Working Paper

STOCHASTIC APPROACHES TO INTERACTIVE MULTI-CRITERIA OPTIMIZATION PROBLEMS

M. Mikhalevich

February 1986
WP-86-10

International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

STOCHASTIC APPROACHES TO INTERACTIVE MULTI-CRITERIA OPTIMIZATION PROBLEMS

M. Mikhalevich

February 1986
WP-86-10

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

AUTHOR

Assistant Professor Mikhail Mikhalevich, from the Kiev State University, U.S.S.R. worked on stochastic and multicriteria optimization and wrote this paper during his three week visit to IIASA.

FOREWORD

A stochastic approach to the development of interactive algorithms for multicriteria optimization is discussed in this paper. These algorithms are based on the idea of a random search and the use of a decision-maker who can compare any two decisions. The questions of both theoretical analysis (proof of convergence, investigation of stability) and practical implementation of these algorithms are discussed. The paper was prepared within the activities of the System and Decision Sciences Program on stochastic optimization and multiple criteria decision making.

A. Kurzhanski
Program Leader
System and Decision Sciences Program.

ACKNOWLEDGEMENTS

I would like to thank Yuri Ermoliev for fruitful discussions and suggestions.

Contents

1. Introduction 1
2. The Basic Algorithm of Search for a most preferable element 3
3. Generalizations of main algorithm 10
4. Stability analysis 13

References 16

STOCHASTIC APPROACHES TO INTERACTIVE MULTI-CRITERIA OPTIMIZATION PROBLEMS

M. Mikhalevich

1. INTRODUCTION

Multicriteria problems often appear in applications, especially in technology, economics, engineering design. In many cases, another aspect is added to the multicriteria character of the problem – it includes elements of uncertainty. For example, the simplest formulation of the problem of system design for hazardous waste treatment has the following criteria:

1) system's malfunction probability

$$f_1(x) = P\{\theta > x\} = 1-p\{\theta \le x\},$$

2) system cost

$$f_2(x) = zx \quad ,$$

where x is the system capacity, θ is a random quantity of the hazardous waste, z is the unit cost of the system. Another example deals with energy system planning, where alternative variants consist of stochastic and deterministic components, such as energy consumption, energy resources and options, results of scientific progress. In a nuclear power design problem there are the following stochastic components: wind direction, weather, environmental situation, etc. In such problems it is rather difficult to define a single function which measures the utility of different components. Therefore, in this work, interactive numerical methods are developed that do not require a utility function construction. These methods use information supplied by the decision-maker (DM) who can compare different variants (decisions) in the course of iterations.

A mathematical model of such DM's ability are preference relations. A preference relation R is introduced on the set X of all decisions as follows: $x\ R\ y$ if and only if the DM says that decision x is no worse than y.

When a DM compare different variants, he makes errors. Therefore the subjective uncertainty to be connected with the DM's errors appears in decision-making problems. It should be noted that these errors are usually random by their character in a more or less serious degree. This stochastic aspect of the decision-making problems will be discussed at the end of this paper. Now let assume that the DM can compare decisions in an absolutely correct manner. It is actually not necessary for the DM to compare all different decisions from the decision set X. Algorithms discussed in this paper require only comparison of two different alternatives, given by the computer at each interaction.

By introducing the relation R, the decision-making problem is reduced to the optimization problem on the preference field (R,X). The latter assumes the following formulation which is rather general. The set of all possible decisions $X = \{x\}$ and its subset of all admissible decisions D are specified. A binary relation R is also specified on the set X. It is necessary to choose such $x^* \in D$ that for $\forall$ $x \in D$, $x^* R\, x$ holds. This problem is called the most preferable element search problem is denoted as follows:

$$x \longrightarrow \text{pref} \qquad (1)$$

$$x \in D$$

If the relation R possesses such properties as reflexivity, completeness, transitivity, continuity [8], a continuous utility index $u(x)$ (called also value function) exists and the problem (1) formally is reduced to the mathematical programming problem: $\min_{x \in D} u(x)$. However, it should be noted that neither analytical form of $u(x)$ nor its values in given points are known. This fact constitutes the basic difficulty when problem (1) is solved. Sometimes, when vector x consists of stochastic components, it is known [1] that $u(x)$ has a form of $Ev(x, \omega)$, where $v(x, \omega)$ is an unknown function and ω-random parameters. In applied problems, assumptions about the analytical form of $u(x)$ are often taken, for example: $u(x)$ is a linear function $\sum_i \alpha_i x_i$. But even in this case a difficult problem of the identification of weights α_i arises.

The lack of information about the objective function $u(x)$ makes problem (1) appears similar to the stochastic programming problems. Therefore, it is natural to use the approach of the stochastic programming to solve the problem (1).

2. THE BASIC ALGORITHM OF SEARCH FOR A MOST PREFERABLE ELEMENT

Let us illustrate this idea for the simplest case when the set D is a compact subset from E^n and the relation R possesses the properties of convexity and regularity [3] besides the properties required above. In this case the utility function $u(x)$ is a continuous quasiconvex function (i.e. the function with the convex level sets $\{y: u(y) \le u(x)\}$. The differentiability of $u(x)$ is an additional assumption for this case.

It should be noted that in this case, at every point x, and for every direction h (except directions from a set of measure zero), there exists $\gamma(x,h) \le 0$ such that either $(x + \gamma(x,h)h)\, Rx$ or $(x - \gamma(x,h)h)\, Rx$ holds. Therefore, a random vector $\xi(x,\theta)$ can be constructed by the formula

$$\xi(x,\theta) = \begin{cases} \theta, & \text{if } (x + \gamma(x,\theta)\theta)\, Rx \\ -\theta, & \text{if} (x - \gamma(x,\theta)\theta) Rx \end{cases} \tag{2}$$

(where θ is a random direction vector uniformly distributed over the n-dimensional unit sphere) which will be, with probability 1, the direction of a decrease of $u(x)$. This fact makes it possible to use, in order to solve problem (1), random search procedures of the stochastic quasigradient type.

Let us give some definitions which will help in the calculation of $E(\xi(x,\theta))$. The support functional to the convex set G at the point x is a bounded vector $l(x)$ such that $(l(x), x-y) \le 0$ holds for any $y \in G$ [10]. The set of all support functionals to set G at point x is the support set $L(x)$. It is known [10] that if $u(x)$ is a quasiconvex function, then the set of all support functionals to set $\{y: u(y) \le u(x)\}$ in point x, which is not optimal, is not empty. If this function is differentiable, then for every $l_1(x) \in L(x)$ and $l_2(x) \in L(x)$, $\frac{l_1(x)}{\| l_1(x) \|} = \frac{l_2(x)}{\| l_2(x) \|}$ holds for every x such that $\{y: u(y) \le u(x)\}$ has more than one point. If the gradient of the function $u(x)$ in point x is not equal 0, then $\frac{l(x)}{\| l(x) \|} = -\frac{u_x(x)}{\| u_x(x) \|}$ holds, where $u_x(x)$ is the gradient of $u(x)$. Hence, the support functional to $\{y: u(y) \le u(x)\}$ is the generalization of the gradient for quasiconvex functions.

Now let us calculate the value of $E(\xi,(x,\theta))$. It is clear for $x \ne x^*$ that i-th component of $E(\xi,(x,\theta))$ is equal to:

$$[E(\xi(x,\theta))]_i = \frac{1}{2v} \underset{\substack{u(x+\gamma(x,t)t) \le u(x) \\ \|t\| \le 1}}{\int \cdots \int} t_i \, dt_1 \cdots dt_n + \frac{1}{2v} \underset{\substack{u(x-\gamma(x,t)t) \le u(x) \\ \|t\| \le 1}}{\int \cdots \int} (-t_i) dt_1 \cdots dt_n$$

where v is the measure of n-dimensional sphere. According to the definition of the support functional:

$$\int_{\substack{u(x+\gamma(x,t)t)\le u(x)\\ \|t\|\le 1}} \cdots \int t_i\, dt_1 \ldots dt_n = \int_{\substack{u(x-\gamma(x,t)t \le u(x)\\ \|t\|\le 1}} \cdots \int (-t_i) dt_1 \cdots dt_n \tag{3}$$

$$= \int_{\substack{\left[\frac{l(x)}{\|l(x)\|},t\right]\ge 0\\ \|t\|\le 1}} \cdots \int t_i\, dt_1 \cdots dt_n$$

where $l(x)$ is the support functional to the set $\{y : u(y) \le u(x)\}$ at the point x.

Rotating the coordinate system for integral (2) in such a way that $\frac{l(x)}{\|l(x)\|}$ will become collinear with one of the axis in new system (see Figure 1) we can obtain:

$$E(\xi(x,\theta)) = k\frac{l(x)}{\|l(x)\|} \tag{4}$$

where $k > 0$ depends only on n.

This statement was proved in details in [2]. The statement (4) justifies the idea of using the direction $\xi(x,\theta)$ as an analogue of the stochastic quasigradient in stochastic quasigradient projection method. In this fashion the following interactive method is obtained.

Take the initial point x^0 to be an arbitrary n-dimensional vector. The algorithm consists of constructing the sequence $\{x^s\}$ by the following rule. Suppose that the approximation x^{s-1} to the solution x^* of the problem (1) is obtained before the beginning of the s-th step of algorithm. On step numbers s computer generates θ^s -- the s-th independent observation of the realization of random vector θ and calculates $x^{s-1} + \gamma\theta^s$. If DM considers that $x^{s-1}+\gamma\theta^s$ is no worse than x^{s-1}, then $\xi(x^{s-1},\theta^s) = \theta^s$; otherwise DM should compare decisions $x^{s-1} - \gamma\theta^s$ and x^{s-1}. If DM considers that $x^{s-1} - \gamma\theta^s$ is no worse than x^{s-1}, then $\xi(x^{s-1}, \theta^s) = -\theta^s$; otherwise the value of γ is decreased and the procedure of comparison of $x^{s-1}+\gamma\theta^s, x^{s-1}, x^{s-1}-\gamma\theta^s$ continues for a new value of γ (but for the old value of θ^s). Having defined $\xi(x^{s-1},\theta^s)$, computer calculates x^s by the formula:

$$x^s = \pi_D(x^{s-1} + \rho_s \xi(x^{s-1},\theta^s)) \tag{5}$$

where $\pi_D(\cdot)$ is the projector on set D, i.e. the point from D being nearest to the argument. The choice of the step ρ_s must satisfy the conditions:

$$\rho_s \ge 0, \sum_{s=0}^{\infty} \rho_s = \infty, \ a.s., \ \sum_{s=0}^{\infty} \rho_s^2 < \infty, \ a.s. \tag{6}$$

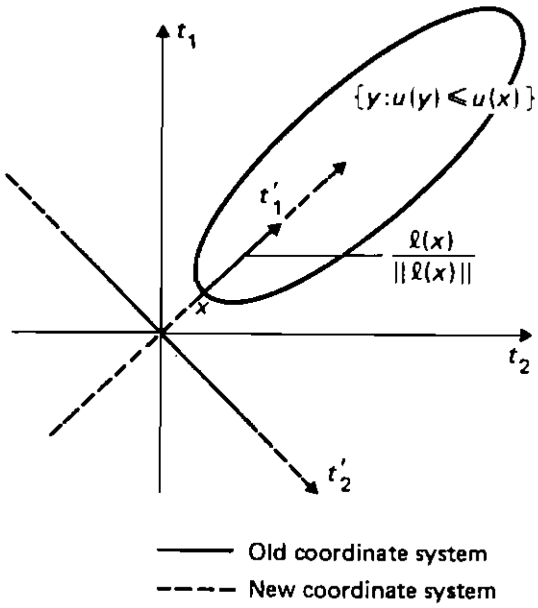

Figure 1. Rotating the coordinate system in the case n =2.

It should be noted that the sequence of x^s is a random sequence determined on some probability space $\{\Omega, A, P\}$. According to (4)

$$E(\xi(x^{s-1}, \theta^s) \,|\, x^0, \ldots, x^{s-1}) = k \frac{l(x^{s-1})}{||l(x^{s-1})||} \tag{7}$$

This fact is the basis for the proof of convergence of the algorithm. But its convergence does not follow immediately from the convergence of an analogous algorithm given in [4]. A quasiconvex function is not always convex hence the proof of the algorithm (5) convergence is an independent problem. However, it is possible to prove that $E(\xi(x^s, \theta^{s+1}) \,|\, x^0, \ldots, x^s)$ possesses the main properties of a quasigradient that are needed in the proof of convergence. Namely it is the direction in which distance between point x^s and the set X^* of optimal points decreases. Let us prove it.

Lemma 1. Let $x^s \to x'(\omega) \bar{\in} X^*$ for some $\omega \in B$, $B \in A$, $P(B)>0$. Then there exists such $\lambda>0$ which depends only on x' and ω that

$$\left[\frac{l(x^s)}{\|l(x^s)\|}, x^s - x^*\right] \le -\lambda < 0$$

for sufficient large numbers s, for any $\omega \in B$ and any $x^* \in X^*$.

Proof:

Because of $x^s \to x'(\omega) \bar{\in} X^*$ for $\omega \in B$, the following statements holds:

(A) for any $\omega \in B$ and $\varepsilon > 0$ such $s_0(\omega, \varepsilon)$ exists, that for $s > s_0$ we have

$$x^s \in \{x : \|x - x'\| < \varepsilon\}$$

(B) for any $\omega \in B$ such $\delta(\omega)$ exists that

$$\min_{x^* \in X^*} \| x' - x^* \| > \delta(\omega)$$

Let in (A) $\varepsilon = \delta(\omega)/2$, then the statement

(C) for any $\omega \in B$ such $\delta_1(\omega) > 0$ exists, that for $s > s_0$ we have

$$x^s \in \{x : \min_{x^* \in X^*} \| x - x^* \| > \delta_1(\omega)\} ,$$

this follows from (A) and (B).

According to (C) and continuity of $u(x)$, there exists such s_2, which depends only on x' and ω, that for $\delta > s_0$ and for all $x^* \in X^*$ the points $x^* - \delta_2 \frac{l(x^s)}{\|l(x^s)\|}$ belong to the set $\{y : u(y) \le u(x^s)\}$

Therefore for any $\omega \in B$, any $x^* \in X^*$ and any $s > s_0$,

$$\left[\frac{l(x^s)}{\|l(x^s)\|}, x^s - x^*\right] = -\delta_2 + \left[\frac{l(x^s)}{\|l(x^s)\|}, x^s - \left[x^* - \delta_2 \frac{l(x^s)}{\|l(x^s)\|}\right]\right]$$

holds. If we define the point $x^{(1)s}$ as follows:

$$x^{(1)s} = x^* - \delta_2 \frac{l(x^s)}{\|l(x^s)\|} \in \{y : u(y) \le u(x^s)\}$$

We obtain the inequality $(l(x^s), x^s - x^{(1)s}) \le 0$ from the definition of the support functional This means that

$$\left[\frac{l(x^s)}{\|l(x^s)\|}, x^s - x^*\right] \le -\delta_2 = -\lambda \le 0$$

Lemma is proved.

Using this lemma, let us prove the following theorem.

Theorem 1.

Let the sequence $\{x^s\}$ be constructed by (5) and the step ρ_s satisfy the conditions (6). Then

$$\min_{x^* \in X^*} \| x^s - x^* \| \underset{s \to \infty}{\to} 0 \quad a.s.$$

Proof:

Let us show that if $x^s \to x'(\omega) \overline{\in} X^*$ for some $\omega \in B$, $P(B) > 0$, then for sufficiently large s

$$\min_{x^* \in X^*} \| x^{s+1} - x^* \|^2 \leq \min_{x^* \in X^*} \| x^s - x^* \|^2 - \lambda' \rho_s + W_s \tag{8}$$

holds, where $\lambda' > 0$ depends only on x' and ω,

$$E(W_s \mid x^0, \ldots, x^s) = 0, \sum_{s=0}^{\infty} E(W_s) < \infty$$

Indeed,

$$\min_{x^* \in X^*} \| x^{s+1} - x^* \|^2 = \| x^{s+1} - x^*(x^{s+1}) \|^2 \leq \| x^{s+1} - x^*(x^s) \|^2 \tag{9}$$

where $x^*(x^\nu) \in X^*$ is such x^*, for which

$$\min_{x^* \in X^*} \| x^\nu - x^* \|^2 = \| x^\nu - x^*(x^\nu) \|^2 \quad \text{holds.}$$

According to (5) and projection properties (4),

$$\| x^{s+1} - x^* \|^2 = \| \pi_D(x^s + \rho_s \, \xi(x^s, \theta^{s+1})) - x^*(x^s) \|^2 \leq$$

$$\leq \| x^s + \rho_s \, \xi(x^s, \theta^{s+1}) - x^*(x^s) \|^2 \leq \| x^s - x^*(x^s) \|^2 +$$

$$+ 2\rho_s \, (\xi(x^s, \theta^{s+1}), x^s - x^*(x^s)) + \rho_s^2 = \min_{x^* \in X^*} \| x^s - x^* \|^2 +$$

$$+ 2\rho_s \, k \left[\frac{l(x^s)}{\| l(x^s) \|}, x^s - x^*(x^s) \right] + \rho_s^2 + 2\rho_s \, (\xi(x^s, \theta^{s+1}) -$$

$$- k \frac{l(x^s)}{\| l(x^s) \|}, x^s - x^*(x^s))$$

According to lemma 1 for $s > s_0$, and $\omega \in B$

$$2\rho_s\ k\ \left[\frac{l(x^s)}{\|l(x^s)\|},\ x^s - x^*(x^s)\right] \le -2\ \rho_s\ \lambda\ k$$

where $\lambda > 0$ depends only on x' and ω; according to the theorem assumptions for sufficiently large s (i.e. for $s > s'$) $\lambda k > \rho_s$ holds.

Therefore if $s > \max(s_0, s')$ then for arbitrary $\omega \in B$

$$\| x^{s+1} - x^*(x^s)\|^2 \le \min_{x^* \in X^*} \| x^s - x^* \|^2 - \lambda k\ \rho_s + W_s \tag{10}$$

holds, where

$$W_s = 2\rho_s \left[\xi(x^s, \theta^{s+1}) - k\frac{l(x^s)}{\|l(x^s)\|},\ x^s - x^*(x^s)\right]$$

According to (7)

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

According to the theorem assumptions $E(W_s^2) \le cE(\rho_s^2)$, where $c < \infty$, so $\sum_{s=0}^{\infty} E(W_s^2) < \infty$.

By combining (9) and (10), the statement (8) is obtained. The conclusion of the proof of the theorem is based on Nurminsky's results (6) and is fully analogous to the proof of theorem 3 from Chapter 4 [4].

Method (5) analysed above is the basis for an interactive decision-making method which uses comparison procedures. There can be different ways of implementing such procedures such as a direct comparison of the values of criteria by the DM; a direct comparison of all components of decisions by DM; simulation.

Several numerical experiments based on these ideas were made with an applied decision-making system for a dynamic planning model of an economic system.

The basic model in this example has the form:

$$(y(1), \ldots, y(T), x(T+1)) \to \text{pref}$$

$$x(t) = ax(t) + b(x(t+1) - x(t)) + y(t),\ \ t = \overline{1,T}$$

$$x(t+1) \ge x(t),\ y(t) \ge 0,\ \ t = \overline{1,T}$$

$$x(1) = x^0$$

where $x(t)$ is the value of the system total product at the time interval t, $y(t)$ is the value of the system final product at the time interval t, x^0 is the known value of the system total product at the beginning, a, b are known parameters.

Information for the comparison procedure was introduced in a combined form: the information about $(y(1), \ldots, y(T))$ was given in graphical form, and the information about $x(T+1)$ (together with information about $\sum_{t=1}^{T} y(t)$) in numerical form (see Figure 2).

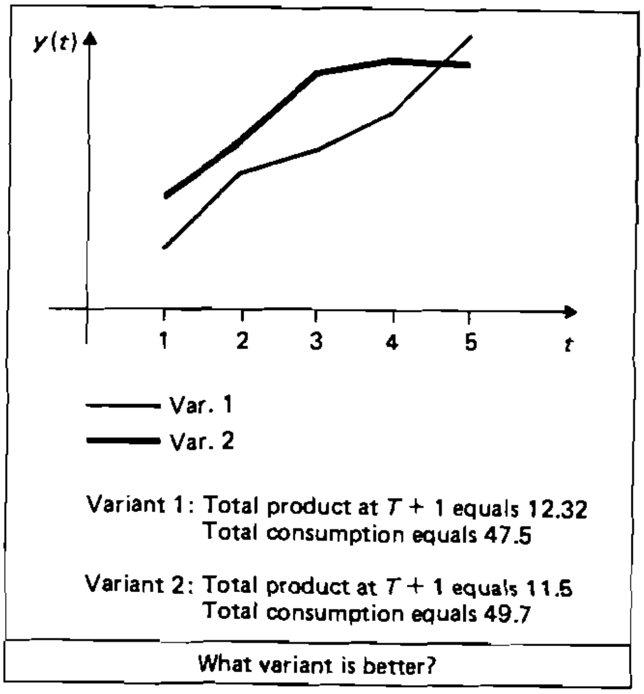

Figure 2. Example of the information given to DM in the dynamic planning model for economic system.

Tests were made for a case when $T = 5$. In the test runs, a known utility function $u(y(1), \ldots, y(T), x(T+1))$ was used as a DM model. Sufficiently good approximates to solutions (with the accuracy from 5% to 1% of x^* measured by the distance $\| x - x^* \|$) were obtained between 41 and 65 iterations of method (5) (these results depend on the accuracy and assumed utility function). Experiments with this algorithm in interactive option were made also.

Another set of numerical experiments were made also with consumption models.

3. Generalizations of main algorithm

It is rather difficult to find the projection of the point on set D if the dimension of x is large and D is given by many constraints. Therefore a method that substitutes the projection by solving a linear problem in each iteration might be useful. The idea of this method is the following: it is known that, if gradient of $u(x)$ is not a zero vector, then $\frac{l(x)}{\|l(x)\|} = \frac{u_x(x)}{\|u_x(x)\|}$. This makes it possible to minimize at each iteration the linear approximation $\left[\frac{l(x^{s-1})}{\|l(x^{s-1})\|}, x - x^{s-1}\right]$ of the function $u(x)$, where x^{s-1} is the point obtained in the previous iteration. The value of $\frac{l(x^{s-1})}{\|l(x^{s-1})\|}$ may be approximated by its statistical estimate z^s which is obtained by the formula: $z^s = z^{s-1} + \delta_s (-\xi(x^{s-1}, \theta^s) - z^{s-1})$, $s = 0, 1, 2, \ldots$

where z^0 is an arbitrary vector, δ_s must possess the properties:

$$\sum_{s=0}^{\infty} \delta_s^2 < \infty, \quad \sum_{s=0}^{\infty} \delta_s = \infty, \quad \frac{\|x^{s-1} - x^{s-2}\|}{\delta_s} \underset{s \to \infty}{\to} 0 \quad a.s.$$

Such approximations were investigated in (4) where it was proved that

$$\| z^s - E(-\xi(x^{s-1}, \theta^s) \mid x^0, \ldots, x^{s-1}) \| \underset{s \to \infty}{\to} 0 \quad a.s.$$

These estimates are used in the so-called stochastic linearization method, which produces the sequence $\{x^s, z^s\}$ according to the rule:

x^0 is an arbitrary admissible decision

z^0 is an arbitrary vector from E^n,

$$z^s = z^{s-1} + \delta_s (-\xi(x^{s-1}, \theta^s) - z^{s-1}) \tag{11}$$

$$(z^s, \bar{x}^s) = \min_{x \in D} (z^s, x),$$

$$x^s = x^{s-1} + \delta_s (\bar{x}^s - x^{s-1}), \quad s = 1, 2, \ldots ,$$

where the steps δ_s, ρ_s satisfy the conditions:

$$\sum_{s=0}^{\infty} \delta_s^2 < \infty, \quad \sum_{s=0}^{\infty} \delta_s = \infty, \quad \rho_s / \delta_s \underset{s \to \infty}{\to} 0, \quad \delta_s \ge 0,$$

$$0 \le \rho_s \le 1, \quad s = 1, 2, \ldots . \quad a.s.$$

This method is convenient for the case when the set D is given by a set of linear equations and inequalities. In this case the decision $\bar{x}^{s-1}$ may be used as the starting approximation of the solution for solving the problem $\min_{x \in D} (z^s, x)$.

For a case when the set D is specified by non-linear inequalities, special penalty function methods are presented in (7). These methods are applicable for cases when some constraints are specified by preference relations or have a form of $E f_i (x, \omega) \le 0$ where $f_i (x, \omega)$ are convex with respect to x for almost every realization of stochastic variable ω.

Another approach to the multicriteria problems is connected with constructing the Pareto-optimal decision set. Let the consequences of decision x in problem (1) be described by the criteria vector $(f_1(x), \ldots, f_m(x))$. Typically, scalarizing functions $G(x , d)$ of known analytical form which depend on additional parameters α are used to parametrize the set of all Pareto decisions. It is done by solving the problem:

$$G(x, \alpha) \underset{x}{\to} \min \qquad (12)$$

$$x \in D .$$

For all parameter values α from a specified set H the Pareto-optimal decision set with respect to the criteria $(f_1(x), \ldots, f_m(x))$ will be either covered or approximated by the set of problem (12) solutions. Examples of such scalarizing functions are $\sum_{i=1}^{m} \alpha_i f_i(x)$, $\max_{i=\overline{1,m}} (\alpha_i f_i(x))$, and many others (see [11]).

Hence, the solution of problem (1) may be obtained as the solution of problem (12) for some "best" value of α". Our idea is to use the search for the most preferable elements to identify this "best" value.

Let $x^*(\alpha)$ be the solution of the problem (12) obtained for a given value α. The new relation R_1 may be constructed on the set H of all possible values of α in the following way:

$\alpha^{(1)} R_1 \alpha^{(2)}$ if, and only if, $x^* (\alpha^{(1)}) R x^* (\alpha^{(2)})$ holds. (Usually, $H = E^m$ or is a unit simplex from E^m.) Now the problem (1) is substituted by the problem:

$$\alpha \underset{R_1}{\to} pref$$

$$\alpha \in H .$$

This problem may be solved by the above mentioned methods, for example, by method (5). The advantages of such an approach is the possibility of using fast deterministic methods (such as variable-metric methods (5) or quadratic approximation methods) for the solving of problem (12) in the case when D is given by non-linear constraints; often, another advantage is the decrease of dimensionality of the problem. This approach gives good possibilities for DM's learning (11). But the assumption about quasiconvexity and differentiability of the function $V(\alpha) = u(x^*(\alpha))$ is not necessarily justified even in the case of quasiconvexity and differentiability of the functions $u(x)$ and $G(x,\alpha)$. Therefore, it is interesting to substantiate the methods of search for the most preferable element such as (5) for the case when we do not require the differentiability and quasiconvexity of $u(x)$.

Let us start with the case when function $u(x)$ is quasiconvex but nondifferentiable. The approach similar to the stochastic smoothing (9) is used for this case. Because of its quasiconvexity, the function $u(x)$ will be differentiable with probability 1 at a randomly chosen point $\tilde{x} = x + \hat{\mu}$, where $\hat{\mu}$ is the random vector with independent components uniformly distributed over $[-\alpha,\alpha]$, $\alpha>0$. Therefore, the vector $\xi(x^{s-1}, \theta^s)$ in (5) may be constructed by the rule:

$$\xi(x^{s-1}, \theta^s) = \begin{cases} \theta^s, & \text{if } (\tilde{x}^{s-1} + \gamma(\tilde{x}^{s-1}, \theta^s)\,\theta^s) R\, \tilde{x}^{s-1} \\ -\theta^s & \text{if } (\tilde{x}^{s-1} - \gamma(\tilde{x}^{s-1}, \theta^s)\,\theta^s) R\, \tilde{x}^{s-1} \end{cases} \tag{13}$$

where $\tilde{x}^{s-1} = x^{s-1} + \mu^s, \mu^s$ is the observation of the random vector with independent components uniformly distributed over $[-\alpha_s, \alpha_s]$.

Because of the differentiability of $u(x)$ at the point $\tilde{x}^{s-1}$, we have:

$$E(-\xi(x^{s-1}, \theta^s) \mid x^0, \ldots, x^{s-1}, \mu^1, \ldots, \mu^s) = k \frac{l(\tilde{x}^{s-1})}{\|l(\tilde{x}^{s-1})\|}$$

and

$$E(-\xi(x^{s-1}, \theta^s) \mid x^0, \ldots, x^{s-1}) = k\, E\left[\frac{l(\tilde{x}^{s-1})}{\|l(\tilde{x}^{s-1})\|} \mid x^0, \ldots, x^{s-1}\right]$$

This fact indicates that $\xi(x^{s-1}, \theta^s)$ determined by (13) is not a stochastic quasigradient of $u(x)$. But it possess the basic property necessary for the convergence of the algorithm (5) - it is the direction in which the distance to X^* decreases, if $\tilde{x}^{s-1}$ is not too close to X^*.

Lemma 2. Let $x^s \to x' \,\overline{\in}\, X^*$ for some $\omega \in B$, $P(B) > 0$, let $\xi(x^s, \theta^{s+1})$ be determined by (13) and $\alpha \underset{s\to\infty}{\to} 0$

Then there exists $\lambda > 0$, which depends only on x' and ω, such that:

$$(E(-\xi\,(x^s, \theta^{s+1}) \mid x^0, \ldots, x^s),\, x^s - x^*) =$$

$$k\left[E\left[\frac{l(\tilde{x}^{s-1})}{\|l(\tilde{x}^{s-1})\|} \mid x^0, \ldots, x^s\right], x^s - x^*\right] \le -\lambda$$

for sufficiently large numbers s, for any $\omega \in B$ and any $x^* \in X^*$.

The proof of this lemma is not principally different from the proof of lemma 1.

Lemma 2 shows that the direction $\xi(x^s, \theta^{s+1})$ being determined by (13) may be used in the frame of method (5) if $\alpha_s \to 0$ from $s \to \infty$ holds. For example, while using lemma 2 instead of lemma 1, it is possible to prove the following theorem (which proof does not differ from the proof of theorem 1):

Theorem 2

Let $\{ x^s \}$ be constructed by algorithm (5), where $\xi(x^{s-1}, \theta^s)$ is specified by (13) and suppose that

$$\sum_{s=0}^{\infty} \rho_s = \infty, \quad \sum_{s=0}^{\infty} \rho_s^2 < \infty, \quad \alpha_s \underset{s\to\infty}{\to} 0 \quad \text{holds } a.s.$$

Then

$$\min_{x^* \in X^*} \| x^s - x^* \| \underset{s\to\infty}{\to} 0 \quad \text{with probability 1.}$$

The convergence of methods (5) and (11) were examined also for the case when function $u(x)$ is non-convex, but differentiable. The convergence of these methods to the set of points which satisfied the Kuhn-Tucker conditions was proved with the additional assumption about the choice of the parameter $\gamma(x^{s-1}, \theta^s)$, namely, that $\gamma(x^{s-1}, \theta^s) \underset{s\to\infty}{\to} 0 \quad a.s.$

4. Stability analysis

The stability of the method (5) was investigated in the following model.

Let a randomization R_e (representing distortion by DM's random errors) be used in (1) instead of relation R and let $p(x, y)$ characterize the distortion by specifying the probability of an erroneous comparison of the decisions x and y.

It was proved in [7] that if $p(x, y) \le p_0 < 1/2$ for all x, y, then the usage of R_e in (2), (13), instead of R does not influence the convergence of the method (5).

In particular, the following theorems were proved in [7].

Theorem 3

Let the function $u(x)$ be convex and differentiable, $p(x, y) = p_0 < 1/2$ for any x, y, and let $\gamma(x, \theta) \cdot Y_s$ where $Y_s \underset{s \to \infty}{\to} 0$ a.s. be used in (2) instead of $\gamma(x, \theta)$ (s denotes here the number of iteration of the method (5)). Let $\{x^s\}$ be constructed by (5) with the use of R_e instead of R. Then

$$\min_{x^* \in X^*} \| x^s - x^* \| \underset{s \to \infty}{\to} 0 \quad \text{with probability 1 (a.s).}$$

Theorem 4

Let the function $u(x)$ be convex and differentiable, $p(x, y) < 1/2$ for any x, y and let $D = E^n$. Let $\{x^s\}$ be constructed by (5) with the use of R_e instead of R. Then

$$\min_{x^* \in X^*} \| x^s - x^* \| \underset{s \to \infty}{\to} 0 \quad a.s.$$

Similar results were proved also for more general cases.

The idea of proofs of these theorems is the following. Let us examine the convergence of a random process of the form:

$$x^{s+1} = \pi_D(x^s + \rho_s \eta^s), \quad s = 0,1,2,\dots.$$

where x^0 is an arbitrary vector $\| x^0 \| \le c \le \infty$, η^s is an arbitrary random vector. The conditions for η^s vectors that ensure the convergence of $\{x^s\}$ to the set X^* of extreme points of some convex differentiable function $u(x)$ can be specified as follows:

(1) if $x^s \to x'(\omega) \overline{\in} X^*$ for some $\omega \in B$, P (B) > 0, implies

$$(E(\eta^s \mid x^0, \dots, x^s), x^s - x^*) \le -\delta(x'(\omega), \omega) < 0 \qquad (14)$$

for $\omega \in B$, for sufficient large s and for every $x^* \in X^*$, and if the step size of ρ^s satisfy the conditions:

$$\sum_{s=0}^{\infty} \delta_s = \infty, \ \sum_{s=0}^{\infty} \delta_s^2 < \infty, \ a.s.$$

then

$$\min_{x^* \in X^*} \| x^s - x^* \| \underset{s \to \infty}{\to} 0 \quad a.s.$$

(2) $x^s \to x'(\omega) \overline{\in} X^*$ for some $\omega \in B$, $P(B) > 0$, implies:

$$(\nabla u(x^s),\ (E(\eta^s \mid x^0, \ldots, x^s)),\ \le -\delta_1 (x'(\omega),\omega) < 0 \qquad (15)$$

for almost every $\omega \in B$, P (B) > 0, for sufficient large s and for every $x^* \in X^*$, and if the step size of ρ_s satisfy the conditions:

$$\sum_{s=0}^{\infty} \delta_s = \infty,\ \sum_{s=0}^{\infty} \delta_s^2 < \infty,\ a.s.$$

then

$$\min_{x^* \in X^*} \| x^s - x^* \| \underset{s \to \infty}{\to} 0 \quad a.s.$$

Therefore we have to estimate the value of

$$E(\xi(x^{s-1}, \theta^s) \mid x^0, \ldots, x^{s-1})$$

assuming that $\xi(x^{s-1}, \theta^s)$ is obtained by (2) with the use of R_e instead of R and to show that, under the adopted assumptions, this value satisfies the condition (14) or the condition (15).

It should be noted that under this assumption the value of

$$E(\xi(x^{s-1}, \theta^s) \mid x^0, \ldots, x^{s-1})$$

may be far from the gradient of $u(x)$, but it possesses the property (14) or the property (15) to ensure the convergence of the algorithm (5).

The conditions (14) and (15) and their generalizations may be successfully used for the analysis of stability of the stochastic quasigradient method for the case of errors in the calculation of stochastic quasigradient [6].

These results make it possible to use the proposed methods even under the assumption that DM might do random errors. These results are very helpful in the cases of nontransitivity of DM's answers (if this nontransitivity could be explained as random errors of DM), for example, in collective decision-making problems.

Computer implementation is being developed on the basis of the above described methods. They include step control procedures to control ρ_s, δ_s and other parameters and special procedures which make it possible for DM to offer his own decisions and check their consequences.

These software packages are oriented towards their incorporation in a system of applied software for decision support.

References

1 Ermoliev, Ju. M. (1976) Methods of Stochastic Programming Nauka, Moscow, 240 p (in Russian)

2 Ermoliev, Ju. M. (1979) A.L. Stochastic Models and Methods in Economical Planning. Nauka, Moscow, 254 p (in Russian)

3 Fishbern. (1970) Utility Theory for Decision-making. Wiley and Sons, New York, 234p

4 Gupal, A.M. (1979) Stochastic Methods of Solution of Non-Smooth Extremum Problems. Naukova Dumka, Kiev, 150 p. (in Russian)

5 Mikhalevich, M.V. (1980) Generalized Stochastic Method of Centers. Cybernetics, No. 2, Vol. 16, pp 293-296

6 Mikhalevich, M.V. (1980) Stability of Stochastic Programming Methods to Stochastic Quasigradient Computing Errors. Cybernetics, No. 3, Vol. 16, pp. 434-438

7 Mikhalevich, M.V., L.B. Koshlai (1983) Some Questions of Stability Theory of the Stochastic Economical Models. Lecture notes in Mathematics Vol. 982, pp 123-135.

8 Nikaido, H. (1968) Convex Structures and Economic Theory. Academic Press (1968) New York, 405 p (in English) Miz, (1972) 517 p (in Russian)

9 Nuzminskii, E.A. (1980) The Numerical Methods for the Solution of Deterministic and Stochastic Minimax Problems - Naukova Dumka, Kiev, 156 p (in Russian)

10 Shor, N.Z. (1978) Methods for non-differentiable Functions Minimization and their applications. Naukova Dumka, Kiev, 231 p (in Russian)

11 Wierzbicki, A.P. (1980) Multiobjective trajectory optimization and model semiregularization. IIASA, Laxenburg, WP-80-181