NON-DIFFERENTIABLE OPTIMIZATION WITH ε - SUBGRADIENT METHODS

E.A. Nurminski

November 1978 WP-78-55

2361
Laxenburg
Austria

International Institute for Applied Systems Analysis

FOREWORD

The development of optimization methods has a significant meaning for systems analysis. Optimization methods provide working tools for quantitative decision making based on correct specification of the problem and appropriately chosen solution methods. Not all problems of systems analysis are optimization problems, of course, but in any systems problem optimization methods are useful and important tools. The power of these methods and their ability to handle different problems makes it possible to analize and construct very complicated systems. Economic planning for instance would be much more limited without linear programming techniques which are very specific optimization methods. LP methods had a great impact on the theory and practice of systems analysis not only as a computing aid but also in providing a general model or structure for the systems problems.

LP techniques, however, are not the only possible optimization methods. The consideration of uncertainty, partial knowledge of the systems structure and characteristics, conflicting goals and unknown exogeneous models and consequently more sophisticated methods to work with these models.

Nondifferentiable optimization methods seem better suited to handle these features than other techniques at the present time. The theory of nondifferentiable optimization studies extremum problems of complex structure involving interactions of subproblems, stochastic factors, multi-stage decisions and other difficulties.

This publication covers one particular, but unfortunately common, situation when an estimation of the outcome from some definite decision needs a solution of a difficult auxiliary, internal, extremum problem. Solution of this auxiliary problem may be very time-consuming and so may hinder the wide analysis of different decisions. The aim of the author is to develop methods of optimal decision making which avoid direct comparison of different decisions and use only easily accessible information from the computational point of view.

1. Introduction

This paper deals with the finite-dimensional unconditional extremum problem

$$\min_{x \in E^n} f(x) \tag{1}$$

where the objective function has no continuous derivatives with respect to the variable $x = (x_1, \ldots, x_n)$. Various methods were discussed and suggested in relevant literature to solve problem (1) with many types of non-differentiable objective functions. Bibliography published in [1] gives a fairly good notion of these works. It should be emphasized, that the non-differentiability of objective function in problem (1) is, as a rule, due to complexity of the function's structure. A representative example is minimax problems where the objective function $f(x)$ is a result of maximization of some function $g(x,y)$ with respect to variables y:

$$f(x) = \max_{y \in Y} g(x,y) \tag{2}$$

In this case even a simple computation of the value of f in some fixed point may be quite a time-consuming task which requires, strictly speaking, an infinite number of operations. With this in mind, it seems to be interesting from the standpoint of theory and practice to investigate the feasibility of solution of problem (1) with an approximate computation of the function $f(x)$ and of its subgradients (if the latter are determind for a given type of nondifferentiability). To the best of our understanding, ε - subgradients of functions of the form (2), introduced by R.T. Rockafellar [2], are quite a convenient object for constructing numerical methods, and so we offer here some results generalizing efforts in this direction [3-5].

2. Weakly Convex Functions

The discussion of a class of the non-differentiable functions broader than the convex functions enables us to gain substantially in generality at the expense of a minor increase in complexity. Properties of the class which will be treated of are described by the following definition [6]:

Definition The continuous function $f(x)$ is called the weakly convex function if for each x there exists at least one vector g such that

$$f(y) \geq f(x) + (g, y-x) + r(x,y) \tag{3}$$

for all y, and the residual term $r(x,y)$ satisfies the condition of uniform smallness with respect to $\|x-y\|$ in each compact subset of E^n, i.e., in any compact set $K \subset E^n$ for any $\varepsilon > 0$ there exists $\delta_k > 0$ such that for $\|x-y\| \leq \delta_k$, $x,y \in K$

$$|r(x,y)| \, \|x-y\|^{-1} \leq \varepsilon$$

Notice that no constraints are imposed on a sign of the residual term $r(x,y)$. Furthermore, strengthening (3) it is possible to add to $r(x,y)$ any expression of the form $\phi(\|x-y\|)$, where

$$\phi(t) \leq 0 \quad , \quad \phi(t)t^{-1} \to 0 \qquad\qquad \text{for } t \to +0$$

The term weakly convex functions is suggested by analogy to the strongly convex functions studied by B.T. Polyak [7].

We will call the vector g, satisfying (3), the subgradient of the function $f(x)$ and will denote a set of subgradients at the point x by $G(x)$.

Describe some simple properties of weakly convex functions and of their subgradients.

Lemma 1. $G(x)$ is convex, closed, bounded and upper semi-continuous with respect to x.

The proof of these properties presents no special problems.

Lemma 2. Let $f(x,\alpha)$ be continuous with respect to α and weakly convex with respect to x for each α belonging to the compact topological space A. That is,

$$f(y,\alpha) - f(x,\alpha) \geq (g_\alpha, y-x) + r_\alpha(x,y) \tag{4}$$

for all y, and here $r_\alpha(x,y)$ satisfies the condition of uniform smallness uniformally with respect to $\alpha \in A$. Then

$$f(x) = \max_{\alpha \in A} f(x,\alpha) \tag{5}$$

is a weakly convex function.

The proof is rather simple.

Let

$$A(x) = \{\alpha : f(x,\alpha) = f(x)\}$$

Then, considering (4) for $\alpha \in A(x)$, we obtain

$$f(y) - f(x) \geq f(y,\alpha) - f(x,\alpha) \geq$$

$$\geq (g_\alpha, y-x) + \bar{r}_\alpha(x,y) \geq$$

$$\geq (g_\alpha, y-x) + \bar{r}(x,y)$$

where

$$-\bar{r}(x,y) = \sup_{\alpha \in A} |r_\alpha(x,y)|$$

It is easily seen that $\bar{r}(x,y)$ satisfies necessary conditions and the lemma is proved.

The proof of Lemm 2 helps in understanding the procedure of calculation of subgradients of the weakly convex functions. Specifically, for functions of the form (5) the vector $g_\alpha \in G_\alpha(x), \alpha \in A(x)$ is the subgradient of the function $f(x)$ at the point x. It follows from Lemma 1 that an arbitrary vector

$$g \in \mathrm{co}\{g_\alpha, \alpha \in A(x)\} = G(x)$$

is also the subgradient.

The finding of even one element of the set $G(x)$ may be a non-trivial problem and, ignoring efforts spent to calculate the subgradient $g_\alpha \in G_\alpha(x)$, it can be said that problems of computing $f(x)$ and of its subgradient $g \in G(x)$ are equal in complexity.

In establishing necessary extremum conditions for weakly convex functions of great importance is the existence of directional derivatives and a formula for their computation in terms of subgradients.

<u>Lemma 3.</u> The weakly convex function $f(x)$ is differentiable in any direction, and

$$\frac{\partial f(x)}{\partial e} = \lim_{h \to +0} \frac{f(x+he) - f(x)}{h} = \max_{g \in G(x)} (g,e)$$

<u>Proof.</u> Let

$$\phi(h) = f(x+he) - f(x)$$

It is easily seen that $\phi(h)$ as a function of h is weakly convex. Denote the set of subgradients of $\phi(h)$ by $G^\phi(h)$. Assume the contrary of what the lemma asserts:

$$\underline{a} = \underline{\lim}_{h \to +0} \frac{\phi(h)}{h} < \overline{\lim}_{h \to +0} \frac{\phi(h)}{h} = \overline{a}$$

and let $\{\tau_k\} = \tau$ and $\{\sigma_k\} = \sigma$ be sequences of values of h such that

$$\lim_{k\to\infty} \frac{\phi(\tau_k)}{\tau_k} = \bar{a} \uparrow$$

$$\lim_{k\to\infty} \frac{\phi(\sigma_k)}{\sigma_k} = \underline{a} \uparrow$$

Furthermore, we have:

$$\phi(\tau_k) \leq g_k^\tau \tau_k + O(\tau_k) \tag{6}$$

where

$$g_k \in G^\phi(\tau_k),\ O(\tau_k)\tau_k^{-1} \to 0 \qquad \text{for} \quad k\to\infty,\ \tau_k \to +0$$

Without loss of generality it may be assumed that

$$\lim_{k\to\infty} g_k^\tau = g^\tau$$

Dividing (6) by τ_k and passing to the limit for $k\to\infty$ we obtain

$$g^\tau \geq \bar{a}$$

By virtue of Lemma 1 $g^\tau \in G^\phi(0)$, therefore

$$\phi(\sigma_k) \geq g^\tau \sigma_k - O(\sigma_k) \tag{7}$$

Dividing (7) by σ_k and passing to the limit when $k\to\infty$ we have a contradiction that proves the differentiablility in any direction. By virtue of the weak convexity of f it is easy to obtain

$$\frac{\partial f(x)}{\partial e} \geq \max_{g\in G(x)} (g,e)$$

Now let

$$x^k = x + t_k e,\ t_k \to +0,\ k \to \infty$$

and

$$g_k \in G(x^k),\ g_k \to g \in (x)$$

Then

$$f(x) - f(x^k) \geq (g_k,\ x - x^k) + r(x^k, x)$$

The division of the above inequality by t_k and the pass to the limit when $k \to \infty$ yield:

$$\frac{\partial f(x)}{\partial e} \leq (\tilde{g}, e) \leq \max_{g \in G(x)} (g, e)$$

and thus the proof is completed.

Lemma 3 implies that the necessary condition for the point be extremal is

$$0 \in G(x^*) \tag{8}$$

however, unlike the case with the convex function, this condition is insufficient.

Local properties of the weakly convex functions do not differ from these of the convex functions but their global properties are radically dissimilar. Specifically, the weakly convex functions lack the salien feature of subgradients that enables us to prove the convergence of subgradient method, i.e., the positivity of scalar product of an arbitrary subgradient at some point X in the direction from the extremum point X^*:

$$(g,\ x - x^*) \geq 0$$

for an arbitrary $g \in G(x)$.

This and the fact, that a shift in the direction of the antigradient does not assure a decrease in value of a function being optimized both for the weakly convex objective functions and the convex functions, complicate tangibly the proof of the subgradient method convergence.

All said above about the complexity of the proof of convergence applies also to the ε - subgradient method of solution of problem (1).

Definition. The vector $g_\varepsilon \in G_\varepsilon(x)$ is called the ε - subgradient of the weakly convex function $f(x)$ if

$$f(y) - f(x) \geq (g_\varepsilon, y - x) + r(x,y) - \varepsilon \qquad (9)$$

for all y and $\varepsilon > 0$.

In (9) it is meant that $r(x,y)$ satisfies the condition of uniform smallness described above.

Properties of $G_\varepsilon(x)$ are obvious:

(i) $G_\varepsilon(x) \supset G(x)$

(ii) $G_\varepsilon(x)$ is convex, closed and bounded.

The property (i) holds out a hope of the definition of ε - subgradient being an easier task than the calculation of subgradient. Indeed, for functions of the type (5) the ε - subgradient of function $f(x)$ is an arbitrary vector

$$g_\alpha \in G_\alpha(x)$$

where

$$\alpha \in A_\varepsilon(x) = \{\alpha : f(x,\alpha) \geq f(x) - \varepsilon\}$$

or an arbitrary vector from the convex hull $co\{G_\alpha, \alpha \in A_\varepsilon(x)\}$. The proof is a standard one: for $\alpha \in A_\varepsilon(x)$

$$f(y) - f(x) \geq f(y,\alpha) - f(x,\alpha) - \varepsilon \geq$$

$$\geq (g_\alpha, y - x) + \bar{r}(x,y) - \varepsilon$$

where notations used in Lemma 2 are preserved.

The demonstrated procedure of computing ε - subgradients also implies that it is inconsistent to employ simultaneously exact computation of the objective function, one-dimensional optimization, etc. Thus, it is safe to say, that the ε - subgradient methods will be the non-relaxation ones for reasons of principles.

Difficulties that present themselves in proving the convergence of non-relaxation algorithms are of common knowledge. However, in a number of cases they pay, opening new possibilities. In the following chapter we will describe certain criteria of convergence of iterative algorithms which made it possible to prove convergence of a number of algorithms whose behaviour is substantially non-monotonic.

3. Convergence of Iterative Methods Of Non-Linear Prgoramming

General conditions of convergence of iterative procedures received attention of a lot of researchers. The most fundamental results appear to belong to W.I. Zangwill who suggested necessary and sufficient conditions of convergence of iterative methods of the mathematical programming [7]. However, the convergence theorems derived by W.I. Zangwill do not exhaust investigations conducted in this field, and many authors formulated other conditions that characterize convergence of iterative procedures. In spite of the fact that the later approaches are less general and universal they proved to be more helpful in investigations of specific algorithms. Take [7-9] as an example. It should be emphasized that in the majority of cases these works deal with convergence of algorithms whose objective function decreases monotonically as a process goes and, therefore, they are not applicable, in principle, to the case in hand. These and other reasons served as the starting point in the elaboration of conditions of convergence of iterative procedures with weakened properties of a monotonous variation of the objective function in the progress of the solution of an extremum problem. The approach set forth below is based on author's paper [12].

We will consider an algorithm of the mathematical programming as a certain rule of construction of a sequence $\{x^s\}$ of points of an n - dimensional Euclidean space E^n. Conditions of convergence of this sequence will be formulated in terms of properties of this sequence and of a certain subset X^* of the space E^n which we will call the solution set. The algorithm will be thought of as the convergent algorithm if each limit point of a sequence generated by it belongs to the set X^*.

The basic convergence theorem is formulated as follows:

<u>Theorem 1.</u> Let the sequence $\{x^s\}$ and the set X^* be such that

A1) If $\quad x^{s_k} \to x^* \in X^*$ then

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

A2) There exists a compact set K such that

$$x^s \in K$$

A3) If $x^{s_k} \to x' \notin X^*$, then there exists $\varepsilon_0 > 0$ such that for all $\varepsilon \le \varepsilon_0$ and any k's there exists a point x^{t_k}, $t_k > s_k$ such that

$$\|x^{t_k} - x^{s_k}\| > \varepsilon$$

We will assume

$$t_k = \min_{t > s_k} t: \|x^t - x^{s_k}\| > \varepsilon$$

A4) There exists a continuous function W(x) such that

$$\overline{\lim_{k\to\infty}} \; W(x^{t_k}) < \lim_{k\to\infty} W(x^{s_k}) = W(x')$$

for arbitrary sequences $\{s_k\}$, $\{t_k\}$ satisfying condition A3.

A5) The function W assumes on X^* an everywhere incomplete set of values.

Then all limit points of the sequence $\{x^s\}$ belong to the set X^*.

This theorem is proved in [12]. A version of conditions given there varies to some extent from the given above, however, proofs of both theorems are practically similar. An assertion weaker that Theorem 1 is also of interest.

Theorem 2. Under the conditions of Theorem 1 A1-A4 there exists a limit point of the sequence $\{x^s\}$ which belongs to the set X^*. The proof of this theorem employs the same arguments than those of the proof of Theorem 1.

4. Minimization Of Weakly Convex Functions

In this chapter we shall study convergence of the recurrent procedure

$$x^{s+1} = x^s - \rho_s g^s, \quad s = 0,1,\ldots \tag{10}$$

for finding the unconditional minimum of the weakly convex function f. In the above relation $\rho_s > 0$ are step multipliers, $g^s \in G_{\varepsilon_s}(x^s)$ is the ε_s - subgradient of the objectiv function f at the point x^s, $\{\varepsilon_s\}$ is some sequence of positive numbers. Requirements placed upon this sequence will be stipulated in what follows.

To prove convergence of procedure (10) requires an auxiliary geometrical lemma. In a simplified form such lemma was first proved in [6].

Lemma 4. Let D be a convex compact set which does not contain a zero and let $\{y^n\}$ be an arbitrary set of vectors from D. By means of a sequence of numbers σ_n such that

$$0 \le \sigma_n \le 1, \ \sigma_n \to 0, \ \Sigma \sigma_n = \infty$$

let us form a sequence of vectors $\{z^n\}$ as follows:

$$z^0 = y^0$$

$$z^{n+1} = z^n + \sigma_n (y^{n+1} - z^n), \quad n = 0, 1, \ldots$$

Denote by $\{n_k\}$ a sequence of indexes such that

$$(z^{n_k}, y^{n_k+1}) \geq \gamma > 0$$

Then for some $\gamma > 0$ such a sequence exists and

$$\sum_{s=n_k}^{n_{k+1}-1} \sigma_s \leq C < \infty$$

Proof. It is obvious that $\{z^n\} \subset D$. Since $0 \notin D$, then constants δ and Δ exist such that

$$0 < \delta \leq \|z^n\| \leq \Delta < \infty$$

Let us consider now the changes in the length of vectors z^n :

$$\|z^{n+1}\|^2 = \|z^n + \sigma_n (y^{n+1} - z^n)\|^2 = \|z^n\|^2 +$$

$$+ \sigma_n^2 \|y^{n+1} - z^n\|^2 + 2\sigma_n (z^n, y^{n+1} - z^n) \leq \|z^n\|^2 +$$

$$+ 4\Delta^2 \sigma_n^2 + 2\sigma_n ((z^n, y^{n+1}) - \|z^n\|^2) \leq \|z^n\|^2 +$$

$$+ 4\Delta^2 \sigma_n^2 + 2\sigma_n ((z^n, y^{n+1}) - \delta^2)$$

If for all n

$$(z^n, y^{n+1}) \leq \frac{1}{2} \delta^2$$

then

$$\|z^{n+1}\|^2 \leq \|z^n\|^2 + 4\Delta^2\sigma_n^2 - \delta^2\sigma_n$$

Since $\sigma_n \to 0$, then for sufficiently large n

$$\|z^{n+1}\|^2 \leq \|z^n\|^2 - \frac{\delta^2}{2}\sigma_n$$

Summing the above inequality with respect to n from N to N + M - 1 we obtain

$$0 \leq \|z^{N+M}\|^2 \leq \|z^N\|^2 - \tag{12}$$

$$- \frac{\delta^2}{2} \sum_{n=N}^{N+M-1} \sigma_n \leq \Delta^2 - \frac{\delta^2}{2} \sum_{n=N}^{N+M-1} \sigma_n$$

The pass to the limit when $M \to \infty$ leads to a contradiction to the supposition (11). It follows that there exists a sequence $\{n_k\}$ such that

$$(z^{n_k}, y^{n_k+1}) \geq \gamma = \frac{1}{2}\delta^2 > 0$$

Further, from (12) it follows for sufficiently large k that

$$0 \leq \Delta^2 - \frac{\delta^2}{2} \sum_{s=n_k}^{n_{k+1}-1} \sigma_s$$

Hence

$$\sum_{s=n_k}^{n_{k+1}-1} \sigma_s \leq 2\frac{\Delta^2}{\delta^2}$$

what complete the proof.

The main result which will be proved here later is the proposition about convergence of procedure (10). At first the solution set will be defined using the necessary extremum conditions:

$$X^* = \{x^* : 0 \in G(x^*)\}$$

The following theorem is valid:

Theorem 3. Let

$$\rho_s, \varepsilon_s \to +0, \; \Sigma \rho_s = \infty$$

and the sequence $\{x^s\}$ be bounded. Then all limit points of this sequence belong to the set X^*.

Proof. In proving this theorem we shall employ the general conditions of convergence described in Section 3.

The objective function f(x) is chosen as W(x) and it is demonstrated that conditions A1-A4 will be also satisfied. For simplicity, we will assume that condition A5 is satisfied.

It is obvious, that the satisfaction of conditions A1,A2 follows directly from the assumptions of the proof.

Let $\{x^{n_k}\}$ be a convergent subsequence and

$$\lim_{k \to \infty} x^{n_k} = x' \overline{\in} X^*$$

In this case $0 \overline{\in} G(x')$ and by virtue of G(x) being upper semi-continuous it is possible to choose so small $\delta > 0$ that

$$0 \overline{\in} \operatorname{co}\{G(x), \|x - x'\| \le \delta\}$$

This is also true for the ε-subgradients. It is always possible to choose so small $\varepsilon, \delta > 0$ that

$$0 \overline{\in} \operatorname{co}\{G_\gamma(x), \|x - x'\| \le \delta, \; \gamma < \varepsilon\} = \tilde{G}_{\varepsilon,\delta}(x')$$

Then, if condition A3 is not satisfied, for k's large enough

$$g^s \in \tilde{G}_{\varepsilon,\delta}(x'), \quad s \geq n_k$$

and by virtue of separation theorems there exists a vector e such that

$$(g^s,e) \leq -C < 0$$

Therewith

$$(x^{s+1},e) = (x^s - \rho_s g^s,e) =$$

$$= (x^s,e) - \rho_s (g^s,e) \geq (x^s,e) + C\,\rho_s$$

The above inequality implies because of our assumptions an unlimited growth of the inner product (x^s,e). This implication obviously contradicts to the assumption and, therefore, proves that condition A3 is satisfied.

Let for some small $\varepsilon > 0$

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

Requirements placed on ε will be refined later.

We meet the dominant difficulty at the following step of the proof; an estimation of a decrease in the objective function when passing from the point x^{n_k}. As the directions $-g^s$ are, generally speaking, not the directions of decrease in the function f(x) the problem of estimation of the function decrease is fairly difficult and rather unwieldy in view of the large number of computation.

Let us fix a sufficiently large k and examine a difference

$$f(x^m) - f(x^{n_k}) \le (g^m, x^m - x^{n_k}) + \varepsilon_m - r(x^{n_k}, x^m), \quad m > n_k$$

Estimate with greater precision the addend on the right side of this inequality.

$$(g^m, x^m - x^{n_k}) = -(g^m, \sum_{s=n_k}^{m-1} \rho_s g^s) =$$

$$= - \sum_{s=n_k}^{m-1} \rho_s (g^m, (\sum_{s=n_k}^{m-1} \rho_s)^{-1} \sum_{s=n_k}^{m-1} \rho_s g^s) =$$

$$= - \sum_{s=n_k}^{m-1} \rho_s (g^m, z_k^{m-1})$$

Vectors z_k^m can be obtained by means of the recurrent formula:

$$z_k^{s+1} = z_k^s + \sigma_s^{(k)} (g^{s+1} - z_k^s), \quad s = n_k, n_k+1, n_k+2, \ldots,$$

with the initial condition

$$z_k^{n_k} = g^{n_k}$$

and coefficients $\sigma_s^{(k)}$ equal to

$$\sigma_s^{(k)} = \rho_s (\sum_{m=n_k}^{s} \rho_m)^{-1}$$

It is easily seen that $0 \le \sigma_s^{(k)} \le 1$

$$\sum_{s > n_k} \sigma_s^{(k)} = \infty, \quad \sigma_s^{(k)} \to 0 \qquad \text{for } s \to \infty$$

Then in virtue of Lemma 4 there exists a sequence $\{s_i^k,\ i = 1,2,\ldots,\}$ of indexes such that

$$(g^{s_i^k},\ z^{s_i^k - 1}) \geq \gamma > 0$$

and here

$$f(x^{s_i^k}) - f(x^{n_k}) \leq -\gamma \sum_{s=n_k}^{s_i^k - 1} \rho_s +$$

$$+ \varepsilon_{s_i^k} - r(x^{n_k},\ x^{s_i^k})$$

Choose from the sequence $\{s_i^k,\ i = 1,\ldots,\}$ a maximum index whose value does not exceed the index m_k and denote it by ν_i^k :

$$\nu_1^k = s_i^k \leq m_k < s_{i+1}^k$$

From the inequality (Lemma 4)

$$\sum_{s=s_i^k}^{s_{i+1}^k - 1} \sigma_s^{(k)} \leq C$$

it follows that for sufficiently large k's

$$1 \geq \prod_{s=S^k}^{s_{i+1}^k - 1} (1 - \sigma_s^{(k)}) \geq p > 0$$

what implies that

$$\left(\sum_{s=n_k}^{\nu_i^k - 1} \rho_s\right) \left(\sum_{s=n_k}^{s_{i+1}^k - 1} \rho_s\right)^{-1} \geq p$$

The above inequality may be put in another form:

$$\sum_{s=\nu_1^k}^{m_k-1} \rho_s \leq q \sum_{s=n_k}^{m_k-1} \rho_s$$

where $q = 1 - p < 1$

Summing up it is possible to say that we have constructed as a result the point $x^{\nu_1^k}$ such that

$$f(x^{\nu_1^k}) - f(x^{n_k}) \leq - \gamma \sum_{s=n_k}^{\nu_1^k-1} \rho_s + \varepsilon_{\nu_1^k} - r\,(x^{n_k}, x^{\nu_1^k}) \quad (12)$$

and therewith

$$\sum_{s=\nu_1^k}^{m_k-1} \rho_s \leq q \sum_{s=n_k}^{m_k-1} \rho_s \quad (13)$$

If in a similar reasoning the point $x^{\nu_1^k}$ is considered as the initial one, than it is possible to show the existence of a point $x^{\nu_2^k}$ such that

$$f(x^{\nu_2^k}) - f(x^{\nu_1^k}) \leq - \gamma \sum_{s=\nu_1^k}^{\nu_2^k-1} \rho_s + \varepsilon_{\nu_2^k} - r(x^{\nu_1^k}, x^{\nu_2^k})$$

and

$$\sum_{s=\nu_2^k}^{m_k-1} \rho_s \leq q \sum_{s=\nu_1^k}^{m_k-1} \rho_s \leq q^2 \sum_{s=n_k}^{m_k-1} \rho_s$$

Let us fix an arbitrary small $\tau > 0$ and repeat this process a required number of times in order to construct a sequence of points $\{x^{\nu_i^k}, i = 1,2,\dots,M\}$ such that for each i inequalities similar to (13)-(14) be satisfied:

$$f(x^{\nu_{i+1}^k}) - f(x^{\nu_i^k}) \le \gamma \sum_{s=\nu_i^k}^{\nu_{i+1}^k - 1} \rho_s + \\ + \varepsilon_{\nu_i^k} - r(x^{\nu_i^k}, x^{\nu_{i+1}^k}) \quad , \tag{15}$$

$$\sum_{s=\nu_i^k}^{m_k - 1} \rho_s \le q^i \sum_{s=n_k}^{m_k - 1} \rho_s$$

and $q^m \le \tau$. It obviously suffices to repeat the above reasonings no more than $M = [\log_q \tau] + 1$ times. Summing (15) with respect to i from zero to $M - 1$ we obtain (assuming $\nu_0^k = n_k$ and denoting $\nu_M^k = t_k$):

$$f(x^{t_k}) - f(x^{n_k}) \le - \gamma \sum_{s=n_k}^{t_k - 1} \rho_s + \sum_{i=1}^{M} \varepsilon_{\nu_i^k} - \sum_{i=0}^{M-1} r(x^{\nu_i^k}, x^{\nu_{i+1}^k})$$

Addends in the right part of the inequality are evaluated separately:

$$\sum_{i=1}^{M} \varepsilon_{\nu_i^k} \le M \sup_{m \ge n_k} \varepsilon_m = M \bar{\varepsilon}_k \to 0 \qquad \text{for } k \to \infty$$

$$\left| \sum_{i=0}^{M-1} r(x^{\nu_i^k}, x^{\nu_{i+1}^k}) \right| \le \sum_{i=0}^{M-1} |r(x^{\nu i^k}, x^{\nu i+1^k}) \le$$

$$\leq M \sup_{\substack{\|x - x^{n_k}\| \leq \varepsilon \\ \|y - x^{n_k}\| \leq \varepsilon}} |r(x,y)| = M \bar{r}_\varepsilon(x^{n_k})$$

For the k's that are large enough $\|x^{n_k} - x'\| \leq \varepsilon$ therefore

$$M \bar{r}_\varepsilon(x^{n_k}) \leq M \sup_{\substack{\|x - x'\| \leq 2\varepsilon \\ \|y - x'\| \leq 2\varepsilon}} |r(x,y)| \leq \varepsilon\,\delta(\varepsilon)$$

where $\delta(\varepsilon) \to 0$ for $\varepsilon \to 0$

Finally we obtain:

$$f(x^{m_k}) - f(x^{n_k}) \leq f(x^{t_k}) - f(x^{n_k}) +$$

$$+ | f(x^{m_k}) - f(x^{t_k}) | \leq - \gamma \sum_{s=n_k}^{t_k - 1} \rho_s + M \bar{\varepsilon}_k + \varepsilon\delta(\varepsilon) +$$

$$+ C \|x^{m_k} - x^{t_k}\| \leq - \gamma \sum_{s=n_k}^{m_k - 1} \rho_s + \gamma\tau \sum_{s=n_k}^{m_k - 1} \rho_s + M \bar{\varepsilon}_k +$$

$$+ \varepsilon\delta(\varepsilon) + C' \sum_{s=t_k}^{m_k - 1} \rho_s \leq - (\gamma - \gamma\tau) \sum_{s=n_k}^{m_k - 1} \rho_s + M \bar{\varepsilon}_k +$$

$$+ \varepsilon\delta(\varepsilon) + C'\tau \sum_{s=n_k}^{m_k - 1} \rho_s \leq - (\gamma - \gamma\tau - C'\tau) \sum_{s=n_k}^{m_k - 1} \rho_s +$$

$$+ M \bar{\varepsilon}_k + \varepsilon\,\delta(\varepsilon) \quad .$$

where τ may be assumed to be so small that

$$\gamma - \gamma\tau - C'\tau > \frac{1}{2}\gamma$$

In doing so we obtain:

$$f(x^{m_k}) - f(x^{n_k}) \leq -\frac{\gamma}{2} \sum_{s=n_k}^{m_k-1} \rho_s + M\bar{\varepsilon}_k + \varepsilon\,\delta(\varepsilon) \qquad (15)$$

Furthermore,

$$\varepsilon < \| x^{m_k} - x^{n_k} \| \leq C \sum_{s=n_k}^{m_k-1} \rho_s$$

Substituting this estimate into (15) we obtain:

$$f(x^{m_k}) - f(x^{n_k}) \leq -\frac{\gamma\varepsilon}{2C} + M\bar{\varepsilon}_k + \varepsilon\,\delta\,(\varepsilon)$$

It may be always assumed that

$$\delta(\varepsilon) \leq \frac{\gamma}{4C}$$

hence

$$f(x^{m_k}) - f(x^{n_k}) \leq -\frac{\gamma\varepsilon}{4C} + M\,\bar{\varepsilon}_k$$

Passing to the limit when $k \to \infty$ we obtain:

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

which is what it was required to prove.

As a result the convergence of algorithm (10) is a sequence of the satisfaction of conditions A1-A5 of Theorem 1.

5. Convex Case

To solve the problem of convex minimization some results can be obtain describing the behaviour of process (10) in the case when $\varepsilon_s = \varepsilon = \text{const}$.

Theorem 4. Let the objective function f(x) be convex

$$\rho_s \to +0 \, , \; \Sigma \rho_s = \infty \, , \quad \varepsilon_s = \varepsilon > 0$$

Then, if the sequence $\{x^s\}$ is bounded, there exists if only one convergent subsequence $\{x^{s_k}\}$ such that

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

and

$$f(\bar{x}) \leq \min_{x\in E^n} f(x) + \varepsilon$$

Proof. The proof will be based on the same formalism as in Theorem 3. Let

$$X^* = \{x^* : f(x^*) = \min f(x), \; x \in E^n\}$$

and

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

Denote

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

In our case the role of a set of solutions will be played by X_ε^*. Let us verify whether conditions A1-A4 from Section 2 can be satisfied. It is obvious, that on no account condition A5 can be satisfied in this case and, therefore, it is possible to prove only a weakened convergence of process (10) in the spirit of Theorem 2.

Conditions A1, A2 are obviously satisfied in assumptions of this theorem: verify whether condition A3 is satisfied. Let be some subsequence:

$$\lim_{k\to\infty} x^{n_k} = x' \,\overline{\in}\, X_\varepsilon^*$$

that is,

$$f(x') > \min_{x\in E^n} f(x) + \varepsilon$$

Assume the contrary to condition A3, that is,

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

Then for an arbitrary $\delta > 0$ for a sufficiently large k

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

for $s > n_k$. Choose $\delta > 0$ in such a way that the set

$$U_{4\delta}(x') = \{x : \|x-x'\| \le 4\delta\}$$

does not intersect with the set X^*_ε : $U_{4\delta}(x') \cap X_\varepsilon^* = \phi$. Then in suppositions of the proof for an arbitary $x^* \in X^*$ and $s > n_k$:

$$\|x^{s+1} - x^*\|^2 = \|x^s - \rho_s g^s - x^*\| =$$

$$= \|x^s - x^*\|^2 + \rho_s^2 \|g^s\|^2 - 2\rho_s(g^s, x^s - x^*) \le \qquad (17)$$

$$\leq \|x^s - x^*\|^2 + C\rho_s^2 - 2\rho_s(g^s, x^s - x^*)$$

since

$$x^s \in U_{4\delta}(x')$$

then

$$\varepsilon < f(x^s) - f(x^*) \leq (g^s,\ x^s - x^*) + \varepsilon$$

whence we have for $s > n_k$

$$(g^s,\ x^s - x^*) \geq \gamma > 0$$

Substituting the above inequality into (17) we obtain

$$W(x^{s+1}) \leq W(x^s) + C\rho_s^2 - 2\gamma\rho_s$$

or for sufficiently large k

$$W(x^{s+1}) \leq W(x^s) - \gamma\rho_s \tag{18}$$

Summing (18) with respect to s from n_k to m-1 we obtain:

$$W(x^m) \leq W(x^{n_k}) - \gamma \sum_{s=n_k}^{m-1} \rho_s \tag{19}$$

Passing in the above inequality to the limit when $m \to \infty$ we have a contradiction to the boundedness of the continuous function W(x) on $U_{4\delta}(x')$. The obtained contradiction proves the fact that condition A3 is satisfied. Let

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

For k's that are large enough

$$U_\delta(x^{n_k}) \subset U_{2\delta}(x') \subset U_{4\delta}(x')$$

therfore the estimate of (19) is also valid for $m = m_k$

$$W(x^{m_k}) \leq W(x^{n_k}) - \gamma \sum_{s=n_k}^{m_k-1} \rho_s$$

However,

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

By means of the above estimate we finally obtain:

$$W(x^{m_k}) \leq W(x^{n_k}) - \frac{\gamma\delta}{C}$$

and passing to the limit when $k \to \infty$

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

that by virtue of Theorem 2 proves our preposition.

In all probability the assertion of this theorem cannot be strengthened unless additional hypotheses concerning the choice of vectors g^s from appropriate sets $G_\varepsilon(x^s)$ of ε - subgradients are involved.

It is also of interest to estimate a deviation of the limit points of the sequence $\{x^s\}$ from the set of solutions X^*_ε. If we denote

$$d = \sup_{x^*_\varepsilon \in X^*_\varepsilon} \inf_{x^* \in X^*} \|x^* - x^*_\varepsilon\|$$

then from geometrical considerations it is easily shown that all limit points of the sequence $\{x^s\}$ occure in the set

$$x^* + d_\varepsilon S$$

where S is a unit ball and the addition is meant in Minkovsky's sense.

6. Appendices and Generalizations

An essential feature that distinguishes the result of Theorem 3 as compared to that obtained earlier in [13] is, as applied to minimax problems of the type

$$\min_x \max_y f(x,y) \tag{20}$$

the possibility to rid oneself of the check of exactness of the solution of an auxiliary problem of finding the internal maximum:

$$\phi(x) = \max_y f(x,y)$$

This enables us to justify the application of Arrow-Gurwitz' method

$$x^{s+1} = x^s - \rho_s f'_x(x^s,y^s) \tag{21}$$

$$y^{s+1} = y^s + \delta_s f'_y(x^s,y^s) \tag{22}$$

in the solution of problem (20) on the basis of broader assumptions than common assumptions of strict convexity-concavity or similar ones. Under some of them concerning the relation between step multipliers it proves to be possible to consider iterative relation (2) as the ε - subgradient method of minimization of the function $\phi(x)$. Convergence of method (21)-(22) is here an implication of Theorem 3. Results obtained in this field are described in more detail in [14]. Of great practical interest is also the development of methods for regulating step multipliers in procedure (10). Basically, Theorem 3 asserts that the ε - subgradient methods converge under the same assumptions as the subgradient methods. In all likelihood, ideas that underlie the subgradient methods are applicable to the ε - subgradient methods when their step multipliers are regulated and, furthermore, the computational effect is also the same.

A non-formal requirement here consists in giving up the exact computation of the objective function as stated earlier in the introduction to this paper. For instance, the generalization on the case of ε - subgradient method of step regulation [11] presents no difficulties.

References

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

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

[3] Bertsecas, D.P. and S.K. Mitter, A Descent Numerical Method for Optimization Problems with Nondifferentiable Cost Functionals, *SIAM Journal Control*, Vol. 11, 4 (1973).

[4] Lemarechal, C., Nondifferentiable Optimization; Subgradient and ε - subgradient Methods, Lecture Notes: Numerical Methods in Optimization and Operation Research, Springer Verlag, August 1975, 191-199.

[5] Rockafellar, R.T., The Multiplier Method of Hesten and Powell Applied to Convex Programming, *JOTA*, Vol. 12, 6 (1974).

[6] Nurminski, E.A., The Quasigradient Method for the Solving of Nonlinear Programming Problems, *Cybernetics*, Vol. 9, 1, (Jan-Feb 1973), 145-150, Plenum Publishing Corporation, N.Y., London.

[7] Zangwill, W.I., Convergence Conditions for Nonlinear Programming Algorithms, *Management Science*, Vol. 16, 1 (1969), 1-13.

[8] Wolf, P., *Convergence Theory in Nonlinear Programming*, North-Holland Publishing Company, 1970, 1-36.

[9] Meyer, G.G.L., A Systematic Approach to the Synthesis of Algorithms, *Numerical Mathematics*, Vol. 24, 4 (1975), 277-290.

[10] Rheinboldt, W.C., A Unified Convergence Theory for a Class of Iterative Processes, *SIAM Journal Numerical Analysis*, Vol. 5, 1 (1968),

[11] Nurminski, E.A. and A.A. Zhelikhovski, Investigation of One Regulating Step, *Cybernetics*, Vol. 10, 6 (Nov-Dec 1974), 1027-1031, Plenum Publishing Corporation, N.Y., London.

[12] Nurminski, E.A., Convergence Conditions for Nonlinear Programming Algorithms, *Kybernetika*, 6 (1972), 79-81 (in Russian).

[13] Nurminski, E.A. and A.A. Zhelikhovski, ε - Quasigradient Method for Solving Nonsmooth External Problems, *Cybernetics*, Vol. 13, 1 (1977), 109-114, Plenum Publishing Corporation, N.Y., London.

[14] Nurminski, E.A. and P.I. Verchenko, Convergence of Algorithms for Finding Saddle Points, *Cybernetics*, Vol. 13, 3, 430-434, Plenum Publishing Corporation, N.Y., London.