International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Nonsmoothness and Quasidifferentiability

Demyanov, V.F., Polyakova, L.N. and Rubinov, A.M.

IIASA Working Paper

WP-84-022

March 1984

NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

NONSMOOTHNESS AND QUASIDIFFERENTIABILITY

V.F. Demyanov
L.N. Polyakova
A.M. Rubinov

March 1984
WP-84-22

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

PREFACE

This paper presents a survery of results related to quasidifferential calculus. First we discuss different classes of directionally differentiable functions (convex functions, maximum functions and quasidifferentiable functions). Several generalizations of the concept of a subdifferential are considered, and the place and role of quasidifferentiable functions are outlined.

ANDRZEJ P. WIERZBICKI
Chairman
System and Decision Sciences

NONSMOOTHNESS AND QUASIDIFFERENTIABILITY

V.F. Demyanov
L.N. Polyakova
A.M. Rubinov

1. Introduction

This is not the place to go into the motivations and origins of nondifferentiability (although these are very important and interesting): for the purpose of this paper it is only necessary to realize that although a nondifferentiable function can often be approximated by a differentiable one, this substitution is usually unacceptable from an optimization viewpoint since some very important properties of the function are lost (see Example 2.1 below). We must therefore find some new analytical tool to apply to the problem.

Define a finite-valued function f on an open set $\Omega \subset E_n$. If function f is directionally differentiable, i.e., if the following limit exists:

$$\frac{\partial f(x)}{\partial g} = \lim_{\alpha \to +0} \frac{1}{\alpha} [f(x+\alpha g) - f(x)] \quad \forall g \in E_n , \tag{1.1}$$

then

$$f(x+\alpha g) = f(x) + \alpha \frac{\partial f(x)}{\partial g} + o(\alpha) .$$

Many important properties of the function can be described using the directional derivative. To solve optimization problems we must be able to (i) check necessary conditions for an extremum; (ii) find steepest-descent or -ascent directions; (iii) construct numerical methods.

In general, we cannot solve these auxiliary problems for an arbitrary function f: we must have some additional information.

In classical differential calculus it is assumed that $\partial f(x)/\partial g$ can be represented in the form

$$\frac{\partial f(x)}{\partial g} = (f'(x),g) ,$$

where $f'(x) \in E_n$ and (a,b) is the scalar product of vectors a and b . The function f is said to be *differentiable* at x and the vector $f'(x)$ is called the *gradient* of f at x . Differentiable functions form a well-known and important class of functions.

The next cases that we shall consider are convex functions and maximum functions. It turns out that for these functions the directional derivative has the form

$$\frac{\partial f(x)}{\partial g} = \max_{v \in \partial f(x)} (v,g) , \tag{1.2}$$

where $\partial f(x)$ is a convex compact set called the *subdifferential* of f at x . Each of these two classes of functions forms a convex cone and therefore their calculus is very limited (only two operations are allowed: addition, and multiplication by a positive number).

The importance of eqn. (1.2) has led to many attempts to extend the concept of a subdifferential to other classes of nondifferentiable functions (see, e.g., [1,15,16,18,22,23,28, 32]).

One very natural and simple generalization was suggested by the authors of the present paper in 1979 [7,13]. We shall say that a function f is *quasidifferentiable* at x if it is directionally differentiable at x and if there exists a pair of compact convex sets $\underline{\partial} f(x) \subset E_n$ and $\bar{\partial} f(x) \subset E_n$ such that

$$\frac{\partial f(x)}{\partial g} = \max_{v \in \underline{\partial} f(x)} (v,g) + \min_{w \in \bar{\partial} f(x)} (w,g) . \tag{1.3}$$

The pair $Df(x) = [\underline{\partial} f(x) , \bar{\partial} f(x)]$ is called a *quasidifferential* of f at x .

It has been shown that quasidifferentiable functions form a linear space closed with respect to all algebraic operations and, even more importantly, to the operations of taking pointwise maxima and minima. This has led to the development of quasidifferential calculus, and many important and interesting properties of these functions have been discovered (including a chain rule, an implicit function theorem, and so on).

One very important property of these functions is that if f is directionally differentiable and its directional derivative $\partial f(x)/\partial g$ at x is a continuous function of direction g (every directionally differentiable Lipschitzian function has this property), then $\partial f(x)/\partial g$ can be approximated to within any prescribed accuracy by a function of form (1.3).

Thus, the quasidifferential is an ideal tool for studying the first-order properties of functions.

A more general approach, involving an extension of quasi-differential calculus, has been presented by Rubinov and Yagubov [29]. They proved that if $\partial f(x)/\partial g$ is continuous in g then it can be represented in the form

$$\frac{\partial f(x)}{\partial g} = \inf \{\lambda > 0 | g \in \lambda U\} + \sup \{\lambda < 0 | g \in \lambda V\}, \qquad (1.4)$$

where U and V are what are known as *star-shaped sets*. If U and V are convex sets then eqn. (1.4) can be rewritten in the form (1.3).

Thus, if f is directionally differentiable it is natural to use this construction (the directional derivative) to study optimization problems. However, if f is not directionally differentiable some other tool must be found. One approach is to generalize the notion of the directional derivative (1.1). We shall mention only the following two generalizations:

1. The *Hadamard upper derivative* of f at x in the direction g, defined as

$$\frac{\partial_H f(x)\uparrow}{\partial g} = \overline{\lim_{\substack{g' \to g \\ \alpha \to +0}}} \frac{1}{\alpha}[f(x+\alpha g') - f(x)] .$$

In the case of a Lipschitzian function this becomes:

$$\frac{\partial_H f(x)\uparrow}{\partial g} = \overline{\lim_{\alpha \to +0}} \frac{1}{\alpha}[f(x+\alpha g) - f(x)] . \qquad (1.5)$$

2. The *Clarke upper derivative* of f at x in the direction g, defined as

$$\frac{\partial_{Cl} f(x)\uparrow}{\partial g} = \overline{\lim_{\substack{x' \to x \\ \alpha \to +0}}} \frac{1}{\alpha}[f(x'+\alpha g) - f(x')] \ . \tag{1.6}$$

Other generalizations and extensions are given in [18,22,28]. Equation (1.5) is a natural generalization of (1.1) and, in the case of a directionally differentiable function, the Hadamard upper derivative (1.5) coincides with the directional derivative (1.1). However, this is not the case for the Clarke upper derivative (1.6). The reason for this is that (1.6) describes not the local properties of f at x but some "cumulative" properties of f in a neighborhood of x . It seems to the authors that for optimization purposes it is better to use the Hadamard derivative (and this idea has been exploited by B.N. Pschenichnyi [23]).

The Hadamard and Clarke upper derivatives are used to study minimization problems: for maximization problems it is necessary to invoke the Hadamard and Clarke lower derivatives. These are defined analogously to (1.5) and (1.6) with the operation $\overline{\lim}$ replaced by $\underline{\lim}$. We shall discuss both these generalizations later in the paper: for now, note only that if the Hadamard upper derivative is continuous (which is always the case if f is Lipschitzian), then it can be approximated by a function of the form (1.3), so that quasidifferential calculus can be used here as well.

In Section 2 we discuss directional differentiability. Section 3 is concerned with convex functions and maximum functions, as well as with the Clarke subdifferential and Pschenichnyi upper convex and lower concave approximations. Quasidifferentiable functions are treated in Section 4.

This should be seen as a survey paper: we hope that it will provide a general introduction to the subject of this Study and enable readers to make use of the results in their own research.

2. Directional differentiability

Let $S \subset E_n$ be an open set and f be defined and finite-valued on S . Fix $x \in S$ and $g \in E_n$. The function f is said to be *differentiable* at x in the direction g if the following finite limit exists:

$$\frac{\partial f(x)}{\partial g} = f'_x(g) \equiv \lim_{\alpha \to +0} \frac{1}{\alpha}[f(x+\alpha g) - f(x)] \ . \tag{2.1}$$

(It is naturally assumed that $x+\alpha g \in S$; since S is open this is the case for all $\alpha \in [0,\alpha_0(g)]$, where $\alpha_0(g) > 0$) . The limit (2.1) is called the (first-order) *directional derivative* of f at x in the direction g .

If f is differentiable in every direction $g \in E_n$ it is said to be *directionally differentiable* at x .

If f is directionally differentiable at x and Lipschitzian in some neighborhood of x , then

$$\lim_{\substack{\alpha \to +0 \\ g(\alpha) \to g}} \frac{1}{\alpha}[f(x+\alpha g(\alpha)) - f(x)] = \frac{\partial f(x)}{\partial g} \ ,$$

i.e., in this case it is sufficient to consider only "line" directions.

It is clear from (2.1) that if f is directionally differentiable then

$$f(x+\alpha g) = f(x) + \alpha \frac{\partial f(x)}{\partial g} + o(\alpha) ,$$

i.e., the directional derivative provides a first-order approximation of f in a neighborhood of x .

Let f be directionally differentiable at x , $x \in S$. A direction g(x) is known as a *steepest-descent direction* of f at x if

$$\frac{\partial f(x)}{\partial g(x)} = \inf_{g \in S_n} \frac{\partial f(x)}{\partial g} ,$$

where $S_n = \{x \in E_n \mid \|g\|=1\}$.

A direction g'(x) is called a *steepest-ascent direction* of f at x if

$$\frac{\partial f(x)}{\partial g'(x)} = \sup_{g \in S_n} \frac{\partial f(x)}{\partial g} .$$

Directions of steepest descent or ascent need not necessarily exist and if they do, they are not necessarily unique.

It is clear that for a point $x^* \in E_n$ to be a minimum point of f it is necessary that

$$\frac{\partial f(x^*)}{\partial g} \geq 0 \quad \forall\, g \in E_n .$$

An analogous necessary condition for a maximum is

$$\frac{\partial f(x^{**})}{\partial g} \leq 0 \qquad \forall\, g \in E_n \ .$$

However, these necessary conditions are in general difficult to verify; they are also trivial reformulations of the definitions of a minimum and a maximum. We therefore have to make use of certain specific properties of the function under consideration.

One very important class is that of differentiable functions. In this case

$$\frac{\partial f(x)}{\partial g} = (f'(x),g) \ , \tag{2.2}$$

where $f'(x)$ is the gradient of f at x .

Applying the concept of a gradient, for example, to the optimization problem, it is possible to:

1. Compute the directional derivative.
2. Derive the following necessary condition for a minimum or a maximum: for a differentiable function f to attain its local minimum (or maximum) value at $x^* \in S$ it is necessary that

$$f'(x^*) = 0 \ . \tag{2.3}$$

   The point x^* at which condition (2.3) is satisfied is called a *stationary point* of f .
3. Find directions of steepest descent and ascent as follows: If $f'(x_0) \neq 0$ then the direction

$$g(x_0) = -\frac{f'(x_0)}{\|f'(x_0)\|} \tag{2.4}$$

is the direction of steepest descent of f at x_0 , and the direction

$$g'(x_0) = \frac{f'(x_0)}{\|f'(x_0)\|}$$

is the direction of steepest ascent of f at x_0 . In this case the directions of steepest descent and ascent both exist and are unique.

4. Construct numerical methods for finding an extremum.

The concept of a gradient (a derivative in the one-dimensional case) has had a profound impact on the development of science. It is impossible to overestimate its importance and influence. From being an art, mathematics became a technical science.

However, differential calculus is only applicable if the functions studied are smooth (i.e., differentiable). For most practical problems tackled in the past (and for many presently under study) it has been sufficient to consider only smooth functions. Nevertheless, an increasing number of problems arising in engineering and technology are of an essentially non-smooth nature. There are two very popular ways to avoid nondifferentiability. First, one tries to replace a non-smooth problem by a smooth one. For example, the problem of minimizing the function

$$f(x) = \max_{i \in I} \phi_i(x) ,$$

where the ϕ_i's are smooth non-negative functions, I=1:N and $x \in E_n$, is often replaced by the minimization of

$$F(x) = \Sigma\, a_i \phi_i(x) \ ,$$

where the a_i are positive coefficients. The function F is smooth but it now describes quite a different problem.

The second possibility is to consider the function

$$F_p(x) = \left(\sum_i [(\phi_i(x)]^p \right)^{1/p}$$

instead of f . It is well-known that $F_p(x) \xrightarrow[p\to+\infty]{} f(x) \ \forall x$. Note that in many cases the computational process by which $F_p(x)$ is minimized becomes unstable. Some very important properties of the original function can thus be lost in the pursuit of smoothness.

We can illustrate this using a very simple example.

Example 2.1. Let $x=(x^{(1)}, x^{(2)}) \in E_2$; $f(x)=|x^{(1)}| - |x^{(2)}|$, $x_0=(0,0)$. The function f is not differentiable at points where $x^{(1)}=0$ or $x^{(2)}=0$. Take a direction $g = (g^{(1)},g^{(2)})$. The function f is directionally differentiable with directional derivative

$$\frac{\partial f(x_0)}{\partial g} \equiv \lim_{\alpha\to+0} \frac{1}{\alpha}[f(x_0+\alpha g) - f(x_0)] = |g^{(1)}| - |g^{(2)}| \ .$$

It is clear that there are two steepest-descent directions of f at x_0 : $g_1=(0,1)$ and $g_1'=(0,-1)$. There are also two steepest-ascent directions: $g_2=(1,0)$, $g_2'=(-1,0)$.

Let us try to smooth the function f . Take $\varepsilon > 0$ and consider the following functions:

$$(1) \qquad f_{1\varepsilon}(x) = \sqrt{(x^{(1)})^2+\varepsilon} - \sqrt{(x^{(2)})^2+\varepsilon}$$

$$(2) \qquad f_{2\varepsilon}(x) = \sqrt{(x^{(1)}+\varepsilon)^2} - \sqrt{(x^{(2)}+\varepsilon)^2}$$

$$(3) \qquad f_{3\varepsilon}(x) = \sqrt{(x^{(1)})^2+\varepsilon} \quad \sqrt{(x^{(2)}+\varepsilon)^2} \quad .$$

It is clear that

$$f_{i\varepsilon}(x) \underset{\varepsilon\to 0}{\Longrightarrow} f(x) \quad \forall\ i \in 1:3 \ .$$

Find the gradients of these functions at x_0 :

$$\frac{\partial f_{1\varepsilon}(x)}{\partial x} = \left(- \frac{x^{(1)}}{\sqrt{(x^{(1)})^2+\varepsilon}} \ , \ \frac{x^{(2)}}{\sqrt{(x^{(2)})^2+\varepsilon}}\right) ;$$

$$\frac{\partial f_{1\varepsilon}(x_0)}{\partial x} = (0,0) \quad \forall\ \varepsilon > 0$$

$$\frac{\partial f_{2\varepsilon}(x)}{\partial x} = \left(- \frac{x^{(1)}+\varepsilon}{\sqrt{(x^{(1)}+\varepsilon)^2}} \ , \ \frac{x^{(2)}+\varepsilon}{\sqrt{(x^{(2)}+\varepsilon)^2}}\right) ;$$

$$\frac{\partial f_{2\varepsilon}(x_0)}{\partial x} = (-1,1) \quad \forall\ \varepsilon > 0$$

$$\frac{\partial f_{3\varepsilon}(x)}{\partial x} = \left(\frac{-x^{(1)}}{\sqrt{(x^{(1)})^2+\varepsilon}} \ ; \ \frac{x^{(2)}+\varepsilon}{\sqrt{(x^{(2)}+\varepsilon)^2}}\right);$$

$$\frac{\partial f_{3\varepsilon}(x_0)}{\partial x} = (0,1) \quad \forall\ \varepsilon > 0 \ .$$

We can then make the following deductions:

For $f_{1\varepsilon}$: x_0 is a stationary point.

For $f_{2\varepsilon}$: the steepest-descent direction at x_0 is $g_3=(\frac{\sqrt{2}}{2},-\frac{\sqrt{2}}{2})$ and the steepest-ascent direction is $g_3'=(-\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2})$.

For $f_{3\varepsilon}$: the steepest-descent direction at x_0 is $g_4=(0,-1)$ and the steepest-ascent direction is $g_4'=(0,1)$.

Thus, all three smoothing functions provide incomplete or even misleading information about stationarity or directions of steepest descent and ascent. The reason is that these smoothing functions are zeroth-order approximations while steepest-ascent and -descent directions reflect first-order properties of the function.

Since it appears that we cannot avoid nondifferentiability, we should rather study the properties of special classes of non-smooth functions with the aim of developing analytical tools to handle these problems.

3. The subdifferential and its generalizations

3.1. *Maximum functions.* Let

$$f(x) = \max_{y\in G} \phi(x,y) \ , \tag{3.1}$$

where $\phi(x,y)$ is continuous in x and y on $S \times G$ and continuously differentiable in x on S ; G is a compact set.

The function f described above is not necessarily continuously differentiable. However, it is directionally differentiable on S and

$$\frac{\partial f(x)}{\partial g} = \max_{y \in R(x)} (\phi_x'(x,y),g) , \tag{3.2}$$

where $R(x)=\{y \in G \mid \phi(x,y)=f(x)\}$.

The set $R(x)$ is closed and bounded. We can rewrite (1.2) in the form

$$\frac{\partial f(x)}{\partial g} = \max_{v \in \partial f(x)} (v,g) , \tag{3.3}$$

where

$$\partial f(x) = \mathrm{co}\{\phi_x'(x,y) \mid y \in R(x)\} . \tag{3.4}$$

It is not difficult to see that the set $\partial f(x)$ described by (3.4) can be used for several purposes [2,6]:

1. To compute the directional derivative (see (3.3)).
2. To derive the following necessary condition for an unconstrained minimum: for $x^* \in S$ to be a local minimum point of f defined by (3.1) it is necessary that

$$0 \in \partial f(x^*) . \tag{3.5}$$

A point $x^* \in S$ at which (3.5) is satisfied is called a stationary point of f (note that S is an open set).
3. If x_0 is not a stationary point then the direction

$$g(x_0) = - \frac{v(x_0)}{\|v(x_0)\|} ,$$

where $v(x_0) \in \partial f(x_0)$, $\|v(x_0)\| = \min\limits_{v \in \partial f(x_0)} \|v\|$, is a steepest-descent direction of f at x_0 . This direction is unique.

If we find $v_1(x_0) \in \partial f(x_0)$ such that $\|v_1(x_0)\| = \max\limits_{v \in \partial f(x_0)} \|v\|$, and if $\|v_1(x_0)\| > 0$, then the direction $g_1(x_0) = \dfrac{v_1(x_0)}{\|v_1(x_0)\|}$ is a steepest-ascent direction of f at x_0 . Note that this direction is not necessarily unique.

The set $\partial f(x)$ can also be used to construct numerical methods for minimizing f on E_n or on a bounded set (see, e.g., [6]).

3.2. *Convex functions.* Let $S \subset E_n$ be a convex open set and f be a convex function defined on S , i.e.,

$$f(\alpha x_1+(1-\alpha)x_2) \leq \alpha f(x_1)+(1-\alpha)f(x_2) \quad \forall\, \alpha \in [0,1] \ , \ \forall\, x_1,x_2 \in S \ .$$

Any finite-valued convex function is necessarily continuous and directionally differentiable on S , and

$$\frac{\partial f(x)}{\partial g} = \max_{v \in \partial f(x)} (v,g) \ , \tag{3.6}$$

where

$$\partial f(x) = \{v \in E_n \mid f(z)-f(x) \geq (v,z-x) \quad \forall\, z \in S\} \ . \tag{3.7}$$

The set $\partial f(x)$ is non-empty, convex and compact, and is called the *subdifferential* of f at x . The subdifferential plays exactly the same role as the set ∂f defined by (3.4) for a maximum function (except that condition (3.5) in the convex case is sufficient as well as necessary). For this reason we shall refer to the set $\partial f(x)$ defined by (3.4) as the subdifferential of the maximum function f described by (3.1). Note that if ϕ is also convex in x for any $y \in G$ then the set $\partial f(x)$ defined by (3.4) coincides with the set $\partial f(x)$ defined by (3.7) (assuming that f is a maximum function of form (3.1)).

Convex functions have been studied and used very widely: their fundamental properties were discovered and exploited by Fenchel [14], Moreau [21], and Rockafellar [27].

Thus we can define the subdifferential mapping ∂f for two very important classes of nondifferentiable functions. We may view the concept of a subdifferential as a generalization of the concept of a gradient (for continuously differentiable functions). If f is differentiable at x (where f is either a maximum function or a convex one), then $\partial f(x) = \{f'(x)\}$.

The properties of convex and maximum functions (and especially eqns. (3.3) and (3.6)) seem to have had a mesmerizing effect on many mathematicians. They have tried to generalize the concept of a subdifferential to other classes of nondifferentiable functions, while trying to somehow preserve eqn. (3.3) [1,15,16,17,18,22,23,28,32].

We shall consider here only two of these generalizations which are particularly relevant to the subject of this Study.

3.3. *The Clarke subdifferential.* Let a function f be Lipschitzian on S . By T(f) we shall denote the subset of S on which f is differentiable. It is well-known that Lipschitzian functions are differentiable almost everywhere.

For $x \in \Omega$, consider the set

$$\partial_{Cl} f(x) = \text{co}\ \partial_{Sh} f(x) ,$$

where

$$\partial_{Sh} f(x) = \{v \in E_n \mid \exists \{x_k\} : x_k \in T(f),\ x_k \to x,\ f'(x_k) \to v\} .$$

The set $\partial_{Sh} f(x)$ was introduced by Shor in [31] and the set $\partial_{Cl} f(x)$ by Clarke in [1]. The latter set will be referred to here as the *Clarke subdifferential* of f at x. It has been shown that $\partial_{Cl} f(x)$ is a non-empty convex compact set.

Clarke also introduced the *Clarke upper derivative* of f at x in the direction $g \in E_n$:

$$\frac{\partial_{Cl} f(x)\uparrow}{\partial g} = \overline{\lim_{\substack{x' \to x \\ \alpha \to +0}}} \frac{1}{\alpha}[f(x'+\alpha g) - f(x')] . \tag{3.8}$$

The most important result related to the Clarke upper derivative is the following:

$$\frac{\partial_{Cl} f(x)\uparrow}{\partial g} = \max_{v \in \partial_{Cl} f(x)} (v,g) . \tag{3.9}$$

It is possible to show that for a point $x^* \in S$ to be a minimum point of f it is necessary that

$$0 \in \partial_{Cl} f(x^*) . \tag{3.10}$$

We shall call any point x^* at which (3.10) is satisfied a *Clarke stationary point*. If x_0 is not such a stationary point then the direction

$$g(x_0) = -\frac{v(x_0)}{\|v(x_0)\|},$$

where

$$\|v(x_0)\| = \min_{v \in \partial_{Cl} f(x_0)} \|v\|$$

is a direction of descent of f at x_0 (but not necessarily a direction of steepest descent).

There are also some very interesting numerical algorithms for minimizing a Lipschitzian function f based on the Clarke subdifferential [19].

Let

$$\frac{\partial_{Cl} f(x)\downarrow}{\partial g} \equiv \varliminf_{\substack{x' \to x \\ \alpha \to +0}} \frac{1}{\alpha}[f(x'+\alpha g) - f(x')] . \tag{3.11}$$

This value is called the *Clarke lower derivative* of f at x in the direction g . It is possible to show that

$$\frac{\partial_{Cl} f(x)\downarrow}{\partial g} = \min_{v \in \partial_{Cl} f(x)} (v,g)$$

and that for x^{**} to be a maximum point of f it is necessary that (3.10) be satisfied at x^{**} , i.e., the necessary conditions for a minimum and a maximum coincide.

Thus, the role played by the Clarke subdifferential with respect to Lipschitzian functions is analogous to that played by the subdifferential for convex and maximum functions.

These results are very attractive from the aesthetic point of view. However, this approach nevertheless has some deficiencies from the optimization standpoint, the main reason for which being the fact that the Clarke upper (lower) directional derivative does not necessarily coincide with the directional derivative (if the latter exists).

Let us consider once again the function f described in Example 2.1:

$$f(x) = |x^{(1)}| - |x^{(2)}| \ , \ x = (x^{(1)}, x^{(2)}) \in E_2 \ , \ x_0 = (0,0) \ .$$

It is not difficult to check that

$$\partial_{Cl} f(x_0) = \text{co}\{(1,1), \ (1,-1) \ , \ (-1,1) \ , \ (-1,-1)\} \ ,$$

i.e., $0 \in \partial_{Cl} f(x_0)$, where (x_0) is a Clarke stationary point but is neither a minimum nor a maximum of f .

The Clarke subdifferential reflects some "cumulative" properties of the function in a neighborhood of a point. For example, if

$$\frac{\partial_{Cl} f(x)}{\partial g} < 0$$

then the direction g is not only a descent direction of f at x : it is also a descent direction of f at every x' in some neighborhood of x .

The Clarke subdifferential enables us to discover some very important properties of the function. However, the Clarke directional derivatives (upper and lower) defined by (3.8) and (3.11) are only very rough approximations of the directional derivative (if it exists).

In our opinion the Clarke subdifferential is not an appropriate tool for solving problems where directional derivatives are used (such as, for example, optimization problems). Nevertheless, the concept of the Clarke subdifferential is very important and can be very powerful in other areas of non-smooth analysis.

Note also that the calculus based on the Clarke subdifferential is incomplete (since the main relations are formulated as inclusions, not equalities) and this makes it unsuitable for computational use.

3.4. *The Pschenichnyi upper convex and lower concave approximations.* Consider first the Hadamard upper derivative

$$F_x(g) = \overline{\lim_{\alpha \to +0}} \frac{1}{\alpha}[f(x+\alpha g) - f(x)] \; , \tag{3.12}$$

where f is a Lipschitzian function and x is fixed. In the case where f is directionally differentiable, $F_x(g)$ coincides with its directional derivative.

The function $F_x(g)$ provides a better local approximation than the Clarke upper directional derivative. However, $F_x(g)$ is not a convex function and therefore it cannot be approximated by a maximum function of linear functions. Pschenichnyi [23] suggested that it should be approximated by a family of convex functions.

Let f be Lipschitzian on S and directionally differentiable at a fixed point $x \in S$. Note that the directional derivative $\partial f(x)/\partial g \equiv f'_x(g)$ is both continuous in g (because f is Lipschitzian) and positively homogeneous, i.e.,

$$f'_x(\lambda g) = \lambda f'_x(g) \qquad \forall\ \lambda \geq 0 .$$

A function p is said to be an *upper convex approximation* (u.c.a.) of f at x if p is sublinear (i.e., convex and positively homogeneous) and if $p(g) \geq f'_x(g) \quad \forall\ g \in E_n$. If p is an u.c.a. of f at x' then

$$f(x+\alpha g) \leq f(x) + \alpha p(g) + o_{x,g}(\alpha) , \tag{3.13}$$

where

$$\frac{o_{x,g}(\alpha)}{\alpha} \xrightarrow[\alpha\to+0]{} 0 .$$

Since p is sublinear there exists a unique convex compact set $\underline{\partial} p \subset E_n$ such that $p(g) = \max\limits_{v\in\underline{\partial} p} (v,g)$.

A function q is said to be a *lower concave approximation* (l.c.a.) of f at x if q is superlinear (i.e., concave and positively homogeneous) and if $q(g) \leq f'_x(g) \quad \forall\ g \in E_n$. Since q is superlinear there exists a unique convex compact set $\bar{\partial} q \in E_n$ such that $q(g) = \min\limits_{w\in\bar{\partial} q} (w,g)$.

Note that an upper convex approximation is not necessarily unique, and therefore a single u.c.a. cannot provide a satisfactory approximation of the function.

The notion of an exhaustive family of upper convex approximations was introduced in [8], where it was defined as follows:

Let Λ be an arbitrary set. A family $\{p_\lambda \mid \lambda \in \Lambda\}$, where p_λ is an u.c.a. of f at x , is called an *exhaustive family* of u.c.a's for f at x if

$$\inf_{\lambda\in\Lambda} p_\lambda(g) = \frac{\partial f(x)}{\partial g} \qquad \forall\, g \in E_n \,, \tag{3.14}$$

i.e., if

$$f(x+\alpha g) = f(x) + \alpha \inf_{\lambda\in\Lambda} p_\lambda(g) + o_{x,g}(\alpha) \qquad \forall\, g \in E_n \,.$$

Analogously, a family $\{q_\lambda \mid \lambda \in \Lambda\}$, where q_λ is a l.c.a. of f at x , is called an exhaustive family of lower concave approximations for f at x if

$$\sup_{\lambda\in\Lambda} q_\lambda(g) = \frac{\partial f(x)}{\partial g} \qquad \forall\, g \in E_n$$

i.e., if

$$f(x+\alpha g) = f(x) + \alpha \sup_{\lambda\in\Lambda} q_\lambda(g) + o_{x,g}(\alpha) \qquad \forall\, g \in E_n \,.$$

The existence of an exhaustive family of u.c.a.'s (or l.c.a.'s) implies that $f'_x(g)$ may be represented in the equivalent forms

$$\frac{\partial f(x)}{\partial g} = \inf_{\lambda\in\Lambda} \max_{v\in\underline{\partial} p_\lambda} (v,g) = \left(\sup_{\lambda\in\Lambda} \min_{w\in\bar{\partial} q_\lambda} (w,g)\right) \tag{3.15}$$

(of course, Λ is not the same for a family of u.c.a.'s and that of l.c.a.'s).

It is possible to show that exhaustive families of u.c.a.'s and l.c.a.'s exist for every directionally differentiable function whose directional derivative is continuous as a function of direction (see [8] for an illustration of the construction of a family of l.c.a.'s).

The concepts of upper convex approximation and lower concave approximation can be applied with some success to the solution of extremal problems. The following properties are of particular use: if x^* is a minimum point of f on S (recall that S is an open set) then for every u.c.a. p(g) it is necessary that $0 \in \underline{\partial} p$.

If $\{p_\lambda | \lambda \in \Lambda\}$ is an exhaustive family of u.c.a.'s of f at x^* then we have the following necessary condition for a minimum:

$$0 \in \underline{\partial} p_\lambda \qquad \forall\ \lambda \in \Lambda . \tag{3.16}$$

If $\{p_\lambda | \lambda \in \Lambda\}$ is an exhaustive family of u.c.a.'s of f at x_0 and (3.16) is not satisfied, find

$$\sup_{\lambda \in \Lambda} \ \min_{v \in \underline{\partial} p_\lambda} \|v\| = \|v_{\lambda_0}\| .$$

The direction $g(x_0) = - \dfrac{v_{\lambda_0}}{\|v_{\lambda_0}\|}$ is then a direction of steepest descent of f at x_0 .

Thus, if we have an exhaustive family of upper convex approximations we can:

1. Compute the directional derivative (see (3.14)).
2. State a necessary condition for a minimum (see (3.16)).
3. Find a steepest-descent direction.

Analogous results can be obtained for maximization problems by using an exhaustive family of l.c.a.'s.

Thus, the essence of this approach is to reduce the optimization problem to one of constructing the required families of u.c.a.'s (or l.c.a.'s).

In what follows we describe a class of functions for which families of upper convex approximations and lower concave approximations can be constructed with relative ease.

4. Quasidifferentiable functions

4.1. *Definitions and properties.* Let f be a finite-valued function defined on an open set $S \subset E_n$. The function f is said to be *quasidifferentiable* at $x \in S$ if it is directionally differentiable at x and if there exist convex compact sets $\underline{\partial} f(x) \subset E_n$ and $\bar{\partial} f(x) \subset E_n$ such that

$$\frac{\partial f(x)}{\partial g} = f'_x(g) = \max_{v \in \underline{\partial} f(x)} (v,g) + \min_{w \in \bar{\partial} f(x)} (w,g) \quad \forall\, g \in E_n . \qquad (4.1)$$

The pair of sets $Df(x) = [\underline{\partial} f(x) , \bar{\partial} f(x)]$ is called a *quasidifferential* of f at x ; sets $\underline{\partial} f(x)$ and $\bar{\partial} f(x)$ are described as a *subdifferential* and a *superdifferential*, respectively, of f at x . It is clear that a quasidifferential at a point is not unique.

If the set of quasidifferentials of f at x contains an element of type $Df(x)=[\underline{\partial} f(x),0]$, then the function f is said to be *subdifferentiable* at x . If there exists a quasidifferential of the form $Df(x)=[0,\bar{\partial} f(x)]$, then this function is said to be *superdifferentiable* at point x .

Some examples of quasidifferentiable functions are given below.

1. If f is continuously differentiable on S then it is quasidifferentiable at every point $x \in S$, and the pair of sets $Df(x)=[f'(x),0]$ (where $f'(x)$ is the gradient of f at x) is a quasidifferential of f at x. It is clear that the pair $Df(x)=[0,f'(x)]$ is also a quasidifferential of f at x. Thus, if a function f is smooth at x it is also both subdifferentiable and superdifferentiable at x.

2. From (3.4) and (3.6) it is clear that both maximum functions (defined by (3.1)) and convex functions are quasidifferentiable at $x \in S$, and that $Df(x)=[\underline{\partial} f(x),0]$, where $\underline{\partial} f(x)=\partial f(x)$ (defined by (3.4) or (3.7), respectively) is a quasidifferential of f at x. In other words, both maximum functions and convex functions are subdifferentiable.

3. In a similar way it can be seen that if f is concave on a convex open set S (i.e., $f_1=-f$ is convex), then f is quasidifferentiable on S, with quasidifferential $Df(x)=[0,\bar{\partial} f(x)]$. Here $\bar{\partial} f(x)=\{w \in E_n \mid f(z)-f(x) \le (w,z-x) \quad \forall z \in E_n\}$ is the superdifferential of the concave function f at x.

Let $D=[A,B]$ be a pair of sets, where $A \subset E_n$, $B \subset E_n$. We define multiplication by a real number λ as follows:

$$\lambda D = \begin{cases} [\lambda A, \lambda B] & \text{if } \lambda \ge 0 \\ \\ [\lambda B, \lambda A] & \text{if } \lambda < 0 \quad . \end{cases} \tag{4.2}$$

Let $D_1=[A_1,B_1]$, $D_2=[A_2,B_2]$, where $A_1, A_2, B_1, B_2 \subset E_n$. We define addition of sets in the following way:

$$D_1 + D_2 = [A,B] \tag{4.3}$$

where $A=A_1+A_2$, $B=B_1+B_2$. It follows from (4.1)-(4.3) that

1. If functions $f_1,\dots,f_N$ are quasidifferentiable at x then the function $f = \sum_{i=1}^{N} c_i f_i$ (where $c_i \in E_1$) is also quasidifferentiable at x and

$$Df(x) = \sum_{i=1}^{N} c_i Df_i(x) . \tag{4.4}$$

2. If functions f_1 and f_2 are quasidifferentiable at x then the function $f=f_1 \cdot f_2$ is also quasidifferentiable at x and

$$Df(x) = f_1(x)Df_2(x) + f_2(x)Df_1(x) . \tag{4.5}$$

3. If functions f_1 and f_2 are continuous and quasidifferentiable at a point x and $f_2(x) \neq 0$ then the function $f=f_1/f_2$ is quasidifferentiable at x and

$$Df(x) = \frac{1}{f_2^2(x)}[f_2(x)Df_1(x) - f_1(x)Df_2(x)] . \tag{4.6}$$

It is clear that (4.4)-(4.6) represent generalizations of well-known relations from classical differential calculus. However, quasidifferentiable functions also have the following very important additional properties (see [7,8,29]):

4. Let functions f_i , $i \in I=1:N$, be quasidifferentiable at $x \in S$. Then the function

$$f(x) = \max_{i \in I} f_i(x)$$

is quasidifferentiable at x and $Df(x)=[\underline{\partial}f(x)\ ,\ \bar{\partial}f(x)]$, where

$$\underline{\partial}f(x) = \text{co}\{\underline{\partial}f_k(x) - \sum_{\substack{i\in R(x)\\ i\neq k}} \bar{\partial}f_i(x)\,|\,k \in R(x)\}$$

$$\bar{\partial}f(x) = \sum_{k\in R(x)} \bar{\partial}f_k(x) \tag{4.7}$$

$$R(x) = \{i \in I\,|\,f_i(x) = f(x)\}\ .$$

5. If functions f_i , $i\in I=1:N$, are quasidifferentiable at $x\in S$ then the function $f(x) = \min_{i\in I} f_i(x)$ is quasidifferentiable at x and $Df(x)=[\underline{\partial}f(x)\ ,\ \bar{\partial}f(x)]$, where

$$\underline{\partial}f(x) = \sum_{k\in Q(x)} \underline{\partial}f_k(x)$$

$$\bar{\partial}f(x) = \text{co}\{\bar{\partial}f_k(x) - \sum_{\substack{i\in Q(x)\\ i\neq k}} \underline{\partial}f_i(x)\,|\,k \in Q(x)\} \tag{4.8}$$

$$Q(x) =\{i \in I\,|\,f_i(x) = f(x)\}\ .$$

Thus, the class of quasidifferentiable functions is a linear space closed with respect to all algebraic operations and, even more importantly, to the operations of taking pointwise maxima and minima.

4.2. *Necessary conditions for an unconstrained extremum.* It is easy to state necessary conditions for extrema of quasidifferentiable functions. We shall limit ourselves to consideration of the unconstrained case; other cases are discussed in detail in [5,25].

Let f be quasidifferentiable on E_n .

Theorem 4.1 (see [24]). *For a point* $x^* \in E_n$ *to be a minimum point of* f *on* E_n *it is necessary that*

$$- \bar{\partial} f(x^*) \subset \underline{\partial} f(x^*) \ . \tag{4.9}$$

Theorem 4.2. *For a point* $x^{**} \in E_n$ *to be a maximum point of* f *on* E_n *it is necessary that*

$$- \underline{\partial} f(x^{**}) \subset \bar{\partial} f(x^{**}) \ . \tag{4.10}$$

A point $x^* \in E_n$ at which condition (4.9) is satisfied is called an *inf-stationary point* of function f on E_n . A point $x^{**} \in E_n$ at which condition (4.10) is satisfied is called a *sup-stationary point* of f on E_n .

Assume that x_0 is not an inf-stationary point (i.e., condition (4.9) does not hold). Find $w_0 \in \bar{\partial} f(x_0)$ and $v_0 \in \underline{\partial} f(x_0)$ such that

$$\max_{w \in \bar{\partial} f(x_0)} \ \min_{v \in \underline{\partial} f(x_0)} \|v+w\| = \min_{v \in \underline{\partial} f(x_0)} \|v+w_0\| = \|v_0+w_0\| \ .$$

It turns out that the direction $g_0 = - \dfrac{v_0+w_0}{\|v_0+w_0\|}$ is a steepest-descent direction of f at the point x_0 . This direction may not be unique.

Analogously, if a point x_0 is not a sup-stationary point of f on E_n then we find $v_1 \in \underline{\partial} f(x_0)$ and $w_1 \in \bar{\partial} f(x_0)$ such that

$$\max_{v \in \underline{\partial} f(x_0)} \ \min_{w \in \bar{\partial} f(x_0)} \|v+w\| = \min_{w \in \bar{\partial} f(x_0)} \|v_1+w\| = \|v_1+w_1\| \ .$$

The direction $g_1 = \dfrac{v_1+w_1}{\|v_1+w_1\|}$ is a steepest-ascent direction of f at x_0 .

The problem of verifying the necessary conditions for a minimum is thus reduced to that of finding the Hausdorff deviation of the set $-\bar{\partial}f(x_0)$ from the set $\underline{\partial}f(x_0)$. Similarly, the verification of the necessary conditions for a maximum is equivalent to finding the Hausdorff deviation of the set $\underline{\partial}f(x_0)$ from the set $-\bar{\partial}f(x_0)$. If the necessary condition for a maximum or for a minimum holds at a point x_0 , then the corresponding Hausdorff deviation is zero. Otherwise the deviation is positive and its absolute value is equal to the rate of steepest ascent (or descent) at point x_0 .

Thus the concept of a quasidifferential is an extension of the idea of a gradient. The main formulae of quasidifferential calculus represent generalizations of relations from classical differential calculus (see (4.4)-(4.6)). A new and important additional operation is allowed in quasidifferential calculus--that of taking pointwise maxima or minima. This brings into play a host of new nondifferentiable functions obtained by combining ordinary "differentiable operations" with the taking of pointwise maxima and minima. A chain rule for quasidifferentiable functions has been discovered and was proved in [8-10], while implicit function and inverse theorems were established in [3,9]. The relation between the quasidifferential and the Clarke subdifferential has also been studied (see [4]): it appears that for a rather wide class of quasidifferentiable functions there exists a very simple relationship between the Clarke subdifferential and the quasidifferential.

The next step is to develop numerical methods for finding extreme points of quasidifferentiable functions. First of all, we should recognize that there may be several directions of steepest descent (or ascent, if we are looking for a maximum). This property requires a new approach to the construction of algorithms. In the convex case, for example, the greatest differences between many algorithms lie in (i) the rule used to find a descent direction and (ii) the step-size rule. In the quasidifferentiable case, however, it is necessary to consider several directions at each step. Some promising results in this area are given in [12,26].

4.3. *The place and role of quasidifferentiable functions in non-smooth optimization.* It follows from (4.1) that

$$\frac{\partial f(x)}{\partial g} = f'_x(g) = \min_{w \in \overline{\partial} f(x)} \left[\max_{v \in \underline{\partial} f(x)} (v+w,g) \right] .$$

It is clear that for every $w \in \overline{\partial} f(x)$ the function

$$p_w(g) = \max_{v \in [w + \underline{\partial} f(x)]} (v,g)$$

is an upper convex approximation of f at x and the set of functions $\{p_w | w \in \overline{\partial} f(x)\}$ is an exhaustive family of upper convex approximations of f at x .

Analogously, for every $v \in \underline{\partial} f(x)$ the function

$$q_v(g) = \min_{w \in [v + \overline{\partial} f(x)]} (w,g)$$

is a lower concave approximation of f at x and the set of functions $\{q_v | v \in \underline{\partial} f(x)\}$ represents an exhaustive family of lower concave approximations of f at x .

Thus quasidifferentiable functions represent one class of functions for which it is possible to construct exhaustive families of upper convex and lower concave approximations.

Note that the most important properties for optimization purposes are those of the directional derivative, because they can be used to check necessary conditions for an extremum and to find directions of steepest descent or ascent. If the directional derivative $f'_x(g)$ is a continuous function (as is always the case for a Lipschitzian, directionally differentiable function), then $f'_x(g)$ can be approximated by the difference of two convex, positively homogeneous functions. This means that the function f can be approximated to within any given accuracy (of $f'_x(g)$) by a quasidifferentiable function, thus ensuring that properties of f which are important from the computational standpoint (e.g., the number of steepest-descent and -ascent directions, etc.) can be derived. The quasidifferential therefore seems to be quite adequate for studying the first-order properties of the function.

Of course, there are many functions which are not quasidifferentiable (see, e.g., [11]), but for the purposes outlined above it is sufficient to consider only those which are.

The main problem is how to approximate $f'_x(g)$ by a quasi-differentiable function, and this is discussed in some detail in papers by Rubinov and Yagubov [29], Shapiro [30] and Melzer [20].

5. Concluding remarks

This paper considers only the finite-dimensional case, although most of the results can be extended to infinite-dimensional spaces (see, e.g., [9]).

Second-order approximation problems seem to present an important and promising area of research, but at present only a few results have been obtained in this field.

REFERENCES

[1] F.H. Clarke, "Generalized gradients and applications", *Transactions of the American Mathematical Society* 205 (1975) 247-262.

[2] J.M. Danskin, *The Theory of Max-Min* (Springer, New York, 1967).

[3] V.A. Demidova and V.F. Demyanov, "A directional implicit function theorem for quasidifferentiable functions", Working Paper WP-83-125, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1983.

[4] V.F. Demyanov, "On a relation between the Clarke subdifferential and the quasidifferential", *Vestnik Leningradskogo Universiteta* 13(1980) 18-24 (translated in *Vestnik Laningrad Univ. Math.* 13(1981) 183-189).

[5] V.F. Demyanov, "Quasidifferentiable functions: necessary conditions and descent directions", Working Paper WP-83-64, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1983.

[6] V.F. Demyanov and V.N. Malozemov, *Introduction to Minimax*, (Wiley, New York, 1974).

[7] V.F. Demyanov and A.M. Rubinov, "On quasidifferentiable functionals", *Doklady Akademii Nauk SSSR* 250(1980) 21-25 (translated in *Soviet Mathematics Doklady* 21(1980) 14-17).

[8] V.F. Demyanov and A.M. Rubinov, "On some approaches to non-smooth optimization problems" (in Russian), *Ekonomika i Matematicheskie Metody* 17(1981) 1153-1174.

[9] V.F. Demyanov and A.M. Rubinov, "Elements of quasidifferentiable calculus" (in Russian), in: V.F. Demyanov, ed., *Nonsmooth Problems of Control Theory and Optimization* (Leningrad University Press, Leningrad, 1982) pp. 5-127.

[10] V.F. Demyanov and A.M. Rubinov, "On quasidifferentiable mappings", *Mathematische Operations Forschung und Statistik, Series Optimization* 14(1)(1983) 3-21.

[11] V.F. Demyanov and I.S. Zabrodin, "Directional differentiability of a continual maximum function of quasidifferentable functions", Working Paper WP-83-58, International Institute for Applied Systems Analysis, Laxenburg, Austria, (1983).

[12] V.F. Demyanov, S. Gamidov and T.I. Sivelina, "An algorithm for minimizing a certain class of quasidifferentiable functions", Working Paper WP-83-122, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1983.

[13] V.F. Demyanov, L.N. Polyakova and A.M. Rubinov, "On one generalization of the concept of subdifferential", in *All Union Conference on Dynamic Control: Abstracts of Reports* (Sverdlovsk, 1979) pp. 79-84.

[14] W. Fenchel, "On conjugate convex functions", *Canadian Journal of Mathematics* 1(1949) 73-77.

[15] A.A. Goldstein, "Optimization of Lipschitzian continuous functions", *Mathematical Programming* 13(1977) 14-22.

[16] J.-B. Hiriart-Urruty, "New concepts in nondifferentiable programming", *Bulletin Société Mathématique de France, Mémoire* 60(1979) 57-85.

[17] A.D. Ioffe, "Nonsmooth analysis: differential calculus of nondifferentiable mappings", *Transactions of the American Mathematical Society* 26(1981) 1-56.

[18] A.Ya. Kruger and B.S. Mordukhovich, "Extremal points - the Euler equation in nonsmooth optimization problems", *Doklady of the Byelorussian Academy of Sciences*, 24(1980) 684-687.

[19] C. Lemarechal and R. Mifflin, eds., *Nonsmooth Optimization* (Pergamon Press, New York, 1977).

[20] D. Melzer, "Expressibility of piecewise linear continuous functions as a difference of two piecewise linear convex functions" (forthcoming).

[21] J.-J. Moreau, "Fonctionelles sous-différentiables", *Comptes Rendus de l'Academie des Sciences de Paris* 257(1963) 4117-4119.

[22] J.-P. Penot, "Calcus sous-differentiél et optimization", *Journal of Functional Analysis* 27(1978) 248-276.

[23] B.N. Pschenichnyi, *Convex Analysis and Extremal Problems* (Nauka, Moscow, 1980).

[24] L.N. Polyakova, "Necessary conditions for an extremum of quasidifferentiable functions" (in Russian), *Vestnik Leningradskogo Universiteta* 13(1980) 57-62 (translated in *Vestnik Leningrad Univ. Math.* 13(1981) 241-247).

[25] L.N. Polyakova, "Minimizing a quasidifferentiable function subject to equality-type quasidifferentiable constraints" (forthcoming).

[26] L.N. Polyakova, "On the minimization of the sum of a convex function and a concave function", this volume, pp.

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

[28] R.T. Rockafellar, *The Theory of Subgradients and Its Applications to Problems of Optimization* (Lecture Notes Series, Montreal University Press, Montreal, 1978).

[29] A.M. Rubinov and A.A. Yagubov, "The space of star-shaped sets and its applications in nonsmooth optimization" (forthcoming).

[30] A. Shapiro, "Quasidifferential calculus and first-order optimality conditions in nonsmooth optimization" (forthcoming).

[31] N.Z. Shor, "On one class of almost-differentiable functions and a method for minimizing functions of this class", *Kibernetika* 4(1972) 65-70.

[32] J. Warga, "Derivative containers, inverse functions and controllability", in: D.L. Russel, ed., *Calculus of Variations and Control Theory* (Academic Press, New York, 1976) pp. 13-45.