THE DIRECT MONOTONE STOCHASTIC OPTIMIZATION METHOD

N.D. Chepurnoj

September 1987
WP-87-81

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

FOREWORD

The monotone method for the solution of a stochastic programming problem of expectation type is concidered in this paper. This method produces a sequence of points x^s with decreasing values of an objective function which distinquishes it from other known methods. The achievement of this method requires estimates of the objective function with accuracy which increases during successive iterations. The paper was prepared during a visit of N. Chepurnoi to the SDS program.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

THE DIRECT MONOTONE STOCHASTIC OPTIMIZATION METHOD

N.D. Chepurnoj

Consider the following optimization problem:

$$\min_{x \in E^n} F(x) = \min_{x \in E^n} E f(x, \omega) \ , \qquad (*)$$

where E^n is an n-dimensional Euclidean space, E is a mathematical expectation symbol, ω is a random element of appropriate probability space.

The ojective of this paper is to develop monotone methods for the solving of the stochastic programming problem of the above type. The proposed method is based on stochastic quasigradient techniques [1]:

$$x^{s+1} = \Pi_x(x^s - \rho_s \xi^s)$$

where ρ_s is the stepsize and ξ^s is a random vector with the following properties:

$$E(\xi^s / x^0, \ldots, x^s) = F_x(x^s)$$

and $F_x(x^s)$ is a gradient of the function $F(x)$. This method can be applied also to nondifferentiable functions $F(x)$ [2] and is characterized by a low amount of effort needed to spend on each iteration. The vector ξ^s can be computed using a very small number of observations, for example the simplest choice is $f_x(x^s, \omega^s)$ where ω^s is an observation of random vector ω. This techniques is used mainly because of the impossibility of performing the mathematical expectation operation which involves multidimensional integration. Therefore it is impossible to apply traditional nonlinear programming methods [3].

The structure of the proposed method is similar to the structure of the monotone method with averaging of the sub-gradients [4] and its stochastic finite-difference analog [5], intended for nondifferentiable optimization problems. Nevertheless between these methods there exists one principal distinction. The point is that in general in the solving of the stochastic optimization problems the exact value of the function $F(x)$ is not known. Therefore it is necessary to develop the special numerical procedure of estimating of the function $F(x)$ value using the values $f(x, \omega^i)$.

The estimating procedure is constructed in such a way that the estimation accuracy would be adaptively increased as the method approaches the minimum.

Let us explain the main idea of the proposed method.

The method performs a double iteration: the "internal" iteration is intended for construction of an appropriate descent direction and the "external" iteration is a minimizing iteration. To determine descent direction the operation of averaging of the current stochastic subgradient with the previous descent direction is used. Thus, the step direction is always a convex combination of the stochastic subgradients, computed in a sufficiently small neighborhood of a current approximation's point. This fact allows to use necessary conditions for a minimum of the function $F(x)$ for the adaptive regulation of the algorithm parameters. The algorithm's parameters are changed if the iterative process gets into a small neighborhood of the minimizer's set. To detect this fact the value of the norm of the step direction vector on the internal iterations is used. If the given value is less than some fixed threshold value, then the algorithm's parameters are changed and an internal iteration is repeated again.

It is necessary to keep in mind that the step direction is a random vector. Therefore using this test we can mistakenly decide that the algorithm arrived in the vicinity of optimum while in fact this is not so. This can lead to the too fast changing of the algorithm parameters. In order to prevent this we start testing the value of the step direction norm after a sufficiently large amount of iterations.

The algorithm description is stated below. At first the general scheme is stated and later the proof is carried out.

In the sequel the following notations will be used:

- $\partial F(x)$ is a set of subgradients of the function $F(x)$ in a point x;
- $g(x)$ is a subgradient of the function $F(x)$ in a point x;
- p is a subscript of the "minimizing" iteration of the algorithm;
- s is a superscript of the "internal" iteration of the algorithm;
- i is a subscript of the numerical sequences of the algorithm;
- $\{x_p\}$ is a minimizing sequence of points;
- $\{x^s\}$ is a sequence of the points on the "internal" iterations;
- $\xi^s(x^s)$ is a stochastic subgradient, computed in the point x^s and its conditional expectation is equaled to one of the subgradients of the function $F(x)$ in the point x^s;

- e^s is a step direction on the internal iterations of the algorithm;
- $\{r_i\}$ is a sequence of the step-size multipliers;
- $\{\epsilon_i\}$ is a sequence of the fixed threshold values to check the test for "getting into" the neigborhood of a solution;
- $\{t_{i+p}\}$ is a sequence used in the algorithm on the internal iterations in order to determine the instants for which the "getting into" test is checked;
- $\{\Pi^{(1)}_{i+s+p}\}$, $\{\Pi^{(2)}_{i+s+p}\}$ are the sequences corresponding to the minimal admissible numbers of the random value ω observations to guarantee a required estimation accuracy of the function $F(x)$ value;
- k, l are superscripts and designate the number of the random value ω observations respectively in the points x^s and x_p;
- $\hat{F}^k(x^s)$, $\hat{F}^l(x_p)$ are the estimates of the function $F(x)$ values respectively in the points x^s and x_p;

γ is an algorithm's parameter.

DESCRIPTION of Algorithm 1.

Let x_0 be an arbitrary initial approximation. Set $e^0 = \xi^0$, where

$$E(\xi^0/x_0) = g(x_0) \ .$$

Put $i = 0$, $s = 0$, $p = 0$.

Step 1 Compute

$$x^{s+1} = x_p - r_i e^s \ ,$$

$$e^{s+1} = e^s + \frac{1}{s+1}(\xi^{s+1} - e^s) \ ,$$

where $E(\xi^{s+1}/x_0, x_1, \ldots, x_p, x^1, \ldots, x^{s+1}) = g(x^{s+1}) \in \partial F(x^{s+1})$.

Step 2 If $s + 1 \leq t_{i+p}$, then $s = s + 1$ and go to Step 1.

Step 3 If $\|e^{s+1}\| \leq \epsilon_i$, then $i = i + 1$, $s = 0$, $e^0 = \xi^0$ and go to Step 1, where $E(\xi^0/x_0, x_1, \ldots, x_p) = g(x_p) \in \partial F(x_p)$.

Step 4 If $k + 1 \leq \Pi^{(1)}_{i+s+p}$, then go to Step 6.

Step 5 Define

$$\hat{F}^{k+1}(x^{s+1}) = \hat{F}^{k}(x^{s+1}) + \frac{1}{k+1}(f(x^{s+1}, \omega^{k+1}) - \hat{F}^{k}(x^{s+1})) \ ,$$

put $k = k + 1$ and go to Step 4.

Step 6 If $l + 1 \le \Pi^{(2)}_{i+s+p}$, then go to Step 8.

Step 7 Define

$$\hat{F}^{l+1}(x^{s+1}) = \hat{F}^{l}(x^{s+1}) + \frac{1}{l+1}(f(x_p, \omega^{l+1}) - \hat{F}^{l}(x_p)) \ ,$$

put $l = l + 1$ and go to Step 6.

Step 8 If

$$\hat{F}^{k+1}(x^{s+1}) \le \hat{F}^{l+1}(x^{p}) - \frac{1}{4}\gamma r_i \epsilon_i^2 \ ,$$

then go to Step 9, else $s = s + 1$ and to to Step 1.

Step 9 Set $x_{p+1} = x^{s+1}$, $e^0 = \xi^0$, $s = 0$, where $E(\xi^0 / x_o, \ x_1, \ldots, x_{p+1}) = g(x_{p+1}) \in \partial F(x_{p+1})$, $p = p + 1$ and go to Step 1.

THEOREM 1

Let $F(x)$ be a convex function, dom $F(x) = E^n$, *the sets $\{x : F(x) \le C\}$ being bounded for any bounded constant C. The set of solutions of the problem $(*)$ will be the set*

$$X^* = \{x^* \in E^n : 0 \in \partial F(x^*)\} \ .$$

Let the algorithm's parameters be such that:

$$o < \gamma < 1 \ ,$$

$$r_i > 0,\ r_i \to 0;\ \epsilon_i > 0,\ \epsilon_i \to 0 \ ;$$

$$t_{i+p} = \frac{1}{\delta^2_{i+p}\nu_{i+p}} \ ,$$

where

$$\sum_{i+p=0}^{\infty} \nu_{i+p} > \infty$$

and

$$\delta_{i+p} > 0,\ \delta_{i+p} \to 0 \ ;$$

$$\Pi^{(1)}_{i+s+p} = \frac{1}{(\delta^{(1)}_{i+s+p})^2 \nu^{(1)}_{i+s+p}} \ ,$$

where

$$\sum_{i+s+p=0}^{\infty} \nu^{(1)}_{i+s+p} < \infty$$

and

$$\delta^{(1)}_{i+s+p} > 0,\ \delta^{(1)}_{i+s+p} \to 0 \ ;$$

$$\Pi^{(2)}_{i+s+p} = \frac{1}{(\delta^{(2)}_{i+s+p})^2 \nu^{(2)}_{i+s+p}} \ ,$$

where

$$\sum_{i+s+p=0}^{\infty} \nu^{(2)}_{i+s+p} < \infty$$

and

$$\delta^{(1)}_{i+s+p} > 0,\ \delta^{(2)}_{i+s+p} \to 0 \ ;$$

$$\delta^{(1)}_{i+s+p} + \delta^{(2)}_{i+s+p} \le \frac{1}{4} \gamma r_i \epsilon_i^2 \ .$$

Let the random trajectory $\{x_p(u)\}$ will be defined on some probability space $<\mathcal{U}, \mathcal{B}, \mathcal{P}>$, where $u \in \mathcal{U}$ is a set of the elementary events, $\mathcal{B}$ is a σ-algebra and $\mathcal{P}$ is a probability measure.

Suppose that for the given trajectory $\{x_p(u)\}$ there exists a constant $C < \infty$ such that $\|\xi^s(u)\| \le C$ for any s.

Then either Algorithm 1 generates the finite number of points $\{x_p(u)\}$ and the last one will belong to the set X^* or all limit points of the trajectory $\{x_p(u)\}$ belong to the set X^*.

PROOF Consider the two possible cases:

- The number of points of the minimizing sequence $\{x_p(u)\}$ is finite;
- the number of points of the minimizing sequence $\{x_p(u)\}$ is infinite.

Let the number of points $\{x_p(u)\}$ be finite and the point $x_{\bar{p}}(u)$ is the last point of the minimizing sequence, generated by Algorithm 1.

Let us denote as $\{s_i\}$ the sequence of the instants, for which the condition

$$||e^{s_i}|| \leq \epsilon_i,\ s_i \geq t_{i+\bar{p}}$$

is fulfilled.

STATEMENT 1 Subscript i is changed the infinite number of times.

Let us assume the opposite. Then there exists some threshold value $\epsilon_i > 0$ such, that

$$||e^s|| > \epsilon_i$$

for all $s \geq t_{i+\bar{p}}$.

Then in this case for some instant $\tilde{s}$ the inequality

$$\hat{F}^k(x^{\tilde{s}}(u)) \leq \hat{F}^l(x_{\bar{p}}(u)) - \frac{1}{4}\gamma r_i \epsilon_i^2$$

is fulfilled and Algorithm 1 goes to Step 9. Hence, the next point $x_{\bar{p}+1}$ will be constructed and the point $x_{\bar{p}}$ is not the last. This fact contradicts the original assumption.

In fact, let us consider the sequence of the random points $x_{\bar{p}}(u)$, $x^1(u)$, $x^2(u),\ldots,x^s(u),\ldots$. For each number s the random values $x^s(u)$ are defined on some σ-algebra $\mathcal{B}s$, induced random vectors x_0, $x_1(u)$, $x_2(u),\ldots,x_{\bar{p}}(u)$, $x^1(u),\ldots,x^s(u)$. The trajectory $\{x^s(u)\}$ by depending on u is defined on σ-algebra $\mathcal{B}$, which contains expanding σ-algebras $\mathcal{B}s$ or, more precisely, on some probability space $(\mathcal{U}, \mathcal{B}, \mathcal{P})$, $u \in \mathcal{U}$ with the measure $\mathcal{P}$.

Later on the dependence x^s from u is omitted.

From the strong law of large numbers for the independent random variables and the same take for the dependent random variables, [6] follows that there exists sufficiently small number $\Delta_i > 0$ such that for the given elementary event $u \in \mathcal{U}$ there exists integer S_1 such, that by $s \geq S_1$, $k \geq K_1(S_1)$, $l \geq L_1(S_1)$ the inequalities

$$||e^s - z^s|| \leq \Delta_i, \text{ where } z^s = \frac{1}{s+1}\sum_{j=0}^{s} g^j(x^j) \ ,$$

$$|\hat{F}^k(x^{s+1}) - F(x^{s+1})| \leq \Delta_i \ ,$$

$$|\hat{F}^l(x_p) - F(x_p)| \leq \Delta_i$$

are realized.

Let us choose Δ_i such, that

$$\Delta_i < \min\left\{\frac{\gamma \epsilon_i^2}{2C}, \epsilon_i, \frac{\gamma}{4} r_i \epsilon_i^2\right\}$$

and in addition the inequality

$$2\Delta_i \epsilon_i - \Delta_i^2 \leq \frac{(1-\gamma)}{2} \epsilon_i^2$$

is fulfilled.

Let us assume, that for all $s \geq S_1$ the ration

$$(g^{s+1}, z^s) \leq \gamma \epsilon_i^2$$

is correct.

Then

$$\|z^{s+1}\|^2 = \|z^s\|^2 + \frac{2}{s+1}((g^{s+1}, z^s) - \|z^s\|^2) + \frac{1}{(s+1)^2}\|g^{s+1} -$$

$$- z^s\|^2 \leq \|z^{S_1}\|^2 - (1-\gamma)\epsilon_i^2 \sum_{j=S_1}^{s} \frac{1}{j+1} + \sum_{j=S_1}^{s} \frac{C}{(j+1)^2} .$$

Since the series $\sum_{j=S_1}^{\infty} 1/(j+1)$ diverges, then passing to the limit for $s \to \infty$ we get a contradiction with the non-negativety of the norm .

Consequently, there exists an instant $\bar{s} \geq S_1$ such, that $(g^{\bar{s}+1}, z^{\bar{s}}) > \gamma \epsilon_i^2$.

In what follows the convexity of the function $F(x)$ is used.

From the inequality

$$F(x_p) - F(x^{\bar{s}+1}) \geq r_i(g^{\bar{s}+1}, z^{\bar{s}}) + r_i(g^{\bar{s}+1}, e^{\bar{s}} - z^{\bar{s}})$$

we have

$$F(x^{\bar{s}+1}) \leq F(x_p) - \frac{\gamma}{2} r_i \epsilon_i^2 .$$

It is easy to prove that

$$\hat{F}^k(x^{\bar{s}+1}) \leq \hat{F}^l(x_p) - \frac{\gamma}{4} r_i \epsilon_i^2 ,$$

but this relation contradicts the original assumption. Statement 1 is proved.

STATEMENT 2 If Algorithm 1 generates the finite number of points $\{x_p\}$, then the last pont $x_{\bar{p}}$ belongs to X^*.

Let us suppose that $x_{\bar{p}} \bar{\in} X^*$. By virtue of the closedness, convexity and upper semi-continuity of the multi-valued mapping $\partial F(x)$ there exists $\theta > 0$ such, that

$$0 \bar{\in} \text{conv } G_\theta(x_{\bar{p}}) \ ,$$

where

$$G_\theta(x_p) = \cup \ \partial F(x), \ \|x - x_p\| \leq \theta \ .$$

Let $\varphi = \min \|\tilde{g}\|$, $\tilde{g} \in G_\theta(x_p)$. Obviously $\varphi > 0$. As $r_i \to 0$ and vector of the step direction on the internal iterations is bounded then there exists an integer $I_1(\theta)$ such, that for $i \geq I_1$ all points x^s belong to the set

$$\{x : \|x - x_p\| \leq \theta\} \ .$$

We next consider the sequence of the random events:

$$A_i = \left\{ u : \max_{s \geq t_{i+\bar{p}}} \|z^s - e^s\| > \delta_{i+\bar{p}} \right\} \ .$$

From the generalized Kolmogorov's inequality for the dependent random variables [6] follows, that

$$P(A_i) \leq \frac{C_1}{t_{i+\bar{p}} \delta^2_{i+\bar{p}}} \ ,$$

where C_1 is some constant. As $t_{i+\bar{p}} = 1/(\delta^2_{i+\bar{p}} \nu_{i+\bar{p}})$, then

$$\sum_{i=0}^{\infty} P(A_i) \leq C_1 \sum_{i+p=0}^{\infty} \nu_{i+p} < \infty \ .$$

It is not difficult to notice that the events $\{A_i\}$ can occur only the finite number of times.

Thus for a fixed elementary event $u \in U$ it is possible to indicate a sufficiently large number $I_2 \geq I_1$ such, that by $i \geq I_2$ and $s \geq t_{i+\bar{p}}$

$$\|z^s - e^s\| \leq \delta_{i+\bar{p}}$$

and hence,

$$\|z^s\| \leq \|e^s\| + \delta_{i+\bar{p}} \ .$$

Then for the instants $\tilde{s}_i$, for which $\|e^{\tilde{s}_i}\| \leq \epsilon_i$ is satisfied the inequality

$$\|z^{\tilde{s}_i}\| \leq \epsilon_i + \delta_{i+\bar{p}}$$

is fulfilled.

Therefore, there exists a sufficiently large number $I_3 \geq I_2$ such, that for $i \geq I_3$ the relation

$$\|z^{\tilde{s}_i}\| \leq \epsilon_i + \delta_{i+\bar{p}} \leq \frac{\varphi}{2}$$

is fulfilled.

We arrived at a contradiction: for one thing $\|z^{\tilde{s}_i}\| \geq \varphi$, but for another

$$\|z^{\tilde{s}_i}\| \leq \frac{\varphi}{2} \ .$$

The statement 2 is proved.

The proof of Statement 2 completes the consideration of the first case.

Let us consider the second case.

Let $\{x_p(u)\}$ be an isolated trajectory for arbitrary fixed elementary event $u \in \mathcal{U}$.

The trajectory $\{x_p(u)\}$ is determined on some σ-algebra $\mathcal{B}$, which contains expanding σ-algebras $\mathcal{B}_p$ induced by the random vectors $x_j(u)$, $j = \overline{1,p}$ or more precisely, on some probability space $(\mathcal{U}, \mathcal{B}, \mathcal{P})$ with the measure $\mathcal{P}$.

Later on the dependence x_p from u is omitted.

STATEMENT 3 For the given trajectory $\{x_p\}$ the subscript i is changed in finite number of times.

Suppose that Statement 3 is false, i.e. the subscript i has changed only the finite number of times.

Consider two sequence of the random events:

$$A_p^{(1)} = \left\{ u : \max_{k \geq \Pi^{(1)}_{i+s+p}} |\hat{F}^k(x^s) - F(x^s)| > \delta^{(1)}_{i+s+p} \right\} ,$$

$$A_p^{(2)} = \left\{ u : \max_{l \geq \Pi^{(2)}_{i+s+p}} | \hat{F}^l(x_p) - F(x)| > \delta^{(2)}_{i+s+p} \right\} .$$

It is easy to prove

$$\sum_{p=0}^{\infty} P(A_p^{(1)}) < \infty \text{ and } \sum_{p=0}^{\infty} P(A_p^{(2)}) < \infty \ .$$

Hence, for the trajectory $\{x_p\}$ the events $\{A_p^{(1)}\}$ and $\{A_p^{(2)}\}$ can occur only the finite number of times.

Thus is is possible to indicate a sufficiently large integer P_1 such, that for $p \geq P_1$, $k \geq \Pi^{(1)}_{i+s+p}$ and $l \geq \Pi^{(2)}_{i+s+p}$ the inequalities

$$|\hat{F}^k(x^s) - F(x^s)| \leq \delta^{(1)}_{i+s+p} \ ,$$

$$|\hat{F}^l(x_p) - F(x_p)| \leq \delta^{(2)}_{i+s+p}$$

are fulfilled.

Since Algorithm 1 generates the infinite number of points of the minimizing sequence $\{x_p\}$, the following inequality is satisfied the infinite numer of times.

$$\hat{F}^k(x^s) \leq \hat{F}^l(x_p) - \frac{1}{4}\gamma r_i \epsilon_i^2$$

As subscript i is changed only the finite number of times it is possible to indicate the integer P_2 such that for all $p \geq P_2$ the inequalities

$$0 < \Delta \leq \frac{1}{4}\gamma r_i \epsilon_i^2$$

are satisfied.

Since $\delta^{(1)}_{i+s+p} \rightarrow 0$ and $\delta^{(2)}_{i+s+p} \rightarrow 0$, then for the selected subscripts p the inequality

$$F(x_{p+1}) \leq F(x_p) - \frac{\Delta}{2}$$

is fulfilled. Taking p to infinity in the inequality

$$F(x_p) \leq F(x_{P2}) - \frac{\Delta}{2}(p - P2)$$

we obtain the contradiction with the boundedness of continuous function on the closed bounded set $\{x : F(x) \leq F(x_{P2})\}$.

Statement 3 is proved.

STATEMENT 4 For the fixed trajectory $\{x_p\}$ it is possible to indicate a subscript $\bar{p}$ such, that for $p \geq \bar{p}$ the inequality

$$F(x_{p+1}) \leq F(x_p)$$

is satisfied. Let us consider the inequality

$$\hat{F}^k(x^s) \le \hat{F}^l(x_p) - \frac{1}{4}\gamma r_i \epsilon_i^2 ,$$

obtained during the proof of Statement 3. This inequality is correct for $p \ge P_1$, $k \ge \Pi^{(1)}_{i+s+p}$ and $l \ge \Pi^{(2)}_{i+s+p}$.

Therefore

$$F(x_{p+1}) \le F(x_p) - \frac{1}{4}\gamma r_i \epsilon_i^2 + \delta^{(1)}_{i+s+p} + \delta^{(2)}_{i+s+p} .$$

The statement 4 follows now from the following inequality:

$$\delta^{(1)}_{i+s+p} + \delta^{(2)}_{i+s+p} \le \frac{1}{4}\gamma r_i \epsilon_i^2 .$$

From Statement 3 follows that it is possible to select the subsequence of points $\{x_{p_i}\}$ such that there exists an instant $s_i \ge t_{i+p_i}$ for which

$$\|e^{s_i}\| \le \epsilon_i .$$

STATEMENT 5 For the fixed minimizing trajectory $\{x_p\}$ the subsequence $\{x_{p_i}\}$ chosen as mentioned above converges to the set X^*.

The proof of Statement 5 is similar to the proof of Statement 2.

The convergence of the sequence $\{x_p\}$ follows from convexity of the function $F(x)$ convexity, convergence the subsequence $\{x_{p_i}\}$ and from the monotonicity of the algorithm.

Let us now define the modification of the algorithm which is the next more general and more acceptable from the practical point of view.

At first, we assume, that

$$E(\xi^s / x_0, x_1, \ldots, x_p, x^1, \ldots, x^s) = g(x^s) + \Delta_{i+p} v^s ,$$

where

$$\|v^s\| \le C_1 < \infty, \Delta_{i+p} \to 0 .$$

Secondly, the function $f(x, \omega)$ values will be used to estimate the function $F(x)$ values in the points x_p and x^s if the $\|x - x_p\|$ and $\|x - x^{s+1}\|$ are sufficiently small.

DESCRIPTION OF ALGORITHM 2.

Let x_0 be an arbitrary initial point.

Set $e^0 = \xi^0$, where $E(\xi^0/x_0) = g(x_0) + \Delta_0 v^0$.

Put $i = 0$, $s = 0$, $p = 0$.

Step 1 Compute

$$x^{s+1} = x_p - r_i e^s \ ,$$

$$e^{s+1} = e^s + \frac{1}{s+1}(\xi^{s+1} - e^s) \ ,$$

where

$$E(\xi^{s+1}/x_0, x_1, \ldots, x_p, x^1, \ldots, x^{s+1}) = g(x^{s+1}) +$$

$$+ \Delta_{i+p} v^{s+1}, \ g(x^{s+1}) \in \partial F(x^{s+1}) \ .$$

Step 2 If $s + 1 \le t_{i+p}$, then take $s = s + 1$ and to to Step 1.

Step 3 If $\|e^{s+1}\| \le \epsilon_i$, then define $i = i + 1$, $s = 0$, $e^0 = \xi^0$ and go to Step 1, where

$$E(\xi^0/x_0, x_1, \ldots, x_p) = g(x_p) + \Delta_{i+p} v^0(x_p), \ g(x_p) \in \partial F(x_p)$$

Step 4 If $k + 1 \ge \Pi^{(1)}_{i+s+p}$, then go to Step 6.

Step 5 Define

$$\hat{F}^{k+1}(x^{s+1}) = \frac{1}{k+1} \sum_{j=1}^{k+1} f(x^j, \omega^j) \ ,$$

set $k = k + 1$ and go to Step 4, where

$$\|x^j - x^{s+1}\| \le \mu^{(1)}_{i+s+p} \ .$$

Step 6 If $l + 1 \ge \Pi^{(2)}_{i+s+p}$, then go to Step 8.

Step 7 Define

$$\hat{F}^{l+1}(x_p) = \frac{1}{l+1} \sum_{j=1}^{l+1} f(x^j, \omega^j) \ ,$$

set $l = l + 1$ and go to Step 6, where

$$\|x^j - x_p\| \le \mu^{(2)}_{i+s+p} \ .$$

Step 8 If

$$\hat{F}^{k+1}(x^{s+1}) \le \hat{F}^{l+1}(x_p) - \frac{1}{4}\gamma r_i \epsilon_i^2 ,$$

then go to Step 9, else set $s = s + 1$ and go to Step 1.

Step 9 Set $x_{p+1} = x^{s+1}$, $e^0 = \xi^0$, $s = 0$, where

$$E(\xi^0 / x_0, x_1, \ldots, x_{p+1}) = g(x_{p+1}) + \Delta_{i+p} v^0$$

$p = p + 1$ and go to Step 1.

THEOREM 2

Let the function $F(x)$ satisfies the conditions of Theorem 1.

Let the trajectories $\{x_p(u)\}$ be defined on some probability space $(\mathcal{U}, \mathcal{B}, \mathcal{P})$ and for the fixed trajectory there exists constant $C < \infty$ such, that

$$\|\xi^s(u)\| \le C \text{ for any } s \ge 0 .$$

Let the algorithm parameters be such that: $0 < \gamma < 1/4$,

$$r_i > 0,\ r_i \to 0;\ \epsilon_i > 0,\ \epsilon_i \to 0 ;$$

$$t_{i+p} = \frac{4}{\delta_{i+p}^2 \nu_{i+p}} ,$$

where

$$\sum_{i+p=0}^{\infty} \nu_{i+p} < \infty$$

and

$$\delta_{i+p} > 0,\ \delta_{i+p} \to 0 ;$$

$$\Delta_{i+p} \le \min \left\{ \frac{\epsilon_i}{2C_1}, \gamma \frac{\epsilon_i^2}{4C}, \frac{\delta_{i+p}}{2C_1} \right\} ;$$

$$\Pi_{i+s+p}^{(1)} = \frac{4}{(\delta_{i+s+p}^{(1)})^2 \nu_{i+s+p}^{(1)}} ,$$

where

$$\sum_{i+s+p=0}^{\infty} \nu_{i+s+p}^{(1)} < \infty,\ \delta_{i+s+p}^{(1)} > 0,\ \delta_{i+s+p}^{(1)} \to 0 ;$$

$$\Pi^{(2)}_{i+s+p} = \frac{4}{(\delta^{(2)}_{i+s+p})^2 \nu^{(2)}_{i+s+p}} \quad ,$$

where

$$\sum_{i+s+p=0}^{\infty} \nu^{(2)}_{i+s+p} < \infty,\ \delta^{(2)}_{i+s+p} > 0,\ \delta^{(2)}_{i+s+p} \to 0 \quad ;$$

$$\delta^{(1)}_{i+s+p} + \delta^{(2)}_{i+s+p} \le \frac{\gamma}{2} r_i \epsilon_i^2 \quad ;$$

$$\mu^{(1)}_{i+s+p} \le \min \left\{ \gamma \frac{r_i \epsilon_i^2}{16L},\ \frac{\delta^{(1)}_{i+s+p}}{2L} \right\} \ ;$$

$$\mu^{(2)}_{i+s+p} \le \min \left\{ \gamma \frac{r_i \epsilon_i^2}{16L},\ \frac{\delta^{(2)}_{i+s+p}}{2L} \right\} \ ,$$

where L is a Lipschitz constant of the function $f(x, \omega)$ with respect to (x, ω).

Then either Algorithm 2 generates the finite number of points $\{x_p(u)\}$ and the last one will belong to the set X^* or all limit points of the isolated trajectory $\{x_p(u)\}$ belong to the set X^*.

The proof of Theorem 2 is similar to the proof of Theorem 1.

REFERENCES

1 Ermoliev, Yu.M.: Stochastic Programming Methods. M.: Nauka, 1976, p. 240.

2 Gupal, A.M.: Stochastic Methods of Solution of Nonsmooth Extremum Problems. Kiev: Naukova dumka, 1979, p. 152.

3 Pschenichnij, B.N. and Yu.M. Danilin: The numerical methods for the extremum problems. M.: Nauka, 1975, p. 320.

4 Chepurnoj, N.D.: A monotone method for minimization of the convex functions. - Dokl. AN USSR, 1982, No. 3, pp. 68-69.

5 Chepurnoj, N.D.: The convergence with probability 1 one random descent algorithm with averaging of the motion direction. Collection: Issledovanie Operacij i ASU. Kiev: Vyshcha shkola, 1982, No. 21, pp. 17-22.

6 Loev, M.: Probability Theory. M.: Izd-vo inostr. lit., 1967, p. 720.