STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume **LIV**, Number 1, 2009

A METHOD WITH RANDOM MODIFICATION OF GRADIENT COMPONENTS FOR CONVEX MODELS

PAVEL BĂLAN

ABSTRACT. A stochastic method is proposed and analyzed, that is a probabilistic generalization of gradient method, for solving convex models with restrictions. A random change of "old" partial derivatives with "new" ones is performed from one iteration to another. Convergence aspects of this method are analyzed for the case when the step is adjusted programmatically. Certain conditions are indicated, that ensure its convergence to the optimal solution with probability 1.

1. INTRODUCTION

Current method has an iterative approach of "Connection-Disconnection" type. Using two random variables a series of indices is generated for next iteration. One is meant to be used for target function and the other one - for function that describes the restriction. Distribution laws that can be used can be a priori set or can be modified within iterations. Usage of different distribution laws may increase convergence speed of the method. Generated indices suggest what components of movement vector should be modified. "Old" components are replaced with corresponding partial derivatives with the same indices from the above mentioned series. The idea of "Connection-Disconnection" mechanism is following: if the restriction inequality is satisfied

Received by the editors: December 19, 2008.

2000 *Mathematics Subject Classification.* 65K05, 90C15.

1998 *CR Categories and Descriptors.* G.1.6 [**Optimization**]: Subtopic – *Convex programming*; Subtopic – *Nonlinear programming*; Subtopic – *Stochastic programming*.

Key words and phrases. stochastic method, differentiable optimization, probabilistic generalization of gradient method, convex model, partial derivatives, movement direction, convergence, with probability 1, almost sure, connect-disconnect type.

at the current iteration, then the vector that determines the movement direction is built using partial derivatives of target function; otherwise this vector contains only partial derivatives of the function from restriction.

2. A method with random modification of gradient components for convex models

The following problem is considered:

$$\left\{\begin{array}{l} F(x) \longrightarrow \min \\ \varphi(x) \leq 0 \\ x \in X \end{array}\right. \tag{1}$$

where X represents a compact and convex set in Euclidian space E^n.

Suppose that problem (1) is solvable.

Let us define $V(X,\varepsilon) = \underset{x\in X}{\cup} V(x,\varepsilon)$ - vicinity with ε radius of X set. By $V(x,\varepsilon)$ is marked the vicinity with radius $\varepsilon > 0$ of the point $x \in E^n$, or, formally:

$$V(x,\varepsilon) = \{y \in E^n : \|x-y\| < \varepsilon\}$$

Let's admit for some $\varepsilon > 0$ that $F(x)$ and $\varphi(x)$ are convex and differentiable functions (with continuous gradients) on $V(X,\varepsilon)$. Therefore, for $\forall x \in X$ are defined the following vectors:

$$(g_{F,1}(x),\ldots,g_{F,n}(x)) = g_F(x) = \operatorname{grad} F(x) = \left(\frac{dF(x)}{dx_1},\ldots,\frac{dF(x)}{dx_n}\right)$$
$$(g_{\varphi,1}(x),\ldots,g_{\varphi,n}(x)) = g_\varphi(x) = \operatorname{grad} \varphi(x) = \left(\frac{d\varphi(x)}{dx_1},\ldots,\frac{d\varphi(x)}{dx_n}\right)$$

Obviously, the norms $\|g_F(x)\| = \sqrt{\sum_{i=1}^{n} g^2_{F,i}(x)}$, $\|g_\varphi(x)\| = \sqrt{\sum_{i=1}^{n} g^2_{\varphi,i}(x)}$ are continuous functions on X compact and, consequently, a constant C exists, so that $\|g_F(x)\| \leq C$, $\|g_\varphi(x)\| \leq C$ for $\forall x$. Hence $\|g_{F,i}(x)\| \leq C$, $\|g_{\varphi,i}(x)\| \leq C$, $\forall i = \overline{1,n}$, $\forall x \in X$.

The numerical method that is proposed to solve the problem (1) has an iterative approach. Assuming that we are positioned on k^{th} iteration the schema is following:

Step 1. Two random variables ξ^k, ψ^k are simulated in series $m_k \geq 1$, $l_k \geq 1$ of independent probes with discrete distribution laws a priori defined:

ξ^k	1	2	...	n
P	$P^k_{\xi,1}$	$P^k_{\xi,2}$	...	$P^k_{\xi,n}$

ψ^k	1	2	...	n
P	$P^k_{\psi,1}$	$P^k_{\psi,2}$	...	$P^k_{\psi,n}$

That is, on every iteration k the sets $I_k = \{i_1, i_2, \ldots, i_{m_k}\}$, $J_k = \{j_1, j_2, \ldots, j_{l_k}\}$ of elements that are independent realizations of ξ^k, ψ^k variables with distribution laws defined above are generated, where

(2)
$$P^k_{\xi,i} \geq \underline{P_\xi} > 0, \forall i = \overline{1,n}, \forall k = 0,1,\ldots$$
$$P^k_{\psi,i} \geq \underline{P_\psi} > 0, \forall i = \overline{1,n}, \forall k = 0,1,\ldots$$

Particularly, we can take $m_k = l_k = 1$, that is a single simulation is accomplished for every random variable ξ^k, ψ^k on every iteration.

Step 2. $g^k_F(x^k)$, $g^k_\varphi(x^k)$ vectors are calculated according to the rule:

(3)
$$g^k_F(x^k) = \left(g^k_{F,1}, \ldots, g^k_{F,i}, \ldots, g^k_{F,n}\right), \; g^k_{F,i} = \left[\begin{array}{l} g^{k-1}_{F,i}, \text{ if } i \notin I_k \\ \frac{dF(x)}{dx_i}, \text{ if } i \in I_k \end{array}\right.$$
$$g^k_\varphi(x^k) = \left(g^k_{\varphi,1}, \ldots, g^k_{\varphi,i}, \ldots, g^k_{\varphi,n}\right), \; g^k_{\varphi,i} = \left[\begin{array}{l} g^{k-1}_{\varphi,i}, \text{ if } i \notin J_k \\ \frac{d\varphi(x)}{dx_i}, \text{ if } i \in J_k \end{array}\right.$$
$$\forall i = \overline{1,n}$$

Step 3. The element x^{k+1} is determined according to relation:

(4)
$$x^{k+1} = \prod_X \left(\tilde{x}^{k+1}\right), \text{ where } \tilde{x}^{k+1} = x^k - \rho_k \eta^k$$

$\prod_X \left(\tilde{x}^{k+1}\right)$ represents the projection of element $\tilde{x}^{k+1}$ on the set X. Starting point x^0 is arbitrary taken from X (it can be indicated under certain considerations for some concrete situations).

Step 4. The numerical sequence $\{\eta^k\}$ is defined in following way:

(5)
$$\eta^k = \begin{cases} \frac{g^k}{\|g^k\|}, \text{ if } g^k \neq \bar{0}, \; \forall k = 1,2,\ldots \\ \bar{0}, \text{ for } g^k = \bar{0} \end{cases}$$

g^0 is considered to be an arbitrary, but bounded vector.

Necessarily, classical requirements are imposed on sequence $\{\rho_k\}$ to ensure the convergence from probabilistic point of view of the iterative process (4) which have the form:

$$\rho_k \geq 0;\ \rho_k \underset{k\to\infty}{\longrightarrow} 0;\ \sum_{k=0}^{\infty} \rho_k = \infty \tag{6}$$

Additionally we will require existence of such a number $\bar{\varepsilon} > 0$, that for $\forall r \in (0, \bar{\varepsilon}]$ and $\forall \tau \in (0,1)$ the convergence of the series occurs [1]:

$$\sum_{k=0}^{\infty} \tau^{L(k,r)} < \infty,\ L\left(k,r\right) = \begin{cases} 0, \text{ if } \rho_k \geq r \text{ or } k = 0 \\ s_k, \text{ if } \sum\limits_{l=k-s_k}^{k} \rho_l < r \text{ and } \sum\limits_{l=k-s_k-1}^{k} \rho_l \geq r \end{cases} \tag{7}$$

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

Remark 1. *Particularly, it is easy to show that numerical sequence* $\rho_k = \frac{R}{(k+1)^{\alpha}}$, $R > 0$, $\alpha \in (0,1]$ *satisfies the conditions (6)-(7).*

B.T. Polyak proposes a schema to solve general convex models [4]. This is a deterministic "connect-disconnect" schema where the vector g^k defines movement direction and is formulated in following way:

$$g^k = g^k(x^k) = \begin{cases} \operatorname{grad} F\left(x^k\right), \text{ if } \varphi\left(x^k\right) \leq 0 \\ \operatorname{grad} \varphi\left(x^k\right), \text{ if } \varphi\left(x^k\right) > 0 \end{cases}$$

But actual method proposes another representation of movement vector g^k:

$$g^k = g^k(x^k) = \begin{cases} g_F^k\left(x^k\right), \text{ if } \varphi\left(x^k\right) \leq \tau_k \\ g_\varphi^k\left(x^k\right), \text{ if } \varphi\left(x^k\right) > \tau_k \end{cases},\ \tau_k > 0 \tag{8}$$

Remark 2. *The iterative process can be modified in following way: different distribution laws can be taken for random variables* ξ^k, ψ^k, *from one iteration to another, with the condition that relation (2) holds. This can favour the increase of convergence speed of the sequence* $\left\{x^k\right\}$.

Applicability of described method can be confirmed, first of all, by establishing convergence, in probabilistic terms, of sequence $\left\{x^k\right\}$ towards optimal domain of solutions X^*. A special interest represents the convergence with probability 1 (only this type of convergence can be accepted with confidence from applied point of view).

Theorem 1. *For fixed $\forall\varepsilon > 0$,* ***all*** *elements of random sequence $\left\{x^k\right\}_{k\geq 0}$, obtained as a result of application of described method, 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:*

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

where $x^k = x^k\left(\theta\right)$ and $\theta = \left(\theta^0, \theta^1, \ldots, \theta^k, \ldots\right)$,
$\theta^k = \left(i^0, i^1, \ldots, i^k\right) \in B^k$- σ- algebra generated by
Cartesian product $\left(\left(I_0 \times J_0\right) \times \left(I_1 \times J_1\right) \times \ldots \times \left(I_k \times J_k\right)\right)$.

Proof. If $X \subset V\left(X^*, 2\varepsilon\right)$ then the statement is obvious.

Let's admit $X\setminus V\left(X^*, 2\varepsilon\right) \neq \varnothing$. A problem of the form:

$$\text{(9)}\qquad \begin{cases} F\left(x\right) \longrightarrow \min \\ \varphi\left(x\right) \leq \tau_k, \tau_k > 0,\ \tau_k \to 0,\ \sum\limits_{k=0}^{\infty} \rho_k \tau_k = \infty,\ \frac{\tau_k}{\rho_k} \to \infty \\ x \in X \end{cases}$$

is associated to initial model (1) on every iteration k.

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 demonstrated, 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 can be indicated such that the following event is produced

$$\text{(10)}\qquad A_1 = \left\{\exists K_q : \forall k \geq K_q,\ \left\|x^k - x^*\right\| \geq \varepsilon,\ \text{or } x^k \notin V_X\left(X^*, \varepsilon\right),\ \forall x^* \in X^*\right\}$$

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

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

Since $F\left(x\right)$, $\varphi\left(x\right)$ are convex and differentiable, the following inequalities are valid:

$$F\left(x^*\right) - F\left(x^k\right) \geq \left(\frac{dF\left(x^k\right)}{dx}, x^* - x^k\right),$$
$$\varphi\left(x^*\right) - \varphi\left(x^k\right) \geq \left(\frac{d\varphi\left(x^k\right)}{dx}, x^* - x^k\right)$$

for $\forall x^* \in X^*$, $\forall x^k \in X$.

Let us denote $\Delta_F = \min\limits_{x \in X_\varepsilon, x^* \in X^*} [F(x) - F(x^*)]$. Evidently, $\Delta_F > 0$, if $\varepsilon > 0$.

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 $\left\| \frac{dF(x)}{dx} \right\| \leq C_2$, $\forall x \in X$.

If $\varphi\left(x^k\right) \leq \tau_k$ and $x^k \in X_\varepsilon$, then $F\left(x^k\right) - F\left(x^*\right) \geq \Delta_F$, or

$$\left(\frac{dF\left(x^k\right)}{dx}, x^k - x^* \right) \geq \Delta_F \tag{11}$$

$$\frac{\left(\frac{dF\left(x^k\right)}{dx}, x^k - x^* \right)}{\left\| \frac{dF\left(x^k\right)}{dx} \right\| \cdot \|x^k - x^*\|} \geq \frac{\left(\frac{dF\left(x^k\right)}{dx}, x^k - x^* \right)}{C_2 C_1} \geq \frac{\Delta_F}{C_1 C_2}$$

Also, if $\varphi\left(x^k\right) > \tau_k$ for $x^k \in X_\varepsilon$, then:

$$\left(\frac{d\varphi\left(x^k\right)}{dx}, x^k - x^* \right) \geq \tau_k \tag{12}$$

$$\frac{\left(\frac{d\varphi\left(x^k\right)}{dx}, x^k - x^* \right)}{\left\| \frac{d\varphi\left(x^k\right)}{dx} \right\| \cdot \|x^k - x^*\|} \geq \frac{\left(\frac{d\varphi\left(x^k\right)}{dx}, x^k - x^* \right)}{C_2 C_1} \geq \frac{\tau_k}{C_1 C_2}$$

Let's consider some numbers δ_F, δ_φ^k from intervals $\left(0, \frac{\Delta_F}{C_1 C_2}\right)$, $\left(0, \frac{\tau_k}{C_1 C_2}\right)$ and label $\bar{\delta}_k = \min\left\{\delta_F, \delta_\varphi^k\right\}$. As a result the following inequalities may be obtained:

$$\begin{array}{l} \left(\frac{dF\left(x^k\right)}{dx}, x^k - x^* \right) \geq 2\bar{\delta}_k \left\| \frac{dF\left(x^k\right)}{dx} \right\| \cdot \left\|x^k - x^*\right\|, \text{ if } \varphi\left(x^k\right) \leq \tau_k \\ \left(\frac{d\varphi\left(x^k\right)}{dx}, x^k - x^* \right) \geq 2\bar{\delta}_k \left\| \frac{d\varphi\left(x^k\right)}{dx} \right\| \cdot \left\|x^k - x^*\right\|, \text{ if } \varphi\left(x^k\right) > \tau_k \end{array} \tag{13}$$

Particularly, δ_F, δ_φ^k may be chosen as centres of intervals $\left(0, \frac{\Delta_F}{C_1 C_2}\right)$, $\left(0, \frac{\tau_k}{C_1 C_2}\right)$:

$$\delta_F = \frac{\Delta_F}{2\left(C_1 C_2\right)}, \delta_\varphi^k = \frac{\tau_k}{2\left(C_1 C_2\right)} \tag{14}$$

The following events are being considered

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

$\overline{A_1^k} = \left\{\exists x^* \in X^* : \left(\eta^k, x^k - x^*\right) < \bar{\delta}_k \left\|x^k - x^*\right\|\right\}$.

Then $D_1 = \left\{\bigcup\limits_{k=K_\delta}^{\infty} \bigcap\limits_{i=k}^{\infty} A_1^i\right\}$, 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 $P\left(A_1\right)$. In order to do this let's represent

$$P\left(A_1\right) = P\left(A_1 \cap D_1\right) + P\left(A_1 \cap \overline{D_1}\right)$$

Both terms from last expression will be estimated.

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

$$\left(\eta^k, x^k - x^*\right) \geq \bar{\delta}_k \left\|x^k - x^*\right\| \tag{15}$$

Taking into consideration (15), for $k \geq K_\delta$ we have the following sequence of relations:

$\left\|x^{k+1} - x^*\right\|^2 \leq \left\|x^k - \rho_k \eta^k - x^*\right\|^2 =$

$= \left\|x^k - x^*\right\|^2 - 2\rho_k \left(x^k - x^*, \eta^k\right) + \rho_k^2 \left\|\eta^k\right\|^2 \leq$

$\leq \left\|x^k - x^*\right\|^2 - 2\rho_k \bar{\delta}_k \left\|x^k - x^*\right\| + \rho_k^2 \leq$

$\leq \left\|x^k - x^*\right\|^2 - 2\rho_k \bar{\delta}_k \varepsilon + \rho_k^2 =$

$= \left\|x^k - x^*\right\|^2 - \rho_k \left(2\bar{\delta}_k \varepsilon - \rho_k\right)$

Because $\rho_k \underset{k \to \infty}{\longrightarrow} 0$, for some K_φ: $\delta_F > \delta_\varphi^k$ or $\bar{\delta}_k = \delta_\varphi^k$, as soon as $k \geq K_\varphi$. According to (9), (14) for some $K_\varepsilon \geq K_\varphi$: $\rho_k \leq \bar{\delta}_k \varepsilon$, as soon as $k \geq K_\varepsilon$. Evidently, for $k \geq \hat{k} = \max\left\{K_\delta, K_\varepsilon\right\}$:

$$\left\|x^{k+1} - x^*\right\|^2 \leq \left\|x^k - x^*\right\|^2 - \rho_k \bar{\delta}_k \varepsilon,$$

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

Due to imposed conditions on τ_k in (9), based on relation (14), we get:

$$\left\|x^{k+1} - x^*\right\|^2 \le \left\|x^{\hat{k}} - x^*\right\|^2 - \frac{\varepsilon}{2\left(C_1 C_2\right)}\sum_{i=\hat{k}}^{k}\rho_i\tau_i \to \infty, \text{ for } k \to \infty \tag{16}$$

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

Let us evaluate $P\left(A_1 \cap \overline{D_1}\right)$. The following event B_1^k is defined:

B_F^k ={at least one time among iterations of the form $j = \overline{k - s_k, k}$ is generated every possible value of the discrete random variable ξ^k}.

B_φ^k ={at least one time among iterations of the form $j = \overline{k - s_k, k}$ is generated every possible value of the discrete random variable ψ^k}.

$B_1^k = B_F^k \cap B_\varphi^k$.

Simulation of variables ξ^k and ψ^k is produced in parallel and independently. Thanks to the fact that B_F^k, B_φ^k events are independent, follows that $P\left(B_1^k\right) = P\left(B_F^k\right) \cdot P\left(B_\varphi^k\right)$.

Let us prove that $P\left(B_1^k\right) \underset{k\to\infty}{\longrightarrow} 1$. Contrary is supposed: $P\left(B_1^k\right) \le p < 1$, for $\forall k$. We have $P\left(\overline{B_F^k}\right) = 1 - P\left(B_F^k\right)$ and $P\left(\overline{B_\varphi^k}\right) = 1 - P\left(B_\varphi^k\right)$.

It is absolutely clear that $P\left(\overline{B_F^k}\right) \le \sum\limits_{i=1}^{n}\left(1-P_{\xi,i}^k\right)^{s_k} \underset{k\to\infty}{\longrightarrow} 0$, $P\left(\overline{B_\varphi^k}\right) \le \sum\limits_{i=1}^{n}\left(1-P_{\psi,i}^k\right)^{s_k} \underset{k\to\infty}{\longrightarrow} 0$. Indeed, if we label $\underline{P} = \min\left\{\underline{P_\xi}, \underline{P_\psi}\right\}$, the following sequence of relations takes place:

$$
\begin{aligned}
&P\left(\overline{B_F^k}\right) \le \sum_{i=1}^{n}\left(1-P_{\xi,i}^k\right)^{s_k} \le n\cdot\max_{1\le i\le n}\left(1-P_{\xi,i}^k\right)^{s_k} \le \\
&\le \left(\sqrt[s_k]{n}\left(1-\underline{P_\xi}\right)\right)^{s_k} \le \left(\sqrt[s_k]{n}\left(1-\underline{P}\right)\right)^{s_k}, \\
&P\left(\overline{B_\varphi^k}\right) \le \sum_{i=1}^{n}\left(1-P_{\psi,i}^k\right)^{s_k} \le n\cdot\max_{1\le i\le n}\left(1-P_{\psi,i}^k\right)^{s_k} \le \\
&\le \left(\sqrt[s_k]{n}\left(1-\underline{P_\psi}\right)\right)^{s_k} \le \left(\sqrt[s_k]{n}\left(1-\underline{P}\right)\right)^{s_k}
\end{aligned} \tag{17}
$$

Because $k\to\infty$, then $s_k\to\infty$ and $\sqrt[s_k]{n}\to 1+0$. For an arbitrary, but fixed value $\tau\in(1-\underline{P},1)$: $\exists K_\tau\in\mathbb{N}$, so that for $k\ge K_\tau$ takes place inequality $\sqrt[s_k]{n}\left(1-\underline{P}\right)\le\tau<1$. Thus, $\left(\sqrt[s_k]{n}\left(1-\underline{P}\right)\right)^{s_k}\le\tau^{s_k}\to 0$. Or $P\left(\overline{B_F^k}\right)\le\tau^{s_k}$ and $P\left(\overline{B_\varphi^k}\right)\le\tau^{s_k}$. It means that $P\left(\overline{B_F^k}\right)\to 0$ and $P\left(\overline{B_\varphi^k}\right)\to 0$, or, $P\left(B_F^k\right)\to 1$ and $P\left(B_\varphi^k\right)\to 1$ for $k\to\infty$. These considerations conclude to the fact that $P\left(B_1^k\right)\underset{k\to\infty}{\longrightarrow} 1$.

The realization of event B_1^k means following: "renovation" of all components of the vectors $g_F^{k-s_k}$ and $g_\varphi^{k-s_k}$ is performed during s_k iterations starting from $k-s_k$ till k (inclusively). In other words, movement vector g^k contains as its components all partial derivatives, all evaluated after iteration $k-s_k$.

The realization of event B_1^k and the fact that partial derivatives of functions $F(x)$, $\varphi(x)$ are continuous, conclude to realization of following event:

$$
\left\{\forall\widetilde{\widetilde{\varepsilon}}>0:\ \begin{array}{l}\left\|g^i-\frac{\partial F\left(x^k\right)}{\partial x}\right\|\le\widetilde{\widetilde{\varepsilon}},\ \text{if } \varphi\left(x^k\right)\le\tau_k \\ \left\|g^i-\frac{\partial\varphi\left(x^k\right)}{\partial x}\right\|\le\widetilde{\widetilde{\varepsilon}},\ \text{if } \varphi\left(x^k\right)>\tau_k\end{array},\forall i=\overline{k-s_k,k}\right\} \tag{18}
$$

Taking into consideration (18), continuity of dot product and satisfaction of conditions (10), (13), we can draw a conclusion that event A_1^k is realized starting with some $k=\widehat{\hat{k}}=\max\left\{K_\tau,\hat{k}\right\}$.

Thus, $B_1^k \subset A_1^k$. In this case $P\left(B_1^k\right) \leq P\left(A_1^k\right)$ and, therefore $P\left(\overline{A_1^k}\right) \leq P\left(\overline{B_1^k}\right)$. But, according to (2), (7) and (17) follows:

$$P\left(\overline{B_1^k}\right) \leq \tau^{L(k,r)}, \text{ that is } \sum_{k=0}^{\infty} P\left(\overline{A_1^k}\right) \leq \sum_{k=0}^{\infty} P\left(\overline{B_1^k}\right) \leq \sum_{k=0}^{\infty} \tau^{L(k,r)} < \infty$$

We are in such situation that the conditions of the Borel-Cantelli lemma [3] are satisfied. It means that $P\left(\overline{D_1}\right) = 0$. Therefore, $q \leq P\left(A_1\right) = P\left(A_1 \cap \overline{D_1}\right) \leq P\left(\overline{D_1}\right) = 0$. Thus, $q = 0$.

A contradiction has been obtained, because we have supposed that $q > 0$. Thus, it 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 sequence $\left\{x^k\right\}$, without just a finite number, belong to set $V_X\left(X^*, 2\varepsilon\right)$ with probability 1.

Following events are defined:

$$\begin{aligned} A_2 &= \left\{\exists \left\{x^{k_l}\right\} \subset \left\{x^k\right\} : \left\{x^{k_l}\right\} \subset V_X\left(X^*, \varepsilon\right)\right\} \\ B_2 &= \left\{\exists \left\{z^{k_m}\right\} \subset \left\{x^k\right\} : \left\{z^{k_m}\right\} \not\subseteq V_X\left(X^*, 2\varepsilon\right)\right\} \end{aligned} \tag{19}$$

Next, $P\left(B_2\right)$ will be appreciated. We will find out that $P\left(B_2\right) = P\left(B_2 \cap A_2\right)$. Indeed,

$P\left(B_2\right) = P\left(\left(B_2 \cap A_2\right) \cup \left(B_2 \cap \overline{A_2}\right)\right) = P\left(B_2 \cap A_2\right) + P\left(B_2 \cap \overline{A_2}\right) = P\left(B_2 \cap A_2\right)$, because $P\left(B_2 \cap \overline{A_2}\right) \leq P\left(\overline{A_2}\right) = 0$.

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

Let us denote by:

K_1 - the first iteration the event $\left\{x^{K_1} \in V_X\left(X^*, \varepsilon\right)\right\}$ is produced,

K_2 - the first iteration the event $\left\{x^{K_2} \in V_X\left(X^*, \frac{3}{2}\varepsilon\right)\right\}$ is produced,

K_3 - the first iteration theinequality $\rho_{K_3} \leq 2\varepsilon\bar{\delta}_k$ is satisfied,

$\bar{K} = \max\left\{K_1, K_{2,} K_3\right\}$.

In case that for some $k \geq \bar{K}$ and $x^k \notin V_X\left(X^*, \frac{3}{2}\varepsilon\right)$ is satisfied inequality that defines event A_1^k, then following sequence of inequalities occurs:

$\left\|x^{k+1} - x^*\right\|^2 \leq \left\|x^k - x^*\right\|^2 - \rho_k\left(2\varepsilon\bar{\delta}_k - \rho_k\right) < \left\|x^k - x^*\right\|^2$, because $\left\|x^k - x^*\right\| > \varepsilon$.

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

$$\left\|x^{k+1} - x^*\right\| < \left\|x^k - x^*\right\| \tag{20}$$

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

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

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

According to (20), $\left\|x^{K^*+1} - x^*\right\| < \left\|x^{K^*} - x^*\right\|$. In the case that $x^{K^*+1} \notin V_X\left(X^*, \frac{3}{2}\varepsilon\right)$, then we have $\left\|x^{K^*+2} - x^*\right\| < \left\|x^{K^*+1} - x^*\right\| < < \left\|x^{K^*} - x^*\right\|$, and so forth, for all $j \geq 0$ that satisfy $x^{K^*+j} \notin V_X\left(X^*, \frac{3}{2}\varepsilon\right)$, takes place

$$\min_{x^*\in X^*} \left\|x^{K^*+j} - x^*\right\| < \min_{x^*\in X^*} \left\|x^{K^*} - x^*\right\| < 2\varepsilon \tag{21}$$

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

$$\min_{x^*\in X^*} \left\|x^j - x^*\right\| < \min_{x^*\in X^*} \left\|x^{k^l} - x^*\right\| < 2\varepsilon \tag{22}$$

Thus, in other words, admitting that for some K elements of type $x^k \notin V_X\left(X^*, \frac{3}{2}\varepsilon\right), k < \infty, k \geq K$ satisfy inequality from event A_1^k, then event B_2 cannot occur with positive probability. Supposition that D_2 is realized means that beyond layer $V_X\left(X^*, \frac{3}{2}\varepsilon\right)$ penetration of layer $X \setminus V_X\left(X^*, 2\varepsilon\right)$ takes place only when infinitely is produced event $\overline{A_1^k}$ considered previously. But $P\left(\overline{D_1}\right) = 0$. So, the conclusion that can be drawn is that the transfer from layer $V_X\left(X^*, 2\varepsilon\right) \setminus V_X\left(X^*, \frac{3}{2}\varepsilon\right)$ into layer $X \setminus V_X\left(X^*, 2\varepsilon\right)$ occurs only a finite number of times. That is, $P\left(D_2\right) = 0$, and implies $P\left(B_2\right) = 0$.

Theorem is proved. □

3. Conclusions

Elaborated method is especially practical for models where modification of gradients is "relatively slow". Such models are often encountered in economical, technical problems etc. It represents a significant generalization of methods meant to solve extremum problems. It can be classified as a direct method of optimization and does not use penalty functions or Lagrange function – common toolkit used to solve such kind of problems.

References

[1] Ermolyev Y.M., Gaivoronski A.A., *Stochastic Methods for Solving Minimax Problems*, Kibernetika, Naukova Dumka, Kiev, 1983

[2] Moiseev N.N, Ivanilov Y.P., Stolyarova E.M., *Methods of optimization*, Nauka, Moskow, 1978

[3] Shiryaev A.V., *Probability*, Nauka, Moskow, 1980

[4] Polyak B.T., *A general method for solving extreme problems*, Dokl. Akad. Nauk SSSR. 174, No.1, p.33-36, 1967

Moldova State University, Chisinau, Republic of Moldova
E-mail address: nalab.levap@gmail.com