TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 281, Number 1, January 1984

APPROXIMATE SUBDIFFERENTIALS AND APPLICATIONS. I: THE FINITE DIMENSIONAL THEORY

BY

A. D. IOFFE

ABSTRACT. We introduce and study a new class of subdifferentials associated with arbitrary functions. Among the questions considered are: connection with other derivative-like objects (e.g. derivatives, convex subdifferentials, generalized gradients of Clarke and derivate containers of Warga), calculus of approximate subdifferentials and applications to analysis of set-valued maps and to optimization.

It turns out that approximate subdifferentials are minimal (as sets) among other conceivable subdifferentials satisfying some natural requirements. This shows that certain results involving approximate subdifferentials are the best possible and, at the same time, marks certain limitations of nonsmooth analysis. Another important property of approximate subdifferentials is that, being essentially nonconvex, they admit a rich calculus that covers the calculus of convex subdifferentials and leads to more precise and sometimes new results for generalized gradients of Clarke.

Introduction. The concept of a subdifferential is usually associated with convex functions for which, to a large extent, subdifferentials proved to be one of the most useful and porweful instruments responsible for the success of convex analysis in the 1960's [**17**]. Many of the good properties of convex subdifferentials were inherited by generalized gradients introduced by Clarke [**3**] for lower semicontinuous functions on Banach spaces. Rockafellar gave an alternative definition which applies to arbitrary functions on arbitrary locally convex spaces [**19**].

Here we study another class of objects called *approximate subdifferentials* denoted by $\partial_a f(x)$. They appeared for the first time in a finite dimensional situation as by-products of certain approximative optimization techniques developed by Mordukhovich [**14**]. Some subsequent attempts to extend the definition to a more general situation were successful only for Banach spaces with an equivalent Gâteaux [**7**] or Fréchet [**12**] differentiable norm. A new definition that applies in arbitrary locally convex spaces was offered by Ioffe in [**8**] where many nice analytic properties of approximate subdifferentials were first announced. An infinite dimensional theory of approximate subdifferentials will be considered in the second ("General theory") and the third ("Banach theory", which is especially rich) parts of the paper approximately corresponding respectively to the first announcement [**8**] and the first version of the paper mimeographically distributed as [**10**]. Here we present the first part where *all spaces are assumed finite dimensional.*

There have been two reasons to consider the finite dimensional case separately. On the one hand, definitions and proofs are much simpler in this case and many

Received by the editors November 20, 1981 and, in revised form, March 8, 1983.
1980 *Mathematics Subject Classification.* Primary 49B27.

technical complications do not even appear. On the other hand, we can prove stronger results if the dimension is finite. The situation is approximately as in convex analysis where certain finite dimensional theorems are valid under assumptions on relative interiors while corresponding general results require assumptions on interiors.

Here is the contents supplied with short comments.

§1. The definition.

§2. Geometric characterizations. (Approximate subdifferentials are defined in a pure analytic fashion as, in a sense, upper limits of so-called Dini subdifferentials (or semidifferentials) [**9**, **16**]. The definition applies to arbitrary functions. With approximate subdifferentials in hand, we define the approximate normal cone (to a set) as the approximate subdifferential of the indicator function of the set. For a closed set this cone coincides with the normal cone introduced by Mordukhovich [**14**] and also with the closed cone generated by the approximate subdifferential of the distance function to the set.)

§3. Connection with other derivative-like objects. (We show that for a convex (strictly differentiable) function the approximate subdifferential coincides with the usual convex subdifferential (resp. with the derivative). Thus, approximate subdifferentials as well as Clarke's generalized gradients (henceforth we abbreviate them C.g.g. and denote them by $\partial_c f(x)$) extend convex subdifferentials to arbitrary functions. But the appearances of the two are very different because $\partial_c f(x)$ is always a convex set while $\partial_a f(x)$ is typically nonconvex. The general relationship between approximate subdifferentials and C.g.g's in the finite dimensional case is defined by the fact that Clarke's normal cone is always the convex closure of the corresponding approximate normal cone. In particular $\partial_a f(x)$ is always smaller than $\partial_c f(x)$. If f is Lipschitz near x, then the latter is the convex hull of the first.

We also show that $\partial_a f(x)$ is smaller than any Warga's derivate container of f at x [**20**]. This result is essentially due to Kruger and Mordukhovich who proved it under somewhat stronger assumptions.)

§4. Approximate subdifferentials of a sum,

§5. Approximate coderivatives and approximate subdifferentials of a composition. (These two sections are central: they contain the calculus of approximate subdifferentials including formulae for sums, compositions, upper and lower bounds. The collection of such formulae for approximate subdifferentials is even richer than for C.g.g's. (Certain formulae for composition and the formula for lower bounds do not have analogues for C.g.g's.) Moreover, in certain cases we prove, as corollaries, new or stronger results on C.g.g.'s. For instance, we show (Corollary 4.4) that the inclusion

$$\partial_c(f_1 + f_2)(x) \subset \partial_c f_1(x) + \partial_c f_2(x) \tag{0.1}$$

(which is the cornerstone of the calculus of C.g.g's) is valid without the assumption that one of the functions is directionally Lipschitz (cf. [**18**]). The assumption disappears also from other theorems of calculus of C.g.g.

It is appropriate to mention that all main theorems of calculus of convex subdifferentials [**17**] follow from corresponding results for approximate subdifferentials collected in §§4 and 5.

The fact that approximate subdifferentials, so heavily nonconvex, admit a developed calculus must be surprising for anyone experienced in convex and nonsmooth analyses. Indeed, known proofs (say, of (0.1)) would not work without convexity. In the theory of approximate subdifferentials the convexity technique does not play a dominating role although it remains an important part of the machinery. What becomes principal is a version of a penalty function method. All necessary technical lemmas will be proved here to make the paper self-contained.

In the second part of the paper we shall axiomatize the techniques by introducing a class of Banach spaces having some good subdifferentiability properties. We refer to [**11**] where such spaces are specially studied mainly in connection with Asplund and weak Asplund spaces.)

§6. Application to surjection and stability theorems.

§7. Application to optimization problems.

(These two sections have been added to demonstrate how the techniques of approximate subdifferentials work. However, the results presented (surjection and stability theorems of §6 and a Lagrange multiplier rule of §7) are new and seem to be the strongest results of such sort.)

§8. Minimality properties. (This is a short section in which we prove that $\partial_a f(x)$ is smaller than any other "subdifferential" $\partial f(x)$ having one or another collection of reasonably good (and, in fact, natural) properties. For instance, each of the collections includes the requirement that

$$0 \in \partial f(x) \tag{0.2}$$

if f attains a local minimum at x, a form of upper semicontinuity property and embrionic calculus requirements.

There are two major implications of the minimality properties. On the one hand, we may be sure that in applications connected with (0.2) (say, those considered in §§6 and 7) approximate subdifferentials provide for the best possible results. On the other hand, we shall see that even these best results are, in certain respects, still not as good as their smooth counterparts. Therefore the minimality theorem actually establishes some natural limitations of nonsmooth optimization and nonsmooth analysis in general.)

By X, Y, Z we denote spaces (all finite dimensional) and by X^*, Y^*, Z^* their duals. It will be convenient for us to discriminate between a space and its dual. If the space is Euclidean we identify X and X^* as sets but retain the notation x^* when x^* plays a role of a linear functional. By B we denote the unit ball, by $\langle x^*, x\rangle$ the pairing between X^* and X (or the inner product if the space is Euclidean) and by $K^0 = \{x^* \mid \langle x^*, x\rangle \leq 0 \ \forall x \in K\}$ the polar of K; the bar always means the closure and $\overline{\operatorname{conv}} A$ is the convex closure of A.

We use the notation $\partial f(x)$ only for convex subdifferentials (except in the last section) while Clarke's generalized gradients and normal cones are denoted by $\partial_c f(x)$ and $N_c(S, x)$ respectively.

1. The definition. Let f be a function on X such that $|f(x)| < \infty$. We set

$$d^-f(x; h) = \liminf_{\substack{u \to h \\ t \searrow 0}} t^{-1}(f(x + tu) - f(x)),$$

$$\partial^- f(x) = \{x^* \in X^* \mid \langle x^*, h \rangle \leq d^-f(x; h), \forall h\}.$$

The function $h \to d^-f(x; h)$ is called the (lower) Dini directional derivative of f at x and the set $\partial^- f(x)$ the Dini subdifferential of f at x (see [**9**, **16**] for more details).

If $|f(x)| = \infty$, we set $\partial^- f(x) = \varnothing$.

We set further

$$U(f, z, \delta) = \{x \mid \|x - z\| < \delta, |f(x) - f(z)| < \delta\}.$$

DEFINITION 1. Let $|f(z)| < \delta$. The set

$$\partial_a f(z) = \bigcap_{\delta > 0} \bigcup_{x \in U(f, z, \delta)} \partial^- f(x) = \limsup_{\substack{x \to z \\ f(x) \to f(z)}} \partial^- f(x)$$

will be called the *approximate subdifferential* of f at x.

If $|f(z)| = \infty$, we set $\partial_a f(z) = \varnothing$.

If f is continuous at z, the condition $x \in U(f, z, \delta)$ may be replaced by $\|x - z\| < \delta$ (with an arbitrary norm $\|\cdot\|$) and the condition $f(x) \to f(z)$ may be excluded from under the sign of upper limit. We note also that for a Lipschitz f

$$d^-f(x; h) = \liminf_{t \searrow 0} t^{-1}(f(x + th) - f(x)).$$

As immediately follows from the definition, approximate subdifferentials have the following upper semicontinuity property:

$$\partial_a f(z) = \limsup_{\substack{x \to z \\ f(x) \to f(z)}} \partial_a f(x). \tag{1.1}$$

Of course, $\partial_a f(z)$ is always a closed set. It is easy to see that

$$\partial_a f(z) \neq \varnothing \quad \text{if } f \text{ is Lipschitz near } z \tag{1.2}$$

and, moreover, $\partial_a f(z)$ belongs to the k-ball about the origin if k is a Lipschitz constant of f near z. Indeed, in this case $d^-f(x; h) \leq k\|h\|$ and all $\partial^- f(x)$ are contained in the ball.

2. Geometric characterizations.

DEFINITION 2. Let $S \subset X$, $z \in S$, and let $\delta(S, x)$ denote the indicator function of S, that is, the one equal to 0 on S and to $+\infty$ outside of S. The set $N_a(S, z) = \partial_a \delta(S, z)$ (which is obviously a cone) will be called the *approximate normal cone to S at z*.

A more detailed description of the approximate normal cone would incorporate the notion of contingent cone. It is easy to see that $d^-\delta(s, x; h)$ may assume only two values, 0 and ∞, and $d^-\delta(S, x; h) = 0$ if and only if there are sequences $u_n \to h$, $t_n \searrow 0$ such that $x + t_n u_n \in S$. The collection of all h having this property is called the contingent cone to S at x [**1**]. This is by definition the set

$$T(S, x) = \limsup_{t \searrow 0} t^{-1}(S - x) = \bigcap_{t > 0} \overline{\bigcup_{0 < r \leq t} r^{-1}(S - x)}.$$

Thus, $d^-\delta(S, x; h)$ is the indicator function of $T(S, x)$:

$$d^-\delta(S, x; h) = \delta(T(S, x), h)$$

so that $\partial^-\delta(S, x) = T^0(S, x)$ is the polar of $T(S, x)$. Thus, we have proved the following result.

PROPOSITION 1.

$$N_a(S, z) = \limsup_{x \to z} T^0(S, x). \tag{2.1}$$

It is possible to reverse definitions and express approximate subdifferentials in terms of approximate normal cones (to the epigraph) as it is often done for subdifferentials of different types.

PROPOSITION 2. *Denote by*

$$\operatorname{epi} f = \{(\alpha, x) \mid \alpha \geqslant f(x)\} \subset R \times X$$

the epigraph of f. If $|f(z)| < \infty$, then

$$\partial_a f(z) = \{x^* \mid (-1, x^*) \in N_a(\operatorname{epi} f, (f(z), z))\}.$$

PROOF. It is easy to see that the contingent cone to epi f at $(f(x), x)$ consists (if $|f(x)| < \infty$) of those (α, h) for which $\alpha \geqslant d^-f(x; h)$. It follows that the inclusion $(-1, x^*) \in T^0(\operatorname{epi} f, (f(x), x))$ is equivalent to the inequality $\langle x^*, h\rangle \leqslant \alpha$ for all (α, h) such that $\alpha \geqslant d^-f(x; h)$ which, in turn, is the same as $x^* \in \partial^-f(x)$. It is an elementary exercise on properties of upper limits to deduce from here that $(-1, x^*) \in N_a(\operatorname{epi} f, (f(z), z))$ if and only if $x^* \in \partial_a f(z)$.

We shall give two more representations to approximate normal cones, one rather more analytic and the other purely geometrical. Let us fix a norm in X, and let ρ denote the corresponding distance function. We set $(z \in S)$

$$N_a^\rho(S, z) = \overline{\bigcup_{\lambda > 0} \lambda \partial_a \rho(S, z)}.$$

PROPOSITION 3. *The set $N_a^\rho(S, z)$ does not depend on ρ. In other words, if ρ, ρ' are two different distance functions on X (corresponding to different norms on X), then $N_a^\rho(S, z) = N_a^{\rho'}(S, z)$.*

PROOF. We note first that $\rho(S, z) = \rho(\bar{S}, z)$ so that S may be assumed closed. If $x \notin S$, $u \in S$ and $\|x - u\| = \rho(S, x)$, then

$$\rho(S, x + h) \leqslant \|x - u\| + \rho(S, u + h).$$

Therefore $d^-\rho(S, x; h) \leqslant d^-\rho(S, u; h)$ which implies together with (1.1) (and the remark following Definition 1) that

$$\partial_a \rho(S, z) = \limsup_{\substack{x \to z \\ x \in S}} \partial^-\rho(S, x). \tag{2.2}$$

If $\|\cdot\|'$ is another norm on X and ρ' is the corresponding distance function, then there are $K \geqslant k > 0$ such that $k\|x\| \leqslant \|x\|' \leqslant K\|x\|$. It follows that $kd^-\rho(S, x; h) \leqslant d^-\rho'(S, x; h) \leqslant Kd^-\rho(S, x; h)$ for all h if $x \in S$ (because $\rho(S, x) = \rho'(S, x) = 0$ for

such x). Together with (2.2) this implies that

$$k\partial_a\rho(S, z) \subset \partial_a\rho'(S, z) \subset K\partial_a\rho(S, z)$$

and the desired equality follows.

One of the implications of Proposition 3 is that the theory we are going to present does not depend on the specific metric structure of X. In particular from now on, *we assume X to be a Euclidean space*; as to ρ, it will always denote the Euclidean metric.

Let S be a closed subset of X and $x \in S$. A vector x^* is said to be a proximate normal to S at x if there are $u \notin S$ and $\lambda > 0$ such that $x^* = \lambda(u - x)$ and $\|u - x\| = \rho(S, u)$. Denote by $N_p(S, x)$ the collection of all proximate normals to S at x. We set

$$N_m(S, z) = \limsup_{\substack{x \to z \\ x \in S}} N_p(S, x).$$

This is just the normal cone introduced by Mordukhovich [**14**].

THEOREM 1. *Let $S \subset X$ be a closed set and $z \in S$. Then $N_a^\rho(S, z) = N_m(S, z) = N_a(S, z)$.*

Before proving the theorem we shall establish a simple but an important property of Dini subdifferentials.

LEMMA 1. *If $0 \in \partial^- f(z)$, then for any $\varepsilon > 0$ the function $f(x) + \varepsilon\|x - z\|$ attains a strict local minimum at z.*

PROOF. Assume the contrary: there are $\varepsilon > 0$ and a sequence $x_n \to z$ such that $f(x_n) + \varepsilon\|x_n - z\| < f(z)$. Setting $t_n = \|x_n - z\|$, $u_n = (x_n - z)/t_n$ and assuming (with no loss of generality since $\|u_n\| = 1$) that the sequence u_n converges to some h, $\|h\| = 1$, we conclude that

$$\begin{aligned} d^- f(z; h) &\leqslant \liminf_{n \to \infty} t_n^{-1}(f(x + t_n u_n) - f(z)) \\ &= \liminf_{n \to \infty} t_n^{-1}(f(x_n) - f(z)) \leqslant -\varepsilon \end{aligned}$$

in contradiction with the assumption that $0 \in \partial^- f(z)$.

PROOF OF THEOREM 1. It is obvious that $\lambda\rho(S, x) \leqslant \delta(S, x)$ for all x and $\lambda > 0$. Therefore $\lambda d^-\rho(S, x; h) \leqslant d^-\delta(S, x; h)$ (if $x \in S$) and as follows from (2.2), $N_a^\rho(S, z) \subset N_a(S, z)$.

Let $x^* \in N_a(S, z)$, $\|x^*\| = 1$. Then for any $\varepsilon > 0$ there are $x \in S$ and $u^* \in \partial^-\delta(S, x)$ such that $\|x - z\| < \varepsilon$, $\|u^* - x^*\| < \varepsilon$ and $\|u^*\| = 1$. By Lemma 1

$$\delta(S, u) - \langle u^*, u - x\rangle + \varepsilon\|u - x\| > 0 \tag{2.3}$$

for all $u \in S$ that are sufficiently close to x. We set $v_\lambda = x + \lambda u^*$, and take $u_\lambda \in S$ such that $\|v_\lambda - u_\lambda\| = \rho(S, v_\lambda)$. Then $u_\lambda \to x$ as $\lambda \to 0$. Let λ be such that $\|u_\lambda - x\| < \varepsilon$ and (2.3) is valid if we substitute u_λ for u. Since $\|u_\lambda - v_\lambda\| \leqslant \|x - v_\lambda\| = \lambda$, we have

$$\begin{aligned} \lambda^2 &\geqslant \langle u_\lambda - v_\lambda, u_\lambda - v_\lambda\rangle = \langle (u_\lambda - x) - \lambda u^*, (u_\lambda - x) - \lambda u^*\rangle \\ &= \|u_\lambda - x\|^2 - 2\lambda\langle u^*, u_\lambda - x\rangle + \lambda^2. \end{aligned}$$

Together with (2.3) this gives $\|u_\lambda - x\| \leqslant 2\varepsilon\lambda$.

The vector $v^* = (v_\lambda - u_\lambda)/\lambda$ belongs to $N_p(S, u_\lambda)$ by definition. On the other hand

$$\lambda\|v^* - u^*\| = \|v_\lambda - u_\lambda - \lambda u^*\| = \|x - u_\lambda\| \leqslant 2\varepsilon\lambda$$

so that $\|v^* - x^*\| \leqslant 3\varepsilon$. Applying the definition of the Mordukhovich normal cone, we conclude that $x^* \in N_m(S, z)$. Thus, $N_a(S, z) \subset N_m(S, z)$.

Finally, let $x^* \in N_p(S, x)$, $\|x^*\| = 1$. Then $x^* = \lambda(u - x)$ for some $\lambda > 0$ and $u \notin S$ such that $\|u - x\| = \rho(S, u)$. We shall show that $x^* \in \partial^-\rho(S, x)$ which will imply the last inclusion we need, $N_m(S, z) \subset N_a^\rho(S, z)$.

Fix an $h \in X$ and let $x_t \in S$ $(t > 0)$ be such that $\|x_t - (x + th)\| = \rho(S, x + th)$. Then $x_t = x + ty_t$ for some y_t such that $\|y_t\| \leqslant 2\|h\|$. (Indeed, $\|x_t - x\| \leqslant \|x_t - (x + th)\| + \|th\| = t\|h\| + \rho(S, x + th) \leqslant 2t\|h\|$.) The equality

$$\begin{aligned}\|x + ty_t - u\|^2 &= \langle x + ty_t - u, x + ty_t - u\rangle = \|x - u\|^2 - 2t\langle u - x, y_t\rangle + t^2\|y_t\|^2 \\ &= \|x - u\|^2 - (2t/\lambda)\langle x^*, y_t\rangle + t^2\|y_t\|^2\end{aligned}$$

shows that $\limsup_{t\searrow 0}\langle x^*, y_t\rangle \leqslant 0$. (Otherwise $x + ty_t$ would be closer to u than x for small t.)

Therefore (recall that $\|x^*\| = 1$)

$$\begin{aligned}d^-\rho(S, x; h) &= \liminf_{t\searrow 0} t^{-1}\rho(S, x + th) = \liminf_{t\searrow 0} t^{-1}\|x + th - (x + ty_t)\| \\ &\geqslant \liminf_{t\searrow 0}\langle x^*, h - y_t\rangle \geqslant \langle x^*, h\rangle\end{aligned}$$

and the proof has been completed.

3. Connection with other derivative-like objects. The following proposition is very easy to prove.

PROPOSITION 4. *Assume that f is strictly differentiable at z, that is, there is x^* such that*

$$\limsup_{\substack{x\to z\\ t\searrow 0}} t^{-1}|f(x + th) - f(x) - t\langle x^*, h\rangle| = 0.$$

Then $\partial_a f(z) = \{x^*\}$.

PROPOSITION 5. *Let f be a convex function. Then*

$$\partial_a f(z) = \partial f(z) = \{x^* \mid f(x) - f(z) \geqslant \langle x^*, x - z\rangle, \forall x\}$$

is the subdifferential of f at z in the sense of convex analysis.

PROOF. If $x^* \in \partial f(z)$, then, obviously, $\langle x^*, h\rangle \leqslant d^-f(z; h)$ for all h so that $x^* \in \partial^-f(z) \subset \partial_a f(z)$.

Conversely, let $x^* \in \partial_a f(z)$ which means that there are sequences $\{x_n\}$ and $\{x_n^*\}$ such that $x_n \to z$, $f(x_n) \to f(z)$, $x_n^* \in \partial^-f(x_n)$ and $x_n^* \to x^*$.

It is obvious that $x_n^* \in \partial f(x_n)$ (since the lower directional derivative is not greater than the directional derivative). In other words, $f(x) - f(x_n) \geqslant \langle x_n^*, x - x_n\rangle$ for all x and n. Therefore $f(x) - f(z) \geqslant \langle x^*, x - z\rangle$ for all x. Q.E.D.

There are several ways to show that approximate subdifferentials are not larger than corresponding Clarke's generalized gradients. This will follow, for instance, from the minimality properties to be established in §8. Here we shall use a more direct approach which is of an independent interest.

Recall that Clarke's tangent cone to S at $z \in S$ is defined as follows:

$$T_c(S, z) = \liminf_{\substack{x \to z \\ x \in S \\ t \searrow 0}} t^{-1}(S - x).$$

In other words, $h \in T_c(S, z)$ if and only if for any sequence $\{x_n\} \subset S$ converging to z there are $t_n \searrow 0$ and $u_n \to h$ such that $x_n + t_n u_n \in S$ for all n.

The following well-known formula (see [**1**]) reveals the basic connection between contingent and Clarke's tangent cones (to a closed set):

$$T_c(S, z) = \liminf_{\substack{x \to z \\ x \in S}} T(S, x) = \liminf_{\substack{x \to z \\ x \in S}} \overline{\operatorname{conv}}\, T(S, x), \tag{3.1}$$

where $\overline{\operatorname{conv}}$ stands for closed convex hull.

We recall further that Clarke's normal cone is defined as the polar of $T_c(S, z)$, $N_c(S, z) = T_c^0(S, z)$, and Clarke's generalized gradient of a function f at z (if $|f(z)| < \infty$) is

$$\partial_c f(z) = \{x^* \mid (-1, x^*) \in N_c(\operatorname{epi} f, (f(z), z))\}.$$

The support function of $\partial_c f(z)$ has been found by Rockafellar [**19**] in the following form (in the case when f is lower semicontinuous):

$$d^{\uparrow} f(z; h) = \sup_{\varepsilon > 0} \limsup_{\substack{x \to z \\ f(x) \to f(z) \\ t \searrow 0}} \inf_{\|u - h\| < \varepsilon} t^{-1}(f(x + th) - f(x)).$$

(We have slightly changed Rockafellar's original notation to adjust it to the accepted style here.)

PROPOSITION 6. *Let f be l.s.c. at z. Then*

$$d^{\uparrow} f(z; h) = \sup_{\varepsilon > 0} \limsup_{\substack{x \to z \\ f(x) \to f(z)}} \inf_{\|u - h\| < \varepsilon} d^- f(x; u).$$

PROOF. If $(\alpha_n, x_n) \in \operatorname{epi} f$ and $\alpha_n \to f(z)$, $x_n \to z$, then $f(x_n) \to f(z)$ since f is l.s.c. On the other hand, for $(\alpha, x) \in \operatorname{epi} f$

$$T(\operatorname{epi} f, (\alpha, x)) \supset T(\operatorname{epi} f, (f(x), x))$$

(provided, of course, that $|f(x)| < \infty$). According to (3.1), it follows that

$$T_c(\operatorname{epi} f, (f(z), z)) = \liminf_{\substack{x \to z \\ f(x) \to f(z)}} T(\operatorname{epi} f, (f(x), x)).$$

By definition

$$T_c(\operatorname{epi} f, (f(z), z)) = \{(\alpha, h) \mid \alpha \geqslant d^{\uparrow} f(z; h)\}$$

and (see the proof of Proposition 2)

$$T(\operatorname{epi} f, (f(x), x)) = \{(\alpha, h) \mid \alpha \geqslant d^- f(x; h)\}.$$

Thus $\alpha \geqslant d^{\uparrow} f(z; h)$ if and only if for any sequence $\{x_k\}$ such that $x_k \to z$, $f(x_k) \to f(z)$, there are sequences $\alpha_k \to \alpha$ and $h_k \to h$ such that $\alpha_k \geqslant d^{-} f(x_k; h_k)$. This amounts to saying that for any $\varepsilon > 0$

$$\alpha + \varepsilon \geqslant \limsup_{k \to \infty} \left(\inf_{\|u-h\|<\varepsilon} d^{-} f(x_k; u) \right).$$

This is true for any sequence $\{x_k\}$ such that $x_k \to z$, $f(x_k) \to f(z)$. Therefore $\alpha \geqslant d^{\uparrow} f(z; h)$ if and only if for any $\varepsilon > 0$

$$\alpha + \varepsilon \geqslant \limsup_{\substack{x \to z \\ f(x) \to f(z)}} \left(\inf_{\|u-h\|<\varepsilon} d^{-} f(x; u) \right)$$

which yields the desired equality.

The inclusion $\partial_a f(z) \subset \partial_c f(z)$ is a corollary of the proposition (though not absolutely trivial). The following lemma communicated to me by Aubin and Wets offers another way to prove the inclusion.

LEMMA 2 (AUBIN AND WETS). *Let T_n be a sequence of closed convex cones in X. Then* $\overline{\operatorname{conv}}(\limsup T_n^0) = (\liminf T_n)^0$.

PROOF. If $x^* \in \limsup T_n^0$ and $x \in \liminf T_n$, then (by definition) there are $x_n \in T_n$, a sequence $\{n_k\}$ of indices and $x_{n_k}^* \in T_{n_k}^0$ such that $x = \lim x_n$ and $x^* = \lim x_{n_k}$. Therefore $\langle x^*, x \rangle = \lim \langle x_{n_k}^*, x_{n_k} \rangle \leqslant 0$ so that $x^* \in (\liminf T_n)^0$. This proves that the left side of the equality belongs to the right one.

To prove the inverse inequality we shall show that

$$\left(\limsup T_n^0\right)^0 \subset \liminf T_n.$$

Let x be such that $\langle x^*, x \rangle \leqslant 0$ for all $x^* \in \limsup T_n^0$, and let x_n be the orthogonal projection of x onto T_n. Then $\|x_n\| \leqslant \|x\|$ (since T_n is a cone), $x_n^* = x - x_n \in T_n^0$ and $\langle x_n^*, x_n \rangle = 0$ (since T_n^0 is a convex cone). Let $\{x_{n_k}\}$ be a converging subsequence of $\{x_n\}$ and u the limit of the subsequence. Then $u^* = x - u$ belongs to $\limsup T_n$ so that $\langle u^*, x \rangle \leqslant 0$. On the other hand, the equality

$$\|x_{n_k}^*\| = \langle x_{n_k}^*, x - x_{n_k} \rangle = \langle x_{n_k}^*, x \rangle$$

implies that $\|u^*\| = \langle u^*, x \rangle = 0$. Thus $u^* = x - u = 0$ and we conclude that any converging subsequences of $\{x_n\}$ converges to x. This means that $x = \lim x_n$ and hence $x \in \liminf T_n$.

We note that the lemma is no longer true in the infinite dimensional case (though the inclusion $\overline{\operatorname{conv}}(\limsup T_n^0) \subset (\liminf T_n)^0$ is always true).

THEOREM 2. *For any closed $S \subset X$ and any $z \in S$, we have*

$$N_c(S, z) = \overline{\operatorname{conv}}\, N_a(S, z) = \overline{\operatorname{conv}} \left(\limsup_{\substack{x \to z \\ x \in S}} T^0(S, x) \right).$$

For any l.s.c. function f and any z such that $|f(z)| < \infty$ we have $\partial_a f(z) \subset \partial_c f(z)$, and $\partial_c f(z) \neq \varnothing \Rightarrow \partial_a f(z) \neq \varnothing$. If, in addition, f is Lipschitz near z, then $\overline{\operatorname{conv}}\, \partial_a f(z) = \partial_c f(z)$.

PROOF. The first equality follows at once from Proposition 1, Lemma 2 and (3.1). The inclusion $\partial_a f(z) \subset \partial_c f(z)$ is an immediate consequence of the equality, as well as the second part of the second statement (see Proposition 2).

So let us assume that f is Lipschitz near z with constant $k > 0$. Then the set $\{(\alpha, h) \mid \alpha \geqslant k\|h\|\}$ belongs to Clarke's tangent cone to epi f at $(f(z), z)$. It is an easy matter to derive from here that $\beta \geqslant \|x^*\|/k$ when $(-\beta, x^*) \in N_c(\text{epi } f, (f(x), x))$.

Let $x^* \in \partial_c f(z)$. Then $(-1, x^*) \in N_c(\text{epi } f, (f(z), z))$ and (by a famous theorem of Carathéodory) there are $(-\beta_i, u_i^*) \in N_a(\text{epi } f, (f(z), z))$, $i = 1, \ldots, r \leqslant \dim X$, such that $\sum \beta_i = 1$ and $\sum u_i^* = x^*$. We may assume that all β_i are positive (because $u_i^* = 0$ if $\beta_i = 0$). Then $x_i^* = u_i^*/\beta_i \in \partial_a f(z)$ by Proposition 2 and $x^* = \sum \beta_i x_i^*$. Q.E.D.

It is very easy to make up an example to show that the convex closure of $\partial_a f(z)$ can be smaller than $\partial_0 f(z)$ and that $\partial_a f(z)$ need not be convex even in the Lipschitz case.

For instance, let us consider the following functions of one variable:

$$f(x) = \begin{cases} 0, & \text{if } x \geqslant 0, \\ (-x)^{1/2}, & \text{if } x < 0. \end{cases}$$

Then $\partial_a f(0) = \{0\}$ while $\partial_c f(0) = (-\infty, 0]$. On the other hand, for the function $f(x) = -|x|$ we have $\partial_a f(0) = \{1, -1\}$, $\partial_c f(0) = [-1, 1]$. The last example can be generalized in the following way.

PROPOSITION 7. *Let f be a concave continuous function. Then*

$$\partial_a f(z) = \limsup_{x \to z} \{\nabla f(x)\}.$$

(Here, as usual, $\nabla f(x)$ denotes the gradient of f at x. We note that a concave continuous function is almost everywhere differentiable due to a theorem of Rademacher.)

PROOF. A concave continuous function is locally Lipschitz and directionally derivable at every point. Therefore $d^- f(x; h)$ coincides with the directional derivative of f at x. This is a concave function of h and the inequality $d^- f(x; h) \geqslant \langle x^*, h \rangle$ may be valid for all h if and only if $d^- f(x; h)$ is linear as a function of h in which case $\partial^- f(x)$ is a singleton. In other words, $\partial^- f(x) \neq \varnothing$ if and only if f is (Gâteaux) differentiable at x and in this case $\partial^- f(x) = \{\nabla f(x)\}$.

The last theorem to be proved in this section implies a comparison theorem approximate subdifferentials and derivate containers of Warga [**20**].

THEOREM 3. *Let $\{f_n(x)\}$ be a sequence of C^2-functions uniformly converging to $f(x)$ on a neighborhood of z. Then*

$$\partial_a f(z) \subset \limsup_{\substack{x \to z \\ n \to \infty}} \{\nabla f_n(x)\}.$$

PROOF. Denote by W the set in the right-hand side of the inclusion

$$W = \bigcap_{n=1}^{\infty} \bigcap_{\delta>0} \overline{\bigcup_{\substack{m \geqslant n \\ \|x-z\| \leqslant \delta}}} \{\nabla f_m(x)\}.$$

Take an $x^* \notin W$. We have to prove that $x^* \notin \partial_a f(z)$. We choose $\delta > 0$, $\varepsilon > 0$ and n such that $\{f_m\}$ converge uniformly on the δ-ball around z and

$$\|x^* - \nabla f_m(x)\| \geqslant \varepsilon \quad \text{if } m \geqslant n, \|x - z\| \leqslant \delta \tag{3.2}$$

(which is possible since $x^* \notin W$). Let $m \geqslant n$ and $\|u - z\| \leqslant \delta/2$ and consider the differential equation

$$\dot{x} = \frac{x^* - \nabla f_m(x)}{\|x^* - \nabla f_m(x)\|} = \varphi_m(x), \qquad x(0) = u. \tag{3.3}$$

Then $\varphi_m(x)$ is a C^1-mapping into X defined on the $\delta/2$-ball about u and $\|\varphi_m(x)\| = 1$ thereon. It follows that (3.3) has a unique solution $x_m(t)$ defined at least on $[0, \delta/2]$ and satisfying the inequality

$$\|x_m(t) - u\| \leqslant t \quad \text{for } t \in [0, \delta/2]. \tag{3.4}$$

We have (using (3.2) and (3.4))

(3.5)

$$\begin{aligned} f_m(x_m(t)) - f_m(u) - \langle x^*, x_m(t) - u \rangle &= \int_0^t \langle \nabla f_m(x_m(\tau)) - x^*, \dot{x}_m(\tau) \rangle \, d\tau \\ &= -\int_0^t \|x^* - \nabla f_m(x_m(\tau))\| d\tau \leqslant -\varepsilon t \leqslant -\varepsilon \|x_m(t) - u\|. \end{aligned}$$

The sequence $\{x_m(\cdot)\}$ is precompact in the space of continuous maps from $[0, \delta/2]$ into X (because $\|\dot{x}_m(t)\| = 1$) and, taking, if necessary, a subsequence, we may view the sequence as converging uniformly to some $x(\cdot)$. For the latter the following inequality immediately results from (3.5):

$$f(x(t)) - f(u) - \langle x^*, x(t) - \rangle \leqslant -\varepsilon t \leqslant -\varepsilon \|x(t) - u\|.$$

The left inequality in (3.6) shows that $x(t) \neq u$ for all $t \in (0, \delta/2]$. Therefore $x(t) = u + \lambda(t) v(t)$, where $\lambda(t) = \|x(t) - u\| \to 0$ as $t \to 0$ and $\|v(t)\| = 1$. It follows that there are sequences $t_k \to 0$ and $\lambda_k \to 0$ such that $v_k = v(t_k)$ converge to some h, $\|h\| = 1$ and $\lambda_k = \lambda(t_k)$. Then (3.6) implies the inequality

$$\begin{aligned} d^- f(u; h) &\leqslant \liminf_{k \to \infty} \lambda_k^{-1} (f(u + \lambda_k v_k) - f(u)) \\ &= \liminf_{k \to \infty} (f(x(t_k)) - f(u)) / \|x(t_k) - u\| \leqslant \langle x^*, h \rangle - \varepsilon \end{aligned}$$

showing that the distance from x^* to $\partial^- f(u)$ is not less than ε. This is true for any u of a neighborhood of z so that x^* cannot belong to $\partial_a f(z)$.

4. Approximate subdifferentials of a sum. In this section we prove the following

THEOREM 4. *Let the functions f_1 and f_2 be lower semicontinuous near z and such that*

$$\operatorname{dom}(d^\uparrow f_1(z; \cdot)) - \operatorname{dom}(d^\uparrow f_2(z; \cdot)) = X. \tag{4.1}$$

Then

$$\partial_a(f_1+f_2)(z) \subset \partial_a f_1(z) + \partial_a f_2(z). \tag{4.2}$$

This is the first theorem of calculus of approximate subdifferentials. As in certain other theories (e.g. convex analysis or analysis of C.g.g.'s), this result will be most instrumental both in theory and applications.

We begin the proof with the following important lemma which is a particular case of a more general result proved in [**10**].

LEMMA 3. *If the functions f_1 and f_2 are lower semicontinuous near z, then for any $\delta > 0$*

$$\delta^-(f_1+f_2)(z) \subset \bigcup_{x_i \in U(f_i, z, \delta)} (\partial^- f_1(x_1) + \partial^- f_2(x_2) + \delta B),$$

where B is the unit ball in X and $U(f, z, \delta) = \{x \mid \|x-z\| < \delta, |f(x)-f(z)| < \delta\}$.

PROOF. The proof of the lemma is based on the penalty function method mentioned in the Introduction. If $\partial^-(f_1+f_2)(z) = \varnothing$, there is nothing to prove. So assume that $x^* \in \partial^-(f_1+f_2)(z)$. Replacing $f_1(x)$ by $f_1(x) - \langle x^*, x-z\rangle$, if necessary, we reduce the problem to the case $x^* = 0$. We have, thus, to show that

$$\bigcup_{x_i \in U(f_i, z, \delta)} (\partial^- f_1(x_1) + \partial^- f_2(x_2)) \cap \delta B \neq \varnothing. \tag{4.3}$$

As follows from Lemma 1, the function $g_\delta(x) = f_1(x) + f_2(x) + \delta\|x-z\|$ attains a strict local minimum at z. We choose an $\varepsilon > 0$ in such a way that $\varepsilon < \delta$ and $g_\delta(x) > g_\delta(z)$, $f_i(x) \geqslant f_i(z) - 1/2$ if $\|x-z\| \leqslant \varepsilon$, $x \neq z$ (which is possible since f_i are l.s.c.), and consider the function

$$p_r(u, v, x) = f_1(u) + f_2(v) + (r/2)\left(\|u-x\|^2 + \|v-x\|^2\right) + \delta\|x-z\|.$$

Let u_r, v_r, x_r be such that $\|u_r - z\| \leqslant \varepsilon$, $\|v_r - z\| \leqslant \varepsilon$, $\|x_r - z\| \leqslant \varepsilon$ and

$$p_r(u_r, v_r, x_r) = \min\left\{p(u, v, x) \mid \|u-z\| \leqslant \varepsilon, \|v-z\| \leqslant \varepsilon, \|x-z\| \leqslant \varepsilon\right\}.$$

We have

$$\begin{aligned} &f_1(z) + f_2(z) + r\left(\|u_r - x_r\|^2 + \|v_r - x_r\|^2\right) + \delta\|x_r - z\| \\ &\quad \leqslant f_1(u_r) + f_2(v_r) + 1 + r\left(\|u_r - x_r\|^2 + \|v_r - x_r\|^2\right) + \delta\|x_r - z\| \\ &\quad = p_r(u_r, v_r, x_r) + 1 \leqslant p_r(z, z, z) + 1 = f_1(z) + f_2(z) + 1. \end{aligned} \tag{4.4}$$

It follows that $r(\|u_r - x_r\|^2 + \|v_r - x_r\|^2) + \delta\|x_r - z\| \leqslant 1$; in particular, the sequences $\{u_r\}$, $\{v_r\}$ and $\{x_r\}$ are bounded and we may assume that they converge to some $\bar{u}$, $\bar{v}$ and $\bar{x}$ such that $\|\bar{x} - z\| \leqslant \varepsilon$, $\|\bar{v} - z\| \leqslant \varepsilon$ and $\|\bar{u} - z\| \leqslant \varepsilon$. The inequality above shows that $\bar{x} = \bar{u} = \bar{v}$.

It follows from (4.4) (in view of the fact that p_r is l.s.c.) that

$$g_\delta(\bar{x}) = f_1(\bar{x}) + f_2(\bar{x}) + \delta\|\bar{x} - z\| \leqslant f_1(z) + f_2(z) = g_\delta(z)$$

which may be true only if $\bar{x} = z$, due to the choice of ε. Thus, $x_r \to z$, $u_r \to z$, $v_r \to z$ and (4.4) implies (as far as f_1 and f_2 are l.s.c.) that $f_1(u_u) \to f_1(z)$, $f_2(v_r) \to f_2(z)$. In

the rest of the proof we assume r so large that $\|x_r - z\| < \varepsilon$, $\|u_r - z\| < \varepsilon$, $\|v_r - z\| < \varepsilon$, $|f_1(u_r) - f_1(z)| < \delta$ and $|f_2(v_r) - f_2(z)| < \delta$ so that, in particular, p_r attains an unconditional local minimum at (u_r, v_r, x_r).

If so then the Dini derivative of p_r at (u_r, v_r, x_r) is nonnegative for all values of its argument. The latter is equivalent to the following inequalities where we have set $u_r^* = -r(u_r - x_r)$ and $v_r^* = -r(v_r - x_r)$ (we recall that $\|\cdot\|$ is the Euclidean norm):

$$\begin{aligned} &d^- f_1(u_r; u) - \langle u_r^*, u\rangle \geqslant 0, && \forall u;\\ &d^- f_2(v_r; v) - \langle v_r^*, v\rangle \geqslant 0, && \forall v;\\ &u_r^*, x + \langle v_r^*, x\rangle + \delta\|x\| \geqslant 0, && \forall x. \end{aligned}$$

The first two relations show that $u_r^* \in \partial^- f_1(u_r)$ and $v_r^* \in \partial^- f_2(v_r)$ and the third that $\|u_r^* + v_r^*\| \leqslant \delta$, that is, that $u_r^* + v_r^* \in \delta B$. As long as $\|u_r - z\| < \delta$, $\|v_r - z\| < \delta$, $|f_1(u_r) - f_1(z)| < \delta$ and $|f_2(v_r) - f_2(z)| < \delta$, this completes the proof.

PROOF OF THEOREM 4. For notational simplicity we denote $Q_i = \operatorname{dom}(d^\uparrow f_i(z; \cdot))$. Let L_i be the linear hull of Q_i. It follows from (4.1) that the relative interiors of Q_1 and Q_2, denoted by $\operatorname{ri} Q_1$ and $\operatorname{ri} Q_2$, meet each other. Otherwise we could separate the sets, which are convex cones according to **[19]**, by a linear functional in contradiction with (4.1). It is also clear that $L_1 + L_2 = X$.

Take an $x^* \in \partial_a(f_1 + f_2)(z)$ (for there is nothing to prove if $\partial_a(f_1 + f_2)(z) = \varnothing$). Then there are sequences $\{x_n\}$ and $\{x_n^*\}$ such that $x_n \to z$, $x_n^* \to x^*$, $(f_1 + f_2)(x_n) \to (f_1 + f_2)(z)$ and $x_n^* \in \partial^-(f_1 + f_2)(x_n)$. By Lemma 3 there are u_n, v_n, u_n^* and v_n^* such that $\|x_n - u_n\| \to 0$, $\|x_n - v_n\| \to 0$, $f_1(u_n) - f_1(x_n) \to 0$, $f_2(v_n) - f_2(x_n) \to 0$, $u_n^* \in \partial^- f_1(u_n)$, $v_n^* \in \partial^- f_2(v_n)$ and $\|x_n^* - u_n^* - v_n^*\| \to 0$.

The theorem will be proved if we show that one of the sequences u_n^* or v_n^* is bounded. Indeed, in this case the other is also bounded and for any accumulation points u^*, v^* of the sequences we would have $u^* \in \partial_a f_1(z)$, $v^* \in \partial_a f_2(z)$. (We note that $f_i(x_n) \to f_i(z)$ because the sum converges and the functions are l.s.c.) We shall also have $u^* + v^* = x^*$ and therefore $x^* \in \partial_a f_1(z) + \partial_a f_2(z)$.

We shall prove that, say, the sequence u_n^* is bounded if we show that for any $h \in X$ there is a sequence of $h_n \to h$ such that $\limsup\langle u_n^*, h_n\rangle$ is finite. (Indeed, assume the contrary: $\|u_n^*\| \to \infty$. Let us set $q_n = u_n^*/\|u_n^*\|$, and let, for simplicity, h be the limit of $\{q_n\}$. (Otherwise we shall take a subsequence.) Then, for any sequence $h_n \to h$, we have, setting $\gamma_n = \|h_n - q_n\| \to 0$.

$$\langle u_n^*, h_n\rangle = \langle u_n^*, q_n\rangle + \langle u_n^*, h_n - q_n\rangle \geqslant \|u_n^*\| - \gamma_n\|u_n^*\| \to \infty.)$$

In fact, this may be proved only for $h \in Q_1 \cup Q_2$ (because $L_1 + L_2 = X$).

Let $h \in Q_1$. By Proposition 6 there is a sequence $h_n \to h$ such that

$$\limsup_{n\to\infty} \langle u_n^*, h_n\rangle \leqslant \limsup_{n\to\infty} d^- f_1(u_n; h_n) \leqslant d^\uparrow f_1(z; h). \tag{4.5}$$

Likewise, for an $h \in Q_2$ there is a sequence $h_n' \to h$ such that

$$\limsup \langle v_n^*, h_n'\rangle \leqslant d^\uparrow f_2(z; h) < \infty.$$

Taking into account that $u_n^* + v_n^* \to x^*$, we see that

$$\liminf_{n\to\infty} \langle u_n^*, h_n'\rangle = \langle x^*, h\rangle - \limsup_{n\to\infty} \langle v_n^*, h_n'\rangle > -\infty. \tag{4.6}$$

Fix an $\bar{h} \in \operatorname{ri} Q_1 \cap \operatorname{ri} Q_2$. If $h \in L_1$, then there is an $\varepsilon > 0$ such that $\bar{h} \pm \varepsilon h \in Q_1$. According to (4.5) and (4.6) there are sequences $h'_n \to \bar{h}$ and $h''_n \to \bar{h} + \varepsilon h$ such that

$$\varepsilon \cdot \limsup \langle u_n^*, h_n \rangle \leqslant \limsup \langle u_n^*, h''_n \rangle - \liminf \langle u_n^*, h'_n \rangle < \infty,$$

where $h_n = \varepsilon^{-1}(h'_n - h''_n) \to h$. Likewise, for any $h \in L_2$ we show in a similar way that there is a sequence $h'_n \to h$ such that $\limsup \langle v_n^*, h'_n \rangle < \infty$. As above, this implies that $\liminf \langle u_n^*, h'_n \rangle > -\infty$. Thus, given an $h \in L_2$, we can find a sequence $h'_n \to -h$ such that for $h_n = -h'_n \to h$, we have

$$\limsup \langle u_n^*, h_n \rangle = -\liminf \langle u_n^*, h'_n \rangle < \infty.$$

This completes the proof.

COROLLARY 4.1 (CF. [17]). *Let f_1 and f_2 be proper convex functions such that* $\operatorname{ri}(\operatorname{dom} f_1) \cap \operatorname{ri}(\operatorname{dom} f_2) \neq \varnothing$. *Then for any z, $\partial(f_1 + f_2)(z) \subset \partial f_1(z) + \partial f_2(z)$.*

PROOF. Denote by L the linear hull of $(\operatorname{dom} f_1) \cup (\operatorname{dom} f_2)$, and let M be the orthogonal complement of L. If φ is a convex function whose domain belongs to L, then $\partial\varphi(x) = \partial\varphi(x) + M$ for any x. This is true for f_1, f_2 and $f_1 + f_2$. Therefore we need to prove the corollary only the restrictions of f_1 and f_2 to L. This, in turn, means that there is no loss of generality in assuming that $L = X$.

If so, then (4.1) follows from the assumption on relative interiors (because for a convex function, the "arrow" directional derivative of Rockafellar coincides with the closure of the directional derivative [19]). Finally, a convex function subdifferentiable at a point is necessarily l.s.c. at the point and if the sum of two convex functions is l.s.c. at a point, then both functions are also l.s.c., so that there is no loss of generality in assuming the two functions lower semicontinuous. Thus, all we need is to apply Theorem 4 and Proposition 5.

COROLLARY 4.2. *Let Q_1 and Q_2 be closed subsets of X such that $z \in Q_1 \cap Q_2$ and*

$$T_c(Q_1, z) - T_c(Q_2, z) = X. \tag{4.7}$$

Then $N_a(Q_1 \cap Q_2, z) \subset N_a(Q_1, z) + N_a(Q_2, z)$.

PROOF. We set

$$f_i(x) = (Q_i, x) = \begin{cases} 0, & \text{if } x \in Q_i, \\ \infty, & \text{otherwise.} \end{cases}$$

Then $d^{\uparrow} f_i(z; h)$ is the indicator function of $T_c(Q_i, z)$ so that we can apply Theorem 4 to the indicator functions.

COROLLARY 4.3. *Under the assumptions of Corollary* 4.2, $N_c(Q_1 \cap Q_2, z) \subset N_c(Q_1, z) + N_c(Q_2, z)$.

PROOF. It follows from Corollary 4.2 and Theorem 2 that $N_c(Q_1 \cap Q_2, z)$ belongs to the closure of $N_c(Q_1, z) + N_c(Q_2, z)$ and we only need to show that the sum is a closed set. But this follows from a general fact of finite dimensional convex analysis: if convex cones K_1 and K_2 are such that $\operatorname{ri} K_1 \cap \operatorname{ri} K_2 \neq \varnothing$ (which is the case if $K_1 - K_2 = X$), then $(K_1 + K_2)^0 = K_1^0 + K_2^0$.

COROLLARY 4.4. *Let the functions f_1 and f_2 satisfy the assumptions of Theorem* 4. *Then* $\partial_c(f_1 + f_2)(z) \subset \partial_c f_1(z) + \partial_c f_2(z)$.

PROOF. Denote by U the space of all triples (α, β, x), $\alpha, \beta \in R$, $x \in X$, and let

$$Q_1 = \{u = (\alpha, \beta, x) \mid \alpha \geqslant f_1(x)\}, \quad Q_2 = \{u = (\alpha, \beta, x) \mid \beta \geqslant f_2(x)\},$$
$$Q = \{u = (\alpha, \beta, x) \mid \alpha + \beta \geqslant f_1(x) + f_2(x)\}.$$

Since f_1 and f_2 are l.s.c., the three sets are closed. It is obvious that $Q_1 \cap Q_2 \subset Q$. We denote $\bar{u} = (f_1(z), f_2(z), z)$. Then $\bar{u} \in Q_1 \cap Q_2$ and, as one can easily verify

$$T_c(Q_1 \cap Q_2, \bar{u}) \subset T_c(Q, \bar{u}). \tag{4.8}$$

We shall prove next that Q_1, Q_2, $\bar{u}$ satisfy (4.7). Take an arbitrary $u = (\alpha, \beta, x) \in U$. Since (4.1) is valid for f_1 and f_2, there are $h_i \in \operatorname{dom}(d^\uparrow f_i(z; \cdot))$ such that $h_1 - h_2 = x$. Then

$$u_1 = \left(d^\uparrow f_1(z; h_1), d^\uparrow f_2(z; h_2) + \beta, h_1\right) \in T_c(Q_1, \bar{u}),$$
$$u_2 = \left(d^\uparrow f_1(z; h_1) - \alpha, d^\uparrow f_2(z; h_2), h_2\right) \in T_c(Q_2, \bar{u})$$

and $u_1 - u_2 = (\alpha, \beta, h_1 - h_2) = u$.

By virtue of (4.8) and Corollary 4.2,

$$T_c^0(Q, \bar{u}) \subset N_c(Q_1, \bar{u}) + N_c(Q_2, \bar{u}). \tag{4.9}$$

It is obvious that

$$\begin{aligned} N_c(Q_1, \bar{u}) &= \{(\lambda, 0, x^*) \mid (\lambda, x^*) \in N_c(\operatorname{epi} f_1, (f_1(z), z))\}, \\ N_c(Q_2, \bar{u}) &= \{(0, \lambda, x^*) \mid (\lambda, x^*) \in N_c(\operatorname{epi} f_2, (f_2(z), z))\}. \end{aligned} \tag{4.10}$$

Finally,

$$T_c^0(Q, \bar{u}) = \{(\lambda, \lambda, x^*) \mid (\lambda, x^*) \in N_c(\operatorname{epi} f, (f(z), z))\}$$

(where we have set $f = f_1 + f_2$). Indeed, by definition

$$T_c(Q, \bar{u}) = \{(\alpha, \beta, h) \mid (\alpha + \beta, h) \in T_c(\operatorname{epi} f, (f(z),))\}$$

and, on the other hand, the polar to the cone $\{(\alpha, \beta, h) \mid (\alpha + \beta, h) \in K\}$ (where $K \subset R \times X$) is $\{(\lambda, \lambda, x^*) \mid (\lambda, x^*) \in K^0\}$.

Let $x^* \in \partial_c(f_1 + f_2)(z)$. Then $(-1, -1, x^*) \in T_c^0(Q, \bar{u})$ and, as follows from (4.8) and (4.9), there are x_i^*, $i = 1, 2$, such that $(-1, x_i^*) \in N_c(\operatorname{epi} f, (f_i(z), z))$ and $x^* = x_1^* + x_2^*$. Q.E.D.

COROLLARY 4.5. *If the functions f_1 and f_2 are finite at z and satisfy the assumptions of Theorem* 4, *then*

$$\partial_a(f_1 \vee f_2)(z) \subset \bigcup_{\substack{\alpha_1 \geqslant 0,\, \alpha_2 \geqslant 0 \\ \alpha_1 + \alpha_2 = 1}} \{x_1^* + x_2^* \mid (-\alpha_i, x_i^*) \in N_a(\operatorname{epi} f_i, (f_i(z), z)), i = 1, 2\}.$$

PROOF. Apply Corollary 4.2 to $Q_1 = \operatorname{epi} f_1$ and $Q_2 = \operatorname{epi} f_2$ having in mind that $\operatorname{epi}(f_1 \vee f_2) = (\operatorname{epi} f_1) \cap (\operatorname{epi} f_2)$ and (see the proof of the preceding corollary) $T_c(Q_i, (f_i(z), z)) = \{(\lambda, h) \mid \lambda \geqslant d^\uparrow f_i(z; h)\}$.

A stronger result for the maximum function will be proved in the next section in the case when the functions f_1 and f_2 are Lipschitz near z.

COROLLARY 4.6. *If, under the assumptions of the preceding corollary,* $\partial_c f_1(z) \neq \varnothing$ *and* $\partial_c f_2(z) \neq \varnothing$, *then*

$$\partial_c(f_1 \vee f_2)(z) \subset \bigcup_{\substack{\alpha_1 \geqslant 0^+,\, \alpha_2 \geqslant 0^+ \\ \alpha_1 + \alpha_2 = 1}} (\alpha_1 \partial_c f_1(z) + \alpha_2 \partial_c f_2(z)),$$

where

$$0^+ \partial_c f(x) = \{x^* \mid u^* + tx^* \in \partial_c f(x), \forall u^* \in \partial_c f(x), \forall t > 0\}.$$

PROOF. This time we apply Corollary 4.3 to $Q_1 = \operatorname{epi} f_1$, $Q_2 = \operatorname{epi} f_2$. If $x^* \in \partial_c(f_1 \vee f_2)(z)$, then $(-1, x^*) \in N_c(Q_1 Q_2, ((f_1 \vee f_2)(z), z))$ and (Corollary 4.3) there are $(-\alpha_i, u_i^*) \in N_c(Q_i, (f_i(z), z))$ such that $(-1, x^*) = (-\alpha_1, u_1^*) + (-\alpha_2, u_2^*)$. It is obvious that $\alpha_i \geqslant 0$, $i = 1, 2$. If, say, $\alpha_1 > 0$, then we set $x_1^* = u_1^*/\alpha_1$ so that $(-1, x_1^*) \in N_c(Q_1, (f_1(z), z))$ and $x_1^* \in \partial_c f_1(z)$ or, equivalently, $u_1^* \in \alpha_1 \partial_c f_1(z)$. If one of α_i, say, α_2, equals zero, then $(0, u_2^*) \in N_c(Q_2, (f_2(z), z))$ which is the same as $u_2^* \in 0^+ \partial_c f_2(z)$ in view of the fact that $\partial_c f_2(z) \neq \varnothing$ and $N_c(Q_2, (f_2(z), z))$ is a convex cone. Q.E.D.

REMARK 4.1. Corollary 4.1 has been included to show that Theorem 4 is sufficiently powerful to imply the strongest version of the corresponding theorem of finite dimensional convex analysis [**17**, Theorem 16.4]. As to Theorem 4 itself and other corollaries, they are new as well as the results to be proved in the next section. We do not assume, any more, one of the functions Lipschitz (as in the first announcement [**8**]) or even directionally Lipschitz (as in Rockafellar's paper [**18**] devoted to calculus of C.g.g.'s or in a forthcoming paper by Kruger [**12**] where a version of approximate subdifferentials is cnsidered). As a matter of fact, neither of the functions is even required to be continuous (or to have the arrow derivative continuous at some point). But the latter is of course possible only in the finite dimensional situation.

REMARK 4.2. It may seem natural to ask under what additional assumptions the inclusions become equalities. A possible set of such assumptions may consist of equalities $d^\uparrow f_i(z; h) = d^- f_i(z; h)$ in the case of functions and $T_c(Q_i, z) = T(Q_i, z)$ in the case of sets (cf. [**4**, **18**]). The proof of the inverses inclusions in this case is almost trivial and follows from the obvious relations

$$d^-(f_1 + f_2)(x; h) \geqslant d^- f_1(x; h) + d^- f_2(x; h),$$
$$T(Q_1 \cap Q_2, x) \subset T(Q_1, x) \cap T(Q_2, x).$$

5. Approximate coderivatives and approximate subdifferentials of a composition. Consider a set-valued map F from X into Y, let $\operatorname{Graph} F = \{(x, y) \mid y \in F(x)\}$ denote the graph of F, and let $w \in F(z)$.

DEFINITION 3. The set-valued map

$$y^* \to D_a^* F(z, w)(y^*) = \{x^* \in X^* \mid (x^*, -y^*) \in N_a(\operatorname{Graph} F, (z, w))\}$$

will be called the *approximate coderivative* of F at (z, w).

We shall also consider the set-valued maps (see [**1**, **2**])

$$y^* \to D_c^* F(z,w)(y^*) = \{x^* \in X^* \mid (x^*, -y^*) \in N_c(\text{Graph } F, (z,w))\}$$

and

$$h \to D^- F(z,w)(h) = \{y \mid (h, y) \in T(\text{Graph } F, (z,w))\}$$

that will be called, respectively, Clarke's coderivative and contingent derivative of F at (z, w). For a single-valued F, we write $D_a^* F(z)$ instead of $D_a^* F(z, F(z))$ etc.

THEOREM 5. *Assume that*:

(i) *F is a continuous single-valued map from a neighborhood of $z \in X$ into Y*;

(ii) *$g(y)$ is a lower semicontinuous function assuming a finite value at $w = F(z)$*;

(iii) $\text{dom}(d^\uparrow g(w; \cdot)) - K_Y = Y$, *where K_Y is the projection of Clarke's tangent cone to* Graph *F at (z, w) onto Y.*

Then for $f(x) = (g \circ F)(x) = g(F(x))$, we have

$$\partial_a f(z) \subset \bigcup_{y^* \in \partial_a g(w)} D_a^* F(z)(y^*) = D_a^* F(z) \circ \partial_a g(w).$$

PROOF. For notational convenience we denote by $\lambda(x, y)$ the function on $X \times Y$ equal to $g(y)$; let $\mu(x, y)$ denote the indicator function of the graph of F (i.e. $\mu(x, y) = 0$ if $y = F(x)$ and $\mu(x, y) = \infty$ if $y \neq F(x)$). Finally, let $\varphi(x, y) = \lambda(x, y) + \mu(x, y)$. Then

$$\begin{aligned} \varphi(x, y) &= f(x), && \text{if } y = F(x), \\ \varphi(x, y) &= \infty \geqslant f(x), && \text{if } y \neq F(x). \end{aligned}$$

This means that $d^-\varphi((x, y); (h, v)) \geqslant d^- f(x; h)$ when $y = F(x)$ or, equivalently, that $\partial^- f(x) \times \{0\} \subset \partial^- \varphi(x, y)$ when $y = F(x)$. The latter implies that $\partial_a f(z) \times \{0\} \subset \partial_a \varphi(z, w)$.

The function $\varphi(x, y)$ is the sum of two l.s.c. functions satisfying the condition of Theorem 4 by virtue of (iii). Indeed,

$$d^\uparrow \lambda((z,w); (h, v)) = d^\uparrow g(z; h), \quad \forall h, v,$$

so that $\text{dom } d^\uparrow \lambda = X \times \text{dom } d^\uparrow g$, while $d^\uparrow \mu$ is the indicator function of Clarke's tangent cone to Graph F at (z, w). For any given $(x, y) \in X \times Y$ we can find (using (iii)) a $v' \in \text{dom}(d^\uparrow g(w; \cdot))$ and $(u'', v'') \in \text{dom}(d^\uparrow \mu((z,w); \cdot))$ such that $v' - v'' = y$. But $(u, v') \in \text{dom}(d^\uparrow \lambda((z,w); \cdot))$ for any u, in prticular for $u' = x + u''$. Therefore we conclude that $\text{dom}(d^\uparrow \lambda((z,w); \cdot)) - \text{dom}(d^\uparrow \mu((z,w); \cdot)) = X \times Y$.

By Theorem 4 $\partial_a \varphi(z,w) \subset \partial_a \lambda(z,w) + \partial_a \mu(z,w)$. It is an easy matter to verify that $\partial_a \lambda(z,w) = \{0\} \times \partial_a g(w)$, and $\partial_a \mu(z,w)$ is by definition the normal cone of Clarke to Graph F at (z, w).

Thus, for any $x^* \in \partial_a f(z)$ there are $y^* \in \partial_a g(w)$ and $(u^*, v^*) \in N_a(\text{Graph } F, (z,w))$ such that $(x^*, 0) = (0, y^*) + (u^*, v^*)$. It follows that $u^* = x^*$, $v^* = -y^*$ and therefore $(x^*, -y^*) \in N_a(\text{Graph } F, (z,w))$ which is the same as $x^* \in D_a^* F(z)(y^*)$. Q.E.D.

COROLLARY 5.1. *Let $Q \subset Y$ be a closed set and G a continuous map from a neighborhood of $z \in X$ into Y such that $w = G(z) \in Q$. We assume that $T_c(Q, w) - K_Y = Y$ where K_Y is the projection of Clarke's tangent cone to* Graph F *at (z, w) onto Y.*

Then for $P = G^{-1}(Q) = \{x \mid G(x) = 0\}$, we have

$$N_a(P, z) \subset \bigcup_{y^* \in N_a(Q, G(z))} D_a^* G(z)(y^*)$$

and

$$N_c(P, z) \subset \bigcup_{y^* \in N_c(Q, G(z))} D_c^* G(z)(y^*).$$

PROOF. The first inclusion follows from Theorem 5 if we set $g(y) = \delta(y, Q)$. We can write the inclusion in the form

$$N_a(P, z) \times \{0\} \subset N_a(\operatorname{Graph} G, (z, w)) + \{0\} \times N_a(Q, G(z))$$

that implies, due to Theorem 2, that $N_c(P, z) \times \{0\}$ belongs to the closure of the sum

$$N_c(\operatorname{Graph} G, (z, w)) + \{0\} \times N_c(Q, G(z)). \tag{5.1}$$

The second inclusion will be proved if we show that the sum is actually a closed set. (Indeed, in this case any $x^* \in N_c(P, z)$ can be represented as a sum $(x^*, 0) = (u^*, v^*) + (0, y^*)$ where $(u^*, v^*) \in N_c(\operatorname{graph} G, (z, w))$ and $y^* \in N_c(Q, w)$ so that $x^* = u^*$, $v^* = -y^*$ and therefore $x^* \in D_c^* G(z)(y^*)$.)

To prove that (5.1) is a closed set, we first note that

$$\{0\} \times N_c(Q, w) = N_c(X \times Q, (z, w))$$

(which is obvious) and that $T_c(X \times Q, (z, w)) = X \times T_c(Q, w)$. The assumptions of the corollary ensure that

$$T_c(\operatorname{Graph} G, (z, w)) - T_c(X \times Q, (z, w)) = X \times Y.$$

The proof that (5.1) is a closed set follows from here precisely as in the proof of Corollary 4.3.

COROLLARY 5.2. *Under the assumptions of the theorem,*

$$\partial_c(g \circ F)(z) \subset \bigcup_{y^* \in \partial_c g(w)} D_c^* F(z)(y^*) = D_c^* F(z) \circ \partial_c g(F(z)).$$

PROOF. We apply Corollary 5.1 to $Q = \operatorname{epi} g$ and G: $(\alpha, x) \to (\alpha, F(x))$ (which is a map from $R \times X$ into $R \times Y$). Then $P = \operatorname{epi}(g \circ F)$ and the assumptions of Corollary 5.1 are clearly satisfied.

It is an easy matter to verify that $(\lambda, x^*) \in D_c^* G(\alpha, x)(\mu, y^*)$ if and only if $\lambda = \mu$ and $x^* \in D_c^* F(x)(y^*)$. By Corollary 5.1 for any x^* such that $(-1, x^*) \in N_c(\operatorname{epi} f, (f(z), z))$ there is $(\beta, y^*) \in N_c(\operatorname{epi} g, (g(w), w))$ (recall that $g(w) = f(z)$) such that $(-1, x^*) \in D_c^* G(f(z), z)(\beta, y^*)$ so that $\beta = -1$ and $x^* \in D_c^* F(z)(y^*)$. Q.E.D.

Certain results on set-valued maps relevant to Theorem 5 will be considered in the next section while here we go on with single-valued compositions. There is another

way to approach the notion of "approximate coderivative" in this case, namely, the one connected with the set-valued map (cf. [7]) $y^* \to \partial_a(y^* \circ F)(z)$.

PROPOSITION 8. *Let F be a continuous single-valued map from a neighborhood of $z \in X$ into Y. Then $\partial_a(y^* \circ F)(z) \subset D_a^* F(z)(y^*)$ and the two sets coincide for all y^* if F is Lipschitz near z.*

PROOF (CF. [9]). If $(h, v) \in T(\text{Graph } F, (x, F(x)))$, then there are $t_n \searrow 0$, $h_n \to h$ such that $t_n^{-1}(F(x + t_n h_n) - F(x)) \to v$. It follows that

$$d^-(y^* \circ F)(x; h) \leqslant \liminf_{n \to \infty} t_n^{-1} \langle y^*, F(x + t_n h_n) - F(x) \rangle = \langle y^*, v \rangle$$

so that

$$(5.2) \quad d^-(y^* \circ F)(x; h) \leqslant \inf\{\langle y^*, y \rangle \mid (h, y) \in T(\text{Graph } F, (x, F(x)))\}.$$

On the other hand, if F is Lipschitz near z (and x is sufficiently close to z) and $t_n \searrow 0$, $h_n \to h$ are such that

$$t_n^{-1} \langle y^*, F(x + t_n h_n) - F(x) \rangle \to d^-(y^* \circ F)(x; h),$$

then the sequence of $t_n^{-1}(F(x + t_n h_n) - F(x))$ is bounded. For any accumulation point v of the sequence we have $(h, v) \in T(\text{Graph } F, (x, F(x)))$ and $\langle y^*, v \rangle \leqslant d^-(y^* \circ F)(x; v)$ so that (5.2) becomes equality.

The inequality (5.2) means that any $x^* \in \partial^-(y^* \circ F)(x)$ satisfies

$$(x^*, -y^*) \in T^0(\text{Graph } F, (x, F(x)))$$

and the two inclusions are equivalent if F is Lipschitz near z. Thus, whenever $x_n \to z$ and $x_n^* \to x^*$ are such that $x_n^* \in \partial^-(y^* \circ F)(x_n)$, we have $(x_n, F(x_n)) \to (z, F(z))$ (since F is continuous) and $(x_n^*, -y^*) \in T^0(\text{Graph } F, (x_n, F(x_n)))$. Therefore,

$$(x^*, -y^*) \in N_a(\text{Graph } F, (z, F(z))).$$

Conversely, let F be Lipschitz with constant k, and let $x_n \to z$, $x_n^* \to x^*$, $y_n^* \to y^*$ be such that $(x_n^*, y_n^*) \in T(\text{Graph } F, (x_n, F(x_n)))$. Then $x_n^* \in \partial^-(y_n^* \circ F)(x_n)$ and, as follows from Lemma 3, there are $u_n', u_n'' \to z$ and $\varepsilon_n \to 0$ such that

$$x_n^* \in \partial^-(y^* \circ F)(u_n') + \partial^-((y_n^* - y^*) \circ F)(u_n'') + \varepsilon_n B.$$

Inasmuch as F is Lipschitz with constant k,

$$\partial^-((y_n^* - y^*) \circ F)(u) \subset \|y_n^* - y^*\| B \to \{0\}$$

and we conclude that $x^* \in \partial_a(y^* \circ F)(z)$. Q.E.D.

Combining Theorem 5 and Proposition 8, we obtain the following result.

COROLLARY 5.3. *If, under the assumptions of Theorem 5, F is Lipschitz near z, then*

$$\partial_a(g \circ F)(z) \subset \bigcup_{y^* \in \partial_a g(F(z))} \partial_a(y^* \circ F)(z).$$

COROLLARY 5.4. *If, under the assumptions of Theorem* 5, *F is strictly differentiable at z, then*

$$\partial_a(g \circ F)(z) \subset \bigcup_{y^* \in \partial_a g(F(z))} F'^*(z)y^* = F'^*(z) \circ \partial_a g(F(z)),$$

where $F'(z)$ is the derivative of F at z.

Corollary 5.4 is an immediate consequence of the theorem. Equally immediate from Corollary 5.2 is the following result.

COROLLARY 5.5. *Under the assumptions of Corollary* 5.4

$$\partial_c(g \circ F)(z) \subset F'^*(z) \circ \partial_c g(F(z)).$$

It is not clear, however, whether it is possible to replace approximate subdifferentials by Clarke's generalized gradients in Corollary 5.3. The reason is that such a replacement would be obviously impossible in case of Proposition 8.

We note also that in Corollary 5.4 both sides of the inclusion coincide if $F'(z)$ is onto. The inverse inclusion can be established by a direct calculation (see also [**16**]).

We conclude the section by proving two simple theorems, one of which is also an easy corollary from Theorem 5.

THEOREM 6. *Assume that the functions $f_1(x)$ and $f_2(x)$ are Lipschitz near z and $f_1(z) = f_2(z)$. Then*

$$\partial_a(f_1 \vee f_2)(z) \subset \bigcup_{\substack{\alpha_1 \geqslant 0,\, \alpha_2 \geqslant 0 \\ \alpha_1 + \alpha_2 = 1}} \partial_a(\alpha_1 f_1 + \alpha_2 f_2)(z).$$

PROOF. This follows from Corollary 5.3 if we define $F\colon X \to R^2$ by $F(x) = (f_1(x), f_2(x))$ and take the function g on R^2 equal to the maximal of components: $g(\xi_1, \xi_2) = \max\{\xi_1, \xi_2\}$. This is a simple convex function whose subdifferential at every point (ξ, ξ) is $\{(\alpha_1, \alpha_2) \mid \alpha_1 \geqslant 0, \alpha_2 \geqslant 0, \alpha_1 + \alpha_2 = 1\}$.

The same inclusion for Clarke's generalized gradients can be obtained by applying together Theorems 6 and 2.

THEOREM 7. *Assume that the functions f_1 and f_2 are lower semicontinuous near z and finite at z. Then $\partial_a(f_1 \wedge f_2)(z) \subset \partial_a f_1(z) \cup \partial_a f_2(z)$.*

PROOF. If, say, $f_1(x) = (f_1 \wedge f_2)(x)$, then $d^- f_1(x; h) \geqslant d^-(f_1 \wedge f_2)(x; h)$ for all h. Thus $\partial^-(f_1 \circ f_2)(x)$ belongs to either $\partial^- f_1(x)$ or $\partial^- f_2(x)$. It remains to apply the definition of approximate subdifferentials.

REMARK 5.1. As in the preceding section, the results we have proved here are new including those for Clarke's generalized gradients. In connection with the latter we note that earlier formulae for the C.g.g. of a composition either were established under the assumption that one of the functions is differentiable (as in [**4**, **18**]) or involved the convexification operation (as in [**7**, Proposition 9.14]).

Another new element is that certain results for approximate subdifferentials do not have analogues for Clarke's generalized gradients (such as Proposition 8, Theorem 7 and, maybe, Corollary 5.3).

6. Application to surjections and stability theorems. We start with the following auxiliary result giving a necessary condition for a minimum in a simple minimization problem. Applications to more complex optimization problems will be considered in the next section.

PROPOSITION 9. *Let $f(x)$ be an l.s.c. function finite at z and $S \subset X$ a closed set containing z. If f attains at z a local minimum on S and if* $\operatorname{dom}(d^{\uparrow} f(z; \cdot)) - T_c(S, z) = X$, *then* $0 \in \partial_a f(z) + N_a(S, z)$.

PROOF. Consider the function $g(x) = f(x) + \delta(x, S)$. (As usual, $\delta(x, S)$ is the indicator function of S.) If f attains at z a local minimum on S, then g attains an unconditional local minimum at z. In this case $0 \in \partial^- g(z) \subset \partial_a g(z)$ (because $d^- g(z; h)$ is nonnegative). It remains to apply Theorem 4.

We shall use this proposition to prove a surjection criterion for set-valued maps.

THEOREM 8. *Let F be a set-valued map with closed graph from X into Y, and let $w \in F(z)$. We assume that there are $\varepsilon > 0$ and $\lambda > 0$ such that $\|x^*\| \geqslant \varepsilon$ when $x^* \in D_a^* F(x, y)(y^*)$ for some x, y and y^* such that $\|x - z\| \leqslant \lambda$, $\|y - w\| \leqslant 2\lambda\varepsilon$ and $\|y^*\| = 1$. Then for any $0 < \alpha < \lambda$ the image under F of the open α-ball around z contains the open $\varepsilon\alpha$-ball around w:*

$$U(w, \varepsilon\alpha) \subset \bigcup_{\|x-z\|<\alpha} F(x) = F(U(z, \alpha)).$$

PROOF. The theorem follows from a similar result for Dini subdifferentials proved in [**9**] but we shall give a direct independent proof based on Theorem 4 (via Proposition 9).

Fix an $\alpha > 0$, $\alpha < \lambda$, and let $\|y - w\| = \xi < \varepsilon\alpha$. We assume that contrary to the statement, $y \notin F(x)$ for any x such that $\|x - z\| < \alpha$. Consider the function $f(x) = \rho(y, F(x)) + k\|x - z\|$, where ρ is the distance from y to $F(x)$ and $\xi/\alpha < k < \varepsilon$. We have $f(z) \leqslant \|y - w\| < k\alpha$; $f(x) \geqslant k\alpha > f(z)$ if $\|x - z\| = \alpha$. Insofar as the graph of F is closed, f is l.s.c. Therefore there is $u \in X$ such that $\|u - z\| < \alpha$ and f attains an unconditional local minimum at u.

Let $q \in F(u)$ be such that $\|y - q\| = \rho(y, F(u))$. Then the function $g(x, v) = \|y - v\| + k\|x - z\|$ attains a local minimum at (u, q) subject to the condition $(x, v) \in \operatorname{Graph} F$.

The function g is Lipschitz and we can apply Proposition 9 to obtain the inclusion

$$0 \in \partial_a g(u, q) + N_a(\operatorname{Graph} F, (u, q)).$$

On the other hand, g is a convex function and the approximate subdifferential of g coincides with its usual convex subdifferential (Proposition 5). Thus, taking into account that $y \neq q$ (because $\|u - z\| < \alpha$ and we have assumed that $y \notin F(x)$ if $\|x - z\| < \alpha$), we conclude that there are x^*, y^*, such that $\|x^*\| \leqslant k < \varepsilon$, $\|y^*\| = 1$ and

$$(x^*, y^*) \in N_a(\operatorname{Graph} F, (u, q))$$

or, equivalently, $x^* \in D_a^* F(u, q)(-y^*)$. But $\|x - z\| \leqslant \lambda$ and

$$\begin{aligned}\|q - w\| &\leqslant \|y - w\| + \|y - q\| \leqslant k\alpha + \rho(y, F(u)) \\ &\leqslant k\alpha + f(u) \leqslant k\alpha + f(z) \leqslant 2k\alpha \leqslant 2\lambda\varepsilon\end{aligned}$$

and we arrive at a contradiction with the assumptions. Q.E.D.

COROLLARY 8.1. *Let F be a set-valued map from X into Y with closed graph. Assume that there is $w \in F(z)$ such that $\|x^*\| > 0$ when $x^* \in D_a^* F(z, w)(y^*)$ and $y^* \neq 0$. Then the image under F of any neighborhood of z contains a neighborhood of w.*

If the assumptions of the theorem are satisfied, then the conclusion of the corollary obviously follows from the theorem. Otherwise, there are sequences $\{x_n\}$, $\{y_n\}$, $\{x_n^*\}$ and $\{y_n^*\}$ such that $x_n \to z$, $y_n \in F(x_n)$, $y_n \to w$, $x_n^* \to 0$, $\|y_n^*\| = 1$ and $x_n^* \in D_a^* F(x_n, y_n)(y_n^*)$. Let y^* be an accumulation point of the sequence $\{y_n^*\}$. Then $y^* \neq 0$ and, as follows from the upper semicontinuity property (1.1), $(0, -y^*) \in N_a(\text{Graph } F, (z, w))$ or, equivalently, $0 \in D_a^* F(z, w)(y^*)$ in contradiction with the assumptions.

COROLLARY 8.2. *Let F be a single-valued Lipschitz mapping from a neighbouhood of $z \in X$ into Y. If $0 \notin \partial_a(y^* \circ F)(z)$ for any $y^* \neq 0$, then f is surjective near z, that is to say, there is a neighborhood U about z such that for any $x \in U$ and any open set V containing x the set $F(V)$ contains a neighborhood of $F(x)$.*

PROOF. As follows from the proof of Corollary 8.1 (in view of Proposition 8) $0 \notin \partial_a(y^* \circ F)(x)$ if $y^* \neq 0$ and x is sufficiently close to z. We prove the corollary by applying Corollary 8.1 to every such x.

In connection with the last corollary we note that a set-valued map satisfying the assumptions of Theorem 8 may fail to have such a surjection property for it may happen that $F(x)$ does not meet a neighborhood of w for certain x arbitrarily close to z. But if we require that $F(z)$ be bounded and every $w \in F(z)$ satisfy the assumption of Theorem 8, then the surjection property for F will easily follow from the theorem (or from Corollary 8.1).

COROLLARY 8.3. *The following statement is true under the assumptions of Theorem 8: for any x such that $\|x - z\| < \lambda/2$ we have*

$$\rho(x, F^{-1}(w)) \leqslant \varepsilon^{-1}\rho(w, F(x)) \tag{6.1}$$

where $F^{-1}(y) = \{x \mid y \in F(x)\}$.

PROOF. Take an x such that $\|x - z\| < \lambda/2$. Then (since $w \in F(z)$) $\rho(x, F^{-1}(w)) = \gamma \leqslant \lambda/2$. We need to prove that $\rho(w, F(x)) \geqslant \varepsilon\gamma$. Assume the contrary. Then there is $v \in F(x)$ such that $\|w - v\| < \varepsilon\gamma$.

We notice further that the $\lambda/2$-neighborhood of x belongs to the λ-ball around z and the $\lambda\varepsilon$-neighborhood of v belongs to the $2\lambda\varepsilon$-ball around w. Therefore the conditions of Theorem 8 are fulfilled with z replaced by x, w by v and λ by $\lambda/2$. It follows that $w \in F(u)$ for some u satisfying $\|u - x\| < \gamma$ so that $\gamma = \rho(x, F^{-1}(w)) < \gamma$. The contradiction proves the claim.

COROLLARY 8.4. *Let F be a single-valued map from X into Y which is Lipschitz near z and $S \subset X$ a closed set containing z. We set $M = \{x \in S \mid F(x) = F(z)\}$. If there are $\varepsilon > 0$ and $\lambda > 0$ such that $\|x^* + u^*\| \geqslant \varepsilon$ whenever $x^* \in \partial_a(y^* \circ F)(x)$, $u^* \in N_a(S, x)$, $\|y^*\| = 1$, $x \in S$ and $\|x - z\| < \lambda$, then*

$$\rho(x, M) \leqslant \varepsilon^{-1}\|F(x) - F(z)\|$$

for any x such that $\|x - z\| \leqslant k\lambda/2$, where $k = \min\{\varepsilon/L, 1\}$, L being the Lipschitz constant of F.

PROOF. We set

$$G(x) = \begin{cases} F(x), & \text{if } x \in S, \\ \varnothing, & \text{if } x \notin S. \end{cases}$$

Then $\operatorname{Graph} G = (\operatorname{Graph} F) \cap (S \times Y)$. Since F is Lipschitz, we can be sure that

$$T_c(\operatorname{Graph} F, (x, F(x))) - T_c(S \times Y, (x, F(x))) = X \times Y$$

for any x sufficiently close to z (because the projection of $T_c(\operatorname{Graph} F, (x, F(x)))$ onto X is all of X and $T_c(S \times Y, (x, F(x))) = T_c(S, x) \times Y$). It follows, due to Corollary 4.2, that

$$N_a(\operatorname{Graph} G, (x, F(x))) \subset N_a(\operatorname{Graph} F, (x, F(x))) + N_a(S, x) \times \{0\}.$$

In other words, any $w^* \in D_a^* G(x, F(x))(y^*)$ (the inclusion implies that $x \in S$) is a sum $w^* = x^* + u^*$ where $x^* \in D_a^* F(x)(y^*)$ and $u^* \in N_a(S, x)$. If $\|y^*\| = 1$ and $\|x - z\| < k\lambda/2$ (which implies that $\|x - z\| < \lambda$ and $\|F(x) - F(z)\| < 2\lambda\varepsilon$), then $\|w^*\| \geqslant \varepsilon$ by the assumption, and applying Corollary 8.3 to G we complete the proof.

We observe finally that the same arguments as in the proof of Corollary 8.1 easily apply to show that conclusions of Corollaries 8.3 and 8.4 with constants not specified beforehand are valid under assumptions imposed only at the point in question. In other words, the following is true.

COROLLARY 8.5. *Let F be a set-valued map with closed graph, let $w \in F(z)$, and let $\|x^*\| > 0$ if $x^* \in D_a^* F(z, w)(y^*)$ for some $y^* \neq 0$. Then there are $\varepsilon > 0$ and $\lambda > 0$ such that the conclusion of Corollary 8.3 is valid.*

Likewise, if F is a Lipschitz map from a neighborhood of $z \in X$ into Y, $S \subset X$ is a closed set containing z and $0 \notin \partial_a(y^* \circ F)(z)$ for any nonzero y^*, then there are $\varepsilon > 0$ and $\lambda > 0$ such that the conclusion of Corollary 8.4 is valid.

The results of this section go back to a famous theorem of Ljusternik containing a description of tangent spaces to level sets of smooth maps. We refer to [**6**] for explanation and further references and also for the first results involving non-smoothness and generalized gradients of Clarke. Theorem 8 is closely connected with corresponding results of [**2**, **7**, **9**]; as a matter of fact, it follows from Theorem 11.9 of [**7**].

7. Application to optimization problems. Consider the following problem:

$$\text{minimize } f(x) \text{ subject to } 0 \in F(x), \tag{7.1}$$

where f is a function on X and F a set-valued map from X into Y. We associate with the problem the set-valued map G from X into $R \times Y$ defined by $G(x) = \{(\alpha, y) \mid \alpha \geqslant f(x),\ y \in F(x)\}$.

Assume that z is a local solution to (7.1), that is, $f(x) \geqslant f(z)$ for any x sufficiently close to z and such that $0 \in F(x)$. This means that for no x (sufficiently close to z) $G(x)$ may contain a point $(\xi, 0)$ with $\xi < f(z)$ or, consequently, that the image under G of no neighborhood of x may contain a neighborhood of $(f(z), 0)$.

This is one of the fundamental principles of the optimization theory. It follows that G cannot meet the assumptions of Corollary 8.1 so that the following first order necessary condition for an extremum in (7.10) is valid.

PROPOSITION 10. *We assume that f is l.s.c. near z and the graph of F is closed. If z is a local solution to* (7.1), *then there is a nonzero* $(\lambda, y^*) \in R \times Y^*$ *such that* $\lambda \leqslant 0$ *and*

$$\in D_a^* G(z, (f(z), 0))(\lambda, y^*).$$

The inequality $\lambda \leqslant 0$ is due to the obvious fact that $\beta \leqslant 0$ for any element (x^*, β, y^*) of the approximate normal cone to Graph G at any point $(x, f(x), y)$.

As usual, more detailed results are available under additional assumptions.

PROPOSITION 11. *We assume, in addition to Proposition* 10, *that f is Lipschitz near z. Then there are* $\lambda \geqslant 0$ *and* $y^* \in Y^*$ *such that* $\lambda + \|y^*\| > 0$ *and* $0 \in \lambda \partial_a f(z) + D_a^* F(z, 0)(y^*)$.

PROOF. Consider the sets

$$Q_1 = \{(x, \alpha, y) \in X \times R \times Y \mid (\alpha, x) \in \operatorname{epi} f\},$$
$$Q_2 = \{(x, \alpha, y) \in X \times R \times Y \mid y \in F(x)\}.$$

Or, if we ignore the order of variables, we may use a more eloquent expression

$$Q_1 = \operatorname{epi} f \times Y, \qquad Q_2 = R \times \operatorname{Graph} F.$$

We have

$$N_a(Q_1, (z, f(z), O_Y)) = N_a(\operatorname{epi} f, (f(z), z)) \times \{O_{Y^*}\} \tag{7.2}$$

(where O_Y is the origin in Y etc.),

$$N_a(Q_2, (z, f(z), O_Y)) = \{O_R\} \times N_a(\operatorname{Graph} F, (z, 0)), \tag{7.3}$$

$$T_c(Q_1, (z, f(z), O_Y)) = T_c(\operatorname{epi} f, (f(z), z)) \times Y, \tag{7.4}$$

$$T_c(Q_2, (z, f(z), O_Y)) = R \times T_c(\operatorname{Graph} F, (z, 0)). \tag{7.5}$$

We notice also that $\operatorname{Graph} G = Q_1 \cap Q_2$ (where G is the same as in the preceding proof).

The projection of $T_c(\operatorname{epi} f, (f(z), z))$ onto X is all of X since f is Lipschitz. Together with (7.4) and (7.5) this implies that

$$T_c(Q_1, (z, f(z), 0)) - T_c(Q_2, (z, f(z), 0)) = X \times R \times Y$$

and therefore (see Corollary 4.2) that

$$(7.6)\quad N_a(Q_1 \cap Q_2, (z, f(z), 0)) \subset N_a(Q_1, (z, f(z), 0)) + N_a(Q_2, (z, f(z), 0)).$$

By Proposition 10 there is a nonzero pair (λ, y^*) such that

$$(0, -\lambda, -y^*) \in N_a(Q_1 \cap Q_2, (z, f(z), 0)),$$

and (7.6), (7.2), (7.3) show that there is an x^* such that

$$(-\lambda, x^*) \in N_a(\text{epi } f, (f(z), z)) \quad \text{and} \quad (-x^*, -y^*) \in N_a(\text{Graph } F, (z, 0)).$$

If $\lambda = 0$, then $x^* = 0$ (since f is Lipschitz, see the proof of Theorem 2) and the second inclusion gives $0 \in D_a^* F(z, 0)(y^*)$. If $\lambda > 0$, then $x^* \in \lambda \partial_a f(z)$ and we have $0 = x^* - x^* \in \lambda \partial_a f(z) + D_a^* F(z, 0)(y^*)$. Q.E.D.

It is almost as easy to derive from Proposition 10 a necessary condition for the "standard" problem of mathematical programming:

$$(7.7)\qquad \text{minimize } f_0(x)$$

subject to

$$(7.8)\qquad \begin{aligned} &f_i(x) \leqslant 0, \quad i = 1, \ldots, m, \\ &f_i(x) = 0, \quad i = m+1, \ldots, n, \\ &x \in S. \end{aligned}$$

But we shall give an independent proof assuming all functions Lipschitz.

Let us consider the spaces $\bar{X} = R^{m+1} \times X$ and $\bar{Y} = R^{n+1}$, the mapping $\bar{F}$: $\bar{X} \to \bar{Y}$,

$$\bar{F}(\alpha_0, \ldots, \alpha_m, x) = (f_0(x) - \alpha_0, \ldots, f_m(x) - \alpha_m, f_{m+1}(x), \ldots, f_n(x))$$

and the set

$$\bar{S} = \{(\alpha_0, \ldots, \alpha_m, x) \mid \alpha_i \leqslant 0, i = 0, \ldots, m; x \in S\} \subset \bar{X}.$$

(We assume for simplicity that $f_0(z) = 0$.)

If z is a local solution to (7.7), (7.8), then $\bar{F}(\bar{S} \cap U)$ does not contain a neighborhood of the origin in $\bar{Y}$ for any neighborhood $U \subset \bar{X}$ of the point $\bar{z} = (f_0(z), \ldots, f_m(z), z)$. Corollary 8.5 says in this case that there is a nonzero $y^* \in \bar{Y}^*$ such that

$$(7.9)\qquad 0 \in \partial_a(y^* \bar{F})(\bar{z}) + N_a(\bar{S}, \bar{z}).$$

If we denote the components of y^* by $\lambda_0, \ldots, \lambda_n$, then $(y^* \circ \bar{F})(\bar{x}) = \Sigma \lambda_i f_i(x) - 1$, where 1 stands for the corresponding linear functional on R^{m+1}, i.e. 1: $(\alpha_0, \ldots, \alpha_m) \to \Sigma \lambda_i \alpha_i$. On the other hand, $N_a(\bar{S}, \bar{z}) = Q \times N_a(S, z)$, where $Q = \{(\beta_0, \ldots, \beta_m) \mid \beta_i \geqslant 0,\ i = 0, \ldots, m;\ \beta_i = 0 \text{ if } f_i(z) < 0\}$. With this notation (7.9) reads

$$0 \in (-1) \times \partial_a(\Sigma \lambda_i f_i)(z) + Q \times N_a(S, z)$$

or, equivalently,

$$(7.10)\qquad 0 \in \partial_a(\Sigma \lambda_i f_i)(z) + N_a(S, z), \qquad 1 \in Q.$$

The last inclusion means that

$$(7.11)\qquad \lambda_i \geqslant 0, \quad i = 0, \ldots, m; \qquad \lambda_i f_i(z) = 0, \quad i = 1, \ldots, m.$$

Thus, we have proved the following result.

PROPOSITION 12. *We assume that the functions f_i, $i = 0, \ldots, n$, are Lipschitz near z and S is a closed set containing z. If z is a local solution to* (7.7), (7.8), *then there are Lagrange multipliers λ_i, $i = 0, \ldots, n$, not all equal to zero and, such that the relations* (7.10) *and* (7.11) *hold true.*

Replacing in (7.10) ∂_a and N_a by ∂_c and N_c, we obtain a result uniting earlier theorems of Clarke and Hiriart-Urruty (see [**5**]). Of course, this result follows from Proposition 12 because approximate subdifferentials are never larger than C.g.g.'s. Certain necessary conditions involving approximate subdifferentials can also be found in [**7**, **13**, **15**].

8. Minimality properties.

THEOREM 9. *Let $\mathfrak{F}$ be a class of l.s.c. functions on X containing convex functions and sums of its elements. Assume that a set $\partial f(x) \subset X^*$ (possibly empty) is associated with every $f \in \mathfrak{F}$ and every $x \in X$ in such a way that the following is true*:

(i) $0 \in \partial f(x)$ *if f attains a local minimum at x*;

(ii) $\partial f(x) = \limsup_{u \to x,\, f(u) \to f(x)} \partial f(u)$;

(iii) *for a convex function $\partial f(x)$ is the usual convex subdifferential of f at x*;

(iv) $\partial(f + g)(x) \subset \partial f(x) + \partial g(x)$, *provided that g is convex continuous.*

Then for any $f \in \mathfrak{F}$ and any $x \in X$, we have $\partial_a f(x) \subset \partial f(x)$.

PROOF. The proof is very simple. If $x^* \in \partial_a f(x)$, then there are sequences x_n and x_n^* such that $x_n \to x$, $x_n^* \to x^*$, $f(x_n) \to f(x)$ and $x_n^* \in \partial^- f(x_n)$. By Lemma 1, the function $g_n(u) = f(u) + (1/n)\|u - x_n\| - \langle x_n^*, x - x_n \rangle$ attains a local minimum at x_n. Then $0 \in \partial g_n(x_n)$, by (i), and (iv) and (iii) imply that $0 \in \partial f(x_n) + (1/n)B - x_n^*$. This is the same as $x_n^* \in \partial f(x_n) + (1/n)B$ and, applying (ii), we see that $x^* \in \partial f(x)$.

The theorem implies the inclusion $\partial_a f(x) \subset \partial f(x)$ for convex l.s.c. functions and $\partial_a f(x) \subset \partial_c f(x)$ for Lipschitz functions.

The upper semicontinuity condition (ii) seems to be the most restrictive of the four (the other three rather agree with any reasonable idea of a subdifferential). But we need such a property in many situations. For instance, Corollaries 8.1 and 8.5 explicitly use it and Lagrangian-type necessary conditions are always based on something like the corollaries. Also, we often need upper semicontinuity to prove that a subdifferential is nonempty (as in (1.2)). (Dini subdifferentials which, in general, are smaller than approximate subdifferentials do not have any upper semicontinuity property. As a result, the relation $\partial^- f(x) \neq \varnothing$ may be not true for them even if f is Lipschitz and no Lagrange multiplier rule can be stated in terms of Dini subdifferentials. But they have the properties (i) and (ii).)

Thus, the condition must be present in one or another form. It can be weakened as, say, in the corollary to follow but at the expense of another assumption which is, however, more along the lines of the other three conditions of the theorem.

COROLLARY 9.1. *Assume that with any finite dimensional space X, any l.s.c. function f on X and any $x \in X$ a set $\partial f(x) \subset X^*$ is associated in such a way that conditions* (i), (iii), *and* (iv) *of Theorem* 9 *are satisfied together with the following two*:

(ii′) *if f is Lipschitz near x, then $\partial f(x) = \limsup_{u \to x} \partial f(u)$*;

(v) $\partial f(x) = \{x^* \mid (-1, x^*) \in N(\text{epi } f, (f(x), x))\}$ *for any l.s.c function f and any x such that $|f(x)| < \infty$; here $N(S, u)$ is the "normal cone" associated with f*:

$$N(S, u) = \overline{\bigcup_{\lambda > 0} \lambda \partial \rho(S, u)}.$$

PROOF. By Theorem 9, $\partial_a f(x) \subset \partial f(x)$ if f is Lipschitz near x. Therefore $\partial_a \rho(S, u) \subset \partial \rho(S, u)$ and an application of Proposition 2 and Theorem 1 concludes the proof.

The corollary implies that $\partial_a f(x) \subset \partial_c f(x)$ for any lower semicontinuous function f. We note in this connection that in examples we know both sets differ only on meager sets. This leads to the following problem that seems to be of an interest.

Problem 1. *Is it true or not that $\partial_a f(x)$ and $\partial_c f(x)$ generically coincide if f is Lipschitz*?

If the answer to the question is positive, then we shall have to conclude that approximate subdifferentials are still large enough, despite being minimal.

But there are some other indications that approximate subdifferentials are not as small as we would like them to be. For instance, the function $f(x) = x \cdot \sin(\log|x|)$ (which is Lipschitz) has the property that $0 \in \partial_a f(0)$ but $f(x) + \varepsilon|x|$ does not attain a local minimum at 0 if $\varepsilon < \sqrt{2}$. We infer from this example that the gap between necessary and sufficient conditions in nonsmooth optimization cannot be closed to such an extent as in the smooth theory. Another example: the inclusion in the composition formula of Theorem 5 cannot be, in general, replaced by equality from which we infer that Lagrangian-type necessary conditions in nonsmooth optimization cannot be invariant with respect to Lipschitz transformations. All this suggests the idea that, maybe, the class of Lipschitz functions is too large for certain properties to be valid that, in turn, leads to the following problem.

Problem 2. *Does there exist a subclass of Lipschitz functions containing continuously differentiable and convex functions together with finite sums and upper and lower bounds of its elements and having one* (*or all*) *of the following properties*:

(a) *if $0 \in \partial_a f(z)$, then $f(x) + \varepsilon\|x - z\|$ attains a local minimum at z for any $\varepsilon > 0$*;

(b) *the inclusion of Theorem* 5 (*or Theorem* 4) *holds as an equality*;

(c) *if $\partial_a f(x) \equiv \partial_a g(x)$, then $f(x) - g(x) \equiv$* const?

REFERENCES

1. J.-P. Aubin, *Contingent derivatives of set-valued maps and existence of solutions to non-linear inclusions and differential inclusions*, Math. Anal. Appl., Advances in Math., Suppl. Studies **7A** (1981), 159–229.

2. _____, *Lipschitz behavior of solutions to convex minimization problems*, Working paper WP-81-76, IIASA, Laxenburg, 1981.

3. F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc. **205** (1975), 247–262.

4. _____, *Generalized gradients of Lipschitz functionals*, Advances in Math. **40** (1981), 52–67.

5. J.-B. Hiriart-Urruty, *Refinements of necessary optimality conditions in non-differential programming*, Appl. Math. Optim. **5** (1979), 63–82.

6. A. D. Ioffe, *Regular points of Lipschitz functions*, Trans. Amer. Math. Soc. **251** (1979), 61–69.

7. _____, *Nonsmooth analysis: differential calculus of nondifferentiable mappings*, Trans. Amer. Math. Soc. **266** (1981), 1–56.

8. _____, *Sous-différentielles approchées de fonctions numériques*, C.R. Acad. Sci. Paris Sér. A-B **292** (1981), 675–678.

9. _____, *Calculus of Dini subdifferentials*, Nonlinear Anal. Theory Appl. (to appear).

10. _____, *Approximate subdifferentials of non-convex functions*, Cahier #8120, Univ. Paris IX "Dauphine", 1981.

11. _____, *On subdifferentiability spaces*, Ann. New York Acad. Sci. (to appear).

12. A. Ya. Kurger, *Calculus of generalized derivatives* (to appear).

13. A. Ya. Kruger and B. Sh. Mordukhovich, *Generalized normals and derivatives and necessary conditions for an extremum in non-differentiable programming*, Deponent #408-80, 494–80, Viniti, 1980. (Russian)

14. B. Sh. Mordukhovich, *Maximum principle in the optimal time control problem with non-smooth constraints*, Prikl. Mat. Meh. **40** (1976), 1014–1023.

15. _____, *Metric approximations and necessary optimality conditions for general classes of non-smooth extremal problems*, Dokl. Akad. Nauk SSSR **254** (1980), 1072–1076.

16. J.-P. Penot, *Calcul sous-différentiel et optimisation*, J. Funct. Anal. **27** (1978), 248–276.

17. R. T. Rockafellar, *Convex analysis*, Princeton Univ. Press, Princeton, N.J., 1970.

18. _____, *Directionally Lipschitzian functions and subdifferential calculus*, Proc. London Math. Soc. **39** (1979), 331–355.

19. _____, *Generalized directional derivatives and subgradients of nonconvex functions*, Canad. J. Math. **32** (1980), 257–280.

20. J. Warga, *Derivate containers, inverse functions and controllability*, Calculus of Variations and Control Theory (D. L. Russel, ed.), Academic Press, New York, 1976.

PROFSOYUSNAYA 85-1-203, MOSCOW 117279 U.S.S.R.