ON ε-DIFFERENTIAL MAPPINGS AND THEIR APPLICATIONS IN NONDIFFERENTIABLE OPTIMIZATION

E.A. Nurminski

December 1978 WP-78-58

2361
Laxenburg
Austria

International Institute for Applied Systems Analysis

Abstract

In Section 1 we give some review of the recent developments in nondifferential optimization and discuss the difficulties of the application of subgradient methods. It is shown that the use of ε-subgradient methods may bring computational advantages.

Section 2 contains the technical results on continuity of ε-subdifferentials. The principal result of this section consists in establishing Lipschitz continuity of ε-subdifferential mappings.

Section 3 gives some results on convergence of weighted sums of multifunctions. These results will be used in the study of the convergence of ε-subgradient method with sequential averages given in Section 4.

Section 4 gives the convergence theory for several modifications of this method. It is shown that in some cases it is possible to neglect accuracy control for the solution of internal maximum problems in the minmax problems. The results when this accuracy is nonzero and fixed are of great practical importance.

On ε-Differential Mappings and Their Applications In Nondifferentiable Optimization

1. INTRODUCTION

During the last years the main progress in nondifferentiable optimization was made due to the development of different schemes, which used some generalization of gradients. Starting with heuristic work [1] the proof of convergence of subgradient method was given primarily in [2] and generalized to functional spaces in [3]. Subgradient methods were successfully applied to many practical problems, especially at the Institute of Cybernetics of the Ukrainian Academy of Sciences, Kiev. N.Z. Shor and his colleagues developed the subgradient methods with space dialation [4]. Ju.M. Ermoliev proposed and investigated subgradient schemes in the extreme problems involving stochastic factors. The results of research by him and his colleagues were summarized in monograph [5]. Many efficient methods of solving NDO problems were developed by V.F. Demyanov and his collaborators [7]. In the 1970s analogous work appeared in Western scientific literature. There were proposed methods which look like numerical algorithm that are successful in the smooth case. The review of the state-of-the-art in the West can be found in [6]. A promising class of descent methods was investigated by C. Lemarechal [8]. R. Mifflin discussed the very general class of semi-smooth functions and developed some methods for their constrained minimization [9,10]. Also A.M. Gupal and V.I. Norkin [11] proposed stochastic methods for minimization of quite general function which can be even discontinuous.

Current theory and numerical algorithms usually use some generalization of the gradient and in some specific but unfortunately common situations these generalizations may be impossible to construct. Let us consider for instance the most elaborately investigated convex case. The subgradient g of a convex function f(x) may be considered as a vector which satisfies infinite system of inequalities:

$$f(y) - f(x) \geq g(y - x) \quad , \qquad \text{for any } y \in E \quad , \tag{1}$$

where E is a Euclidean space. We denote the set of vectors g satisfying (1) as $\partial f(x)$. Because (1) is as nonconstructive as the definition of standard derivative of smooth function, we need some kind of differential calculus to compute subgradients. Naturally we need an additional hypothesis about function f(x) in that case. Quite often function f(x) has a special structure:

$$f(x) = \sup_{u \in U} f(x,u) \quad , \tag{2}$$

where the functions $f(x,\cdot)$ have known differential properties in relation to the variables $x \in E$. These functions may be convex and differentiable in the usual sense. The supremum (2) for a given x, takes place on the set $U(x)$:

$$U(x) = \{u \in U : f(x) = f(x,u)\} \quad ,$$

any vector

$$g_x \in \partial_x f(x,u) \quad , \qquad u \in U(x)$$

belongs to the set $\partial f(x)$, and

$$\partial f(x) = \overline{co} \{\partial_x f(x,u), u \in U(x)\} \quad .$$

Unfortunately, the finding of any $u \in U(x)$ may be a rather complicated problem and a time-consuming operation. Strictly speaking, it may take an infinite amount of time.

R.T. Rockafellar [12] proposed a notion of ε-subgradient. These may be easier to construct. Formally the ε-subdifferential or the set of ε-subgradient $\partial_\varepsilon f(x)$ is the set of vectors g which satisfy an inequality

$$f(y) - f(x) \geq g(y - x) - \varepsilon \quad , \tag{3}$$

for given $\varepsilon \geq 0$ and any $y \in E$. Obviously,

$$\partial_\varepsilon f(x) \supset \partial f(x)$$

and so we may hope that to find some $g \in \partial_\varepsilon f(x)$ will be easier in comparison to the problem of computing some $g \in \partial f(x)$. In fact for the function of type (2), it is easy to see that any vector

$$g \in \partial_x f(x,u) \quad ,$$

where $u \in U_\varepsilon(x) = \{U : f(x,u) \geq f(x) - \varepsilon\}$ satisfies (3). The use of $U_\varepsilon(x)$ instead of $U(x)$ has many advantages. First of all for some problems there exists no u such that

$$f(x,u) = \sup_{u \in U} f(x,u) \quad .$$

In any case $U_\varepsilon(x)$ does exist always. Furthermore $U_\varepsilon(x)$ has some continuity properties [13] and it gives the corresponding continuity properties to $\partial_\varepsilon f(x)$. In the following we will discuss the continuity of the point-to-set mapping $\partial_\varepsilon f(x)$.

2. CONTINUITY PROPERTIES OF ε-SUBDIFFERENTIAL MAPPING

The study of the continuity properties of ε-subdifferentials started with the establishment of some properties of ε-subdifferentials which are the same as the properties of subdifferentials of the convex function. In [14] upper-semicontinuity of the mapping $\partial_\varepsilon f(x) : R^+ \times E \to 2^E$, where R^+ is a non-negative semiaxis and 2^E is a family of all subsets E, was proved, as well as the convexity and boundness of the set of ε-subgradients. It is important to say that this result was obtained in the assumption that $\varepsilon \geq 0$. If we assume that ε is strictly positive then it is possible to get more ingenious results. The continuity of ε-subdifferential mapping when $\varepsilon > 0$ was proved directly in the author's work [15]. After that the author became familiar with the article [16], where the reference to the unpublished theorem by A.M. Geoffrion was given. This theorem states that for function $f(x,y)$ which is convex with respect to variable $y \in Y$ and has finite infinum

$$v(x) = \inf_{y \in Y} f(x,y) \quad , \tag{4}$$

the set of y-solutions which solve (4) within some positive accuracy $\varepsilon > 0$,

$$Y_\varepsilon(x) = \{y : f(x,y) \leq v(x) + \varepsilon \ , \quad y \in Y\} \quad ,$$

is continuous with respect to variable x.

As we have for the set of ε-subgradients the representation given by R.T. Rockafellar:

$$\partial_\varepsilon f(x) = \{g : f(x) + f^*(g) - xg \leq \varepsilon\} \quad ,$$

where f* - is a conjugate function; and due to Fenhel's theorem,

$$\inf_g \{f(x) + f^*(g) - xg\} = 0 \quad ,$$

the result of the work [15] immediately follows from Geoffrion's theorem.

The author's opinion is that the establishment of continuity properties of ε-subdifferential mappings is of important principal significance but for practical purposes it is necessary to get a more exact estimation of the degree of continuity of these mappings. Such estimation will be given in what follows. The main result of this section is that in every compact set K, ε-subdifferentials are Lipschitz continuous in the Hausdorf metric which we denote by

$$\Delta(A,B) = \max \{\sup_{a\in A} \inf_{b\in B} \|a - b\|, \sup_{b\in B} \inf_{a\in A} \|a - b\|\} \quad .$$

So for $\varepsilon > 0$ there exists such $L_k < \infty$ such that for any $x,y \in K$,

$$\Delta(\partial_\varepsilon f(x), \partial_\varepsilon f(y)) \leq L_k \|x - y\| \quad . \tag{5}$$

We start with a study of the continuity properties of $\partial_\varepsilon f(x)$ with respect to ε for fixed x in some compact set K. It is rather easy to show that $\partial_\varepsilon f(x)$ is Lipschitz continuous with respect to ε. In fact, consider the support function of $\partial_\varepsilon f(x)$ for fixed x

$$v_p(\varepsilon) = \sup_{f(x) + f^*(y) - xg \leq \varepsilon} pg \tag{6}$$

where $p \in S$ - unit ball.

It is well known that $v_p(\varepsilon)$ is a concave function of ε and as far as $v_p(0) < \infty$ consequently $v_p(\varepsilon)$ satisfies the Lipschitz condition on $[\varepsilon', \varepsilon'']$, where $0 < \varepsilon' \leq \varepsilon'' < \infty$. As the Lagrange multiplier in (6) is uniformly bounded for $p \in S$ the following inequality holds:

$$\Delta(\partial_{\varepsilon+\gamma} f(x), \partial_\varepsilon f(x)) = \sup_{p \in S} (v_p(\varepsilon+\gamma) - v_p(\varepsilon)) < L_\varepsilon \gamma \qquad (7)$$

for some constant $L_\varepsilon < \infty$. Unfortunately this consideration does not allow the possibility to estimate L_ε.

In B.N. Pshenichiy's review of the original version of the paper, he remarked that the Lipschitz continuity of $\partial_\varepsilon f(x)$ with respect to $\varepsilon > 0$ for fixed x follows from his Lemma 4.1 in [18]. This lemma states that any convex* closed bounded set-valued mappings satisfy the Lipschitz condition within its domain. This lemma is applicable to $\partial_\varepsilon f(x) : R^+ \to 2^X$ because

$$\partial_\varepsilon f(x) = \{g : f(y) - f(x) \geq g(y - x) - \varepsilon \text{ , for any } y\}$$

is convex as described by the infinite system of linear inequalities.

The second reviewer of the paper, R.M. Chaney, gave a remarkable short direct proof of the statement which is worth mentioning,

Let us denote for given $\varepsilon > 0$ an estimate

$$\Delta(\partial_\varepsilon f(x), \partial f(x)) < C_k$$

for any $x \in K$ - compact set.

Theorem 1. For any $x \in K$ and $0 < \gamma \leq \varepsilon$,

$$\Delta(\partial_{\varepsilon-\gamma} f(x), \partial_\varepsilon f(x)) \leq \frac{2C_k}{\varepsilon} \gamma \quad .$$

Proof: Let $g_\varepsilon \in \partial_\varepsilon f(x)$ and choose $g \in \partial f(x)$ such that

---

*Set-valued mapping $Y(x) : x \to 2^Y$ is called convex if and only if the subset of the $X \times Y$:

$$\text{graf } Y = \{(x,y), \ y \in Y(x)\}$$

is convex.

$$\|g - g_\varepsilon\| \le 2C_k \quad .$$

Then for any y,

$$f(y) \ge f(x) + g_\varepsilon(y - x) - \varepsilon \quad , \tag{8}$$

and

$$f(y) \ge f(x) + g(y - x) \quad . \tag{9}$$

After multiplying (8) by $(\varepsilon - \gamma)/\varepsilon$ and (9) by γ/ε and summing, we get

$$f(y) \ge f(x) + g_\gamma(y - x) - (\varepsilon - \gamma)$$

where

$$g_\gamma = \frac{\varepsilon - \gamma}{\varepsilon} g_\varepsilon + \frac{\gamma}{\varepsilon} g \in \partial_{\varepsilon-\gamma} f(x)$$

by definition. As far as

$$\|g_\gamma - g_\varepsilon\| \le \frac{\gamma}{\varepsilon} \|g_\varepsilon - g\| \le \frac{2C_k}{\varepsilon} \gamma \quad ,$$

the theorem is proved.

From this statement the Lipschitz continuity of $\partial_\varepsilon f(x)$ with respect to x in every compact subset follows. R.M. Chaney also gave a short version of the further proofs which replaced the original incommunicable ones given by the author.

Theorem 2. Let x and y be in the compact set K; C_k be as defined above; and $\varepsilon_1 > 0$. Then there exists $A_k < \infty$ such that

$$\Delta(\partial_\varepsilon f(x), \partial_\varepsilon f(y)) \le \frac{2C_k A_k}{\varepsilon} \|x - y\| \quad ,$$

if

$$\|x - y\| \le \frac{\varepsilon}{A_k} , \qquad 0 < \varepsilon \le \varepsilon_1 .$$

Proof: From the Lipschitz continuity f(x) on K,

$$|f(x) - f(y)| \le A\|x - y\| ,$$

for some $A < \infty$. Also from the boundness of $\partial_\varepsilon f(x)$ on K, $\|g\| \le B$ for $g \in \partial_\varepsilon f(x)$, $x \in K$, $0 \le \varepsilon \le \varepsilon_1$.

Now let $A_k = A + B$ and $\|x - y\| \le \frac{\varepsilon}{A_k}$, $\gamma = \|x - y\|A_k \le \varepsilon$. For any $g \in \partial_{\varepsilon-\gamma} f(x)$ and any z,

$$f(z) - f(x) \ge g(z - x) - (\varepsilon - \gamma) ,$$

or

$$\begin{aligned} f(z) - f(x) &\ge f(x) - f(y) + g(z - y) + g(y - x) - (\varepsilon - \gamma) \\ &\ge g(z - y) - \varepsilon + \gamma - |f(x) - f(y)| - \|g\|\|x - y\| \\ &\ge g(z - y) - \varepsilon , \end{aligned}$$

that is $g \in \partial_\varepsilon f(y)$. So for γ defined above and x close enough to y,

$$\partial_{\varepsilon-\gamma} f(x) \subset \partial_\varepsilon f(y) .$$

Then

$$\begin{aligned} \sup_{g\in\partial_\varepsilon f(x)} \inf_{g'\in\partial_\varepsilon f(y)} \|g - g'\| &\le \sup_{g\in\partial_\varepsilon f(x)} \inf_{g'\in\partial_{\varepsilon-\gamma} f(x)} \|g - g'\| \\ &= \Delta(\partial_\varepsilon f(x), \partial_{\varepsilon-\gamma} f(x)) \le \frac{2C_k}{\varepsilon} \gamma \end{aligned}$$

$$= \frac{2C_k A_k}{\varepsilon} \|x - y\| \quad .$$

Due to the symmetry between x and y the same estimate is valid for

$$\sup_{g \in \partial_\varepsilon f(y)} \inf_{g' \in \partial_\varepsilon f(x)} \|g - g'\| \quad ,$$

which proves the theorem.

The idea of the proof for the final statement also belongs to R.M. Chaney.

Theorem 3. For $0 < \varepsilon \leq \varepsilon'$ and any x,y in the compact set K, there exists a constant, B_k, such that

$$\Delta(\partial_\varepsilon f(x), \partial_\varepsilon f(y)) \leq \frac{B_k}{\varepsilon} \|x - y\| \quad .$$

Proof: In accordance with Theorem 2, for any x' and y' in K such that

$$\|x' - y'\| \leq \frac{\varepsilon}{A_k} \quad ,$$

$$\Delta(\partial_\varepsilon f(x'), \partial_\varepsilon f(y')) \leq \frac{2C_k A_k}{\varepsilon} \|x' - y'\| \quad .$$

Without the loss of any generality, we may assume that

$$\|x - y\| > \frac{\varepsilon}{A_k} \quad .$$

Let us consider the finite covering of the set K by the open balls of the radii $\frac{\varepsilon}{4A_k}$. The number of these balls will be denoted by N_k and their centers by $\{x_i, i = 1,\ldots,N_k\}$. Then the sequence of points $\{x_{i_k}, k = 1,\ldots,M\}$ exists such that

$$x_{i_1} = x \quad , \qquad x_{i_M} = y \quad ,$$

and

$$\|x_{i_k} - x_{i_{k+1}}\| \le \frac{\varepsilon}{2A_k} \quad .$$

Therefore

$$\begin{aligned}
\Delta(\partial_\varepsilon f(x),\partial_\varepsilon f(y)) &\le \sum_{k=1}^{M} \Delta(\partial_\varepsilon f(x_{i_k}),\partial_\varepsilon f(x_{i_{k+1}})) \\
&\le \frac{2C_k A_k}{\varepsilon} \sum_{k=1}^{M} \|x_{i_k} - x_{i_{k+1}}\| \\
&\le M \frac{2C_k A_k}{\varepsilon} \cdot \frac{\varepsilon}{2A_k} \\
&\le N_k \frac{C_k A_k}{\varepsilon} \|x - y\| \\
&= \frac{B_k}{\varepsilon} \|x - y\| \quad .
\end{aligned}$$

Thus the theorem is proved.

3. WEIGHTED SUMS OF MULTIFUNCTIONS

The rather strong property of Lipschitz continuity of ε-subdifferentials proved in the previous section of the article makes it possible to establish some useful features of the sums:

$$z^{s+1} = z^s - \delta_s(z^s - g^s) \quad , \qquad s = 0,1,\ldots \tag{10}$$

where

$\delta_s > 0$ is a numerical sequence

and

$g^s \in \partial_{\varepsilon_s} f(x^s)$ is a sequence

of ε_s-subgradients of the function $f(x)$, calculated at some points $\{x^s\}$ taken from the compact set $X \subset E$.

Instead of (10) we can consider quite general recurrent relation of the kind

$$z^{s+1} = z^s - \delta_s(z^s - y^s) \quad , \qquad s = 0,1,\ldots$$

where

$$y^s \in Y(x^s)$$

and $Y(x)$ is a given multifunction.

Under the assumption of Lipschitz continuity of $Y(x)$ and some other assumptions specified below, the following general result is valid.

Lemma 1. Let $Y(x)$ be a bounded continuous convex-valued multifunction which satisfies Lipschitz condition with constant L uniformly on the set X and $\{x^n\}$ be a sequence of points from compact set X and $\{\delta_n\}$ be a numerical sequence such that

(i) $\dfrac{\|x^{n+1} - x^n\|}{\delta_n} \to 0$ *when* $n \to \infty$, *and*

(ii) $0 \le \delta_n < 1$, $\Sigma\, \delta_n = \infty$.

Then for the sequence of points z^s such that

$$z^{s+1} = z^s - \delta_s(z^s - y^s) \quad , \qquad s = 0,1,\ldots \quad , \qquad (11)$$

$$y^s \in Y(x^s) \quad ,$$

it is true that

$$\lim_{s\to\infty} \inf_{y \in Y(x^s)} \|z^s - y\| = 0$$

Proof: Let us denote

$$d_s = \inf_{y\in Y(x^s)} \|z^s - y\|^2$$

$$\bar{d}_{s+1} = \inf_{y\in Y(x^s)} \|z^{s+1} - y\|^2$$

and develop some useful inequalities.

Let

$$d_s = \inf_{y\in Y(x^s)} \|z^s - y\|^2 = \|z^s - y_s^*\|^2 \quad .$$

Then

$$\bar{d}_{s+1} = \inf_{y\in Y(x^s)} \|z^{s+1} - y\|^2 \leq \|z^{s+1} - y_s^*\|^2$$

$$= (1 - \delta_s)^2\|z^s - y_s^*\|^2 + \delta_s^2\|y^s - y_s^*\|^2$$

$$+ 2\delta_s(1 - \delta_s)(z^s - y_s^*)(y^s - y_s^*)$$

$$\leq (1 - \delta_s)^2\|z^s - y_s^*\|^2 - 2\delta_s(1 - \delta_s)(z^s - y_s^*)(y_s^* - y^s) + \delta_s^2 K \quad ,$$

where K is the upper estimate for $\|y^s - y^*\|^2$. Further we shall use the notation K for those constants the exact value of which has no meaning for the consideration.

Due to the convexity of the set $Y(x^s)$,

$$(z^s - y_s^*)(y_s^* - y^s) \geq 0 \quad ;$$

then

$$\bar{d}_{s+1} \leq (1 - \delta_s)^2 d_s + \delta_s^2 K \quad . \tag{12}$$

Furthermore, under the conditions of the lemma,

$$\Delta(Y(x^{s+1}), Y(x^s)) \leq L\|x^{s+1} - x^s\| = \gamma_s \quad ;$$

hence due to Cauchy's inequality

$$d_{s+1} = \inf_{y \in Y(x^{s+1})} \|z^{s+1} - y\|^2$$

$$\leq \inf_{y \in Y(x^{s+1}) \cap \{y_s^* + \gamma_s S\}} \|z^{s+1} - y\|^2$$

$$\leq 2 \sup_{z \in \gamma_s S} \|z\| \|z^{s+1} - y_s^*\| + \|z^{s+1} - y_s^*\|^2 + \sup_{z \in \gamma_s S} \|z\|^2 \quad .$$

Finally we get

$$d_{s+1} \leq \bar{d}_{s+1} + \gamma_s^2 + 2\gamma_s \sqrt{\bar{d}_{s+1}} \quad .$$

As long as $\bar{d}$ and d_s are bounded it is possible to choose D such that

$$\bar{d}_s \leq \frac{1}{2} D \quad , \qquad d_s \leq \frac{1}{2} D^2 \quad .$$

Then

$$d_{s+1} \le \bar{d}_{s+1} + \gamma_s D + \gamma_s^2 \quad , \tag{13}$$

which together with (12) gives

$$d_{s+1} \le (1 - \delta_s)^2 d_s + \gamma_s^2 + \gamma_s D + \delta_s^2 K \quad . \tag{14}$$

Now if we suppose that $d_s \geq \delta > 0$ for any $s \geq N$ where N is some large number, then from (14) we get

$$\begin{aligned} d_{s+1} &\le d_s - 2\delta_s d_s + \delta_s^2 d_s + \gamma_s^2 + \gamma_s D + \delta_s^2 K \\ &\le d_s - 2\delta_s \delta + \delta_s^2 \frac{1}{2} D + \gamma_s^2 + \delta_s^2 K + \gamma_s D \\ &\le d_s - 2\delta_s \delta \left(1 - \frac{\gamma_s D}{2\delta_s \delta}\right) + \delta_s^2 \left(K + \frac{D}{2} + \frac{\gamma_s^2}{\delta_s^2}\right) \quad . \end{aligned}$$

As $\frac{\gamma_s}{\delta_s} \to 0$ and $\delta_s \to 0$ when $s \to \infty$ then for s large enough,

$$1 - \frac{\gamma_s D}{2\delta_s \delta} \geq \frac{1}{2} \quad ,$$

$$K + \frac{D}{2} + \frac{\gamma_s^2}{\delta_s^2} \le C \quad ,$$

and

$$1 - \frac{C}{\delta} \delta_s \geq \frac{1}{2} \quad .$$

Hence we get

$$d_{s+1} \leq d_s - \frac{\delta}{2}\,\delta_s \quad . \tag{15}$$

Summing (15) on s from N to M > N we have

$$d_{M+1} \leq d_N - \frac{\delta}{2} \sum_{s=N}^{M} \delta_s \to -\infty$$

when M goes to infinity. This contradicts the positiveness of d_s for any s and therefore we should have at least one subsequence $\{s_k\}$ such that

$$d_{s_k} \to 0 \qquad \text{when } k \to \infty \quad .$$

Nevertheless, let us suppose

$$\overline{\lim_{s\to\infty}}\, d_s = \delta > 0 \quad .$$

As $|d_{s+1} - d_s| \to 0$ when $s \to \infty$ and as we proved

$$\underline{\lim_{s\to\infty}}\, d_s = 0 \quad ,$$

then a subsequence of d_{p_k} exists such that

$$d_{p_k} > \delta/2$$

and

$$d_{p_{k+1}} \geq d_{p_k} \quad .$$

But this contradicts (15) and completes the proof.

<u>Remark</u>:

It is important to note that the Lemma is valid even when we have a sequence of multi-functions $\{Y_s(x)\}$ and $y^s \in Y_s(x^s)$. The lemma remains correct if

(i) $Y_s(x)$ uniformly bounded on $x \in X$ for $s = 0,1,\ldots;$

(ii) $Y_s(x)$ satisfy the Lipschitz condition with constants L_s and $\frac{L_s \| x^{s+1} - x^s \|}{\delta_s} = \frac{\gamma_s}{\delta_s} \to 0$ when $s \to \infty$.

4. MULTISTAGE ALGORITHMS

The results stated above give an opportunity to build up algorithms in which the directions of movement on every iteration are not directions of anti-subgradients or anti-ε-subgradients, but are weighted sums of ε-subgradients computed on the previous iterations. Such weighted sums may have more smooth properties (see Figure 1) and can bring some computational advantages. So in this part we will investigate iterative processes of the kind:

$$x^{s+1} = x^s - \rho_s z^s$$

$$z^{s+1} = z^s - \delta_s (z^s - g^s) \quad , \qquad s = 0,1,\ldots \quad ,$$

where

$$g^s \in \partial_{\varepsilon_s} f(x^s)$$

and ε_s, ρ_s, and γ_s are numerical sequences with properties that will be specified later on.

We have in mind that such processes will be applied when seeking the solution of the unconstrained extremum problem

$$\min_{x \in E} f(x) \tag{17}$$

with the convex function $f(x)$. When proving the convergence of such algorithms we will use the conditions of convergence

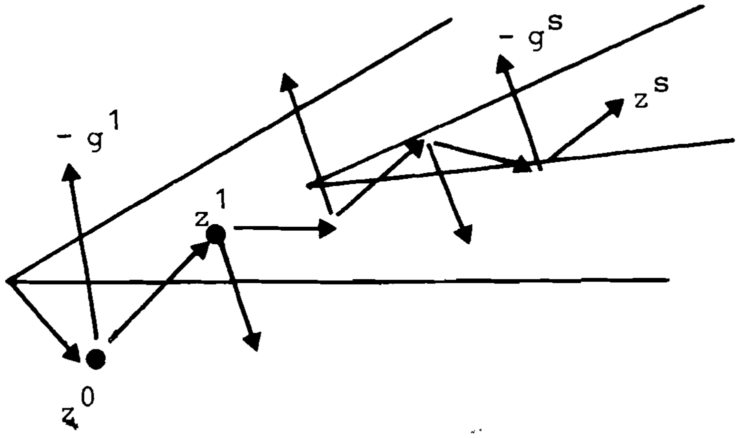

Figure 1

for infinite iterative algorithms of nonlinear programming given in [17]. We will use a slightly different version of these conditions which can be proved in the same manner.

An algorithm of nonlinear programming is considered by definition as a rule for constructing the sequence of points $\{x^s\}$ which should give useful information about the solution of a given extremum problem, such as problem (17). With every such problem we may associate a set of "solutions" X^* and consider the given algorithm as convergent if the limit points of the sequence $\{x^s\}$ generated by this algorithm lie in the set X^*.

If we adopt such definitions the following quite general conditions stated in terms of sequence $\{x^s\}$ and solution set X^* happen to be sufficient for convergence of the algorithm.

These conditions are the following:

1. If $x^{s_k} \to x^* \in X^*$, then

$$\|x^{s_k} - x^{s_k+1}\| \to 0$$

where $k \to \infty$.

2. There exists such compact set K such that

$$x^s \in K \quad , \qquad \text{for any } s.$$

3. If $x^{n_k} \to x' \bar{\in} X^*$, then for any $\varepsilon > 0$ sufficiently small and for any k, indices m_k defined below

$$m_k = \min_{m>n_k} m : \|x^m - x^{n_k}\| > \varepsilon$$

are finite.

4. A continuous function W(x) exists such that, for any sequences x^{n_k}, x^{m_k} mentioned in condition 3,

$$\overline{\lim_{k\to\infty}} W(x^{m_k}) < \lim_{k\to\infty} W(x^{n_k}) \quad .$$

5. Set

$$W^* = \{W(x^*),\ x^* \in X^*\}$$

has such property that in every subinterval (a,b) of the real axis the point $a < c < b$ exists such that

$$c \bar{\in} W^* \quad .$$

Under these conditions all limit points of the sequence $\{x^s\}$ belong to the set X^*. It is important to note that under conditions 1-4 when condition 5 is deleted a weaker result can be proved. Namely, under conditions 1-4 a limit point,

$$\bar{x} = \lim_{k\to\infty} x^{s_k} \quad ,$$

exists such that $\bar{x} \in X^*$.

Let us consider now the algorithm (16) for solving extremum problem (17).

Theorem 4. If sequence of $\{x^s\}$ *generated by (16) is bounded, and*

$$(i) \quad \varepsilon_s \to +0 \ , \quad \gamma_s \to +0 \ , \quad \Sigma \rho_s = \infty \ ,$$

$$(ii) \quad \frac{\rho_s}{\varepsilon_s \delta_s} \to 0 \ ,$$

then every limit point of the sequence $\{x^s\}$ *is a solution of problem (17).*

Proof: It is easy to see that conditions 1 and 2 are satisfied by definition of the algorithm and supposition of the theorem.

Let us define the function $W(x)$ as follows:

$$W(x) = \inf_{x^* \in X^*} \|x - x^*\|^2 \ .$$

Obviously $W^* = \{0\}$, so condition 5 is satisfied and it remains to prove that conditions 3 and 4 are satisfied as well.

Let us suppose that for some subsequence $\{x^{n_k}\}$ such that

$$\lim_{k \to \infty} x^{n_k} = x' \notin X^* \ ,$$

condition 3 is invalid. It is easy to see in that case the sequence $\{x^s\}$ has the same limit:

$$\lim_{s \to \infty} x^s = x' \ ,$$

and for arbitrarily small $\delta > 0$ there exists N_δ such that for $s > N_\delta$,

$$\|x^s - x'\| \le 4\delta \ .$$

Choose $\theta, \delta > 0$ small enough that $s > N_\delta$ for arbitrary $x^* \in X^*$ and $g_s \in \partial_{\varepsilon_s} f(x^s)$,

$$0 \le \theta \le f(x^s) - f(x^*) \le g^s(x^s - x^*) + \varepsilon_s$$

so

$$\theta \le g^s(x^s - x^*) + \varepsilon_s \le g^s(x' - x^*) + \|g^s\|\|x^s - x^*\| + \varepsilon_s$$

$$\le g^s(x' - x^*) + K\delta + \varepsilon_s \quad .$$

As $\varepsilon_s \to 0$ when $s \to \infty$ and $\delta > 0$ arbitrarily small, then, for s sufficiently large,

$$g^s(x' - x^*) \ge \theta - K\delta - \varepsilon_s \ge \frac{1}{2}\theta > 0 \quad .$$

Due to Lemma 1 of Section 3, for some

$$g^s \in \partial_{\varepsilon_s} f(x^s) \quad ,$$

$$\|z^s - g^s\| \to 0 \qquad \text{when } s \to \infty \quad ,$$

and hence, for s sufficiently large,

$$z^s(x' - x^*) \ge g^s(x' - x^*) - \|z^s - g^s\|\|x' - x^*\| \ge \frac{1}{4}\theta > 0 \quad .$$

Hence, for $m > N_\delta$,

$$(x^{m+1} - x^{N_\delta})(x' - x^*) = -\sum_{s=N_\delta}^{m} \rho_s z^s(x' - x^*)$$

$$\le -\frac{1}{4}\theta \sum_{s=N_\delta}^{m} \rho_s \to -\infty \quad ,$$

$$\text{as } m \to \infty \quad .$$

This contradicts the boundness of the sequence $\{x^s\}$ and consequently proves condition 3.

Let us denote

$$m_k = \min_{m>n_k} m : \|x_m - x^{n_k}\| > \delta > 0 \quad ,$$

where $\delta > 0$ will be specified below.

First of all let us choose $\delta > 0$ and $\theta > 0$ small enough that for any x such that

$$\|x - x'\| \leq 4\delta \quad ,$$

and for any $x^* \in X^*$,

$$f(x) - f(x^*) \geq \theta > 0 \quad . \tag{18}$$

Then we can see that, for sufficiently large k,

$$\|x^{n_k} - x'\| \leq \delta$$

and

$$\delta < \|x^{m_k} - x^{n_k}\| \leq 2\delta$$

(see Figure 2). Hence

$$\|x^{m_k} - x'\| \leq \|x^{m_k} - x^{n_k}\| + \|x' - x^{n_k}\| \leq 4\delta \quad ,$$

and inequality (18) is valid for $x = x^s : n_k \leq s \leq m_k$. In this case for $n_k \leq s \leq m_k$ and any $g_s \in \partial_{\varepsilon_s} f(x^s)$,

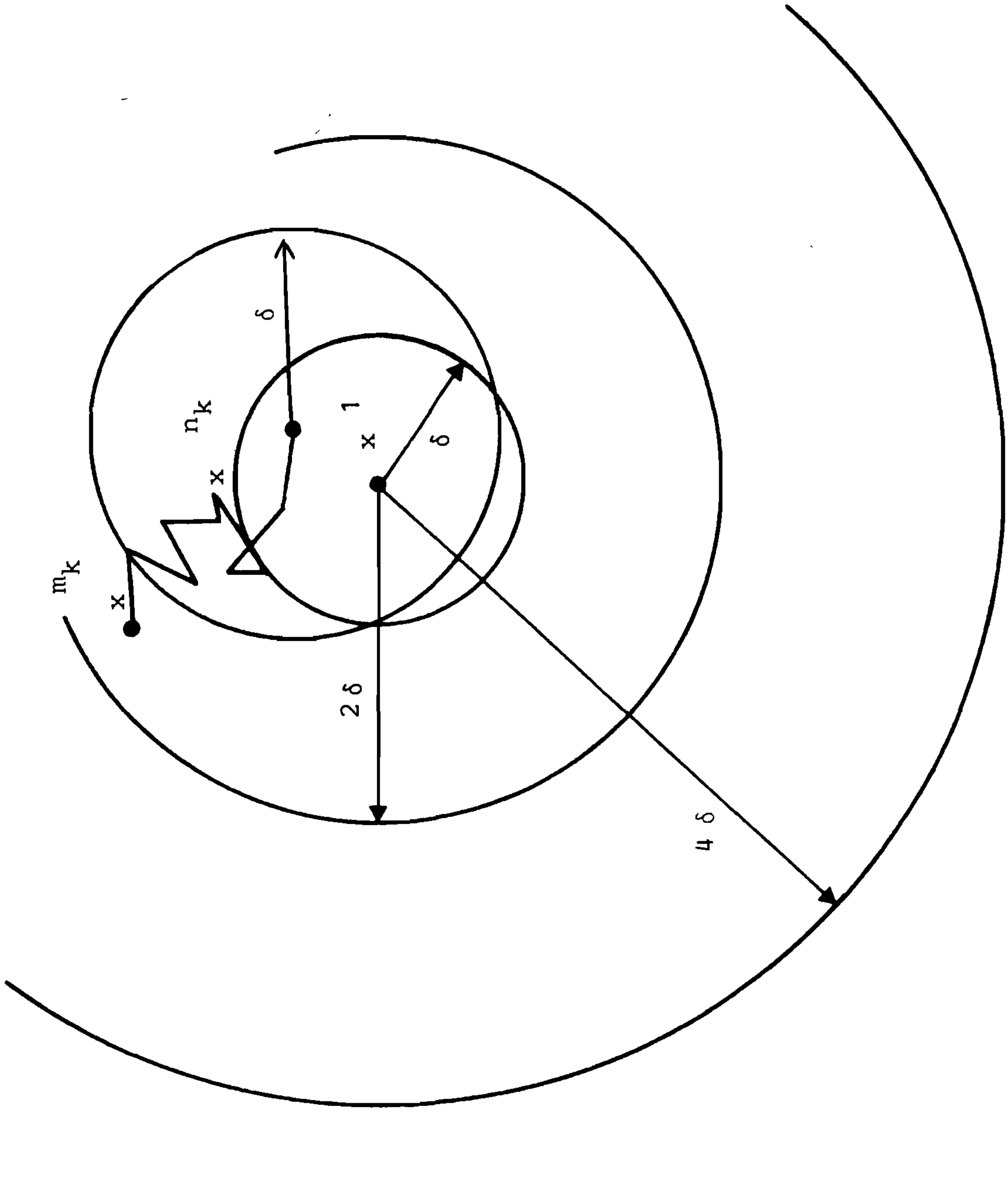

Figure 2

$$0 < \theta \le f(x^s) - f(x^*) \le g^s(x^s - x^*) + \varepsilon_s$$

$$\le g^s(x' - x^*) + \|g^s\|\|x^s - x'\| + \varepsilon_s$$

$$\le z^s(x' - x^*) + \|z^s - g^s\|\|x' - x^*\| + K\delta + \varepsilon_s \quad .$$

As we can always suppose that

$$\|z^s - g^s\| \to 0 \quad ,$$

where $s \to \infty$, then, for n_k sufficiently large,

$$z^s(x' - x^*) \ge \frac{1}{2}\,\theta > 0 \quad , \tag{19}$$

for $n_k \le s \le m_k$. Then

$$W(x^{m_k}) = \inf_{x^*\in X^*} \|x^{m_k} - x^*\|^2$$

$$= \inf_{x^*\in X^*} \{\|x^{m_k} - x^{n_k}\|^2 + \|x^{n_k} - x^*\|^2 + 2(x^{n_k} - x^*)(x^{m_k} - x^{n_k})\}$$

$$\le 4\delta^2 + \inf_{x^*\in X^*} \{\|x^{n_k} - x^*\|^2 - 2\sum_{s=n_k}^{m_k-1} \rho_s z^s(x^{n_k} - x^*)\}$$

$$\le 4\delta^2 + \inf_{x^*\in X^*} \{\|x^{n_k} - x^*\|^2 - 2\sum_{s=n_k}^{m_k-1} \rho_s z^s(x' - x^*)$$

$$+ 2\sum_{s=n_k}^{m_k-1} \rho_s \|z^s\|\|x' - x^{n_k}\|\}$$

$$\leq 4\delta^2 + \inf_{x^*\in X^*} \{\|x^{n_k} - x^*\|^2 - \theta \sum_{s=n_k}^{m_k-1} \rho_s + K\delta \sum_{s=n_k}^{m_k-1} \rho_s\}$$

$$\leq \inf_{x^*\in X^*} \|x^{n_k} - x^*\|^2 + 4\delta^2 - \frac{1}{2}\theta \sum_{s=n_k}^{m_k-1} \rho_s$$

$$= W(x^{n_k}) + 4\delta^2 - \frac{1}{2}\theta \sum_{s=n_k}^{m_k-1} \rho_s \quad .$$

As

$$\|x^{m_k} - x^{n_k}\| > \delta \quad ,$$

then

$$\delta < \sum_{s=n_k}^{m_k-1} \|z^s\|\rho_s \leq K \sum_{s=n_k}^{m_k-1} \rho_s$$

and

$$\sum_{s=n_k}^{m_k-1} \rho_s \geq \frac{\delta}{K} \quad . \tag{20}$$

Substituting (20) into the previous inequality we get

$$W(x^{m_k}) \leq W(x^{n_m}) + 4\delta^2 - \frac{1}{2}\theta \frac{\delta}{K} \quad .$$

Since we may consider $\delta \leq \frac{\theta}{16K}$ it follows that

$$W(x^{m_k}) \leq W(x^{n_k}) - \frac{\theta\delta}{4K} \quad , \tag{21}$$

and then passing to the limit in (21)

$$\overline{\lim_{k\to\infty}} W(x^{m_k}) < \lim_{k\to\infty} W(x^{n_k}) \quad ,$$

which completes the proof.

This theorem shows that we need some control of accuracy in computing ε-subgradients but the monotonical increasing of $\partial_\varepsilon f$ with respect to ε makes it possible to get rid of this limitation.

Theorem 5. Let all but condition (ii) of Theorem 4 be satisfied, and

$$\frac{\rho_s}{\delta_s} \to 0 \qquad \textit{when } s \to \infty \quad .$$

Then every limit point of the sequence $\{x^s\}$ *generated by (16) belongs to the set* X^*.

Proof: To prove this theorem it is enough to apply Theorem 4 to algorithm (16) with

$$\varepsilon'_s = \max \left\{ \left(\frac{\rho_s}{\delta_s}\right)^{1/2}, \varepsilon_s \right\} \quad .$$

Indeed on the one hand $\varepsilon'_s \geq \varepsilon_s$ so

$$\partial_{\varepsilon'_s} f(x^s) \supset \partial_{\varepsilon_s} f(x^s) \quad ,$$

and hence, if

$$g^s \in \partial_{\varepsilon_s} f(x^s) \quad ,$$

then

$$g^s \in \partial_{\varepsilon'_s} f(x^s) \quad .$$

On the other hand,

$$\frac{\rho_s}{\delta_s \varepsilon'_s} \leq \frac{\rho_s}{\delta_s} \frac{\gamma_s^{1/2}}{\rho_s^{1/2}} = \left(\frac{\rho_s}{\delta_s}\right)^{1/2} \to 0 \quad ,$$

so all assumptions of Theorem 4 are satisfied and sequence $\{x^s\}$ has all limit points belonging to the set X^*.

From the practical point of view, it is useful to get the results on convergence of algorithm (16) when $\varepsilon_s = \varepsilon =$ constant.

Theorem 6. Let conditions of Theorem 5 be satisfied, but

$$\varepsilon_s = \varepsilon > 0 \quad ;$$

then there exists a subsequence $\{x^{s_k}\}$ *such that*

$$\lim_{k \to \infty} x^{s_k} = \bar{x} \quad ,$$

$$f(\bar{x}) \leq \min_{x \in E} f(x) + \varepsilon \quad .$$

Proof: Let us denote

$$X^*_\varepsilon = \{x^*_\varepsilon : f(x^*_\varepsilon) < \min_{x \in E} f(x) + \varepsilon\}$$

and let

$$W_\varepsilon(x) = \inf_{x^* \in X^*_0} \|x - x^*\|^2 \quad . \tag{22}$$

We will show that conditions 1-4 are satisfied for the solution set X_ε^* and the function $W_\varepsilon(x)$. Of course condition 5 cannot be satisfied when $\varepsilon > 0$.

The proof will follow the same line as the proof of Theorem 4. Conditions 1 and 2 are satisfied under the assumption of the theorem and due to the properties of the algorithm. Let us suppose that 3 is invalid for some subsequence $\{x^{s_k}\}$ such that

$$\lim_{k\to\infty} x^{s_k} = x' \notin X_\varepsilon^* \quad .$$

It is easy to see that

$$\lim_{s\to\infty} x^s = x' \quad .$$

Then choose $\delta > 0$ small enough that

$$f(x) - f(\bar{x}) \geq \theta > 0$$

for any $\bar{x} \in X_\varepsilon^*$ and $\|x - x'\| \leq 4\delta$. This means that

$$f(x) - f(x^*) \geq \theta + \varepsilon$$

for any $x^* \in X_0^*$. In the same way as during the proof of Theorem 4, we can show that for k large enough and $\delta > 0$ sufficiently small,

$$\frac{1}{4}\theta \leq z^s(x' - x^*) \quad , \qquad s \geq N_\delta$$

for any $x^* \in X_0^*$ and hence for any k there exists $m_k > n_k$ such that

$$m_k = \min_{m>n_k} m : \|x^m - x^{n_k}\| > \delta \quad .$$

Then the same argument can be applied as when inequality
(19) was developed for $n_k \leq s \leq m_k$ and the same estimate as
(21) can be achieved:

$$W_\varepsilon(x^{m_k}) \leq W_\varepsilon(x^{n_k}) - \frac{\theta\delta}{4K} .$$

Passing to the limit we get

$$\overline{\lim_{k\to\infty}} W(x^{m_k}) < \lim_{k\to\infty} W(x^{n_k}) ,$$

and hence the theorem is proved.

ACKNOWLEDGEMENTS

The author is grateful to Professor R.M. Chaney, Dr. C. Lemarechal and Dr. B.N. Pshenichniy, whose helpful remarks improved this paper in a decisive way. The discussions with Professor Yu. M. Ermoliev were also very useful.

REFERENCES

[1] Shor, N.Z., *On the Structure of Algorithms for the Numerical Solution of Problems of Optimal Planning and Design*, Dissertation, Kiev, 1964 (in Russian).

[2] Ermoliev, Ju.M., and N.Z. Shor, On the Minimization of Nondifferentiable Functions, *Kibernetika*, 3, 1 (1967), 101-102 (in Russian); English translation in *Cybernetics*, 3, 1 (1972).

[3] Poliak, B.T., Minimization of Unsmooth Functions, *Zurnal Vycislitel'noy Mathematiki i Mathematicheskoy Fiziki*, 9 (1969), 509-521; English translation in *USSR Computational Mathematics and Mathematical Physics*, 9 (1969), 14-29.

[4] Shor, N.Z., and N.G. Zhurbenko, A Minimization Method Using Space Dilation in the Direction of the Difference of Two Successive Gradients, *Kibernetica*, 7 (1971), 51-59 (in Russian); English translation in *Cybernetics*, 7, 3 (1973), 450-459.

[5] Ermoliev, Ju.M., Stochastic Programming Methods, *Nauka*, Moscow, 1976 (in Russian).

[6] Balinski, M.L., and P. Wolfe, eds., Nondifferentiable Optimization, *Mathematical Programming, Study 3*, North-Holland, Amsterdam, 1975.

[7] Demyanov, V.F., ed., *The Questions of the Theory and Software Package for the Solution of Minmax Problems*, Leningrad State University Press, 1977 (in Russian).

[8] Lemarechal, C., *Nonsmooth Optimization and Descent Methods*, RR-78-4, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1978.

[9] Mifflin, R., An Algorithm for Constrained Optimization with Semismooth Functions, *Mathematics of Operations Research*, 2, 2 (1977), 191-207.

[10] Mifflin, R., *Semismooth and Semiconvex Functions in Constrained Optimization*, RR-76-21, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1976.

[11] Gupal, A.M., and V.I. Norkin, A Minimization Algorithm for Discontinuous Functions, *Kibernetica*, 2 (1977), 73-75 (in Russian).

[12] Rockafellar, R.T., *Convex Analysis*, Princeton University Press, Princeton, N.J., 1970.

[13] Hogan, W.W., Point-to-Set Maps in Mathematical Programming, *SIAM Review*, 15, 3 (1975), 591-603.

[14] Lemarechal, C., *An Estimation of Davidson Methods to Non-differentiable Problems*, in M.L. Balinski and P. Wolfe, eds., Nondifferentiable Optimization, *Mathematical Programming, Study 3*, North-Holland, Amsterdam, 1975, 95-109.

[15] Nurminski, E.A., On the Continuity of ε-Subdifferential Maps, *Kibernetica*, 5 (1977), 148-149 (in Russian).

[16] Bridgland, T.F., Trajectory Integrals of Set-Valued Functions, *Pacific J. Math.*, 33 (1970), 43-68.

[17] Nurminski, E.A., Convergence Conditions for Nonlinear Programming Algorithms, *Cybernetics*, 8, 6 (1972), 959-962.

[18] Pshenichiy, B.N., Necessary Extremum Conditions for Differential Inclusions, *Kibernetica*, 6 (1976), 60-73 (in Russian).