INFORMATICA, 2015 Vol. 0, No. 0, 1

SAMPLE APPROXIMATIONS OF MULTIOBJECTIVE STOCHASTIC OPTIMIZATION PROBLEMS

Bogdan NORKIN

Glushkov Institute of Cybernetics, National Academy of Sciences of Ukraine,
Glushkov avenue 40, 03187 Kiev, Ukraine
E-mail: bogdan.norkin@gmail.com

Abstract. The article describes approximation technique for solving multiobjective stochastic optimization problems. As a generalized model of a stochastic system to be optimized a vector "input – random output" system is used. Random outputs are converted into a vector of deterministic performance/risk indicators. The problem is to find those inputs that correspond to Pareto-optimal values of the output indicators. The problem is approximated by a sequence of deterministic multicriterion optimization problems, where, for example, the objective vector function is a sample average approximation of the original one and the feasible set is a discrete sample approximation of the feasible inputs. Approximate optimal solutions are defined as weakly Pareto efficient ones within some vector tolerance. Convergence analysis includes establishing convergence of the general approximation scheme and establishing conditions of convergence with probability one under proper regulation of sampling parameters. Parallel computations are used for statistical evaluation of the performance indicators, as well as to accelerate sampling in the space of the system inputs. The proposed solution technique can also be interpreted as an interactive parallel Monte Carlo method with selection of approximately Pareto-nondominated points. The proposed technique is illustrated by an example of multicriterion optimization in insurance.

Key words: Multiobjective stochastic optimization, sample approximation, approximate Pareto optimality, random search, parallel Monte Carlo method, optimization of insurance business.

1. Introduction. Contemporary approach to optimal decision making is based on systems modeling and systems optimization. Any complex system can be discribed by an "input-output" model $y = g(x, \omega)$, where x denotes the input parameter vector from some feasible set X, ω is a vector of uncertain parameters from a set Ω and y is an output vector from a set Y, g is some mapping of $\mathrm{X} \times \Omega$ into Y. The model g may be given by a simulation computer program or as an output result of an optimization solver. The optimization (e.g., maximization) is applied to some utility functional $f(x, y)$ under conditions $y = g(x, \omega)$, $x \in \mathrm{X}$.

The vector ω of uncertain parameters can be either deterministic or random, with distribution P. In the first case, the optimization problem reads as: $\min_{\omega \in \Omega} f(x, g(x, \omega)) \to \max_{x \in X}$ that corresponds to the so-called minimax decision-making approach. In the second case, the problem relates to the stochastic

programming and, in particular, can be formulated as: $F(x) = \mathrm{E} f(x, g(x, \omega)) \to \max_{x \in X}$, where E denotes the mathematical expectation operation (Shapiro et al., 2014).

However, efficient and unambiguous choice of the utility function f is not always possible. Although, a preference relation can exist on the set of "input-output" pairs $\mathrm{X} \times \mathrm{Y}$ or on Y, that allows to consider only non-dominated "input-output" pairs. If parameter ω can be reasonably chosen and fixed, then such situation relates to the deterministic multi-criteria optimization, $g(x, \omega) \to \max_{x \in X}$. If the value of ω is uncertain, then this situation is qualified as an uncertain programming and can be formalized in different ways (Liu, 2005). Note that the stochastic programming problem already contains a vector criterion $f(x, g(x, \omega))$, $\omega \in \Omega$, with a large number of components that, in many cases, are reduced to one or more aggregate indicators. The most commonly used indicator is the average value $F(x) = \mathrm{E} f(x, g(x, \omega))$, along with the variance functions, probability, quantile (VaR) and other risk indicators (Gutjahr and Pichler, 2013), (Shapiro et al., 2014), (Stancu-Minasian, 1984). Optimization of such indicators requires substantial computational resources, and in particular, usage of parallel computing. In this paper we consider the usage of parallel calculations for solving multiobjective stochastic optimization problems of actuarial mathematics.

2. Literature review. Unlike standard one-criterion stochastic programming problems (Shapiro et al., 2014) the problem of output vector optimization can contain non-convex, non-smooth and discontinuous functions, so traditional stochastic programming methods like gradient type procedures might not be applicable. In this case random search methods, for example, evolutionary or hybrid algorithms should be applied (Branke et al., 2008), (Coello Coello, 2006), (Gutjahr, 2012), (Konak et al.,2006), (Marler and Aurora, 2004). In case of small dimension n of the set $X \subset \mathrm{R}^n$ a simple random search method can appear to be competitive, moreover it allows natural parallelization. This method randomly generates a cloud of points in the feasible region, and then non-dominated points are selected from the cloud. In the vicinity of the selected points new random points are generated, again non-dominated points are chosen and so on. Efficiency is boosted due to the fact that the new points are generated in perspective areas. However, in solving multidimensional combinatorial problems the simple random search method gives the way for more sophisticated algorithms, see e.g. (Kleeman and Lamont, 2008), (Žilinskas, 2014). In case of stochastic multicriterion optimization additional difficulty consists in proper numerical evaluation of the vector objective function.

In (Norkin, 2014c) the results of previous author's works (Norkin, 2012), (Norkin, 2014a) on simulation of insurance business are extended towards dynamic stochastic multicriterion optimization. In the present paper we analyze an interactive multicriterion controlled random search method, where the region for new random search is defined by a decision maker on the basis of visual analysis of the set of non-dominated points from the previous iteration. Probabilistic per-

formance indicators are estimated by means of parallel Monte Carlo simulations. Weakly Pareto-optimal (within some vector tolerance) set is searched. As compared to (Norkin, 2014c) in the present paper we focus on convergence analysis of the method. First we prove convergence of a general multicriterion approximation scheme. The original problem is approximated by a sequence of deterministic multicriterion optimization problems, where the original multiobjective function is approximated by a sequence of graphically convergent vector functions. The latter concept extends the concept of epi-convergence (Rockafellar and Wets, 1998; sec. 7B) from scalar to a vector case. In particular, the objective vector function can be replaced by sample average approximations and the feasible set is substituted by its discrete sample approximations. We establish conditions of convergence with probability one under proper regulation of sampling parameters.

Other non-scalarization approaches to stochastic multicriterion optimization are reviewed in (Gutjahr and Pichler, 2013). Overview of computer systems that implement interactive deterministic multi-criteria optimization is available in (Poles et al., 2008). Usage of parallel calculations in multi-criteria optimization is discussed in (Tabli et al.,2008). In (Fliege and Xu, 2011) another (scalarization) approach to vector stochastic optimization was studied: the vector performance indicators such as mathematical expectations are evaluated by means of Monte Carlo method and then are aggregated into scalar indicators by means of nonlinear utility functions. In (Jevne et al., 2012) evolutionary multiobjective optimization algorithms are applied to search for Pareto-optimal financial portfolios, other applications are considered in (Stancu-Minasian, 1984), (Ben Abdelaziz, 2012), (Gutjahr and Pichler, 2013).

3. Multiobjective stochastic optimization. A recent paper (Gutjahr and Pichler, 2013) provides a good review of stochastic multiobjective optimization problem settings and corresponding solution techniques. In context of "input-output" models, when the utility function is unknown, we have to deal directly with the vector model $y = g(x)$, which maps the set of inputs X into a set of outputs $Y = g(X) = \{y = g(x),\, x \in X\} \subseteq \mathrm{Y}$. At the same time some kind of preference relation may be available on the set of outputs. In case of m-dimensional output space $\mathrm{Y} \subseteq \mathrm{R}^m$ this relation is usually defined by the cone of nonnegative vectors $\mathrm{R}^m_+ = \{y = (y_1, ..., y_m) \in \mathrm{R}^m :\ y_i \geq 0,\, 1 \leq i \leq m\}$ or by the cone of positive vectors $\mathrm{R}^m_{++} = \{y \in \mathrm{R}^m :\ y_i > 0,\, 1 \leq i \leq m\}$. We say that $y_1 \prec y_2$ if and only if (iff) $(y_2 - y_1) \in \mathrm{R}^m_{++}$. The set of $\prec$-optimal vectors Y^* is a subset of $Y = g(X)$ such that there is no $y \in Y$, $y^* \prec y$. Corresponding set of inverse images of $X^* = \{x \in X :\ g(x) \in Y^*\}$ is called the set $\prec$-optimal solutions. Vector optimization problem consists in finding or approximating the sets Y^* and X^*.

If the model contains uncertain or stochastic parameters ω and instead of a utility function only a preference relation $\prec$ is known on the set of multiple outputs, then we are dealing with the problem of a vector (or multicriterion) stochastic programming (Stancu-Minasian, 1984), (Gutjahr and Pichler, 2013). In this case

for each ω, there is a multiple-output set $Y(\omega) = f(X,\omega) = \{y = g(x,\omega),\ x \in X\}$ and its $\prec$-optimal subset $Y^*(\omega)$ and we have to determine what set could be considered as a common optimal for all ω. To do this, we have to define additional preference relation on the set of random vectors $g(x,\omega)$.

In case of one-dimensional stochastic model, when $y = g(x,\omega) \in \mathrm{R}$, the required preference relation can be a stochastic dominance of the first $\prec_{(1)}$ or higher orders $\prec_{(i)}$, $i \geq 2$, (Müller and Stoyan, 2002). Let us remind that $g(x_1,\omega) \prec_{(1)} g(x_2,\omega)$, if for distribution functions $F_1(t)$ and $F_2(t)$ of random variables $g(x_1,\omega)$ and $g(x_2,\omega)$, it fulfills $F_1(t) \geq F_2(t)$ for all $t \in \mathrm{R}$. First order relation $\prec_{(1)}$ is easily verifiable in case of a finite discrete distribution of ω. The second order relation $g(x_1,\omega) \prec_{(2)} g(x_2,\omega)$ holds if $\mathrm{E}u\left(g(x_1,\omega)\right) \leq \mathrm{E}u\left(g(x_2,\omega)\right)$ for all concave non-decreasing functions $u(\cdot)$ such that mathematical expectations exist. This definition can be extended to vector random variables in many ways (Gutjahr and Pichler, 2013), but verification of the stochastic dominance relation may be rather difficult as it involves checking an infinite number of inequalities.

We can consider the preference relation $\prec_{\mathrm{E}}$ on the basis of expectations, $g(x_1,\omega) \prec_{\mathrm{E}} g(x_2,\omega)$ iff $\mathrm{E}g(x_1,\omega) < \mathrm{E}g(x_2,\omega)$ (componentwise). As the optimal set one can take $\prec_{\mathrm{E}}$-optimal subset of $\{\mathrm{E}g(x,\omega),\ x \in X\}$. It is known that the set of $\prec_{\mathrm{E}}$-optimal points in the convex case can be obtained by optimizing the set of linear $\langle w, \mathrm{E}g(x,\omega)\rangle$, $w \in W$, or nonlinear $U(\mathrm{E}g(x,\omega))$, $U \in \mathrm{U}$ scalar convolutions of vector criterion $\mathrm{E}g(x,\omega)$ (Jahn, 2011). In (Fliege and Xu, 2011) expectations $\mathrm{E}g(x,\omega)$ were replaced by their empirical estimates $\mathrm{E}_N g(x,\omega)$ and the convergence of the set of optimal points $U\left(\mathrm{E}_N g(x,\omega)\right) \to \max_{x\in X}$, $U \in \mathrm{U}$, to $\prec_{\mathrm{E}}$-optimal set in case of $N \to \infty$ was studied.

In a similar way other preference relations $\prec_P$ and $\prec_Q$ can be introduced, based on comparison of sets of probabilities $\Pr\{g_i(x,\omega) \geq t_i\}$, $i = 1,...,m$, and quantiles $Q_i(x,q_i) = \inf\{t:\ \Pr\{g_i(x,\omega) \geq t\} \leq q_i\}$, $i = 1,...,m$, respectively, where $\vec{t} = (t_1,...,t_m)$ and $\vec{q} = (q_1,...,q_m)$ are given vectors.

Along with the average values $\mathrm{E}g(x,\omega) = (\mathrm{E}g_1(x,\omega),...,\mathrm{E}g_m(x,\omega))$ it is worthwhile to consider the standard deviations $\sigma_i(x) = \left(\mathrm{E}(g_i(x,\omega) - \mathrm{E}g_i(x,\omega))^2\right)^{1/2}$ or standard semi-deviations $\sigma_i^-(x) = \left(\mathrm{E}\max^2\{0, \mathrm{E}g_i(x,\omega) - g_i(x,\omega)\}\right)^{1/2}$ of random indicators $g_i(x,\omega)$ from their average values $\mathrm{E}g_i(x,\omega)$, $i = 1,...,m$, as well as vector indicators $(\mathrm{E}g_1(x,\omega) + \alpha_1\sigma_1(x),...,\mathrm{E}g_m(x,\omega) + \alpha_m\sigma_m(x))$, $\alpha = (\alpha_1,...,\alpha_m) \in \mathrm{R}^m_+$. In (Caballero et al., 2001) relationships between Pareto optimal sets of various deterministic formalizations of the multiobjective stochastic optimization problem were studied, including the problems:

$\mathrm{E}g(x,\omega) \to \max_{x\in X}$,
$\sigma(x) = (\sigma_1(x),...,\sigma_m(x)) \to \min_{x\in X}$,
$\{\mathrm{E}g_1(x,\omega) - \alpha_1\sigma_1(x),...,\mathrm{E}g_m(x,\omega) - \alpha_m\sigma_m(x)\} \to \max_{x\in X}$,
$\{\Pr\{g_i(x,\omega) \geq t_i\},\ i = 1,...,m\} \to \max_{x\in X}$,
$\{Q_i(x,q_i),\ i = 1,...,m\} \to \max_{x\in X}$.

In this context the present paper deals with the vector optimization problem

$\vec{F}(x) = \{f_1(x), ..., f_\mu(x)\} \to \max_{x \in X}$,
where component functions $f_i(x)$ can be any ones from the collection

$\left\{\mathrm{E}g_i(x,\omega),\ -\sigma_i(x), -\sigma_i^-(x),\ \Pr\left\{g_i(x,\omega) \geq t_i\right\},\ Q_i(x,q_i),\ i = 1, ..., m\right\}$.

Here performance indicators $\mathrm{E}g_i(x,\omega)$ serve as utility measures, but $\sigma_i(x)$, $\sigma_i^-(x)$ and $\Pr\{g_i(x,\omega) \geq t_i\}$ represent risk measures for the decision vector x.

4. $\vec{\epsilon}$-dominance and $\vec{\epsilon}$-efficiency. The following concept appears to be useful for control of accuracy, strength and direction of dominance.

Definition 1. ($\vec{\epsilon}$-dominance and $\vec{\epsilon}$-efficiency/optimality). Vector $\vec{z}_1 \in \mathrm{R}^m$ $\vec{\epsilon}$-dominates the vector $\vec{z}_2 \in \mathrm{R}^m$, if $\vec{z}_1 > \vec{z}_2 + \vec{\epsilon}$ (componentwise), where $\vec{\epsilon} \in \mathrm{R}^m$. Subset $Z^*(\vec{\epsilon})$ of the set $Z \subset \mathrm{R}^m$ is called $\vec{\epsilon}$-efficient/optimal/nondominated if for any $\vec{z} \in Z^*(\vec{\epsilon})$ there is no $\vec{z}' \in Z$ such that $\vec{z}' > \vec{z} + \vec{\epsilon}$.

The concept of $\vec{\epsilon}$-efficiency was introduced in (Kutateladze, 1979). In case of $\vec{\epsilon} > 0$, it generalizes the standard notion of ϵ-optimality of scalar optimization. It also includes the notion of weak Pareto optimality, that corresponds to $\vec{\epsilon} = 0$. Further various generalizations of the $\vec{\epsilon}$-efficiency are discussed in (Gutierre et al., 2006), (Gutierre et al., 2012), (Gutjahr, 2012). By adding $\vec{\epsilon}$ to a vector $\vec{z}$ importance of components of $\vec{z}$ can be controlled (e.g. importance of criteria in vector optimization), namely, increase of ϵ_i component decreases importance of z_i component. Moreover, in contrast to (Gutierre et al., 2006), (Kutateladze, 1979) we allow $\vec{\epsilon} \notin \mathrm{R}^m_+$. If $\vec{\epsilon}$ contains negative components, then $\vec{\epsilon}$-dominance of $\vec{z}_1$ over $\vec{z}_2$ admits that some components of $\vec{z}_1$ can be somewhat smaller than the corresponding components of $\vec{z}_2$. Remark that if point $\vec{z}^* \in Z$ is $\vec{\epsilon}_k$-efficient for some sequence $\{\mathrm{R}^m \ni \vec{\epsilon}_k \to 0,\ k = 1, 2, ...\}$, then $\vec{z}^*$ is called a generalized efficient point (see. (Mordukhovich, 2006; Def. 5.53)).

Let us remind some notation and definitions (Rockafellar and Wets, 1998; Sec. 4A), that concern convergence of a sequence of sets $\{Z_i \subset \mathrm{R}^n,\ i = 1, 2, ...\}$:

$\limsup_i Z_i = \{z : \exists\ z_{i_k} \in Z_{i_k},\ z = \lim_k z_{i_k}\}$,

$\liminf_i Z_i = \{z : \exists\ z_i \in Z_i,\ z = \lim_i z_i\}$,

$\lim_i Z_i = \liminf_i Z_i = \limsup_i Z_i$.

Lemma 1. *(Properties of the $\vec{\epsilon}$-optimal mapping). Let a sequence of sets $\{Z_i \in \mathrm{R}^m\}$ converges to a compact set $\{Z \subset \mathrm{R}^m\}$, $\lim_i Z_i = Z$. Denote $Z_i^*(\vec{\epsilon})$, $Z^*(\vec{\epsilon})$ subsets of $\vec{\epsilon}$-nondominated points in Z_i and Z, respectively. Let $\lim_i \vec{\epsilon}_i = \vec{\epsilon}$. Then for any $\vec{\epsilon}' \leq \vec{\epsilon}$, $\vec{\epsilon}' \neq \vec{\epsilon}$, it holds true*

$$Z^*(\vec{\epsilon}') \subseteq \liminf_i Z_i^*(\vec{\epsilon}_i) \subseteq \limsup_i Z_i^*(\vec{\epsilon}_i) \subseteq Z^*(\vec{\epsilon}), \tag{1}$$

where the last inclusion, in particular, indicates that the set $Z^*(\vec{\epsilon})$ is closed and the mapping $\vec{\epsilon} \to Z^*(\vec{\epsilon})$ is upper semicontinuous.

Proof. Let us prove the first inclusion $Z^*(\vec{\epsilon}') \subseteq \liminf_i Z_i^*(\vec{\epsilon}_i)$, i.e. that for each point $\vec{z}^* \in Z^*(\vec{\epsilon}')$ there exists a sequence $\{Z_i^*(\vec{\epsilon}_i) \ni \vec{z}_i^* \to \vec{z}^*\}$. Let us fix $\vec{z}^* \in$

$Z^*(\vec{\epsilon}')$ and assume that $\vec{z}^* \notin \liminf_i Z_i^*(\vec{\epsilon}_i)$. Since $\vec{z}^* \in Z^*(\vec{\epsilon}') \subseteq Z = \lim_i Z_i$, there exists a sequence $Z_i \ni \vec{z}_i \to \vec{z}^*$. Due to assumption $\vec{z}^* \notin \liminf_i Z_i^*(\vec{\epsilon}_i)$ there is an infinite subsequence $\{\vec{z}_{i_k} \notin Z_{i_k}^*(\vec{\epsilon}_{i_k})\}$. So there are $\vec{z}'_{i_k} \in Z_{i_k}(\vec{\epsilon}_{i_k})$ such that $\vec{z}'_{i_k} > \vec{z}_{i_k} + \vec{\epsilon}_{i_k}$. Since Z is a compact set then, without loss of generality, we can consider that $\vec{z}'_{i_k} \to \vec{z}' \in Z$ and hence $\vec{z}' = \lim_k \vec{z}'_{i_k} \geqslant \lim_k \vec{z}_{i_k} + \lim_k \vec{\epsilon}_{i_k} = \vec{z}^* + \vec{\epsilon} \geq \vec{z}^* + \vec{\epsilon}'$ and $\vec{z}' \neq \vec{z}^* + \vec{\epsilon}'$. This means that point $\vec{z}^*$ is $\vec{\epsilon}'$-dominated, a contradiction.

Let us prove the last inclusion, $\limsup_i Z_i^*(\vec{\epsilon}_i) \subseteq Z^*(\vec{\epsilon})$. Assume the contrary, that there is a subsequence $\{Z_{i_k}^*(\vec{\epsilon}_{i_k}) \ni \vec{z}_{i_k}^* \to \vec{z}' \notin Z^*(\vec{\epsilon})\}$. Then there is a point $\vec{z}'' \in Z$ and a sequence $\{\vec{z}_i \in Z_i\}$ such that $\vec{z}'' > \vec{z}' + \vec{\epsilon}$ and $\vec{z}_i \to \vec{z}''$. So we obtain $\lim_k \vec{z}_{i_k} = \vec{z}'' > \vec{z}' + \vec{\epsilon} = \lim_k \left(\vec{z}_{i_k}^* + \vec{\epsilon}_{i_k}\right)$, that contradicts $\vec{\epsilon}_{i_k}$-efficiency of points $\vec{z}_{i_k}^*$ for sufficiently large k.

By choosing $Z_i = Z$ from (1) we obtain upper semicontinuity of $Z^*(\cdot)$ and hence closedness of $Z^*(\vec{\epsilon})$. The proof is complete.

The next simple example illustrates Lemma 1. Define sets $Z = Z_i = [0,1] \subset \mathrm{R}^1$, $i = 1, 2, ...$ Let $\vec{\epsilon} = \vec{\epsilon}_i = \epsilon \in \mathrm{R}^1$ and $\vec{\epsilon}' = \epsilon'$, where $0 < \epsilon' < \epsilon < 1$. Then $\vec{\epsilon}' < \vec{\epsilon}$ and $Z^*(\vec{\epsilon}') = [1 - \epsilon', 1] \subset \lim_i Z_i^*(\vec{\epsilon}_i) = Z^*(\vec{\epsilon}) = [1 - \epsilon, 1]$.

5. Multicriterion random search (MRS) algorithm and its convergence. Consider a general multicriterion optimization problem:

$$\vec{F}(x) = \{f_1(x), ..., f_m(x)\} \to \max_{x \in X \subset \mathrm{R}^n}, \tag{2}$$

where functions $f_i(x),\ i = 1, ..., m$, are assumed to be (semi)continuous on a compact set $X \subset \mathrm{R}^n$, and the preference relation in the criteria space R^m is set out by the nonnegative cone $\mathrm{R}^m_{++} = \{x \in \mathrm{R}^m : x_i > 0,\ i = 1, ..., m\}$. The problem is to find the weak Pareto-optimal set X^* and the subset $X^*(\vec{\epsilon})$ of $\vec{\epsilon}$-efficient points of $\vec{F}(X)$, $\vec{\epsilon} \in \mathrm{R}^m$.

It is easy to see that the mapping $\vec{\epsilon} \to X^*(\vec{\epsilon})$ is upper semicontinuous for upper semicontinuous vector function $\vec{F}(\cdot)$, i.e. $\limsup_i X^*(\vec{\epsilon}_i) \subseteq X^*(\vec{\epsilon})$ for any sequence $\vec{\epsilon}_i \to \vec{\epsilon}$. Indeed, let us assume the contrary. Then for some sequence $\vec{\epsilon}_i \to \vec{\epsilon}$ there exist $X^*(\vec{\epsilon}_{i_k}) \ni x_{i_k}^* \to x' \notin X^*(\vec{\epsilon})$. Since x' is $\vec{\epsilon}$-dominated then there is a point $x'' \in X$ such that $\vec{F}(x'') > \vec{F}(x') + \vec{\epsilon} = \lim_k \vec{F}(x_{i_k}^*) + \vec{\epsilon}$, that contradicts to $\vec{\epsilon}$-efficiency of x_{i_k} for sufficiently large k.

Let us consider also approximations for problem (2):

$$\vec{F}^i(x) = \{f_1^i(x), ..., f_m^i(x)\} \to \max_{x \in X_i \subset \mathrm{R}^n}, \quad i = 1, 2, ..., \tag{3}$$

where the sequence of sets $\{X_i\}$ converges to the set X, $\lim_i X_i = X$, and the sequence of vector functions $\{\vec{F}^i(x),\ x \in X_i\}$ converges to the vector function $\vec{F}(x),\ x \in X$ (in the sense of Definition 2 or 3). Denote $X_i^*(\vec{\epsilon})$ the set of $\vec{\epsilon}$-nondominated points of problem (3).

Concerning $\vec{F}$ and $\{\vec{F}^i\}$ we shall assume certain continuity and convergence properties, outlined in the following definitions.

DEFINITION 2. (Continuous convergence of a sequence of vector functions). A sequence of vector functions $\{\vec{F}^i(x),\ x \in X_i\}$ is called continuously convergent to a vector function $\vec{F}(x),\ x \in X$, if a) $\lim_i X_i = X$, b) for any sequence $X_i \ni x_i \to x$ it holds $\limsup_i \vec{F}^i(x_i) = \vec{F}(x)$ (componentwise).

DEFINITION 3. (Graphical convergence from below of a sequence of vector functions). A sequence of vector-functions $\{\vec{F}^i(x),\ x \in X_i\}$ is called graphically convergent from below to a vector function $\vec{F}(x),\ x \in X$, if a) $\lim_i X_i = X$, b) for each sequence $X_i \ni x_i \to x$ it holds $\limsup_i \vec{F}^i(x_i) \leqslant \vec{F}(x)$ (componentwize), and c) for any point $x \in X$ there is a sequence $X_i \ni x_i \to x$ such that $\lim_i \vec{F}^i(x_i) = \vec{F}(x)$ (componentwize).

The concepts of continuous and graphical convergence of multivalued mappings and functions (in the latter case, epi- and hypo-convergence) were comprehensively studied in (Rockafellar and Wets, 1998; Sec. 6E, 6G, 7B). Definitions 2 and 3 differ from the corresponding notions in (Rockafellar and Wets, 1998) by that in Definitions 2, 3 domains X_i, X of functions $\vec{F}^i$, $\vec{F}$ are explicitly exposed, but in (Rockafellar and Wets, 1998; Def. 5.41) extended functions are considered on a common domain X or R^n. And besides, the Definition 3 extends the definition of the graphical convergence of scalar functions (Rockafellar and Wets, 1998; 7(3), 7(9), Def. 7.1) to vector functions.

Examples of graphically convergent from below sequences of vector functions.

E1. Obviously, if a sequence $\{\vec{F}^i(x),\ x \in X_i\}$ converges continuously to $\{\vec{F}(x),\ x \in X\}$, i.e. $\lim_i X_i = X$ and $\lim_i \vec{F}^i(x_i) = \vec{F}(x)$ for any sequence $X_i \ni x_i \to x$, then $\{\vec{F}^i(\cdot)\}$ converges to $\vec{F}(\cdot)$ graphically.

E2. Obviously, if all, except the first, scalar components of $\{\vec{F}^i(\cdot)\}$ converge continuously to the corresponding scalar components of $\{\vec{F}(\cdot)\}$, and the first component of $\{\vec{F}^i(\cdot)\}$ hypo-converges (Rockafellar and Wets, 1998; formula 7(9)) to the first component of $\vec{F}(\cdot)$, then $\{\vec{F}^i(\cdot)\}$ graphically converges from below to $\vec{F}(\cdot)$.

E3. If $\vec{F}^i(x) = \vec{F}(x, y_i),\ x \in X,\ Y \ni y_i \to y$, where $\vec{F}(x, y)$ is componentwise upper semicontinuous on $X \times Y$ and is continuous at $y \in Y$ for any $x \in X$, then $\{\vec{F}^i(\cdot, y_i)\}$ graphically converges from below to $\vec{F}(\cdot, y)$. In particular, this case includes a stationary sequence of upper semicontinuous vector functions $\vec{F}^i(x) = \vec{F}(x),\ x \in X_i = X$.

E4. Example of construction of a continuously convergent sequence of functions. Let function $\vec{F}(x),\ x \in X$ be continuous on a closed set X, and a sequence of functions $\{\vec{F}^i(x),\ x \in X_i \subseteq X\}$ is such that $\Delta_i := \sup_{x \in X_i} \|\vec{F}^i(x) - \vec{F}(x)\| \to 0$ with $i \to \infty$. Then, for any sequence $(X_i \ni) x_i \to x$, we obtain

$$\|\vec{F}^i(x_i) - \vec{F}(x)\| \leqslant \|\vec{F}^i(x_i) - \vec{F}(x_i)\| + \|\vec{F}(x_i) - \vec{F}(x)\| \leqslant \Delta_i + \|\vec{F}(x_i) - \vec{F}(x)\| \to 0.$$

E5. If the objective vector function of problem (2) has a form of expectation, $\vec{F}(x) = \mathrm{E}\vec{F}(x,\omega)$, $x \in X$, then the sample average approximations $\vec{F}^i(x) = (1/M_i)\sum_{k=1}^{M_i} \vec{F}(x,\omega_k)$ can be used instead of $\vec{F}(x)$, where $\{\omega_k,\ k = 1,2,...\}$ are i.i.d. observations of the random parameter ω. Terms of uniform and, therefore, continuous convergence of empirical estimates $\vec{F}^i(x)$ to $\vec{F}(x)$ on the set X can be found in (Shapiro et al., 2014; Sec. 7.2.5). Next we present sufficient conditions of continuous convergence of discretely defined empirical functions $\vec{F}^i$ to a continuous expectation function $\vec{F}$. Assume that functions $\vec{F}(x,\omega)$ are uniformly bounded on X, $\|\vec{F}(x,\omega)\| \leqslant M$ and at each point $x \in X_i$ of a discrete set $X_i \subset X$ (with the number of elements N_i) an empirical estimate $\vec{F}^i(x) = (1/M_i)\sum_{k=1}^{M_i} \vec{F}(x,\omega_k)$ is independently constructed such that

$$\Pr\left\{\|\vec{F}^i(x) - \vec{F}(x)\| > \delta\right\} \leqslant C\exp\left\{-2M_i\delta^2/M^2\right\} \quad \forall\, \delta > 0.$$

Such estimates follow, e.g., from Hoeffding inequality (Shapiro et al., 2014; Sec. 7.2.8) with $C = 2m$. Then $\Delta_i = \max_{x\in X_i} \|\vec{F}^i(x) - \vec{F}(x)\| \to 0$ with probability one, if, for example, $M_i \geqslant \alpha N_i$, $\alpha > 0$, and the numerical sequence $\{N_i\}$ strictly monotonically increases to infinity. Indeed, the assertion follows from the fact that for any $\delta > 0$ it holds true

$$\sum_{i=1}^{\infty} \Pr\{\Delta_i > \delta\} \leqslant \sum_{i=1}^{\infty} CN_i \exp\left\{-2M_i\delta^2/M^2\right\} \leqslant \sum_{i=1}^{\infty} CN_i \exp\left\{-2N_i\alpha\delta^2/M^2\right\} < +\infty.$$

Theorem 1. *(Convergence of solutions of approximate problems (3)). Let a sequence of sets $\{X_i\}$ converges to a compact set X, $\lim_i X_i = X$, and a sequence of functions $\{\vec{F}^i(x),\ x \in X_i\}$ graphically converges from below to a vector function $\vec{F}(x),\ x \in X$. Let $\lim_i \vec{\epsilon}_i = \vec{\epsilon}$. Then for each $\vec{\epsilon}' \leq \vec{\epsilon}$, $\vec{\epsilon}' \neq \vec{\epsilon}$, we have*

$$X^*(\vec{\epsilon}') \subseteq \liminf_i X_i^*(\vec{\epsilon}_i) \subseteq \limsup_i X_i^*(\vec{\epsilon}_i) \subseteq X^*(\vec{\epsilon}).$$

Proof. Let us show, that $\limsup_i X_i^*(\vec{\epsilon}_i) \subseteq X^*(\vec{\epsilon})$. Assume the contrary, that $X_{i_k}^*(\vec{\epsilon}_{i_k}) \ni x_{i_k}^* \to x' \notin X^*(\vec{\epsilon})$. Since $x_{i_k}^* \in X_{i_k}^* \subseteq X_{i_k}$ and $\lim_i X_i = X$, then $x' \in X$. As $x' \notin X^*(\vec{\epsilon})$, then there exists $x'' \in X$ such that $\vec{F}(x'') > \vec{F}(x') + \vec{\epsilon}$ and, due to graphical convergence from below of $\{\vec{F}^i(\cdot)\}$ to $\vec{F}(\cdot)$, it holds $\vec{F}(x'') > \vec{F}(x') + \vec{\epsilon} \geqslant \lim_k \vec{F}^{i_k}(x_{i_k}^*) + \vec{\epsilon}$. Since $\lim_i X_i = X$ then there is a sequence $\{x_i \in X_i\}$ such that $x_i \to x''$, $\vec{F}^i(x_i) \to \vec{F}(x'')$, and thus $x_{i_k} \to x''$, $\vec{F}^{i_k}(x_{i_k}) \to \vec{F}(x'')$. Hence, $\lim_k \vec{F}^{i_k}(x_{i_k}) = \vec{F}(x'') > \vec{F}(x') + \vec{\epsilon} \geqslant \lim_k \left(\vec{F}^{i_k}(x_{i_k}^*) + \vec{\epsilon}_{i_k}\right)$ and for sufficiently large k points x_{i_k} $\vec{\epsilon}_{i_k}$-dominate points $x_{i_k}^*$. This contradiction proves the required assertion, $\limsup_i X_i^*(\vec{\epsilon}_i) \subseteq X^*(\vec{\epsilon})$.

Let us now prove that $X^*(\vec{\epsilon}') \subseteq \liminf_i X_i^*(\vec{\epsilon}_i)$, i.e. for each point $x^* \in X^*(\vec{\epsilon}')$ there exists a sequence $X_i^*(\vec{\epsilon}_i) \ni x_i^* \to x^*$. Let us fix $x^* \in X^*(\vec{\epsilon}')$ and suppose the contrary, that $x^* \notin \liminf_i X_i^*(\vec{\epsilon}_i)$. Since $x^* \in X^*(\vec{\epsilon}') \subseteq X$, then there is a sequence $X_i \ni x_i \to x^*$ such that $\lim_i \vec{F}^i(x_i) = \vec{F}(x^*)$. By the contrary assumption,

$x_{i_k} \notin X^*_{i_k}(\vec{\epsilon}_{i_k})$ for some infinite subsequence $\{x_{i_k}\}$. Hence there are $x'_{i_k} \in X_{i_k}$ such that $\vec{F}^{i_k}(x'_{i_k}) > \vec{F}^{i_k}(x_{i_k}) + \vec{\epsilon}_{i_k}$. Since X is compact, without loss of generality, we can consider that $x'_{i_k} \to x' \in X$ and conclude

$$\vec{F}(x') \geqslant \limsup_k \vec{F}^{i_k}(x'_{i_k}) \geqslant \lim_k \vec{F}^{i_k}(x_{i_k}) + \lim_k \vec{\epsilon}_{i_k} = \vec{F}(x^*) + \vec{\epsilon} \geq \vec{F}(x^*) + \vec{\epsilon}'$$

and $\vec{F}(x') \neq \vec{F}(x^*) + \vec{\epsilon}'$, that contradicts $\vec{\epsilon}'$-nondominance of x^*. The proof is complete.

The next **multicriterion random search (MRS) algorithm** uses random discrete approximations $X_i = \cup_{k=1}^{i} \tilde{X}_k$ of the feasible set X and estimates $\vec{F}^i(x),\ x \in X_i$, of the objective function of (2). The algorithm generates a random sequence of approximate solutions X^*_i, $i = 1, 2, ...$, of task (2) as follows.

At the first iteration a first generation of N_1 points $\tilde{X}_1$ is randomly generated in the set X, the estimates $\vec{F}^1(x)$ of the objective function $\vec{F}(x)$ are built for all points $x \in \tilde{X}_1$ and in the set $\{\vec{F}^1(x),\ x \in \tilde{X}_1\}$ a subset $\{\vec{F}^1(x),\ x \in X^*_1(\vec{\epsilon})\}$ of all $\vec{\epsilon}_1$-nondominated points is chosen.

Suppose that at iteration i we already have built the set X^*_i. Then (preferably in a vicinity of the set X^*_i, as, for example, in (Žilinskas, 2014)) a new generation of N_i random points $\tilde{X}_i$ is generated, estimates $\vec{F}^i(x)$ of the objective function $\vec{F}(x)$ for all $x \in X_i = \cup_{k=1}^{i} \tilde{X}_k$ are built, and from the set $\{\vec{F}^i(x),\ x \in X_i\}$ the subset $\{\vec{F}^i(x),\ x \in X^*_i\}$ of $\vec{\epsilon}_i$-nondominated points is chosen, and then we proceed to iteration $i + 1$. The process continues indefinitely long or ends at reaching the limit of iterations.

The asymptotic convergence of the algorithm is almost obvious, formal arguments follow. Convergence analysis of the described random search algorithm is based on the deterministic approximation scheme of Theorem 1. Stochastic part of the proof is collected in the next Lemma 2.

Lemma 2. *Let a sequence of random sets* $\{\tilde{X}_i,\ i = 1, 2, ...\}$ *be such that* $\tilde{X}_i \subseteq X$ *with probability one and with non-zero probability* $p_i(x, \delta) > 0$ *the set* $\tilde{X}_i$ *intersects with any* δ*-vicinity of any point* $x \in X$*, and it holds* $\sum_i p_i(x, \delta) = +\infty$*. Then with probability one* $\limsup_i \tilde{X}_i = X$ *and thus* $\lim_i \cup_{k=1}^{i} \tilde{X}_k = X$*.*

Proof. By assumption of the lemma, $\limsup_i \tilde{X}_i \subseteq X$. Let us show that with probability one the reverse inclusion is satisfied, $X \subseteq \limsup_i \tilde{X}_i$. Let us choose a countable everywhere dense subset X' in X and show that with probability one $\limsup_i \tilde{X}_i \supseteq X'$, from which the required assertion follows. Let us fix an arbitrary point $x' \in X'$. By the conditions of the lemma, with probability one the sequence $\tilde{X}_i$ hits any δ-vicinity of x' the infinite number of times, so with probability one, there exists a subsequence $\{\tilde{X}_{i_k} \ni x_{i_k} \to x'\}$, i.e., with probability one $x' \in \limsup_i \tilde{X}_i$. The countability of X' guarantees that with probability one $X' \subseteq \limsup_i \tilde{X}_i$, and by virtue of the density of X' in X and closeness of the

set $\limsup_i \tilde{X}_i$ it follows that $X \subseteq \limsup_i \tilde{X}_i$ with probability one. The proof is complete.

Lemma 3. *Let $\tilde{X}_i \subseteq X$, $\limsup_i \tilde{X}_i = X$, $\lim_i \vec{\epsilon}_i = \vec{\epsilon}$, $\vec{F}(x)$ is continuous on X and $\Delta_i = \sup_{x \in X_i} \|\vec{F}^i(x) - \vec{F}(x)\| \to 0$. Then, for each $\vec{\epsilon}' \leq \vec{\epsilon}$, $\vec{\epsilon}' \neq \vec{\epsilon}$, we have*

$$X^*(\vec{\epsilon}') \subseteq \liminf_i X_i(\vec{\epsilon}) \subseteq \limsup_i X_i(\vec{\epsilon}) \subseteq X^*(\vec{\epsilon}).$$

Proof. Note, that for $X_i = \cup_{k=1}^i \tilde{X}_k$ we get $\lim_i X_i = X$, so the assertion of the lemma follows from Theorem 1.

As a consequence of Lemmas 2, 3 we obtain the following result on convergence of the multicriterion random search algorithm to the $\vec{\epsilon}$-nondominated set of problem (2).

Theorem 2. *(Convergence of the MRS algorithm). Let vector function $\vec{F}(x)$ be continuous on a compact set $X \subset \mathrm{R}^m$, random sets $\tilde{X}_i$ with positive probability $p_i(x, \delta) > 0$ intersect with any δ-vicinity of each point $x \in X$, and $\sum_i p_i(x, \delta) = +\infty$. Denote $X_i = \cup_{k=1}^i \tilde{X}_k$. Let $\{\vec{F}^i(x),\, x \in X_i \subset X\}$ be a sequence of random vector functions such that with probability one $\Delta_i = \sup_{x \in X_i} \|\vec{F}^i(x) - \vec{F}(x)\| \to 0$. Then, with probability one a) all cluster points of $\{X_i^*(\vec{\epsilon}_i)\}$ belong to $X^*(\vec{\epsilon})$ and b) for each point $x^* \in X^*(\vec{\epsilon}')$, where $\vec{\epsilon}' \leq \epsilon$, $\vec{\epsilon}' \neq \epsilon$, there is a sequence of points $\{x_i \in X_i^*(\vec{\epsilon}_i)\}$ convergent to x^*.*

In particular, if $\vec{\epsilon} > 0$, then for each weakly Pareto optimal point $x^* \in X^* \subseteq X$ there is a sequence of points $\{x_i \in X_i(\epsilon)\}$, convergent to x^*. And by virtue of upper semicontinuity of mapping $X^*(\vec{\epsilon})$ at $\vec{\epsilon} = \vec{0}$, for a sufficiently small vector $\vec{\epsilon}$ the set $X^*(\vec{\epsilon})$ will be in an arbitrarily small neighborhood of the weakly Pareto optimal set $X^* = X^*(\vec{0})$.

6. Multiobjective stochastic optimal control of insurance business. We apply methodology of multiobjective stochastic optimization to solution of multiobjective actuarial stochastic optimal control problems. Consider a controlled vector stochastic process of the following form (Gihman and Skorohod, 1977):

$$\begin{aligned} y^{t+1} \;=\; & f_t(y^0, y^1, ..., y^t; \xi^0, \xi^1 ..., \xi^t; u_t(y^0, ..., y^t; \xi^0, ..., \xi^t)), \\ & y^0 = x_0, \;\; t = 0, 1, ..., T, \end{aligned} \tag{4}$$

where $t = 0, 1, ..., T$ denotes the discrete time; $\{y^0, y^1, ...\}$ is a sequence of states of the process; $\{\xi^0, \xi^1, ...\}$ is an uncontrolled sequence of random variables that affect the state of the process; $\{u_t(\cdot) \in U_t\}$ is a sequence of random controls selected from sets of admissible controls U_t; $y^0 = x_0$ is an initial state of the process; $\{f_t(\cdot)\}$ is the process model.

As an example we can consider the simplified case, in which the evolution of a capital y^t of an insurance firm in a discrete time $t = 0, 1, ...$ can be described by

equation (Schmidli, 2008):

$$y^{t+1} = \begin{cases} y^t - u_t + Z^t, & y^t \geq 0 \\ y^t, & y^t < 0, \end{cases}, \tag{5}$$

where y^0 is a seed capital, u_t is dividend amount, Z^t is a random insurance premium or claim at time t.

Suppose that at each step t of process (4) the decision maker (DM), evaluates the process by means of functions $r_{ti}(y^0, ..., y^t; u_0(\cdot), ..., u_k(\cdot); \xi^0, ..., \xi^k)$, $i = 1, ..., m$, and the total discounted estimates of the process for the $T+1$ periods of time are given by

$$I_i\left(u_0(\cdot), u_1(\cdot), ..., u_T(\cdot)\right) = \mathrm{E}\sum_{k=0}^{T} \gamma^k r_{ki}(y^0, ..., y^k; u_0(\cdot), ..., u_k(\cdot); \xi^0, ..., \xi^k), \tag{6}$$

where $\gamma \in (0, 1]$ is discount factor, $i = 1, ..., m$. By choosing different functions $r_{ki}(y^0, ..., y^k; u_0(\cdot), ..., u_k(\cdot); \xi^0, ..., \xi^k)$ evaluation of different aspects of the model is possible. If, for example,

$$r_{ki^*}(y^0, ..., y^k; u_0(\cdot), ..., u_k(\cdot); \xi^0, ..., \xi^k) = \begin{cases} 1, & \exists y^k \in A, \ k \leq t, \\ 0, & y^k \notin A \ \forall k \leq t, \end{cases} \quad \gamma = 1,$$

then the corresponding indicator $I_{i^*}(\cdot)$ represents the probability of the process hitting A.

Thus, the problem of stochastic multicriterion (Pareto-optimal) control has the form:

$$\left[I_i\left(u_0(\cdot), u_1(\cdot), ..., u_T(\cdot)\right), \ i = 1, ..., m\right] \to \mathrm{extr}_{\{u_t(\cdot) \in U_t, \ t=0,...,T\}}. \tag{7}$$

The essential difficulty in solving the problem (7) is calculation or estimation of mathematical expectation (6) over all possible trajectories of the process (4). In general, this can only be done by Monte Carlo method. Another difficulty is that problem (7) is infinite-dimensional.

Problem (7) is considerably simplified if to restrict the set, where we search for optimal controls, to a class of parametrically defined functions $U_t = \left\{u_t(y^0, ..., y^t; \xi^0, ..., \xi^t; x^t), \ x^t \in X^t\right\}$, where x^t is a finite-dimensional parameter. Then the functional I_i becomes a function of the finite-dimensional parameters, and the optimization problem (7) is transformed into a finite-dimensional vector stochastic programming problem.

Another way of simplification is to consider (vector) Markov processes:

$$y^{t+1} = f(y^t; \xi^t; u(y^t; \xi^t)), \ \ y^0 = x_0, \ \ t = 0, 1, ..., T, \tag{8}$$

where the model f and control u are not changed with time, and the next state y^{t+1} depends only on the current state y^t of the system and on the current state

of the environment ξ^t. In many cases, the optimal control $u_t(\cdot)$ can be found in the class of functions that depend only on $(y^t;\xi^t)$ or y^t (Gihman and Skorohod, 1977). In particular, optimal control of dividends in dynamic models of insurance companies (5) often takes on the form of the so-called barrier strategy (Albrecher and Thonhauser, 2009), (Avanzi, 2009), (Schmidli, 2008), where dividends are paid only if the company's capital y^t exceeds a certain threshold b (barrier). If the functional form of the control is chosen and depends only on n-dimensional parameter $x \in \mathrm{R}^n$, i.e. $u(y,\xi,x)$, then (7) again becomes the problem of a finite dimensional multicriterion stochastic programming,

$$\left[I_i^T(x_0,x) = \mathrm{E}\sum_{k=0}^{T} \gamma^k r_{ki}(y^k,\xi^k,u(y^k,\xi^k,x)),\ i=1,...,m \right] \to \mathrm{extr}_{x\in X}. \qquad (9)$$

A separate problem is the computation of indicator values $I_i(x_0,x)$ for some fixed parameter vector (x_0,x), because they are mathematical expectations over random trajectories of process (8). Even if the random variables ξ^k are discrete with known distributions, finding the expectation in (9) requires summation over all possible paths of the process (8), which appears to be problematic. A universal method for estimating integrals (6) is the Monte Carlo method, which, however, may require a very large number of trials to achieve acceptable accuracy. Note that simulating trajectories of the process (4) can be performed in *parallel* that significantly reduces the computation time of the Monte Carlo method. An alternative approach of finding indicators $I_i^T(x_0,x)$ is solving Bellman type integral equations (Norkin, 2014b). It appears that under certain conditions on Markov process (8), indicators $I_i^t(x_0,x)$, $t=0,1,...,T$, satisfy relations (Norkin, 2014b):

$$I_i^t(x_0,x) = \mathrm{E}r_{(T-t)i}(x_0,\xi,u(x_0,\xi,x)) + \gamma \mathrm{E} I_i^{t-1}\left(f(x_0;\xi;u(x_0,\xi,x)),x\right),$$

$$I_i^0(x_0,x) = \mathrm{E}r_{Ti}(x_0,\xi,u(x_0,\xi,x)), \quad t=1,...,T. \qquad (10)$$

If the random variable $\xi \in \{\xi_1,...,\xi_S\}$ is discrete with known probabilities p_s of its realizations ξ_s, then (10) are turn into deterministic relations,

$$I_i^t(x_0,x) = \sum_{s=1}^{S} r_{(T-t)i}(x_0,\xi_s,u(x_0,\xi_s,x))p_s + \gamma \sum_{s=0}^{S} I_i^{t-1}\left(f(x_0;\xi_s;u(x_0,\xi_s,x)),x\right)p_s,$$

$$I_i^0(x_0,x) = \sum_{s=1}^{S} r_{Ti}(x_0,\xi_s,u(x_0,\xi_s,x))p_s, \quad t=0,...,T-1,$$

which can be used to calculate $I_i^T(x_0,x)$. To do this, we have to define a grid over a one-dimensional variable x_0 and successively compute $I_i^0(\cdot,x)$, $I_i^1(\cdot,x)$, ..., $I_i^T(\cdot,x)$ at the grid nodes, and do interpolation in between. Note that this iterative process possesses natural parallelism property. The values of $I_i^t(\cdot,x)$ at

different grid points are calculated independently of each other on the basis of the discrete approximation of the function $I_i^{t-1}(\cdot, x)$, obtained in the previous iteration. Therefore, these calculations are similar to (Norkin, 2011; equation (7)) and can be parallelized.

7. Implementation of the multiobjective stochastic optimization. On the basis of the presented methodology, a new version of multicriterion insurance decision support system was developed. The previous version was described in (Norkin, 2012), (Norkin, 2014a), where the main optimization tool was studying dependence of some scalar performance indicator on a scalar parameter. The new version of the system differs significantly from the previous one: it provides multicriterion optimization over vector parameters, in particular computation and visualization of Pareto-optimal solution set.

Evolution of the reserves y^t of an insurance company in discrete time (a year or a quarter) $t = 0, 1, ..., T$ is modeled by the following risk process (Norkin, 2012), (Norkin, 2014a):

$$y^{t+1} = \begin{cases} y^t + x_1 - \xi^t x_1 + \rho^t x_2 x_1 + \zeta^t x_3 x_1 + \theta^t x_4 x_1 \\ \qquad -x_4 x_1 - x_5 x_1 - x_7 \max\{0, y^t - x_6\}, & y^t \geq x_8, \\ y^t, & y^t < x_8, \end{cases}$$

where $\{\xi^t \geqslant 0\}$ are random realizations of normalized (per unit of premiums) insurance clames; $\{\rho^t \geqslant 0\}$ are deposite rates; $\{\zeta^t \geqslant -1\}$ are investment interest rates (random); $\{\theta^t \geqslant 0\}$ are reinsurance profitability rates; $x = (x_0, x_1, ...)$ is a vector of deterministic parameters.

Components of the parameter vector x are following:

$x_0 = y^0$ is seed capital;

x_1 is aggregate premium (per unit of time);

x_2 is level of investment in bank deposits, i.e. share of premiums that is placed in deposit (per unit of time),$0 \leq x_2 \leq 1$;

x_3 is investment level, i.e. share of premium x_1 that is invested, $0 \leq x_3 \leq 1$;

x_4 is reinsurance level, i.e. share of premium x_1, that is used for reinsurance, $0 \leq x_4 \leq 1$;

x_5 is level of mandatory payments as proportion of premium x_1, including taxes and other overheads, $0 \leq x_5 \leq 1$;

x_6 is dividend barrier (dropping the capital below it suspends dividend payments), $0 \leq x_4 \leq x_0$;

x_7 is part of capital $y^t - x_6$, that is payed as dividends, $0 \leq x_7 \leq 1$;

x_8 is insolvency threshold;

$x_9 = T$ is time horizon;

$x_{10} = \gamma$ is discounting coefficient, $0 < x_9 \leqslant 1$.

Here the last expression $D(y^t, x) = x_7 \max\{0, y^t - x_6\}$ defines a proportional-barrier strategy of dividend payments. Thus, the financial portfolio (x_2, x_3, x_4, x_5)

in this model, expresses the structure of distribution of premium x_1, is combined of four parts: investment in bank deposits x_2, risky investment x_3, reinsurance costs x_4 and mandatory payments x_5, note that $x_2 + x_3 + x_4 + x_5 = 1$. In this aggregate model proportional reinsurance model is used, investing $x_4 x_1$ in the reinsurance contract give compensatory payments $\theta^t x_4 x_1$ in time t, where θ^t is random return from reinsurance investments that correlates with claims ξ^t. We assume, that the joint distribution of random variables $\{\xi^t, \rho^t, \zeta^t, \theta^t\}$ is given by series of equi-probable empirical observations. Other models of dividend and reinsurance payments surely could be used.

The following performance indicators are considered:

$I_1(x) = \mathrm{E} \sum_{k=0}^{x_9} (x_{10})^k x_7 \max\{0, y^k - x_6\}$ defines expected discounted dividends;

$I_2(x) = \mathrm{E}(x_9)^{x_9} \max\{0, y^{x_9}\}$ defines expected capital at the end of the planning horizon, in case of non-bankruptcy;

$I_3(x) = \mathrm{E} \sum_{\{k:\, y^k \ge x_8,\, y^{k+1} < x_8,\, 0 \le k < x_9\}} 1$ defines (discounted) probability of insolvency;

$I_4(x) = \mathrm{E} \max_{0 \le t \le x_9} \left\{ t : \min_{0 \le k < t} y^k \ge 0 \right\}$ defines expected ruin time.

Auxiliary functionals, such as quantiles, conditional means and variances of various random characteristics of the process (collected dividends, residual capital, lifetime) are used.

Software system provides the following *features*:

Saving and loading of the project (data and model parameters);

loading and analyzing statistical data;

defining reference intervals of the variable parameters of the model;

configuring statistical modeling parameters (setting number of points in the parameter space and number of simulations per point);

graphical representation of dependence of any indicator I_i on any parameter x_j;

accuracy estimation;

displaying the simulation results (optimization cloud) in planes "indicator i - indicator j" for any i, j;

discrete approximation of the Pareto-optimal points of the problem for any subset of indicators;

visualizing characteristics of the Pareto-optimal points;

viewing and comparing the preferred Pareto-optimal points;

saving computation results.

Real-world statistical data from www.forinsurer.com was used for demonstrative computations.

Backend of the system is capable of using Intel IvyBridge *hardware* random number generator (RdRand).

The system supports the following *multicriteria optimization algorithm*:

Load data (Excel file);

Specify parameters ranges (box constraints);

Sample parameter points (random search in the box);

Evaluate performance indicators (parallel Monte Carlo simulation of trajectories);

Plot & explore clouds of indicator points;

Specify new constraints, bounds on parameters and indicators;

Specify dominance (set of indicators and direction of optimization) and tolerances $\vec{\epsilon}$;

Excluded $\vec{\epsilon}$-dominated points;

Iterate...

8. Numerical illustration of the multicriterion optimization of insurance business. Let us demonstrate some results of calculations using the following test data:

$$x = (x_0, x_1, ..., x_{10}) = (0,\ 1,\ [0.4 \div 0.6],\ [0 \div 1],\ [0 \div 5],\ 0,\ 0,\ 0,\ 50,\ [0.8 \div 1])\,;$$

$$\xi = (\xi_1, ..., \xi_5) = (0.419,\ 0.514,\ 0.650,\ 0.658,\ 0.502)\,, P\{\xi = \xi_i\} = 1/5.$$

Fig. 1 shows clouds of $N = 1000$ indicator points in planes (I_1, I_3) (left), (I_2, I_3) (right) for randomly selected parameter values from $x_2 \in [0.4 \div 0.6]$, $x_3 \in [0 \div 1]$, $x_4 \in [0 \div 5]$ and under fixed values of other parameters. The right figure displays the Pareto optimal boundary in the plain "residual capital – ruin probability".

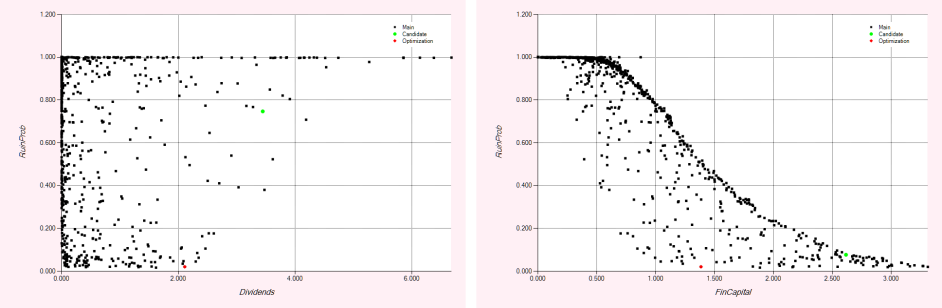

Fig. 1. Optimization clouds in plain (I_1, I_3) and (I_2, I_3).

Fig. 2 shows us an opitmization cloud of $N = 10000$ indicator points in plain (I_1, I_2) before (left fig.) and after (right fig.) (I_1, I_2, I_3)-Pareto optimization.

Fig. 3 show the same results of (I_1, I_2, I_3)-Pareto optimization, but in plains (I_1, I_3) (left) and (I_2, I_3) (right).

Further choice of a compromise in terms of (I_1, I_2, I_3) is carried out by visual comparison and analysis of the graphs such as on Fig. 2, 3. System provides a visual aid: pointing to some point highlights it on all graphs and displays values of all corresponding indicators and parameters.

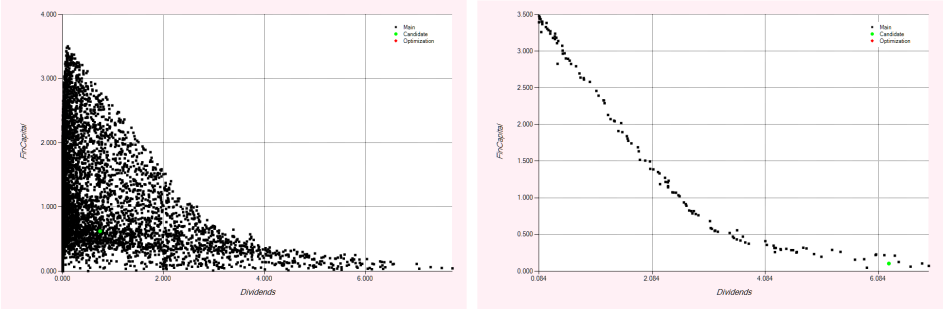

Fig. 2. Optimization cloud in plain (I_1, I_2) before and after Pareto optimization.

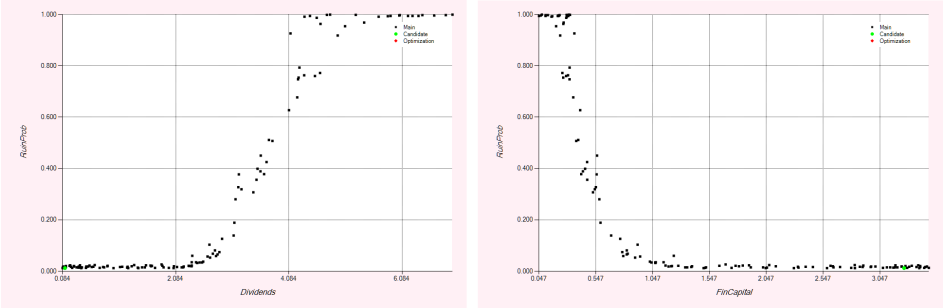

Fig. 3. (I_1, I_2, I_3)-Pareto-optimal set in plains (I_1, I_3) and (I_2, I_3).

9. Conclusions. The article describes an approximation technology for multicriterion stochastic optimization of "input – random output" systems. In practice, these models are highly nonlinear and non-convex. Their functioning can generally be evaluated by deterministic vector indicators such as means, quantiles, probabilities of reaching/exiting specified areas, etc. Collapsing the vector indicator into a scalar one with the view of its optimization is not always possible. So, the task is to find such inputs, that correspond to the Pareto-optimal performance vector indicators. This paper proposes a methodology for solving such problems by visualized interactive parallel random search with selection of Pareto-optimal points. The technology is illustrated by an insurance multicriterion optimization support system.

REFERENCES

H. Albrecher, S. Thonhauser (2009). Optimality results for dividend problems in insurance, *Rev. R. Acad. Cien. Serie A. Mat.* **103**(2), 295-320.

B. Avanzi (2009). Strategies for dividend distribution: A review. *North American Actuarial Journal*. **13**(2), 217-251.

F. Ben Abdelaziz (2012). Solution approaches for the multiobjective stochastic programming, *European J. of Operational Research*, **216**, 116.

J. Branke, K. Deb, K. Miettinen, R. Slowiński (Eds.) (2008). *Multiobjective Optimization. Interactive and Evolutionary Approaches.* Springer-Verlag, Berlin, Heidelberg. 470 p.

R. Caballero, E. Cerda, M.M. Munoz, L. Rey (2001). Efficient solution concepts and their relations in stochastic multiobjective programming. *J. Optim. Theory Appl.*, **110**, 53-74.

C.A. Coello Coello (2006). Evolutionary multi-objective optimization: a historical view of the field *Computational Intelligence Magazine, IEEE*, **1**(1), 28-36.

J. Fliege, H. Xu (2011). Stochastic multiobjective optimization: sample average approximation and applications. *J. of Optimization Theory and Applications*, **151(1)**, 135-162.

I.I. Gihman, A.V. Skorohod: Controlled Stochastic Processes. Berlin, Heidelberg, New York, Springer-Verlag, 1979. – 237 p.

C. Gutiérrez, B. Jiménez, and V. Novo (2006). A unified approach and optimality conditions for approximate solutions of vector optimization problems. *SIAM J. Optim.*, **17(3)**, 688710.

C. Gutiérrez, B. Jiménez, and V. Novo (2012). Equivalent ϵ-efficiency notions in vector optimization. *TOP*, **20**, 437455.

W.J. Gutjahr (2012). Runtime analysis of an evolutionary algorithm for stochastic multi-objective combinatorial optimization. *Evolutionary computation.* **20(3)**, 395-421.

W. Gutjahr, A. Pichler (2013). Stochastic multi-objective optimization: a survey on non-scalarizing methods. *Annals of Operations Research.*, 1-25.

J. Jahn (2011). *Vector Optimization. Theory, Applications, and Extensions. Second Edition.* Springer-Verlag, Berlin, Heidelberg. 481 p.

H.K. Jevne, P.C. Haddow, A.A. Gaivoronski (2012). *Evolving constrained mean-VaR efficient frontiers. In Evolutionary Computation (CEC)*, 2012 IEEE Congress on Computational Intelligence. IEEE, Brisbane, Australia, 1-8.

M.P. Kleeman, G.B. Lamont (2008). *Evolutionary Multi-Objective Optimization in Military Applications. In: Multi-Objective Optimization in Computational Intelligence: Theory and Practice / L.T. Bui and S. Alam, eds. Hershey*, Information Science Reference (an imprint of IGI Global), New York, 388-429.

A. Konak, D.W. Coit, A.E. Smith (2006). Multi-objective optimization using genetic algorithms: A tutorial *Reliability Engineering & System Safety*, **91**(9), 992-1007.

S.S. Kutateladze (1979). Convex e-programming. *Soviet Math. Dokl.* **20**(2), 391-393.

B. Liu (2005) *Theory and Practice of Uncertain Programming.* Series: Studies in Fuzziness and Soft Computing, **239** 2nd ed. Springer 431 p.

R.T. Marler, J.S. Arora (2004). Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization*, **26**(6), 369-395.

B.S.Mordukhovich (2006). *Variational Analysis and Generalized Differentiation II. Applications.* Springer, Berlin. 610 p.

A. Müller, D. Stoyan (2002). *Comparison Methods for Stochastic Models and Risks.* John Wiley & Sons, Chichester, UK, 330 p.

B. Norkin (2011). Parallel computations in insurance business optimization. *Proceedings of the 1-st International Conference on High Performance Computing (HPC-UA 2011, October 12-14, 2011, Kyiv, Ukraine).* National Academy of Sciences of Ukraine, Kyiv. 33-39.

B.V. Norkin (2012). On performing actuarial calculations on GPU. *Visnyk NTUU KPI. Informatika, upravlinnya ta obchislyuvalna tehnika.* Collection of scientific papers, Vek+, Kyiv, **56**, 113-119.

B.V. Norkin (2014a). Systems simulation analysis and optimization of insurance business. *Cybernetics and Systems Analysis.* **50**(2), 260-270.

B.V. Norkin (2014b). Stochastic optimal control of risk processes with Lipschitz payoff functions. *Cybernetics and Systems Analysis.* **50**(5), 774-787.

B. Norkin (2014c). Random Search of Multicriterion Optimum in Insurance. In *Theoretical and Applied Aspects of Cybernetics. Proceedings of the 4th International Scientific Conference of Students and Young Scientists.* Bukrek, Kyiv, 1-12.

S. Poles, M. Vassileva, D. Sasaki (2008). Multiobjective Optimization Software / J. Branke et al. (Eds.) *Multiobjective Optimization.* Springer-Verlag, Berlin Heidelberg, 329-348.

R.T. Rockafellar, R.J-B. Wets (1998). *Variational Analysis. Grundlehren der Mathematischen Wissenschaften*, Springer, (3rd Printing in 2009). 734 p.

H. Schmidli (2008). *Stochastic control in insurance*. Springer-Verlag, London. 254 p.

A. Shapiro, D. Dentcheva, A. Ruszczyński (2014). *Lectures on stochastic programming: Modeling and theory*. Second Edition. SIAM, Philadelphia. 494 p.

I.M. Stancu-Minasian (1984). *Stochastic programming with multi-objective functions*. Edeisteurrvae Academiei, Bucharest. 334 p.

E. Talbi, S. Mostaghim, T. Okabe, H. Ishibuchi, G. Rudolph, C.A. Coello Coello (2008). Parallel Approaches for Multiobjective Optimization / J. Branke et al. (Eds.): *Multiobjective Optimization*. Berlin Heidelberg: Springer-Verlag, 349-372.

A. Žilinskas (2014). A statistical model-based algorithm for black-box multi-objective optimization. *International Journal of Systems Science*, **45**(1), 8293.

Received November 2014

B. Norkin graduated from Kyiv National University (2002), received Candidate degree (2006) from Glushkov Institute of Cybernetics, Kyiv, Ukraine. Presently he is a research scholar at this institute and a senior lecturer at the Department of Informatics of Kyiv-Mohila Academy. His research interests include risk theory, actuarial and financial mathematics, systems analysis, stochastic programming, high performance computing.