BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 1(62), 2010, Pages 33–46
ISSN 1024–7696

A probabilistic method for solving minimax problems with general constraints

Anatol Godonoaga, Pavel Balan

Abstract. The method proposed in paper solves a convex minimax problem with a set of general constraints. It is based on a schema elaborated previously, but with constraints that can be projected on quite elementary. Such kind of problems are often encountered in technical, economical applied domains etc. It does not use penalty functions or Lagrange function – common toolkit for solving above mentioned problems. Movement directions have a stochastic nature and are built using estimators corresponding to target function and functions from constraints. At the same time every iteration admits some tolerance limits regarding non-compliance with constraints conditions.

Mathematics subject classification: 49M37, 90C15, 90C25, 90C30, 90C47, 49K35, 49K45.

Keywords and phrases: Minimax problems, stochastic, convex, nondifferentiable, optimization, subgradient, constraints, probability repartition, estimator, almost certain, with probability 1, convergence, Borel-Cantelli.

The following problem is considered:

$$\begin{cases} F(x) = \max\limits_{y \in Y_f} f(x,y) \to \min \\ \Phi(x) = \max\limits_{y \in Y_\varphi} \varphi(x,y) \le 0 \\ \quad x \in X \end{cases} \tag{1}$$

where X represents a compact and convex set in Euclidian space E^m, the sets Y_f, Y_φ are compact sets in E^{m_1} and E^{m_2}correspondingly. Suppose that the set of optimal solutions $X^* \neq \emptyset$.

Let us define:

$$\begin{aligned} &V(x,\varepsilon) = \{z \in E^n : \|x - z\| < \varepsilon\}, \\ &V(X,\varepsilon) = \bigcup_{x \in X} (x,\varepsilon), \\ &V_X(X^*,\varepsilon) = V(X^*,\varepsilon) \bigcap X, \\ &W_X(\tilde{x},r) = (V(\tilde{x},r) \bigcap X) \setminus V(X^*,\varepsilon), r > 0, \\ &W_Y(y,r) = V(y,r) \bigcap Y, r > 0. \end{aligned} \tag{2}$$

The functions $f(x,y_f)$ and $\varphi(x,y_\varphi)$ are supposed to be convex on $V(X,\varepsilon^*)$ for some $\varepsilon^* > 0$ and continuous on $V(X,\varepsilon^*) \times Y_f$ and $V(X,\varepsilon^*) \times Y_\varphi$ correspondingly.

Let's admit that on the sets Y_f, Y_φ probability repartitions $P_f(\cdot)$, $P_\varphi(\cdot)$ are defined that satisfy the conditions:

$$\int\limits_{Y_f} P_f(dy) = 1, \ \int\limits_{Y_\varphi} P_\varphi(dy) = 1. \tag{3}$$

For $\forall r > 0 \ \exists \gamma > 0$:

$$\begin{array}{llll} \int\limits_{W_Y(y,r)} P_f(dz) \geq \gamma, & \text{if} \quad Y = Y_f & \text{for every} & y \in Y_f \ , \\ \int\limits_{W_Y(y,r)} P_\varphi(dz) \geq \gamma, & \text{if} \quad Y = Y_\varphi & \text{for every} & y \in Y_\varphi. \end{array} \tag{4}$$

1 Method description

Starting element $x^0 \in X$ is arbitrary taken. The sequence $\{x^k\}_{k\geq 1}$ is built. Let's admit that the approximate solution of order k – the element x^k – is already obtained. The approximation x^{k+1} is determined in the following way:

(A1) Two random variables $\xi \in Y_f$, $\psi \in Y_\varphi$ are simulated in series $m_k \geq 1$, $l_k \geq 1$ of independent probes with distribution laws P_f and P_φ correspondingly. More specifically, the sets $M_k = \{\xi_1, \xi_2, \ldots, \xi_{m_k}\}$, $L_k = \{\psi_1, \psi_2, \ldots, \psi_{l_k}\}$ are generated on each iteration k that contain independent realizations of random vectors $\xi(y_f) = y_f \in Y_f$, $\psi(y_\varphi) = y_\varphi \in Y_\varphi$.

(A2) The elements $y_f^k(x) = \xi_i \in M_k, 1 \leq i \leq m_k$, $y_\varphi^k(x) = \psi_j \in L_k, 1 \leq j \leq l_k$ are indicated:

$$\begin{aligned} f\left(x^k, y_f^k(x)\right) &= \max_{y \in M_k} f\left(x^k, y\right), \\ \varphi\left(x^k, y_\varphi^k(x)\right) &= \max_{y \in L_k} \phi\left(x^k, y\right). \end{aligned} \tag{5}$$

(A3) $y_f^k \in \left\{y_f^{k-1}, y_f^k(x)\right\}$, $y_\varphi^k \in \left\{y_\varphi^{k-1}, y_\varphi^k(x)\right\}$ *aredetermined* where:

$$\begin{aligned} f\left(x^k, y_f^k\right) &= \max\left\{f\left(x^k, y_f^{k-1}\right), \ f\left(x^k, y_f^k(x)\right)\right\}, \quad \text{where } y_f^0 = y_f^0(x), \\ \varphi\left(x^k, y_\varphi^k\right) &= \max\left\{\varphi\left(x^k, y_\varphi^{k-1}\right), \ \varphi\left(x^k, y_\varphi^k(x)\right)\right\}, \quad \text{where } y_\varphi^0 = y_\varphi^0(x). \end{aligned} \tag{6}$$

Definition 1. $f\left(x^k, y_f^k\right)$, $\varphi\left(x^k, y_\varphi^k\right)$ are called *estimators* of the functions $F(x)$ and $\Phi(x)$, correspondingly, for $x = x^k$.

(A4) The new element x^{k+1} is built using the relation:

$$x^{k+1} = \prod\nolimits_X \left(\tilde{x}^{k+1}\right), \; \tilde{x}^{k+1} = x^k - \rho_k \eta^k \tag{7}$$

where $\prod_X (\tilde{x})$ represents the projection of the element $\tilde{x} \in E^m$ on the set X, that is $\prod_X (\tilde{x})$ represents the closest element from X regarding $\tilde{x}$; ρ_k is the step value corresponding to iteration k.

(A5) The sequence of vectors $\{\eta^k\}$ is defined in the following way:

$$\eta^k = \begin{cases} \dfrac{g^k}{\|g^k\|}, & \text{if} \quad g^k \neq \bar{0}, \; k = 0, 1, 2, \ldots \\ \bar{0}, \; \text{for } g^k = \bar{0}. \end{cases} \tag{8}$$

(A6) The vector g^k is built as follows:

$$g^k = g^k\left(x^k\right) = \begin{cases} \partial f\left(x, y_f^k\right) & \text{for } \; x = x^k, \; \text{ if } \; \varphi\left(x^k, y_\varphi^k\right) \leq \tau_k, \\ \partial \varphi\left(x, y_\varphi^k\right) & \text{for } \; x = x^k, \; \text{ if } \; \varphi\left(x^k, y_\varphi^k\right) > \tau_k. \end{cases} \tag{9}$$

Here $\partial f\left(x^k, y_f^k\right)$ denotes the **subgradient** of the function $f\left(x, y_f^k\right)$ [2], and, respectively, $\partial \varphi\left(x^k, y_\varphi^k\right)$ is the **subgradient** of the function $\varphi\left(x, y_\varphi^k\right)$ for $x = x^k$. The vector g^0 is considered to be an arbitrary, but bounded vector.

At the same time we consider that the numerical sequence $\{\rho_k\}$ satisfies classical requirements that ensure the convergence of the methods with programmable modification of the step:

$$\rho_k > 0, \; \rho_k \to 0, \; \sum_{k=0}^{\infty} \rho_k = \infty. \tag{10}$$

Additionally, for any number $\tau \in (0, 1)$ we require the existence of a sequence $\{\bar{\varepsilon}_k\}$ with properties:

$$\bar{\varepsilon}_k \to 0, \;\; \frac{\bar{\varepsilon}_k}{\rho_k} \to \infty \tag{11}$$

so that for $\forall r_k \in \left[\frac{\bar{\varepsilon}_k}{2}, \bar{\varepsilon}_k\right]$ occurs the convergence of the series:

$$\sum_{kk=0}^{\infty} \tau^{L(k, r_k)} < \infty \tag{12}$$

where

$$L(k, r_k) = \begin{cases} 0, \; \text{if } \rho_k \geq r_k \quad \text{or} \quad k = 0, \\ s_k, \;\; \text{if} \;\; \sum\limits_{l=k-s_k}^{k} \rho_l < r_k \;\; \text{and} \;\; \sum\limits_{l=k-s_k-1}^{k} \rho_l \geq r_k. \end{cases} \tag{13}$$

In other words s_k is the biggest integer number among all numbers $j \geq 0$ that satisfies the relation $\sum\limits_{l=k-j}^{k} \rho_l < r_k$.

We will show that such numerical sequences $\{\rho_k\}$ and $\{\bar{\varepsilon}_k\}$ exist that conforms to the requirements (10)–(13). Above mentioned are justified by the following lemma:

Lemma 1. *The sequences of the form* $\rho_k = \dfrac{c}{k^\alpha + d}$, $c > 0$, $d \geq 0$, $\alpha \in (0, 1]$ *and* $\bar{\varepsilon}_k = \dfrac{p}{k^\beta + q}$, $p > 0$, $q \geq 0$, $\beta \in (0,\ \alpha)$ *satisfy the (10)–(13) requirements.*

Proof. It is obvious that $\lim\limits_{k\to\infty} L(k, r_k) = \lim\limits_{k\to\infty} s_k = \infty$. For consecutive values of $k = 0, 1, 2, \ldots$ the resulting values of $L(k, r_k)$ have the form:

$$\underbrace{0, \ldots, 0}_{0 \leq C_0 \text{ times}},\ \underbrace{1, \ldots, 1}_{0 \leq C_1 \text{ times}},\ \ldots, \underbrace{s_k, s_k, \ldots, s_k}_{0 \leq C_i \text{ times}}, \tag{14}$$

$$\underbrace{(s_k + 1), \ldots, (s_k + 1)}_{0 \leq C_{i+1} \text{ times}}, \underbrace{(s_k + 2), \ldots, (s_k + 2)}_{0 \leq C_{i+2} \text{ times}}, \ldots \tag{15}$$

In other words $L(k, r_k)$ takes the value 0 for C_0 times, the value 1 for C_1 times etc., the value s_k for C_i times, where $i = s_k$. We find out that the sequence $\{C_i\}$, $i = 0, 1, \ldots$, is bounded. If we suppose the contrary, it means that exists a value $C_j \in \{C_i\}$ that can be however big. This implies that starting from some $k \geq k'$ all $L(k, r_k) = s_{k'}$. As a result, starting from k' all the values ρ_l from (13) contradict the requirement (11). Thus, there exists a number $C < \infty$ so that $C_i < C$, $\forall i = 0, 1, \ldots$ So, we can conclude that the sequence $\{s_k\}$ can take values however big ($s_k \to \infty$).

Further we take an arbitrary, but fixed number $\tau \in (0, 1)$. The numerical series:

$$\begin{aligned}
&\Bigg(\underbrace{\tau^0 + \ldots + \tau^0}_{C \text{ times}}\Bigg) + \Bigg(\underbrace{\tau^1 + \ldots + \tau^1}_{C \text{ times}}\Bigg) + \ldots \\
&+ \Bigg(\underbrace{\tau^{s_k} + \ldots + \tau^{s_k}}_{C \text{ times}}\Bigg) + \Bigg(\underbrace{\tau^{s_k+1} + \ldots + \tau^{s_k+1}}_{C \text{ times}}\Bigg) + \ldots = \\
&= C\tau^0 + C\tau^1 + \ldots + C\tau^{s_k} + C\tau^{s_k+1} + \ldots = \\
&= C\left(\tau^0 + \tau^1 + \ldots + \tau^{s_k} + \tau^{s_k+1} + \ldots\right) = C \sum_{k=0}^{\infty} \tau^k = \frac{C}{1 - \tau} < \infty.
\end{aligned} \tag{16}$$

But, on the other hand:

$$\sum_{k=0}^{\infty} \tau^{L(k, r_k)} \leq C \sum_{k=0}^{\infty} \tau^k. \tag{17}$$

That leads us to the satisfaction of the (12) requirement. Lemma is proved. □

Now let's get back to the method of computation of the sequence $\left\{x^k\right\}$. It is the moment to remark that the iterative process can be modified, namely different distribution laws are applied for definition and simulation of random variables ξ, ψ for every new iteration. This can favor the increase of convergence speed in a certain sense of the sequence $\left\{x^k\right\}$.

The idea of using the subgradients of target function $F(x)$, in case that $\Phi\left(x^k\right) \leq 0$, and subgradients of the function $\Phi\left(x\right)$, if $\Phi\left(x^k\right) > 0$, for solving a convex model, is launched for the first time by B. Polyak in paper [1].

The stochastic subgradient method for solving a convex problem is defined in the following way:

$$\begin{cases} F\left(x\right) = \max\limits_{y\in Y_f} f\left(x,y\right) \rightarrow \min \\ \quad x \in X \end{cases}$$

is realized and argued in [5]. Paper [4] describes this method that is developed using the operation of normalization of subgradients and the convergence is established in the same probabilistic terms. The proof of the convergence is based on two principal stages. We will use and develop the mathematical mechanism used in [4] for arguing the method (A1)-(A6) when solving the problem (1). Thus, the following affirmation takes place:

Theorem 1. *Let's suppose that along with conditions mentioned above following take place:*

$$\tau_k > 0, \quad \tau_k \rightarrow 0, \quad \sum_{k=0}^{\infty} \rho_k \tau_k = \infty, \quad \frac{\tau_k}{\rho_k} \rightarrow \infty. \tag{18}$$

Then, for $\forall \varepsilon > 0$ *fixed, all elements of the random sequence* $\left\{x^k\right\}_{k\geq 0}$, *obtained as a result of application of the described method (A1) -(A6) , are localized almost certain (with probability 1) in vicinity* $V\left(X^*, 2\varepsilon\right)$, *but excepting a finite number of elements. Formally this can be represented in the following way:*

$$P\left\{\lim_{k\rightarrow\infty} \min_{x^*\in X^*} \left\|x^k - x^*\right\| = 0\right\} = 1,$$

where $x^k = x^k\left(\theta^0, \theta^1, \ldots, \theta^{k-1}\right)$, $\quad \theta^k \in \Theta^k = \left(M_k \times L_k\right)$.

Proof. If $X \subset V\left(X^*, 2\varepsilon\right)$ then the statement is obvious. Let's admit $X \backslash V\left(X^*, 2\varepsilon\right) \neq \emptyset$. We mention here that on every iteration k for the initial model (1) is associated the following problem:

$$\begin{cases} F\left(x\right) = \max\limits_{y\in Y_f} f\left(x,y\right) \rightarrow \min, \\ \quad \Phi\left(x\right) \leq \tau_k, \\ \quad x \in X \end{cases} \tag{19}$$

or, the group $\left\{f\left(x^k, y_f^k\right), \varphi\left(x^k, y_\varphi^k\right), \tau_k, X\right\}$ corresponds to the iteration k, in order to determine the direction η^k that will lead to obtaining the next element–x^{k+1}.

Two stages for proof development will be accentuated.

Stage 1. Firstly, the existence of a subsequence $\left\{x^{k_l}\right\} \subset \left\{x^k\right\}_{k\geq 0}$ that almost certain is contained in $V_X\left(X^*, \varepsilon\right)$ will be proved, i.e. $P\left\{\exists \left\{x^{k_l}\right\} \subset \left\{x^k\right\}_{k\geq 0} : x^{k_l} \in V_X\left(X^*, \varepsilon\right)\right\} = 1$.

Let's suppose the contrary. In this case for some $q \in (0,1)$ a natural number $K_q < \infty$ can be indicated such that the following event is produced

$$A_1 = \left\{\exists K_q : \forall k \geq K_q, \left\|x^k - x^*\right\| \geq \varepsilon, \text{ or } x^k \notin V_x\left(X^*, \varepsilon\right), \forall x^* \in X^*\right\} \tag{20}$$

with probability $P\left(A_1\right) \geq q$.

Let's denote $X_\varepsilon = X \backslash V\left(X^*, \varepsilon\right)$.

Since the functions $F(x)$, $\Phi(x)$ and their estimators $f\left(x, y_f^k\right)$, $\varphi\left(x, y_\varphi^k\right)$ are convex, the following inequalities are valid [2]:

$$\begin{aligned} & F\left(x^*\right) - F\left(x^k\right) \geq \left(\partial F\left(x^k\right), x^* - x^k\right),\ f\left(x^{k+1}, y_f^k\right) - f\left(x^k, y_f^k\right) \geq \\ & \geq \left(\partial f\left(x^k, y_f^k\right), x^{k+1} - x^k\right), \\ & \Phi\left(x^*\right) - \Phi\left(x^k\right) \geq \left(\partial \Phi\left(x^k\right), x^* - x^k\right),\ \varphi\left(x^{k+1}, y_\varphi^k\right) - \varphi\left(x^k, y_\varphi^k\right) \geq \\ & \geq \left(\partial \varphi\left(x^k, y_\varphi^k\right), x^{k+1} - x^k\right) \end{aligned} \tag{21}$$

for $\forall x^* \in X^*,\ \forall x^k, x^{k+1} \in X$.

Taking into consideration all properties enumerated above, two constants $C_1 > 0,\ C_2 > 0$ may be chosen, such that $\|x' - x''\| \leq C_1,\ \forall x', x'' \in X$ and $\|\partial F(x)\| \leq C_2$, $\|\partial \Phi(x)\| \leq C_2$, $\left\|\partial f\left(x, y_f^k\right)\right\| \leq C_2$, $\left\|\partial \varphi\left(x, y_\varphi^k\right)\right\| \leq C_2$, $\forall x \in X, \forall y_f \in Y_f, \forall y_\varphi \in Y_\varphi$.

Let's consider the case $\varphi\left(x^k, y_\varphi^k\right) \leq \tau_k$ and $x^k \in X_\varepsilon$. Since the function $F(x)$ is convex, results that exists the number $\Delta_F = \Delta(\varepsilon) > 0$ such that

$$\inf_{x \in X_\varepsilon, x^* \in X^*} \left(F(x) - F\left(x^*\right)\right) = 2\Delta_F \tag{22}$$

or, on basis of (21):

$$\left(\partial F\left(x^k\right), x^k - x^*\right) \geq 2\Delta_F, \tag{23}$$

$$\frac{\left(\partial F\left(x^k\right), x^k - x^*\right)}{\left\|\partial F\left(x^k\right)\right\| \cdot \left\|x^k - x^*\right\|} \geq \frac{\left(\partial F\left(x^k\right), x^k - x^*\right)}{C_2 \cdot C_1} \geq \frac{2\Delta_F}{C_1 \cdot C_2}.$$

From (22) it follows that for $\forall \tilde{x} \in X_\varepsilon$:

$$f\left(\tilde{x}, y_f\left(\tilde{x}\right)\right) - F\left(x^*\right) \geq 2\Delta_F \tag{24}$$

where $y_f\left(\tilde{x}\right)$ is **such an element from** Y_f **that** $f\left(\tilde{x}, y_f\left(\tilde{x}\right)\right) = F\left(\tilde{x}\right)$.

Taking into consideration the last inequality and the continuity of the function $f\left(x, y_f\right)$ regarding $\left(x, y_f\right) \in X \times Y_f$, we conclude that for $\forall \tilde{x} \in X_\varepsilon$ a number $r_0\left(\tilde{x}\right) > 0$ corresponds, so that:

$$f\left(x, y_f\right) \geq F\left(x^*\right) + \frac{3}{2}\Delta_f \tag{25}$$

as soon as $x \in W_X\left(\tilde{x}, r_0\left(\tilde{x}\right)\right)$ and $y_f \in W_{Y_f}\left(y_f\left(\tilde{x}\right), r_0\left(\tilde{x}\right)\right)$.

The set X_ε is compact. Therefore, there exists the number

$$r_0 = \min\left\{\min_{\tilde{x}\in X_\varepsilon} r_0\left(\tilde{x}\right),\ \varepsilon\right\} > 0. \tag{26}$$

Hence, the inequality (25) is satisfied for all $\forall \tilde{x} \in X_\varepsilon$, $x \in W_X\left(\tilde{x}, r_0\right)$, $y_f \in W_{Y_f}\left(y_f\left(\tilde{x}\right), r_0\right)$.

Similarly, in case that $\varphi\left(x^k, y_\varphi^k\right) > \tau_k$ and $x^k \in X_\varepsilon$, it follows

$$\Phi\left(x\right) - \Phi\left(x^*\right) \geq 2\tau_k \tag{27}$$

or, on basis of inequality from (21):

$$\left(\partial\Phi\left(x^k\right), x^k - x^*\right) \geq 2\tau_k, \tag{28}$$

$$\frac{\left(\partial\Phi\left(x^k\right), x^k - x^*\right)}{\left\|\partial\Phi\left(x^k\right)\right\| \cdot \left\|x^k - x^*\right\|} \geq \frac{\left(\partial\Phi\left(x^k\right), x^k - x^*\right)}{C_2 \cdot C_1} \geq \frac{2\tau_k}{C_1 \cdot C_2}.$$

From (27) it follows that for $\forall \tilde{x} \in X_\varepsilon$:

$$\varphi\left(\tilde{x}, y_\varphi\left(\tilde{x}\right)\right) - \Phi\left(x^*\right) \geq 2\tau_k \tag{29}$$

where $y_\varphi\left(\tilde{x}\right)$ **is such an element from** Y_φ **that** $\varphi\left(\tilde{x}, y_\varphi\left(\tilde{x}\right)\right) = \Phi\left(\tilde{x}\right)$.

Taking into consideration the last inequality and the continuity of the function $\varphi\left(x, y_\varphi\right)$ regarding $\left(x, y_\varphi\right) \in X \times Y_\varphi$, we conclude that for $\forall \tilde{x} \in X_\varepsilon$ a number $r_0\left(\tilde{x}\right) > 0$ corresponds so that:

$$\varphi\left(x, y_\varphi\right) \geq \Phi\left(x^*\right) + \frac{3}{2}\tau_k \tag{30}$$

as soon as $x \in W_X\left(\tilde{x}, r_0\left(\tilde{x}\right)\right)$ and $y_\varphi \in W_{Y_\varphi}\left(y_\varphi\left(\tilde{x}\right), r_0\left(\tilde{x}\right)\right)$.

As was specified previously, the set X_ε is compact. Therefore, there exists the number

$$r_0 = \min\left\{\min_{\tilde{x}\in X_\varepsilon} r_0\left(\tilde{x}\right),\ \varepsilon\right\} > 0. \tag{31}$$

Hence, the inequality (30) is satisfied for all $\forall \tilde{x} \in X_\varepsilon$, $x \in W_X(\tilde{x}, r_0)$, $y_\varphi \in W_{Y_\varphi}(y_\varphi(\tilde{x}), r_0)$.

Let's consider some numbers δ_F, δ_Φ^k from intervals $\left(0, \dfrac{2\Delta_F}{C_1 \cdot C_2}\right)$, $\left(0, \dfrac{2\tau_k}{C_1 \cdot C_2}\right)$ and label $\tilde{\delta}_k = \min\left\{\delta_F, \delta_\Phi^k\right\}$. Particularly, δ_F, δ_Φ^k can be taken as midpoints of the intervals $\left(0, \dfrac{2\Delta_F}{C_1 \cdot C_2}\right)$, $\left(0, \dfrac{2\tau_k}{C_1 \cdot C_2}\right)$:

$$\delta_F = \frac{\Delta_F}{C_1 \cdot C_2}, \ \delta_\Phi^k = \frac{\tau_k}{C_1 \cdot C_2} \tag{32}$$

As a result the following is obtained:

$$\begin{aligned} &\left(\partial F\left(x^k\right), x^k - x^*\right) \geq 2\tilde{\delta}_k \left\|\partial F\left(x^k\right)\right\| \cdot \left\|x^k - x^*\right\|, \text{ if } \varphi\left(x^k, y_\varphi^k\right) \leq \tau_k, \\ &\left(\partial \Phi\left(x^k\right), x^k - x^*\right) \geq 2\tilde{\delta}_k \left\|\partial \Phi\left(x^k\right)\right\| \cdot \left\|x^k - x^*\right\|, \text{ if } \varphi\left(x^k, y_\varphi^k\right) > \tau_k. \end{aligned} \tag{33}$$

The following events are being considered:

1. $A_1^k = \left\{\left(\eta^k, x^k - x^*\right) \geq \tilde{\delta}_k \left\|x^k - x^*\right\|, \forall x^* \in X^*\right\}$. Obviously, the opposite event with regards to A_1^k has the following form:
$\overline{A_1^k} = \left\{\exists x^* \in X^* : \left(\eta^k, x^k - x^*\right) < \tilde{\delta}_k \left\|x^k - x^*\right\|\right\}$;

2. $D_1 = \left\{\bigcup\limits_{k=K_\delta}^{\infty} \bigcap\limits_{i=k}^{\infty} A_1^i\right\}$, or, in other words, occurs all A_1^k $(k \geq K_q)$, without, perhaps, a finite number. It is obvious that $\overline{D_1} = \left\{\bigcap\limits_{k=K_\delta}^{\infty} \bigcup\limits_{i=k}^{\infty} \overline{A_1^i}\right\}$, or, in other words, an infinite number of events $\overline{A_1^k}$ are produced.

Let us evaluate $P(A_1)$. In order to do this let's represent

$$P(A_1) = P\left(A_1 \bigcap \left(D_1 \bigcup \overline{D_1}\right)\right) = P\left(A_1 \bigcap D_1\right) + P\left(A_1 \bigcap \overline{D_1}\right).$$

Both terms from the last expression will be estimated.

From the realization of event $A_1 \bigcap D_1$ follows the existence of such a natural number $K_\delta < \infty$ that for all $k \geq K_\delta$ and $\forall x^* \in X^*$ the following inequality occurs

$$\left(\eta^k, x^k - x_k^*\right) \geq \tilde{\delta}_k \left\|x^k - x_k^*\right\|. \tag{34}$$

Taking into consideration (34), for $k \geq K_\delta$ we have the following sequence of relations:

$$\left\|x^{k+1}-x^*\right\|^2 \leq \left\|x^k-\rho_k\eta^k-x^*\right\|^2 = \left\|x^k-x^*\right\|^2 - 2\rho_k\left(x^k-x^*,\eta^k\right)+\rho_k^2\left\|\eta^k\right\|^2 \leq$$
$$\leq \left\|x^k-x^*\right\|^2 - 2\rho_k\tilde{\delta}_k\left\|x^k-x^*\right\| + \rho_k^2 \leq \left\|x^k-x^*\right\|^2 - 2\rho_k\tilde{\delta}_k\varepsilon+\rho_k^2 =$$
$$= \left\|x^k-x^*\right\|^2 - \rho_k\left(2\tilde{\delta}_k\varepsilon-\rho_k\right).$$

Because $\rho_k \underset{k\to\infty}{\to} 0$, for some K_Φ: $\delta_F > \delta_\Phi^k$ or $\tilde{\delta}_k=\delta_\Phi^k$, as soon as $k \geq K_\Phi$. According to (18), (32) for some $K_\varepsilon \geq K_\Phi$: $\rho_k \leq \tilde{\delta}_k\varepsilon$, as soon as $k \geq K_\varepsilon$. Evidently, for $k \geq \hat{k} = \max\{K_\delta, K_\varepsilon\}$:

$$\left\|x^{k+1}-x^*\right\|^2 \leq \left\|x^k-x^*\right\|^2 - \rho_k\tilde{\delta}_k\varepsilon,$$
$$\left\|x^k-x^*\right\|^2 \leq \left\|x^{k-1}-x^*\right\|^2 - \rho_{k-1}\tilde{\delta}_k\varepsilon \leq \left\|x^{k-2}-x^*\right\|^2 -$$
$$-\varepsilon\left(\rho_{k-2}\tilde{\delta}_{k-2}+\rho_{k-1}\tilde{\delta}_{k-1}\right), \ldots$$
$$\left\|x^{k+1}-x^*\right\|^2 \leq \left\|x^{\hat{k}}-x^*\right\|^2 - \varepsilon\sum_{i=\hat{k}}^{k}\rho_i\tilde{\delta}_i,$$
$$\text{or} \quad \left\|x^{k+1}-x^*\right\|^2 \leq \left\|x^{\hat{k}}-x^*\right\|^2 - \varepsilon\sum_{i=\hat{k}}^{k}\rho_i\delta_\varphi^i.$$

Due to imposed conditions on τ_k in (18), based on relation (32), we get:

$$\left\|x^{k+1}-x^*\right\|^2 \leq \left\|x^{\hat{k}}-x^*\right\|^2 - \frac{\varepsilon}{C_1\cdot C_2}\sum_{i=\hat{k}}^{k}\rho_i\tau_i \to -\infty, \quad \text{for} \quad k\to\infty. \tag{35}$$

We obtain a contradiction because the norm of any vector, moreover its square value, cannot be negative. Therefore, the realization of event $A_1 \bigcap D_1$ implies realization of an event that is practically unrealizable, $F_1 = \left\{\left\|x^{k+1}-x^*\right\|^2 < 0,\ k\to\infty\right\}$. That is $P\left(A_1\bigcap D_1\right) \leq P\left(F_1\right)=0$. It means that $P\left(A_1\right)=P\left(A_1\bigcap\overline{D_1}\right)$.

Let us evaluate $P\left(A_1\bigcap\overline{D_1}\right)$. Let's take an arbitrary number r_k from the interval $\left[\frac{\bar{\varepsilon}_k}{2},\bar{\varepsilon}_k\right]$, where $\bar{\varepsilon}_k=\min\left\{r_0,\frac{\Delta_F}{2C_2},\frac{\tau_k}{2C_2}\right\}$. The following events are defined:

1. B_f^k={at least one time among the iterations of the form $j=\overline{k-s_k,k}$ an element from the set $W_{Y_f}\left(y_f\left(x^j\right),r_k\right)$ is generated, where $x^j\in W_X\left(x^k,r_k\right)$, for s_k defined in (13)};
2. B_φ^k={at least one time among the iterations of the form $j=\overline{k-s_k,k}$ an element from the set $W_{Y_\varphi}\left(y_\varphi\left(x^j\right),r_k\right)$ is generated, where $x^j\in W_X\left(x^k,r_k\right)$, for s_k defined in (13)};
3. $B_1^k=B_f^k\bigcap B_\varphi^k$.

The simulation of the variables ξ and ψ on iteration k is executed in parallel and independently. Since the events B_f^k, B_φ^k are independent, it follows that $P\left(B_1^k\right) = P\left(B_f^k\right) \cdot P\left(B_\varphi^k\right)$.

The realization of the event B_f^k implies: for some iteration $j_k \in \overline{k - s_k, k}$ the generated element $y_f^{j_k}\left(x^{j_k}\right) = \xi_{t_k} \in M_{j_k}, 1 \le t_k \le m_{j_k}$ has the property $y_f^{j_k}\left(x^{j_k}\right) \in W_{Y_f}\left(y_f\left(x^{j_k}\right), r_k\right)$, that is, according to (25):

$$f\left(x^{j_k}, y_f^{j_k}\right) \ge f\left(x^{j_k}, y_f^{j_k}\left(x^{j_k}\right)\right) \ge F\left(x^*\right) + \frac{3}{2}\Delta_F. \tag{36}$$

Let's admit that j_k is an arbitrary element from the set of iterations $\{k - s_k, \ldots, k-1\}$. We will show that $f\left(x^k, y_f^k\right) \ge F\left(x^*\right) + \Delta_F$. Indeed, taking into consideration the convexity of the estimator $f\left(x, y_f\right)$ for $\forall y_f \in Y_f$ and the way of computation of the sequence $\left\{x^k\right\}$, we get:

$$\begin{aligned} & f\left(x^{k+1}, y_f\right) - f\left(x^k, y_f\right) \ge \left(\partial f(x^k, y_f), x^{k+1} - x^k\right) \ge \\ & \ge -\left\|\partial f(x^k, y_f)\right\| \cdot \left\|\textstyle\prod_X\left(x^k - \rho_k \eta^k\right) - x^k\right\| \ge -C_2\rho_k. \end{aligned} \tag{37}$$

From (36) and (37) it follows:

$$\begin{aligned} & f\left(x^{j_k+1}, y_f^{j_k+1}\right) \ge f\left(x^{j_k+1}, y_f^{j_k}\right) \ge f\left(x^{j_k}, y_f^{j_k}\right) - C_2\rho_{j_k}, \\ & \ldots \\ & f\left(x^{j_k+i}, y_f^{j_k+i}\right) \ge f\left(x^{j_k}, y_f^{j_k}\right) - C_2\sum_{l=0}^{i-1}\rho_{j_k+l} \ge F\left(x^*\right) + \frac{3}{2}\Delta_F - C_2 r_k \ge \\ & \ge F\left(x^*\right) + \frac{3}{2}\Delta_F - C_2\frac{\Delta_F}{2C_2} = F\left(x^*\right) + \Delta_F \end{aligned} \tag{38}$$

for all i that $\sum\limits_{l=0}^{i-1}\rho_{j_k+l} \le r_k$.

But, $\sum\limits_{l=0}^{k-j_k}\rho_{j_k+l} = \rho_{j_k} + \rho_{j_k+1} + \ldots + \rho_k \le \sum\limits_{l=k-s_k}^{k}\rho_l \le r_k$. Therefore,

$$f\left(x^k, y_f^k\right) \ge F\left(x^*\right) + \Delta_F. \tag{39}$$

But if $j_k = k$, then the last inequality is satisfied even more. As a consequence to (39) we have the following chain of inequalities

$$-\Delta_F \ge F\left(x^*\right) - f\left(x^k, y_f^k\right) \ge f\left(x^*, y_f^k\right) - f\left(x^k, y_f^k\right) \ge \left(\partial f\left(x^k, y_f^k\right), x^* - x^k\right)$$

or,

$$\left(\partial f\left(x^k, y_f^k\right), x^k - x^*\right) \ge \Delta_F. \tag{40}$$

Taking into consideration (40) and the way the number $\tilde{\delta}_k$ is chosen, we get:

$$\frac{\left(\partial f\left(x^k, y_f^k\right), x^k - x^*\right)}{\left\|\partial f\left(x^k, y_f^k\right)\right\| \cdot \|x^k - x^*\|} \geq \tilde{\delta}_k$$

or, in other words, the event A_1^k is realized.

The realization of the event B_φ^k implies: for some iteration $j_k \in \overline{k - s_k, k}$ the generated element $y_\varphi^{j_k}\left(x^{j_k}\right) = \psi_{t_k} \in L_{j_k}, 1 \leq t_k \leq l_{j_k}$ has the property $y_\varphi^{j_k}\left(x^{j_k}\right) \in W_{Y_\varphi}\left(y_\varphi\left(x^{j_k}\right), r_k\right)$, that is, according to (30):

$$\varphi\left(x^{j_k}, y_\varphi^{j_k}\right) \geq \varphi\left(x^{j_k}, y_\varphi^{j_k}\left(x^{j_k}\right)\right) \geq \Phi\left(x^*\right) + \frac{3}{2}\tau_k. \tag{41}$$

Let's admit that j_k is an arbitrary element from the set of iterations $\{k - s_k, \ldots, k-1\}$. We will show that $\varphi\left(x^k, y_\varphi^k\right) \geq \Phi\left(x^*\right) + \tau_k$. Indeed, taking into consideration the convexity of the estimator $\varphi\left(x, y_\varphi\right)$ for $\forall y_\varphi \in Y_\varphi$ and the way of computation of the sequence $\left\{x^k\right\}$, we get:

$$\begin{aligned} &\varphi(x^{k+1}, y_\varphi) - \varphi\left(x^k, y_\varphi\right) \geq \left(\partial\varphi(x^k, y_\varphi), x^{k+1} - x^k\right) \geq \\ &\geq -\left\|\partial\varphi(x^k, y_\varphi)\right\| \cdot \left\|\textstyle\prod_X\left(x^k - \rho_k\eta^k\right) - x^k\right\| \geq -C_2\rho_k. \end{aligned} \tag{42}$$

From (41) and (42) it follows:

$$\begin{aligned} &\varphi\Big(x^{j_k+1}, y_\varphi^{j_k+1}\Big) \geq \varphi\left(x^{j_k+1}, y_\varphi^{j_k}\right) \geq \varphi\left(x^{j_k}, y_\varphi^{j_k}\right) - C_2\rho_{j_k}, \\ &\ldots \\ &\varphi\Big(x^{j_k+i}, y_\varphi^{j_k+i}\Big) \geq \varphi\left(x^{j_k}, y_\varphi^{j_k}\right) - C_2\sum_{l=0}^{i-1}\rho_{j_k+l} \geq \Phi\left(x^*\right) + \frac{3}{2}\tau_k - C_2 r_k \geq \\ &\geq \Phi\left(x^*\right) + \frac{3}{2}\tau_k - C_2\frac{\tau_k}{2C_2} = \Phi\left(x^*\right) + \tau_k \end{aligned} \tag{43}$$

for all i that $\sum\limits_{l=0}^{i-1}\rho_{j_k+l} \leq r_k$.

But, $\sum\limits_{l=0}^{k-j_k}\rho_{j_k+l} = \rho_{j_k} + \rho_{j_k+1} + \ldots + \rho_k \leq \sum\limits_{l=k-s_k}^{k}\rho_l \leq r_k$. Therefore,

$$\varphi\left(x^k, y_\varphi^k\right) \geq \Phi\left(x^*\right) + \tau_k. \tag{44}$$

But if $j_k = k$, then the last inequality is satisfied even more. As a consequence to (44) we have the following chain of inequalities

$$-\tau_k \geq \Phi\left(x^*\right) - \varphi\left(x^k, y_\varphi^k\right) \geq \varphi\left(x^*, y_\varphi^k\right) - \varphi\left(x^k, y_\varphi^k\right) \geq \left(\partial\varphi\left(x^k, y_\varphi^k\right), x^* - x^k\right)$$

or,

$$\left(\partial\varphi\left(x^k, y_\varphi^k\right), x^k - x^*\right) \geq \tau_k. \tag{45}$$

Taking into consideration (45) and the way the number $\tilde{\delta}_k$ is chosen, we get:

$$\frac{\left(\partial\varphi\left(x^k, y_\varphi^k\right), x^k - x^*\right)}{\left\|\partial\varphi\left(x^k, y_\varphi^k\right)\right\| \cdot \left\|x^k - x^*\right\|} \geq \tilde{\delta}_k$$

or, in other words, the event A_1^k is realized.

The realization of the events B_f^k and B_φ^k implies the realization of the event B_1^k. At the same time the following implication takes place: $B_1^k \subset A_1^k$. Therefore, we get $P\left(B_1^k\right) \leq P\left(A_1^k\right)$, or, $P\left(\overline{A_1^k}\right) \leq P\left(\overline{B_1^k}\right)$. But, accordingly to (4), (13) follows: $P\left(\overline{B_1^k}\right) \leq \alpha^{L\left(k, r^k\right)}$ where $\alpha = 1 - \gamma$. We get following set of inequalities:

$$\sum_{k=0}^{\infty} P\left(\overline{A_1^k}\right) \leq \sum_{k=0}^{\infty} P\left(\overline{B_1^k}\right) \leq \sum_{k=0}^{\infty} \alpha^{L\left(k, r^k\right)} < \infty.$$

We are in the situation that the conditions of the Borel–Cantelli lemma are met [3]. It means that $P\left(\overline{D_1}\right) = 0$. Therefore,

$$q \leq P\left(A_1\right) = P\left(A_1 \bigcap \overline{D_1}\right) \leq P\left(\overline{D_1}\right) = 0.$$

Thus, $q = 0$.

A contradiction has been obtained, because we have supposed that $q > 0$. Thus, there exists a subsequence $\left\{x^{k_l}\right\} \subset \left\{x^k\right\}_{k \geq 0}$ that almost certainly is contained in $V_X\left(X^*, \varepsilon\right)$.

Stage 2. Further will be proved that all elements of the sequence $\left\{x^k\right\}$, without just a finite number, belong to the set $V_X\left(X^*, 2\varepsilon\right)$ with probability 1.

The following events are defined:

$$\begin{aligned} A_2 &= \left\{\exists \left\{x^{k_l}\right\} \subset \left\{x^k\right\} : \ \left\{x^{k_l}\right\} \subset V_X\left(X^*, \varepsilon\right)\right\}, \\ B_2 &= \left\{\exists \left\{z^{k_m}\right\} \subset \left\{x^k\right\} : \ \left\{z^{k_m}\right\} \not\subset V_X\left(X^*, 2\varepsilon\right)\right\}. \end{aligned} \tag{46}$$

Next, $P\left(B_2\right)$ will be appreciated. We will find out that $P\left(B_2\right) = P\left(B_2 \bigcap A_2\right)$. Indeed, $P\left(B_2\right) = P\left(\left(B_2 \bigcap A_2\right) \bigcup \left(B_2 \bigcap \overline{A_2}\right)\right) = P\left(B_2 \bigcap A_2\right) + P\left(B_2 \bigcap \overline{A_2}\right) = P\left(B_2 \bigcap A_2\right)$, because $P\left(B_2 \bigcap \overline{A_2}\right) \leq P\left(\overline{A_2}\right) = 0$.

Further, the following event will be considered: $D_2 = A_2 \bigcap B_2$. Suppose that $P\left(D_2\right) > 0$. Realization of the event D_2 means that the transfer from $V_X\left(X^*, \varepsilon\right)$ to $X \backslash V_X\left(X^*, 2\varepsilon\right)$ and vice versa takes place infinitely.

Let us denote by:

1. K_1 – the number of first iteration when the event $\left\{x^{K_1} \in V_X\left(X^*, \varepsilon\right)\right\}$ is produced;

2. K_2 – the number of first iteration when the event $\left\{x^{K_2} \in V_X\left(X^*, \dfrac{3}{2}\varepsilon\right)\right\}$ is produced;

3. K_3 – the number of first iteration when the inequality $\rho_{K_3} \leq 2\varepsilon\tilde{\delta}_{K_3}$ is satisfied;
4. $\bar{K} = \max\{K_1, K_2, K_3\}$.

In case for some $k \geq \bar{K}$ and $x^k \notin V_X\left(X^*, \dfrac{3}{2}\varepsilon\right)$ the inequality that defines the event A_1^k is satisfied, then the following sequence of inequalities occurs: $\left\|x^{k+1} - x^*\right\|^2 \leq \left\|x^k - x^*\right\|^2 - \rho_k(2\varepsilon\tilde{\delta}_k - \rho_k) < \left\|x^k - x^*\right\|^2$, because $\left\|x^k - x^*\right\| > \varepsilon$.

That is, as soon as $k \geq \bar{K}$ and $x^k \notin V_X\left(X^*, \dfrac{3}{2}\varepsilon\right)$ it follows:

$$\left\|x^{k+1} - x^*\right\| < \left\|x^k - x^*\right\|. \tag{47}$$

Since $\rho_k \underset{k\to\infty}{\to} 0$, a number $K^* \geq \bar{K}$ will appear with the property: $x^{K^*} \in V_X\left(X^*, 2\varepsilon\right) \backslash V_X\left(X^*, \dfrac{3}{2}\varepsilon\right)$. This will happen certainly. Particularly, for $\rho_k < \dfrac{\varepsilon}{2}$:

$$\left\|x^{k+1} - x^k\right\| \leq \left\|x^k - \rho_k\eta^k - x^k\right\| \leq \rho_k < \frac{\varepsilon}{2}.$$

Therefore, there exists a number k that satisfies $x^k \in V_X\left(X^*, 2\varepsilon\right) \backslash V_X\left(X^*, \dfrac{3}{2}\varepsilon\right)$.

According to (47), $\left\|x^{K^*+1} - x^*\right\| < \left\|x^{K^*} - x^*\right\|$. In case $x^{K^*+1} \notin V_X\left(X^*, \dfrac{3}{2}\varepsilon\right)$, then $\left\|x^{K^*+2} - x^*\right\| < \left\|x^{K^*+1} - x^*\right\| < \left\|x^{K^*} - x^*\right\|$, and so forth, for all $j \geq 0$ that satisfy $x^{K^*+j} \notin V_X\left(X^*, \dfrac{3}{2}\varepsilon\right)$, takes place:

$$\min_{x^*\in X^*}\left\|x^{K^*+j} - x^*\right\| < \min_{x^*\in X^*}\left\|x^{K^*} - x^*\right\| < 2\varepsilon. \tag{48}$$

Let us denote $\left\{x^{k^l}\right\}_{l\geq 1}$ the sequence of all elements $\left\{x^k\right\}$ with the property that $k^l \geq K^*,\ x^{k^l} \in V_X\left(X^*, 2\varepsilon\right) \backslash V_X\left(X^*, \dfrac{3}{2}\varepsilon\right)$ and $x^{k^l-1} \in V_X\left(X^*, \dfrac{3}{2}\varepsilon\right)$. Then, for $l \geq 1,\ k^l < j < k^{l+1}$ and $x^j \notin V_X\left(X^*, \dfrac{3}{2}\varepsilon\right)$ the following inequality occurs:

$$\min_{x^*\in X^*}\left\|x^j - x^*\right\| < \min_{x^*\in X^*}\left\|x^{k^l} - x^*\right\| < 2\varepsilon. \tag{49}$$

Thus, in other words, admitting that for some K elements of type $x^k \notin V_X\left(X^*, \dfrac{3}{2}\varepsilon\right)$,

$k < \infty$, $k \geq K$ satisfy the inequality from the event A_1^k, then the event B_2 cannot occur with positive probability. The supposition that D_2 is realized means that beyond the layer $V_X\left(X^*, \dfrac{3}{2}\varepsilon\right)$ the penetration of the layer takes place only when infinitely the event $\overline{A_1^k}$ considered previously is produced. But, $P\left(\overline{D_1}\right) = 0$. So, the conclusion that can be drawn is that the transfer from the layer

$V_X(X^*, 2\varepsilon) \backslash V_X\left(X^*, \frac{3}{2}\varepsilon\right)$ into the layer $X \backslash V_X(X^*, 2\varepsilon)$ occurs only a finite number of times. That is, $P(D_2) = 0$, and it implies $P(B_2) = 0$. Theorem is proved. $\square$

Remark 1. In case the set of optimal solutions $X^* = \emptyset$, application of the above described method for solving the problem (1) leads us to the solution of the following problem:

$$\begin{cases} \Phi(x) = \max\limits_{y \in Y_\varphi} \varphi(x, y) \to \min \\ x \in X. \end{cases}$$

References

[1] Polyak B.T. *A general method for solving extreme problems.* Dokl. Akad. Nauk SSSR, 1967, **174**, No. 1, 33–336 (in Russian).

[2] Shor N.Z. *Nondifferentiable Optimization and Polynomial Problems.* Kluwer Academic Publishers, Netherlands, 1998.

[3] Shiryaev A.V. *Probability.* Moscow, Nauka, 1980 (in Russian).

[4] Godonoaga A. *Stohasticheskie shemy otyskanija minimaksa.* Sbornik nauchnyh trudov: Verojatnost' i ee prilozhenija, Kishinev, 1990, 3–13 (in Russian).

[5] Ermol'ev Yu. M., Gaivoronskii A.A. *Stochastic methods for solving minimax problems.* Kibernetika, 1983, No. 4, 92–97 (in Russian; English translation in: Cybernetics, **19**, No. 4, 550–559).

Anatol Godonoaga
Academy of Economic Studies
Chisinau, Moldova
E-mail: *anatgo@ase.md*

Received May 23, 2009

Pavel Balan
Moldova State University
Chisinau, Moldova
E-mail: *nalab.levap@gmail.com*