Application of the Monte-Carlo Method to Nonlinear Stochastic Optimization with Linear Constraints

Leonidas SAKALAUSKAS
Institute of Mathematics and Informatics
Akademijos 4, 08663 Vilnius, Lithuania
e-mail: sakal@ktl.mii.lt

Received: April 2003

Abstract. We consider a problem of nonlinear stochastic optimization with linear constraints. The method of ε-feasible solution by series of Monte-Carlo estimators has been developed for solving this problem avoiding "jamming" or "zigzagging". Our approach is distinguished by two peculiarities: the optimality of solution is tested in a statistical manner and the Monte-Carlo sample size is adjusted so as to decrease the total amount of Monte-Carlo trials and, at the same time, to guarantee the estimation of the objective function with an admissible accuracy. Under some general conditions we prove by the martingale approach that the proposed method converges a.s. to the stationary point of the problem solved. As a counterexample the maximization of the probability of portfolio desired return is given, too.

Key words: Monte-Carlo method, portfolio optimization, stochastic programming, ε-feasible solution.

1. Introduction

Optimal decisions in business and finance are frequently provided by solving nonlinear stochastic programming problems with linear constraints:

$$F(x) \equiv Ef(x,\xi) \to \max_{x \in X}, \tag{1}$$

where the objective function is an expectation of a random function $f\colon \Re^n \times \Omega \to \Re$ depending on a random vector $\xi \in \Omega$ from a certain probability space (Ω, Σ, P), and the feasible set $x \in X \subset \Re^n$ is a bounded and convex linear set in general:

$$X = \{x | Ax = b,\ x \geqslant 0\}, \tag{2}$$

$b \in R^m$, A is the $n \times m$-matrix, $X \neq \varnothing$.

The methods of stochastic approximation were first proposed to solve stochastic optimization problems. The convergence in stochastic approximation is ensured by varying certain step-length multipliers in a scheme of stochastic gradient search (Mikhalevitch *et al.*, 1987; Kushner, 1997; Han-Fu-Chen, 2002; Ermoliev *et al.*, 2003; etc.). However, the rate of convergence of stochastic approximation slows down for constrained problems (Polyak, 1987; Uriasyev, 1990), besides, the gradient-type projection method, usually applied here, can no converge when constraints are linear due to "zigzagging" or "jamming" (Bertsekas, 1982; Polyak, 1987; etc.).

The Monte-Carlo method is a tool also applied very often in solving problems of stochastic optimization appearing here, particularly, in that of stochastic linear programming (Prekopa, 1999; Ermoliev *et al.*, 2003). Kjellstrom (1969) was the first who suggested using series of Monte-Carlo estimators for the iterative improvement of convergence behavior in nonlinear stochastic optimization. Further this approach has found applications to technical design of electronic devices (Beliakov *et al.*, 1985; Sakalauskas, 1997). Application of this method in stochastic optimization is based on replacement of the objective function, being mathematical expectation, by averaged means, provided the during Monte-Carlo simulation (see, e.g., Shapiro, 1989). The issues remain important in approaching such programs to stochastic optimization related with a great amount of computations usually required for the performance, and in the evaluation of uncertainty of the Monte-Carlo estimators obtained. On the other hand, the Monte-Carlo approach also has some properties that could be helpful for enhancement of stochastic programs, namely, via the Monte-Carlo simulation rather often we can estimate both functions with their derivatives without essential additional costs (see, i.e., Rubinstein, 1983; Shapiro, 1986; Sakalauskas, 2002), besides, sampled Monte-Carlo estimators usually have the Gaussian distribution in asymptotic (Bentkus and Gotze, 1999) that offers a way of applying the standard theory of normal statistics (Krishnajah and Lee, 1988) to a simple computation of confidence intervals of estimators and testing of optimality hypotheses, etc.

The properties mentioned have been used in the development of the approach to unconstrained stochastic optimization by Monte-Carlo estimators (Sakalauskas, 2000), where the optimality of portfolio is tested in a statistical manner and the rule for Monte-Carlo sample size adjustment has been introduced in order to decrease the total amount of Monte-Carlo trials and, at the same time, to guarantee the solution of an optimization task with an admissible accuracy. Further this approach was extended to constrained optimization with one probabilistic constraint using the method of the Lagrange function (Sakalauskas, 2002). However, in many applications the stochastic optimization with linear constraints is connected with a strict validity of constraints in each iteration, which pose the above mentioned problems of "jamming" or "zigzagging". In this paper, we develop a method for stochastic optimization with linear constraints by Monte-Carlo ε-feasible estimators, which avoids the later problem and focuses on a rational performance of computations as well as on the control of computational error.

The paper is organized as follows. In the next section we describe the stochastic optimization procedure and analyze its convergence. The termination rules based on the asymptotic properties of Monte-Carlo estimators are introduced in Section 3, and a counterexample of portfolio VAR optimization with log-normal returns is considered in Section 4.

2. Optimization Procedure and Convergence Analysis

For simplicity, assume the distribution of market uncertainty factors to be absolutely continuous and described by the density function p: $\Omega \rightarrow \Re_+$ that are supposed to be smoothly differentiable, $p(0) > 0$. Thus the objective function can be expressed as a multivariate integral:

$$F(x) = \int_{R^n} f(x, y) \cdot p(y)\,\mathrm{d}y. \tag{3}$$

The differentiability of integrals of this kind has been studied rather well, and there exists a technique for stochastic differentiation to express such an objective function and its gradient both together as expectations in the same probability space (Rubinstein, 1983; Prekopa, 1999; Ermolyev *et al.*, 2003; Uriasyev, 1994; etc.): $\nabla F(x) = \int_{R^n} g(x, y) \cdot p(y)\,\mathrm{d}y$, where g: $\Re^n \times \Re \times \Omega \rightarrow \Re^n$ is a certain function (for explicit formulas see in the given above references). Thus, differentiability of the objective function (3) can be assumed for a wide class of optimization problems and, consequently, both the objective function and its gradient can be estimated using the Monte-Carlo method. Thus a gradient-type nonlinear optimization method by Monte-Carlo estimators can be developed, using ε-feasible solutions as the standard way to guarantee the validity of linear constraints in each iteration and avoid "jamming" or "zigzagging".

Following the standard approach to determine the optimality condition, let us define a *set of feasible directions* for some solution $x \in X$ as:

$$V(x) = \{g \in \Re^n | Ag = 0,\ \forall_{1 \leqslant i \leqslant n} (g_j \geqslant 0,\ \text{if } x_j = 0)\}. \tag{4}$$

Further we denote the projection of the vector g to a certain set Q by g_Q.

Thus the necessary condition of optimality (Bertsekas, 1982) for the solution $x \in X$ is written now as

$$\nabla F(x)_V = 0. \tag{5}$$

Assume a certain multiplier $\widehat{\rho} > 0$ to be given. Let us define the function ρ_x: $V(x) \rightarrow \Re_+$

$$\rho_x(g) = \begin{cases} \min\Big\{\widehat{\rho},\ \min\limits_{\substack{g_j < 0,\\ 1 \leqslant j \leqslant n}} \big(-\frac{x_j}{g_j}\big)\Big\}, & g \neq 0, \\ \widehat{\rho}. \end{cases} \tag{6}$$

Thus $x + \rho \cdot g \in X$, when $\rho = \rho_x(g)$, for any $g \in V$, $x \in X$.

Let a certain small value $\widehat{\varepsilon} > 0$ be given. Now, let us introduce an *ε-feasible set*

$$V_\varepsilon(x) = \big\{g | Ag = 0,\ \forall_{1 \leqslant i \leqslant n} \big(g_j \geqslant 0,\ \text{if } \big(0 \leqslant x_j \leqslant \varepsilon_x(g)\big)\big)\big\}, \tag{7}$$

where the function ε_x: $V(x) \rightarrow \Re_+$ is denoted as $\varepsilon_x(g) = \widehat{\varepsilon} \cdot \max_{\substack{1 \leqslant j \leqslant n \\ g_j \leqslant 0}} \big\{ \min\{x_j, -\widehat{\rho} \cdot g_j\}\big\}$, $\forall x \in X$.

It is a well-known fact that in stochastic optimization only the first order procedures are working and ensuring the best rate of convergence (Polyak, 1987; Ermolyev, 2003; etc). On the other hand, it has been also theoretically studied that a stochastic method of the first order method should converge if the variance of the stochastic error of the gradient estimate is proportional to the square norm of the gradient (Polyak, 1987; Sakalauskas, 2000). Since the error of Monte-Carlo estimators depends, first of all, on the sample size, we confine ourselves to the gradient-type methods introducing the corresponding rule for size regulation in Monte-Carlo estimators.

Thus, let the initial approximation of the solution $x^0 \in X$, some initial Monte-Carlo sample size N^0 be given and Monte-Carlo estimators of the objective function and the gradient would be computed. We define the sequence $\{x^t, N^t\}_0^\infty$ in an iterative way by setting

$$x^{t+1} = x^t + \rho_t \cdot \widetilde{G}^t, \tag{8}$$

$$N^{t+1} \geqslant \frac{\widehat{\rho} \cdot C}{\rho^t \cdot |\widetilde{G}^t|^2}, \tag{9}$$

where $C > 0$ is a certain constant, $\rho^t = \rho_{x^t}(\widehat{G}^t)$, $\widetilde{G}^t$ is an ε-feasible direction at the point x^t (i.e., projection of the gradient estimate to the ε-feasible set (4)). The following theorem provides conditions for the convergence of the method (8), (9).

Theorem 1. *Let the function F: $X \to \Re$ be differentiable, the gradient of this function be Lipshitzian with the constant $L > 0$, $\sup_{x\in X} |\nabla F(x)| < \infty$, $\sup_{x\in X} F(x) < \infty$.*

Assume the set $X = \{x \in \Re^n | Ax = b,\ x \geqslant 0\}$ to be bounded and having more than one element, $b \in R^m$, A is the $n \times m$-matrix.

Let it be possible to generate Monte-Carlo samples and corresponding estimates $\frac{1}{N}\sum_{j=1}^N n_j$, $\frac{1}{N}\sum_{j=1}^N \gamma_j$ to compute for any size $N > 1$, when $E\eta_j = F(x)$, $E\gamma_j = \nabla F(x)$, $E|\eta_j| < \infty$, $E|\gamma_j| < \infty$, $E|\gamma_j - \nabla F(x)|^2 < K$, $\forall x \in X$.

Then, starting from any initial approximation $x^0 \in X$ and $N^0 > 1$, formulae (8), (9) *define the sequence $\{x^t, N^t\}_0^\infty$ so that $x^t \in X$, and there exist values $\bar{\rho} > 0$, $\varepsilon_0 > 0$, $\overline{C} > 0$ such that*

$$\lim_{t\to\infty} \left|\nabla F(x^t)_{V^t}\right|^2 = 0(\mathrm{mod}\,(P)), \tag{10}$$

for $0 < \widehat{\rho} \leqslant \bar{\rho}$, $0 < \varepsilon \leqslant 0$, $C \geqslant \overline{C}$.

The proof of the theorem is given in Appendix.

Thus, we see that the application of an ε-feasible solution enables us to avoid "jumping" due to the statistical nature of Monte-Carlo estimators.

Note that for numerical implementation, the next rule similar to (8) is sometimes rather convenient:

$$N^{t+1} = \frac{\widehat{\rho} \cdot \Phi_\gamma}{\rho^t \cdot \widetilde{G}^t \cdot (Q^t)^{-1} \cdot \widetilde{G}^t}, \tag{11}$$

where Q^t is the sampling matrix of vectors γ_j and Φ_γ is the corresponding quantile of Fisher's distribution (see, also Sakalauskas, 2002).

The Monte-Carlo sample size regulation according to (8) enables us to construct reasonable, from the computational standpoint, stochastic methods for stochastic optimization. Namely, the method can start from a small initial size $N^0 = 20 - 50$, because there is no great necessity to evaluate estimators with a high accuracy at the beginning of optimization, when it suffices only to estimate an approximate direction leading to the optimum. Further the sample size is increased with respect to (8) or (11), gaining the values, sufficient to evaluate the estimators with an admissible accuracy only at the final stage of optimization, when the gradient becomes small in the neighbourhood of optimum. The numerical experiments and testing corroborate such a conclusion.

3. Termination Procedure

It is convenient to use the fact of asymptotic normality of Monte-Carlo estimators to evaluate the uncertainty of estimators and test the hypotheses of optimality (Sakalauskas, 2002). Thus, iteration by (8)–(9) or (8)–(11) should be terminated when:

a) the statistical criterion does not contradict the hypothesis on the criticality of the point of the current iteration (9) with the significance $1 - \sigma$:

$$(N^t - n_t)(\widetilde{\nabla} F^t)' \cdot (Q^t)^{-1} \cdot \widetilde{\nabla} F^t \leqslant \Phi_\sigma, \tag{12}$$

where Q^t is the covariance sampling matrix of vectors γ_j, Φ_σ is the quantile of the Fisher distribution with degrees $N^t - n_t$, and n_t, n_t is the dimension of the ε-feasible set;

b) the objective function has already been evaluated with an admissible confidence interval δ:

$$2\eta_\beta \cdot \frac{\widetilde{D}^t}{\sqrt{N^t}} \leqslant \delta, \tag{13}$$

where η_β is the normal β-quantile and $\widetilde{D}^t = \widetilde{D}(x^t, R)$ is the sampling standard deviation of sample η_j.

4. Counterexample

Financial planning in the case of uncertainty is often reduced to stochastic nonlinear optimization with linear constraints (Duffie and Pan, 1997; Mansini *et al.*, 2003). Let us to consider an application of the developed approach to the optimization of portfolio of the Lithuanian Stock Market with $n = 4$ securities.

We make the analysis for daily returns of the following assets:

Table 1

	ENRG	MAZN	ROKS	RST	μ_i	σ_i
	Correlations					
ENRG	1	0.0120	0.0010	0.1621	0.5029	0.7439
MAZN	0.0120	1	−0.0310	0.0954	0.4447	0.6414
ROKS	0.0010	−0.031	1	0.0572	0.2609	0.3320
RST	0.1621	0.0954	0.0572	1	0.3327	0.3555

- ENRG – joint stock company “Lietuvos energija” (power industry);
- MAZN – joint stock company “Mazeikiu Nafta” (oil refinery);
- ROKS – joint stock company “Rokiskio suris” (dairy products);
- RST – joint stock company “Rytu skirstomieji tinklai” (power industry).

A brief description of the data is given in Table 1, where empirical data were fitted by a lognormal model according to the Kolmogorov–Smirnov criterion. The data source is `www.nse.lt/nvpb/index_en.php`, time period – 2002.01–2003.10.

Thus, the portfolio return function is as follows,

$$r(x,\xi) = \sum_{i=1}^{n} x_i \cdot \mathrm{e}^{\xi_i},$$

$\xi \succ N(\mu, \Sigma)$, $\mu = (\mu_1, \mu_2, ..., \mu_n)$, $\Sigma = [\sigma_{ij}]_1^n$. Selection of portfolio weighs has been considered to maximize a probability of portfolio return to exceed the desired threshold R:

$$F(x) = P\big(r(x,\xi) \geqslant R\big) \to \max_{x \in X}, \tag{14}$$

subject to a simple set of constitutional constraints $X = \big(x | x_i \geqslant 0,\ \sum_{i=1}^{n} x_i = 1\big)$.

Selection of portfolio according to this objective function by the method developed is shown in Table 2. The gradient of the objective function (14) was expressed, using the transformation to polar variables described by Sakalauskas (1998). The parameters of the method were as follows: $\rho = 2.0$, $\delta = 1\%$, $\gamma = \sigma = \beta = 0.95$, $\varepsilon = 0.7$.

We see that, after $t = 10$ iterations and total 17753 Monte-Carlo trials, the probability of the desired portfolio increased from 78.12% (67.92 87.33) to 84.29% (83.79 84.79) (third column), changing the strategies of portfolio sharing with respect to (8) (second column) and choosing the Monte-Carlo sample size with respect to (11) (last column). The total amount of trials $\sum_{i=1}^{t} N_i$ exceeded the Monte-Carlo sample size N_t at the time of the stopping decision only by 1.79 times.

Table 2

t	x_1	x_2	x_3	x_4	Estimate $\widetilde{F}_t$ (Confidence)	Hotelling statistics (12) (Fisher quantile F_σ)	N_t
1	25.0	25.0	25.0	25.0	78.12% (68.92 87.33)	2.04 (2.57)	50
2	39.6	28.1	18.7	13.6	80.83% (73.59 88.08)	2.21 (2.53)	63
3	35.5	42.2	12.4	9.9	78.50% (71.39 85.61)	0.20 (2.51)	72
4	37.3	44.8	11.2	6.7	82.94% (81.14 84.73)	5.55 (2.38)	870
5	40.2	46.1	8.8	4.9	85.12% (82.58 87.67)	1.96 (2.40)	376
6	41.6	48.8	7.3	2.3	83.66% (81.25 86.07)	3.46 (2.39)	459
7	44.3	50.4	5.3	0.0	82.84% (79.92 85.76)	2.58 (2.63)	319
8	49.3	47.7	0.3	0.0	83.14% (80.28 86.00)	0.16 (2.63)	326
9	50.3	49.2	0.5	0.0	84.00% (83.30 84.69)	0.84 (2.61)	5318
10	50.7	49.3	0.0	0.0	84.29% (83.79 84.79)	0.18 (3.00)	9900
						$\Sigma N_t =$	17753

5. Conclussion

The method for stochastic programming with linear constraints by ε-feasible Monte-Carlo estimators has been developed. The method distinguishes itself by two peculiarities: the optimality of the solution is tested with respect to statistical criteria and the Monte-Carlo sample size is adjusted in an iterative way so as to guarantee the estimation of the objective function with an admissible confidence after a finite number of series. The theoretical study and a counterexample demonstrate the applicability of the approach proposed in the stochastic portfolio optimization.

Appendix

We need several lemmas to prove the theorem.

Lemma 1. *Let $g \in V(x), x \in X$. Denote $\varepsilon = \varepsilon_x(g)$ according to* (7) *and by g_ε the projection of the vector g to an ε-feasible set. Then*

A) *each nonzero vector g contains negative components; moreover, there exists a value $a < \infty$ such that uniformly*

$$\frac{|g|}{\max_{\substack{1 \leqslant k \leqslant n \\ g_k < 0}} \{|g_k|\}} \leqslant a, \quad g \in V(x),\ x \in X;$$

B) $g \neq 0 \Rightarrow \varepsilon > 0$*;*

C) $g = 0 \Leftrightarrow g_\varepsilon = 0$*;*

D) *if j is such that* $\frac{x_j}{|g_j|} = \min_{\substack{1\leqslant i\leqslant n\\ g_i<0}} \left(-\frac{x_i}{g_i}\right)$, *then*

$$|g_j| \geqslant \widehat{\varepsilon}\cdot \max_{\substack{1\leqslant k\leqslant n\\ g_k<0}} \{|g_k|\}.$$

Proof. Note that X is a bounded convex and closed set from some linear space of nonzero dimensions and $V(x)$ are convex closed cones (Rockafellar, 1996). It is easy to get sure that every nonzero vector $g \in V$ contains negative components. Indeed, in the opposite case, we have that $x + \rho \cdot g \in X$ for any $\rho \geqslant 0$, which contradicts the assumption on the finiteness of the set of institutional restrictions X. The estimate in A follows from the closeness and finiteness of X, too.

We have that $\{g \in V \cap g \neq 0\} \Rightarrow \exists_{1\leqslant j\leqslant n}(g_j < 0 \cap x_j > 0)$ by definition of the feasible set, because all the components of any $x \in X$ cannot be zero at the same time and any nonzero vector $g \in V$ contains negative components. This implies B.

Let us study the structure of an ε-feasible set (7). According to the definition, this set is an intersection of a finite number of linear half-spaces. It is concave, because

$$\begin{aligned}
&\max_{1\leqslant j\leqslant n}\Big\{\min\big\{x_j, \max(0, -\lambda\cdot\widehat{\rho}\cdot g_j^1)\big\}\Big\} \\
&\qquad + \max_{1\leqslant j\leqslant n}\Big\{\min\big\{x_j, \max(0, -(1-\lambda)\cdot\widehat{\rho}\cdot g_j^2)\big\}\Big\} \\
&\quad \geqslant \lambda \max_{1\leqslant j\leqslant n}\Big\{\min\big\{x_j, \max(0, -\widehat{\rho}\cdot g_j^1)\big\}\Big\} \\
&\qquad +(1-\lambda)\max_{1\leqslant j\leqslant n}\Big\{\min\big\{x_j, \max(0, -\widehat{\rho}\cdot g_j^2)\big\}\Big\},
\end{aligned}$$

when $0 \leqslant \lambda \leqslant 1$, $\forall x \in X$. The ε-feasible set $V_\varepsilon(x)$ is a subset of the feasible set $V(x)$. It is easy to get convinced that it contains the zero vector $g = 0$ in a close vicinity from $V(x)$. The latter conclusions imply proposition C.

Now let the index j be such that $\frac{x_j}{|g_j|} = \min_{\substack{g_i<0,\\ 1\leqslant i\leqslant n}} \left(\frac{x_i}{|g_i|}\right)$, $g \in V_\varepsilon$. Then

$$|g_j| \geqslant |g_i|\frac{x_j}{x_i} \geqslant \widehat{\varepsilon}\cdot|g_i|\frac{\max\limits_{\substack{1\leqslant k\leqslant n\\ x_k+\widehat{\rho}\cdot g_k\leqslant 0}}\{x_k\}}{x_i} \geqslant \widehat{\varepsilon}\cdot|g_i|, \quad \forall_{1\leqslant i\leqslant n}(x_i + \widehat{\rho}\cdot g_i \leqslant 0).$$

However,

$$|g_j| \geqslant \frac{x_j}{\widehat{\rho}} \geqslant \widehat{\varepsilon} \max_{\substack{1\leqslant k\leqslant n\\ x_k+\widehat{\rho}\cdot g_k>0\\ g_k<0}} \{|g_k|\}, \quad \forall_{1\leqslant i\leqslant n}(x_i + \widehat{\rho}\cdot g_i > 0,\ g_i < 0), \quad \text{too.}$$

Both last estimates imply D.

The lemma is proved.

Lemma 2. *If vectors g and g^1 are ε-feasible at the point $x \in X$, then for a certain $a > 0$*

$$\left|\rho_x(g^1)\cdot g^1 - \rho_x(g)\cdot g\right| \leqslant \widehat{\rho}\cdot\left(\frac{a}{\widehat{\widehat{\varepsilon}}} + 1\right)\cdot|g^1 - g|.$$

Proof. Denote indices j and i such that $\frac{x_j}{|g^1_{\varepsilon,j}|} = \min_{\substack{1\leqslant k\leqslant n \\ g^1_{\varepsilon,k}<0}} \left(\frac{x_k}{|g^1_{\varepsilon,k}|}\right)$ and $\frac{x_i}{|g_{\varepsilon,i}|} = \min_{\substack{1\leqslant k\leqslant n \\ g_{\varepsilon,k}<0}} \left(\frac{x_k}{|g_{\varepsilon,k}|}\right)$. Let, for the sake of simplicity, $\rho_x(g^1) \geqslant \rho_x(g)$. Then

$$\begin{aligned}\rho_x(g^1) - \rho_x(g) &= \min\left(\widehat{\rho}, \frac{x_j}{|g^1_j|}\right) - \min\left(\widehat{\rho}, \frac{x_i}{|g_i|}\right) \leqslant \left|\frac{x_i}{|g^1_i|} - \frac{x_i}{|g_i|}\right| \\ &\leqslant \rho_x(g^1)\cdot\left|\frac{g^1_i}{g_i} - 1\right| \leqslant \widehat{\rho}\cdot\frac{|g^1-g|}{g_i}.\end{aligned}$$

Thus, by virtue of A and D of Lemma 1:

$$\begin{aligned}&\left|\rho_x(g^1)\cdot g^1 - \rho_x(g)\cdot g\right| = \left|\rho_x(g^1)\cdot(g^1-g) + \left(\rho_x(g^1)-\rho_x(g)\right)\cdot g\right| \\ &\leqslant \widehat{\rho}\cdot|g^1-g|\cdot\left(1+\frac{|g|}{g_i}\right) \leqslant \widehat{\rho}\cdot|g^1-g|\cdot\left(1+\frac{\max_{\substack{g_i<0 \\ 1\leqslant i\leqslant n}}|g|}{g_i}\cdot\frac{|g|}{\max_{\substack{g_i<0 \\ 1\leqslant i\leqslant n}}|g|}\right) \\ &\leqslant \widehat{\rho}\cdot|g^1-g|\cdot\left(1+\frac{a}{\widehat{\varepsilon}}\right).\end{aligned}$$

The lemma is proved.

Lemma 3. *Assume the conditions of theorem to be valid and let* $\widehat{\rho} > 0$ *and* $\widehat{\varepsilon} > 0$ *be some small values. Then:*

$$EF(x+\widetilde{\rho}\cdot\widetilde{G}) \geqslant F(x) + E\left(\widetilde{\rho}\cdot|\widetilde{G}|^2\right)\cdot\left(1-\frac{\widehat{\rho}\cdot L}{2}\right) - \frac{\widehat{\rho}\cdot\left(1+\frac{a}{\widehat{\varepsilon}}\right)\cdot K}{N}, \quad \forall(x\in X),$$

where $\widetilde{G}$ *is the projection of the estimate* $\widetilde{\nabla}F$ *to the* ε*-feasible set,* $\widetilde{\rho} = \rho_x(\widetilde{G})$ *is the corresponding step length chosen according to* (6)*.*

Proof. We have from the Lagrange formula (Diedonne, 1960) that

$$\begin{aligned}F(x+\widetilde{\rho}\cdot\widetilde{G}) &= F(x) + \widetilde{\rho}\cdot\widetilde{G}'\cdot\int_0^1 \nabla F(x+\widetilde{\rho}\cdot\tau\cdot\widetilde{G})\,\mathrm{d}\tau \\ &= F(x) + \widetilde{\rho}\cdot\widetilde{G}'\cdot\widetilde{\nabla}F - (\widetilde{\rho}\cdot\widetilde{G} - \rho\cdot G)'\cdot(\widetilde{\nabla}F - \nabla F) \\ &\quad +\rho\cdot G'\cdot(\widetilde{\nabla}F-\nabla F) + \widetilde{\rho}\cdot\widetilde{G}'\cdot\int_0^1 (\nabla F(x+\widetilde{\rho}\cdot\tau\cdot\widetilde{G}) - \nabla F(x))\,\mathrm{d}\tau,\end{aligned}$$

where $\rho = \rho_x(G)$, G is the ε-feasible projection of gradient ∇F. Thus, the proof of the lemma is complete taking the expectation of both sides of this expression and applying further the Lipshitz condition and Lemma 2.

The lemma is proved.

Proof of Theorem 1

Denote a stream of σ-algebras generated by the sequence $\{x^t, N^t\}_{t=0}^{\infty}$ by $\{\Im_t\}_{t=0}^{\infty}$. Let us introduce a random sequence

$$X_t = F(x^t, R) - \frac{\widehat{\rho} \cdot \left(1 + \frac{a}{\widehat{\varepsilon}}\right) \cdot K}{N^t}.$$

Assume $0 < \widehat{\rho} \leqslant \bar{\rho} = \frac{1}{L}$. Then by virtue of Lemma 3 we have that

$$\begin{aligned} E(X_{t+1}|\Im_{t-1}) &\geqslant X_t + \left(1 - \frac{\widehat{\rho} \cdot L}{2}\right) \cdot E\left(\rho^t \cdot |\widetilde{G}^t|^2 |\Im_{t-1}\right) \\ &\quad - \widehat{\rho} \cdot \left(1 + \frac{a}{\widehat{\varepsilon}}\right) \cdot K \cdot E\left(\frac{1}{N^{t+1}} |\Im_{t-1}\right) \\ &\geqslant X_t + \left(\frac{1}{2} - \frac{\left(1 + \frac{a}{\widehat{\varepsilon}}\right) \cdot K}{C}\right) \cdot E\left(\rho^t \cdot |\widetilde{G}^t|^2 |\Im_{t-1}\right), \quad t = 1, 2, \ldots. \end{aligned} \tag{1A}$$

It follows that X_t is a submartingale for $C > \frac{\left(1 + \frac{a}{\widehat{\varepsilon}}\right) \cdot K}{4 \cdot L}$.

By summing up unconditional expectations on both sides of inequality (1A) and setting $C \geqslant \overline{C} = \frac{\left(1 + \frac{a}{\widehat{\varepsilon}}\right) \cdot K}{4 \cdot L}$, one can get:

$$\frac{1}{4} \sum_{k=0}^{t} E\left(\rho^k \cdot \left|\widetilde{G}(x^k)\right|^2\right) \leqslant EF(x^{t+1}) - F(x^0) + \widehat{\rho} \cdot \left(1 + \frac{a}{\widehat{\varepsilon}}\right) \cdot K. \tag{2A}$$

The left-hand side of this inequality is bounded, and therefore the series on the left converges as $t \to \infty$.

Now, say $\lim_{t\to\infty} |\widetilde{G}(x^t)|^2 \neq 0$. Then a certain small value $\delta^2 > 0$ could be found that a converging infinite subsequence $\{x^{t_k}\}_{k=0}^{\infty}$ exists such that $|\widetilde{G}^{t_k}|^2 > \delta^2$ for any term of this subsequence. Denote the limit of this subsequence by $\hat{w}$. Let us fix a vicinity of $\hat{w}$ such that $|\widetilde{G}^{t_k}|^2 > \frac{\delta^2}{2}$ for all points of the subsequence from this vicinity. It follows by virtue of B) of Lemma 1 and the continuity of $\varepsilon_x(\cdot)$ that there exists a certain ε_1 so that $\varepsilon^{t_k} \geqslant \varepsilon_1 > 0$ for all points of the subsequence hitting this vicinity. Hence, by virtue of (15), we have: $\rho^{t_k} \cdot |\widetilde{G}^{t_k}|^2 \geqslant \min\left(\widehat{\rho}, \frac{\varepsilon_1}{|\widetilde{G}^t|}\right) \cdot |\widetilde{G}^t|^2 \geqslant \min(\widehat{\rho}, \varepsilon_1) \cdot \frac{\delta}{2}$. Consequently, we should have an infinite number of terms in (2A) exceeding $\min(\widehat{\rho}, \varepsilon_1) \cdot \frac{\delta}{2} > 0$ which contradicts the convergence in (2A). This implies:

$$\lim_{t\to\infty} \left|\widetilde{G}(x^t)\right|^2 = 0 \quad (\mathrm{mod}(P)). \tag{3A}$$

Next, by virtue of (19), (3A), we obtain

$$\lim_{t\to\infty} \frac{1}{N^t} = 0 \quad (\mathrm{mod}(P)). \tag{4A}$$

Further

$$\lim_{t\to\infty} |G^t|^2 \leqslant \lim_{t\to\infty} \left|\widetilde{G}^t\right|^2 + \lim_{t\to\infty} \left|\widetilde{G}^t - G^t\right|^2 = 0 \quad (\mathrm{mod}(P)),$$

because $\lim_{t\to\infty} |\widetilde{G}^t - G^t|^2 = 0$ by virtue of (4A) and the law of large numbers. It remains to apply the proposition C of Lemma 1 to establish (10).

The proof of the theorem is completed.

References

Bentkus, V., and F. Gotze (1999). Optimal bounds in non-Gaussian limit theorems for U-statistics. *Annals of Probability*, **27**(1), 454–521.

Bertsekas, D.P. (1982). *Constrained Optimization and Lagrange Multiplier Methods. Computer Science and Applied Mathematics.* Academic Press, Inc. Harcourt Brace Jovanovich, N.Y., London.

Dieudonne, J. (1960). *Foundations of Modern Analysis*. Academic Press, N.Y., London.

Duffie, D., and J. Pan (1997). An overview of value-at-risk. *J. Deriv.*, **4**, 7–49.

Ermoliev, Y.M., K. Marti and G. Pflug (2003). *Dynamic Stochastic Optimization*. Springer-Verlag.

Han-Fu-Chen (2002). *Stochastic Approximation and Its Applications* (Nonconvex Optimization and Its Applications, 64). Kluwer Academic Publishers.

Kjellstrom, G. (1969). Network optimization by random variation of component values. *Ericson Techniques*, **3**, 133–151.

Krishnaiah, P.R. (1988). *Handbook of Statistics*. *Analysis of Variance*, Vol. 1. North-Holland, N.Y.–Amsterdamm.

Kushner, H.J., and G.G. Jin (1997). *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, N.Y.

Mikhalevitch, V.S., A.M. Gupal and V.I. Norkin (1987). *Methods of Nonconvex Optimization*. Nauka, Moscow (in Russian).

Mansini, R., W. Ogryczak and M.G. Speranza (2003). LP solvable models for portfolio optimization. *Informatica*, **14**(1), 37–62.

Polyak, B.T. (1987). *Introduction to Optimization. Translations Series in Mathematics and Engineering*. Optimization Software, Inc., Publications Division, New York.

Prekopa, A. (1999). *Stochastic Programming*. Kluwer Academic Publishers.

Rockafellar, R.T. (1996). *Convex Analysis*. Princeton University Press.

Rubinstein, R. (1983). Smoothed functionals in stochastic optimization. *Mathematical Operations Research*, **8**, 26–33.

Sakalauskas, L. (1998). Portfolio management by the Monte-Carlo method. In *Proceedings of the 23rd Meeting of the EURO Working Group on Financial Modelling*, Krakow. Progress&Business Publ.. pp. 179–188.

Sakalauskas, L. (2000). Nonlinear stochastic optimization by the Monte-Carlo method. *Informatica*, **11**(4), 455–468.

Sakalauskas, L. (2002). Nonlinear stochastic programming by Monte-Carlo estimators. *European Journal on Operational Research*, **137**, 558–573.

Shapiro, A. (1989). Asymptotic properties of statistical estimators in stochastic programming. *The Annals of Statistics*, **17**(2), 841–858.

Shyryajev, A. (1989). *Probability*. Nauka, Moscow (in Russian).

Uriasyev, S.P. (1990). *Adaptive Algorithms of Stochastic Optimization and Theory of Games*. Nauka, Moscow (in Russian).

Uryasev, S. (1994). Derivatives of probability functions and integrals over sets given by inequalities. *J. Comput. Appl. Math.*, **56**(1–2), 197–223.

L. Sakalauskas has graduated from the Kaunas Polytechnical Institute (1970), received the PhD degree from this Institute (1974) and the degree of DrHabil. from the Institute of Mathematics and Informatics (2000), Vice-President of Lithuanian Operational Research Society (2001), elected member of the International Statistical Institute (2002), presently is a head of the Operational Research Division of the Institute of Mathematics and Informatics and professor of the Department of Information Technologies of the Vilnius Gediminas Technical University. His research interest include stochastic modeling and optimization with applications.

Monte-Karlo metodo taikymas netiesiniam stochastiniam programavimui su tiesiniais ribojimais

Leonidas SAKALAUSKAS

Darbe nagrinėjama stochastinio netiesinio programavimo su tiesiniais ribojimais problema. Sukurtas leistinų sprendinių metodas šiai problemai spręsti panaudojus Monte-Karlo imčių serijas, kuris leidžia išvengti "užsikirtimo" arba "zigzagavimo". Metodas pasižymi dviem pagrindinėm savybėm: sprendinio optimalumas yra testuojamas pasinaudojus statistiniais kriterijas bei Monte-Karlo imčių tūris yra reguliuojamas taip, kad sumažinti skaičiavimų apimtį, reikalingą uždaviniui išspręsti, bei užtikrinti metodo konvergavimą. Pritaikius martingalų metodą prie gana bendrų sąlygų įrodytas sukurto optimizavimo metodo konvergavimas b.v. į stacionarų sprendžiamos problemos tašką. Skaitmeninio vertybinių popierių portfolio optimizavimo pavyzdys pateikiamas metodo veikimui pademonstruoti.