzhkov. *On the differentiability almost everywhere of functions of many variables.* Vestnik Moskovskogo Universiteta, Matematika, Vol 31, No 6 (1976), 67–70 = Moscow University Mathematics Bulletin, Vol 31, No 5–6, (1976), 129–131.

[40] P-J. Laurent. *Approximation et Optimisation,* Hermann (1972).

[41] G. Lebourg. *Valeur moyenne pour gradient généralisé.* Note aux Comptes Rendus Acad. Sc. Paris, Série A, (1975), 795–797.

[41]* G. Lebourg. *Generic differentiability of Lipschitzian functions.* Cahiers de Mathématiques de la Décision (1977), No 7704.

[42] C. Lescarret. *Sous-différentiabilité de fonctions composées.* Travaux du Séminaire d'Analyse Unilatérale, Vol. 1, Exposé No 2 (1968), Université de Montpellier.

[43] E. S. Levitin, A. A. Milyutin and N. P. Osmolovskii. *On necessary and sufficient conditions for a local minimum in a problem with constraints.* Soviet. Math. Dokl., Vol 14, No 3 (1973), 853–857.

[44] O. L. Mangasarian. *Nonlinear programming.* Mc. Graw-Hill, New York (1972).

[45] H. Massam and S. Zlobec. *Various definitions of the derivative in mathematical programming.* Math. Programming 7 (1974), 144–151.

[46] P. Michel. *A propos du problème des inégalités et de ses applications à la programmation et au contrôle optimal.* Publications Mathématiques de l'Université de Bordeaux I, 1973–1974 (3), 31–44.

[47] A. Nijenhuis. *Strong derivatives and inverse mappings.* Amer. Math. Monthly, 81 (1974), 969–980.

[48] E. A. Nurminskii. *The quasi-gradient method for the solving of the nonlinear programming problems.* Kibernetika, 1, (1973), 122–125 = Cybernetics, Vol 9, No 1 (1974), 145–150.

[49] E. A. Nurminskii and A. A. Zhelikhovskii. *ϵ-quasi-gradient method for solving nonsmooth extremal problems.* Kibernetika, 1, (1977), 109–113 = Cybernetics, Vol 13, No 1, (1977), 109–114.

[50] J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables.* Acad. Press, New York and London (1970).

[51] R. Pallu De La Barrière. *Fonctions sous-linéarisables et principe de Pontryagin.* Publications Mathématiques de l'Université de Bordeaux I, 1973–1974 (3), 65–72.

[52] J. P. Penot. *Sous-différentiels de fonctions numériques non convexes.* Note aux Comptes Rendus Acad. Sc. Paris, t. 278, Serie A (1974), 1553–1555.

[53] J-P. Penot. *Calcul sous-différentiel et Optimisation.* Publications Mathématiques de l'Université de Pau (1974).

[54] J-P. Penot (to appear).

[55] B. N. Pshenichnyi. *Necessary conditions for an extremum.* Marcel Dekker Inc., New York (1971).

[56] B. N. Pshenichnyi. *Leçons sur les jeux différentiels.* Cahiers de l'I.R.I.A., No 4, (1971), 145–226.

[57] L. Rigby. *Contribution to Dubovitskii and Milyutin's optimization formalism.* Optimization Techniques, Modeling and Optimization in the Service of Man. Part 2. Lecture Notes in Computer Science (J. Ceá, Ed.) Springer-Verlag (1976), 438–453.

[58] R. T. Rockafellar, *Convex Analysis.* Princeton University Press (1970).

[59] N. Z. Shor. *A class of almost-differentiable functions and a minimization method for functions of this class.* Kibernetika, 8, (1972), 65–70 = Cybernetics, Vol 8, No 4, (1974), 509–606.

[60] E. M. Stein. *Singular integrals and differentiability properties of functions.* Princeton Math. Ser. No 30, Princeton Univ. Press, N. J. (1970).

[61] L. Thibault. *Propriétés des sous-différentiels des fonctions localement Lipschitzcennes définies sur un espace de Banach séparable. Applications.* Thèse de Doctorat de Spécialité, Université de Montpellier (1976).

[62] J. Warga. *Derivate containers, Inverse functions and Controllability.* Calculus of Variations and Control Theory (D. L. Russell, Ed.) Acad. Press (1976).

[63] Y. I. Zabotin, A. I. Korablev and R. F. Khabibullin. *Conditions for an extremum of a functional in case of constraints.* Cybernetics, Vol 9, No 6, (1975), 982–988.

[64] W. I. Zangwill. *Nonlinear programming: a unified approach.* Prentice-Hall, Inc., N. J. (1969).

Université de Clermont II
Complexe Scientifique des Cézeaux
Département de Mathématiques Appliquées
B. P. No 45
63170. AUBIERE (FRANCE)

Presently on leave at:

The Department of Mathematics
The University of Kentucky
Lexington, Kentucky 40506