Liet. mat. rink. LMD darbai, **48/49**, 2008, 320–326

Epsilon-projection method for two-stage SLP

Leonidas SAKALAUSKAS (MII), Kęstutis ŽILINSKAS (ŠU)
e-mail: sakal@ktl.mii.lt, kest.zil@gmail.com

Abstract. In this paper, a stochastic adaptive method with the ε-projection for stochastic gradient has been developed to solve two-stage stochastic linear problems. The method is based on the adaptive regulation of the size of Monte-Carlo samples and a statistical termination procedure taking into consideration the statistical modelling accuracy. To avoid "jamming" or "zigzagging" in solving the constraint problem the procedure for stochastic gradient epsilon-projection has been implemented. Four algorithms of the epsilon-projection are tested by the computer simulation. The recommendations to solve two-stage stochastic linear problems are made under the numerical study and results of solving the test examples.

Keywords: stochastic programming, Monte Carlo method, optimization, stochastic gradient, statistical criteria, epsilon-feasible direction, epsilon-projection.

1. Introduction

Stochastic programming deals with a class of optimization models in which some data may be a subject to significant uncertainty. Such models are appropriate when data evolve over time and decisions have to be made prior to observing the entire data streams. Although widespread applicability of stochastic programming models has attracted considerable attention of researchers, stochastic linear models remain one of more challenging optimization problems. Methods based on approximation and decomposition are often applied to solve stochastic programming tasks (see, e.g., [4–7], etc.); however, they can lead to very large-scale problems, and, thus, require very large computational resources. In this paper, we have developed an adaptive approach for solving stochastic linear problems by the Monte-Carlo method based on asymptotic properties of Monte-Carlo sampling estimators. To avoid "jamming" or "zigzagging" in solving a constraint problem we implement the procedure for stochastic gradient ε-projection. This approach is grounded on the treatment of a statistical simulation error in a statistical manner and the rule for iterative regulation of the size of Monte-Carlo samples [9,10].

Let us consider a two-stage stochastic optimization problem with a complete recourse:

$$F(x) = cx + E\{Q(x,\xi)\} \rightarrow \min_{x \in D \subset \Re^n_+}, \tag{1}$$

subject to feasible set

$$D = \left\{x \mid Ax = b, x \in \Re^n_+\right\}, \tag{2}$$

where

$$Q(x,\xi) = \min_y \left[qy \mid Wy + Tx \leqslant h,\ y \in \Re_+^m \right], \tag{3}$$

vectors b, q, h and full rank matrices A, W, T are of appropriate dimensionality [11]. Assume vectors q, h and matrices W, T be random in general, and distributed by absolute probability law with the density function $p(\cdot)\colon \Omega \to \Re_+$.

2. Stochastic differentiation and Monte-Carlo estimators

Let us consider the analytical approach (AA) to estimate the gradient of the objective function in two-stage stochastic programming with recourse (1). Indeed, by the duality of linear programming we obtain that

$$F(x) = cx + E\left\{ \max_u \left[(h - Tx)u \mid uW^T + q \geqslant 0,\ u \in \Re_+^n \right] \right\}. \tag{4}$$

It's easy to see, that objective function is piecewise linear. Thus, by Rademacher theorem, this function is differentiable almost everywhere and, by Lebesgue theorem, the gradient of two-stage stochastic linear problems objective function can be expressed as

$$\nabla_x F(x) = E\{g^i(x,\xi)\}, \tag{5}$$

where $g(x,\xi) = c - Tu^*, u^*$ is given by a set of solutions of a dual problem ([8,13], etc.)

$$(h - Tx)^T u^* = \max_u \left[(h - Tx)^T u \mid uW^T + q \geqslant 0,\ u \in \Re^n \right], \tag{6}$$

Under the assumption on the continuity of the measure of the second stage variables the objective function is smoothly differentiable.

Let us define the set of *feasible directions* as follows:

$$V(x) = \left\{ g \in \Re^n \mid Ag = 0, \forall_{1 \leqslant i \leqslant n} (g_j \leqslant 0,\ \text{if } x_j = 0) \right\}, \quad x \in D. \tag{7}$$

Since the objective function is differentiable, the solution $x \in D$ is optimal if [1]:

$$\nabla F(x)_V = 0, \tag{8}$$

where g_U is a projection of the vector g onto the set U.

Let us assume that Monte-Carlo samples $Y = (y^1, y^2, \ldots, y^N)$ of a certain size N are provided for any $x \in D$, where y^i are independent random variables of the second stage, identically distributed at density $p(\cdot)$. Then the sampling estimator $\tilde{F}(x)$ and the sampling standard deviation $\tilde{D}(x)$ can be computed from the random sample: $\{f(x,y^1), f(x,y^2), \ldots, f(x,y^N)\}$. Next, assume the stochastic gradient $g(x,y^j)$ can be computed for any $x \in D$ and y^j. Thus, the corresponding sampling gradient estimator $\tilde{G}(x) = \frac{1}{N}\sum_{j=1}^N g(x,y^j)$ and the sampling covariance matrix $Z(x) = \frac{1}{N-n}\sum_{j=1}^N [g(x,y^j) - \tilde{G}] \cdot [g(x,y^j) - \tilde{G}]'$ can be estimated by same random sample, too.

3. Stochastic procedure for optimization

Let us introduce the stochastic procedure for the optimization. The gradient search approach with projection to a feasible set would be a chance to create optimizing sequence; however, the problems of "jamming" or "zigzagging" are typical in this case. To avoid them the ε-feasible direction approach is applied.

Assume a certain multiplier $\tilde{\rho} > 0$ is given. Define the function $\rho_x\colon V(x) \to \Re_+$ by

$$\rho_x(g) = \min\left\{\tilde{\rho},\ \min_{\substack{g_j>0,\\ 1\leqslant j\leqslant n}} \left(\frac{x_j}{g_j}\right)\right\}, \quad \exists_{1\leqslant j\leqslant n}(g_j > 0), \tag{9}$$

$$\rho_x(g) = \hat{\rho}, \quad \text{if } \forall_{1\leqslant j\leqslant n}(g_j \leqslant 0).$$

Thus, $(x + \rho \cdot g) \in D$, when $\rho = \rho_x(g)$, for any $g \in V(x)$, $x \in D$, Now, let a certain small value $\tilde{\varepsilon} > 0$ be given. Then we introduce the function $\varepsilon_x\colon V(x) \to \Re_+$

$$\varepsilon_x(g) = \tilde{\varepsilon} \cdot \max_{\substack{1\leqslant j\leqslant n\\ g_j>0}} \left\{\min\{x_j, \tilde{\rho}\cdot g_j\}\right\}, \quad \exists_{1\leqslant j\leqslant n}(g_j > 0),$$

$$\varepsilon_x(g) = 0, \quad \text{if } \forall_{1\leqslant j\leqslant n}(g_j \leqslant 0),$$

and define the ε-*feasible set*

$$V_\varepsilon(x) = \left\{g \in \Re^n \mid Ag = 0, \forall_{1\leqslant i\leqslant n}\big(g_j \leqslant 0,\ \text{if } (0 \leqslant x_j \leqslant \varepsilon_x(g))\big)\right\}. \tag{10}$$

Denote the projector to feasible area by $P = I - B^T \cdot (BB^T)^{-1} \cdot B$, where B is a non-degenerate matrix.

Assume that initial point $x^0 \in D$ is given, the Monte-Carlo sample of a certain size N^0 is generated and Monte Carlo estimators are calculated. Iterative stochastic procedure of the gradient search is used further:

$$x^{t+1} = x^t - \rho^t \cdot \tilde{G}_\varepsilon(x^t), \tag{11}$$

where $\rho^t = \rho_{x^t}(\tilde{G}^t_{\varepsilon^t})$ is a step-length multiplier defined by (9), and $\tilde{G}^t_\varepsilon$ is a feasible direction. To find this direction the projection of the gradient estimator to linear subset, describing by constraint matrix A, P_A is calculated. Then the set of the indexes of the variables is determined for those steps of the gradient search will be done. In this process, indexes of the zero components of the point x^{i}, which are corresponding to the positive components of the gradient estimator, are eliminating from the set of all variables. The indexes are eliminating one by one, beginning with components corresponding to the maximal positive component of the gradient. Then the index is eliminated, the corresponding column of the constraint matrix is replaced by the column with zero values and projection matrix Z' and projection of the gradient g' are recalculated:

$$Z' = Z - \frac{Z^{<i>} \cdot (Z^{<i>})^T}{Z_{i,j}}, \quad g' = g - \frac{g^{<i>} \cdot (Z^{<i>})^T}{Z_{i,j}},$$

where i – the eliminating index.

Four rules for the choice of eliminating index were tested:

- finding the minimal component of the gradient between almost admissible directions g_{jmin} – (R1);
- finding the minimal projection of the gradient to admissible directions – (R2);
- finding the minimal component of the point x between almost admissible directions x_j – (R3);
- finding the minimal ratio $\frac{x_j}{g_j}$ between almost admissible directions – (R4).

A possible decision on finding an optimal solution and stopping of the algorithm were tested by statistical criteria. The optimality hypothesis is accepted for some point x_t with significance $1-\mu$, if the following condition is met:

$$\frac{1}{n} \cdot (N^t - n) \cdot \tilde{G}(x^t) \cdot \left(Z(x^t)\right)^{-1} \cdot \tilde{G}(x^t) \leqslant Fish(\mu, n, N^t - n). \tag{12}$$

Next, again it is accepted that the objective function is estimated with permissible accuracyδ, if its confidence bound does not exceed this value:

$$\frac{2\eta_\beta \cdot \tilde{D}(x^t)}{\sqrt{N^t}} \leqslant \delta, \tag{13}$$

where $\tilde{D}(x^t)$ is a sampling variance of the objective function, η_β is the β-quantile of a standard normal distribution [12]. The iterations of the algorithm are being repeated while both conditions are met.

4. Computer simulation of stochastic gradient estimators

In this section, a computer simulation study on the gradient estimators considered in Section 3 is presented by using testing examples.

Example (Two-stage stochastic linear optimization problem). The data of the problem is taken from the database [2,3] at `http://www.math.bme.hu/~deak/twostage/l1/20x20.1` (accessed on 2006-01-20).

The dimensions of the task are as follows: the first stage has 10 rows and 20 variables; the second stage has 20 rows and 30 variables. The estimate of the optimal value of the objective function given in the database is 182.94234 $\pm$0.066. The application of the considered approach allows us to improve the estimate of the optimal value up to 182.59248 $\pm$0.033.

The computer study for various rules of ε-projection was done by PC computer (Intel Core2 Quad CPU 2.0GHz, 1 GB of RAM, Windows XP). Algorithms implemented by original C++ application.

The example was solved 400 times by the approach developed for four different e-projection approaches described above. In the Figs. 1–5, the number of the gradient ε-projection steps, the number of iterations, the time of the computation, the ratio and the total amount of the Monte Carlo trials needed to solve the problem under the ε-projection constant ε for various rules of ε-projection are presented. $\varepsilon = 0$ means that any ε-projection was done.

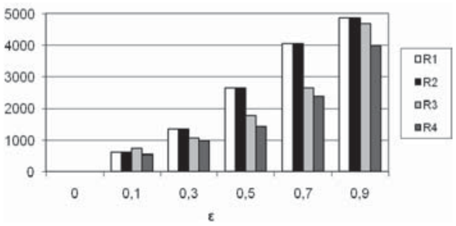

Fig. 1. Number of the gradient ε-projection steps under the ε-projection constant ε for various rules of ε-projection.

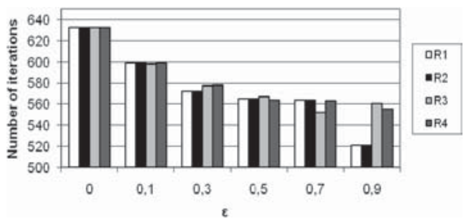

Fig. 2. Number of iterations under the ε-projection constant ε for various rules of ε-projection.

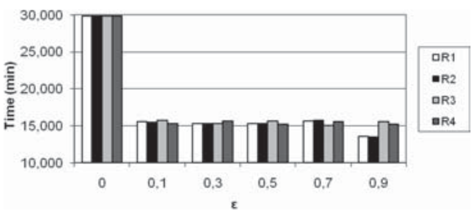

Fig. 3. Time of the computation under the ε-projection constant ε for various rules of ε-projection.

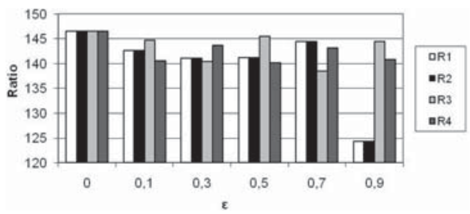

Fig. 4. Ratio $\sum_{j=1}^{t} \frac{N^j}{N^t}$ under the ε-projection constant ε for various rules of ε-projection.

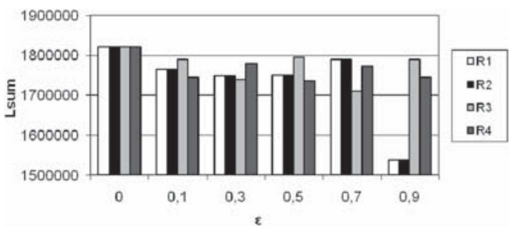

Fig. 5. Total amount of the computations under the ε-projection constant ε for various rules of ε-projection.

The results of computer simulation showed the theoretical presumption that constant epsilon, chosen properly (0.8–0.9), guarantees convergence of the algorithm a. s. to an optimal solution [10].

5. Conclusions

Thus, the stochastic iterative method has been developed to solve two-stage stochastic linear programming problems by a finite sequence of Monte-Carlo sampling estimators. The method of the ε-projection of the stochastic gradient by feasible directions was implemented. The proposed method was applied to solve the examples taken from standard database of two-stage stochastic programming tests.

The approach presented in this paper is grounded on the stopping procedure and the rule for adaptive regulation of the size of Monte-Carlo samples, taking into account statistical modelling accuracy. Several stochastic gradient estimators were compared by computer simulation by studying the workability of the estimators for testing the optimality hypothesis by statistical criteria. The regulation of a sample size in case this size is taken inversely proportional to the square of the norm of the gradient of the Monte-Carlo estimator allows us to solve stochastic linear programming problems rationally from a computational viewpoint and guarantees convergence a. s. The numerical study corroborates theoretical conclusions on the convergence method and shows that the developed procedures make it possible to solve stochastic problems with sufficiently agreeable accuracy by the means of an acceptable amount of computations.

It follows from the computer study that the ε-projection approach enables us to solve the optimization problem avoiding "zigzagging" or "jamming" and decreasing the amount of computations as compared with that without the ε-projection. The conclusion also might be done that the ε-projection constant should be chosen not small (about 0.8–0.9). The rules for the choice of eliminating index are more preferable as follows: find minimal component of the gradient between almost admissible directions (R1) and find minimal projection of the gradient to admissible set (R2).

References

1. D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, London (1982).
2. I. Deak, Successive regression approximations for solving equations, *Pure Mathematics and Applications*, **12**, 25–50 (2001).
3. I. Deak, Solving stochastic programming problems by successive regression approximations. Numerical results, in: K. Marti, Y. Ermoliev, G. Pflug (Eds.), *Dynamic Stochastic Optimization*, Springer LNEMS, vol. 532 (2003), pp. 209–224.
4. Yu. Ermolyev, R. Wets, *Numerical Techniques for Stochastic Optimization*, Springer-Verlag, Berlin (1988).
5. K. Marti, Differentiation formulas for probability functions: The transformation method, *Mathematical Programming*, **75**(2), 201–220 (1996).
6. K. Marti, *Stochastic Optimization Methods*, Springer, N.Y. (2005).
7. A. Prekopa, *Stochastic Programming*, Kluwer (1995).
8. R. Rubinstein, A. Shapiro, *Discrete Events Systems*: *Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, Wiley & Sons, N.Y. (1993).
9. L. Sakalauskas, Nonlinear stochastic programming by Monte-Carlo estimators, *European Journal on Operational Research*, **137**, 558–573 (2002).
10. L. Sakalauskas, Application of the Monte-Carlo method to nonlinear stochastic optimization with linear constraints, *Informatica*, **15**(2), 271–282 (2004).
11. L. Sakalauskas, K. Zilinskas, Application of the Monte-Carlo method to stochastic linear programming, *Computer Aided Methods in Optimal Design and Operations / Series on Computers and Operations Researck*, **7**, 39–48 (2006).
12. L. Sakalauskas, K. Zilinskas, Application of statistical criteria to optimality testing in stochastic programming, *Technological and Economic Development of Economy*, **XII**(4), 314–320 (2006).
13. A. Shapiro, Stochastic Programming by Monte-Carlo simulation methods, *Stochastic Programming E-Print Series* (2000).

REZIUMĖ

L. Sakalauskas, K. Žilinskas. Epsilon-projektavimo metodas dviejų etapų stochastiniam tiesiniam programavimui

Straipsnyje nagrinėjamas stochastinis adaptyvus metodas su stochastinio gradiento epsilon-projektavimu dviejų etapų stochastiniams tiesiniams uždaviniams spręsti. Metodas pagrįstas Monte Karlo imčių tūrio adaptyviu reguliavimu ir stabdymo procedūra, vertinant statistinio modeliavimo tikslumą statistiniais kriterijais. Epsilon-projektavimo metodas yra sudarytas „užstrigimo“ arba „zigzagavimo“ problemoms išvengti sprendžiant uždavinius su ribojimais. Keturi epsilon-projektavimo algoritmai aprašyti ir ištirti kompiuteriniu modeliavimu. Remiantis modeliavimo rezultatais, pateiktos praktinio realizavimo rekomendacijos.

Raktiniai žodžiai: stochastinis programavimas, Monte-Karlo metodas, optimizavimas, stochastinis gradientas, epsilon-leistina kryptis, epsilon-projekcija.