

Minorant methods for stochastic global optimization

Norkin V.I., Onischenko B.O.

Institute of Cybernetics of the Ukrainian Academy of Sciences
03187 Kiev, Glushkov avenue 40, Ukraine

Abstract. Branch and bound method and Pijavskii's method are extended for solution of global stochastic optimization problems. These extensions employ a concept of stochastic tangent minorants and majorants of the integrand function as a source of global information on the objective function. A calculus of stochastic tangent minorants is developed.

Keywords. Stochastic global optimization, stochastic tangent minorant, branch and bound method.

Problem of stochastic optimization consists in minimization of a mathematical expectation or a probability function. Difficulty of the problem is that the objective function cannot be calculated precisely, but only statistical estimators of its values and, probably, of its gradients are known. The task is in the search of local and global minima of the problem with the use of these estimators. Extensive literature is devoted to the solution of convex stochastic optimization problems [4]. Problems and methods of searching local minima in nonconvex stochastic optimization problems are discussed in [1]. A number of global stochastic optimization problems and a stochastic branch and bound method for their solution are studied in [10 - 12], where estimations of branches (subtasks) are obtained by means of the so-called interchange relaxation, i.e. by interchange of minimization and integration (mathematical expectation or probability) operators. In works [5, 6, 9] the specified stochastic branch and bound method is applied to global optimization of probabilities with application to the control of environmental contamination, and in [2, 3] it is applied to problems of optimal routing and to project management.

The basic results of the present work are developed in [7, 8, 13 - 17] and consist in extension of deterministic Pijavski's global optimization method and a classical branch and bound method on problems of global stochastic optimization (with

2 Norkin V.I., Onischenko B.O.

mathematical expectation and probability objective functions).

A common feature of considered methods is the use of tangent minorants of the objective function as a source of global information on the function behavior. Thus the key problem is how to construct tangent minorants. For example, as such minorants tangent cones or tangents paraboloids to the function graph can be used. The use tangent paraboloids instead of cones considerably increases efficiency of the method [13]. In [7] a calculus of tangent minorants for complex nonconvex criterion functions is developed. In the present paper we give new rules for calculation of tangent minorants, in particular, for minimum and maximum functions and also stochastic tangent minorants for mathematical expectation and probability functions. Consideration of stochastic minorants is similar to generalization of a deterministic gradient method to a stochastic quasigradient method for solution of stochastic programming problems. The search of deterministic minorants as well as gradients of mathematical expectation functions can be problematic, however calculation and use of stochastic minorants and stochastic quasigradients is quite possible.

Let's consider a problem of global stochastic optimization:

$$\min_{x \in X} [F(x) = Ef(x, \theta)],$$

or

$$\min_{x \in X} [P(x) = P\{f(x, \theta) \geq 0\}],$$

where θ is a random parameter; E is a symbol of a mathematical Expectation over θ , $f(x, \theta)$ is some continuous in x and integrable in θ function; $\theta \in \Theta$; (Θ, Σ, P) is a problem probability space, $P\{\cdot\}$ is a symbol of probability; X is a continuous or discrete set.

We shall assume, that for each θ functions $f(\cdot, \theta)$ admit tangent at points y minorants $\varphi(x, y, \theta)$ and, thus, it is implicitly assumed [7], that functions $f(\cdot, \theta)$ are maximum ones: $f(x, \theta) = \max_{y \in X} \varphi(x, y, \theta)$. Actually, we shall consider global optimization problems of the form:

$$\begin{aligned} \min_{x \in X} [F(x) = E \max_{y \in Y} \psi(x, y, \theta)], \\ \left(\min_{x \in X} [F(x) = E \min_{y \in Y} \psi(x, y, \theta)] \right), \\ \min_{x \in X} \left[P(x) = P \left\{ \max_{y \in Y} \psi(x, y, \theta) \geq 0 \right\} \right], \\ \left(\min_{x \in X} \left[P(x) = P \left\{ \min_{y \in Y} \psi(x, y, \theta) \geq 0 \right\} \right] \right), \end{aligned}$$

where Y is some finite or infinite set. These are stochastic minimax (minimin) problems.

Definition 1 [7]. Let X be a topological space, functions $F(x)$, $x \in X$, and $\varphi(x, y)$ $x \in X$, $y \in X$ are connected by conditions:

- (i) $F(x) \geq \varphi(x, y)$ for all $x \in X$ $y \in X$;
- (ii) $F(y) = \varphi(y, y)$ for all $y \in X$;
- (iii) functions $\left\{ \phi_y(x) = \varphi(x, y) \right\}_{y \in X}$ are equicontinuous in x .

Then functions $\left\{ \varphi(\cdot, y), y \in X \right\}$ are called tangent (at points y) minorants for $F(x)$.

Minorant bounds on optimal values of the objective function. We designate $F_* = \min_{x \in X} F(x)$. Let $\left\{ \varphi(x, y) \right\}_{y \in X}$ be a family of tangent at points $y \in X$ minorants for $F(x)$, $\left\{ y \in Z \subseteq X \right\}$ be some finite or infinite set of points from X . It is obvious, that function $\phi(x) = \max_{y \in Z} \varphi(x, y)$ is a minorant for $F(x)$, tangent at all points $y \in Z$, and quantity $F_1 = \min_{x \in X} \phi(x)$ is an estimate from below for F_* .

With the purpose of the solution of stochastic programming problems we shall introduce a concept of stochastic tangent minorant.

Definition 2. Functions $\left\{ \phi(\cdot, y, \theta), y \in X, \theta \in \Theta \right\}$, where Θ is a carrier of some probability space (Θ, Σ, P) , are called stochastic tangent minorants for $F(x)$

if functions $\phi(x, y, \theta)$ are measurable in θ , and for every $y \in X$ mathematical expectations $\varphi(x, y) = E\phi(x, y, \theta)$ exist and are tangent at point y minorants for $F(x)$ (in the sense of definition 1).

As stochastic **tangent minorants of a mathematical expectation function** $F(x) = Ef(x, \theta)$ it is possible to take tangent minorants of subintegral function $f(x, \theta)$.

Lemma 1. Assume that $f(\cdot, \theta)$ has tangent minorants $\phi(x, y, \theta)$:

- 1) $f(x, \theta) \geq \phi(x, y, \theta)$ for all $x, y \in X$;
- 2) $f(y, \theta) = \phi(y, y, \theta)$ for all $y \in X$;
- 3) $\phi(x, y, \theta)$ is continuous in (x, y) a.s. in θ ;
- 4) $\phi(x, y, \theta)$ is measurable in θ for all $x, y \in X$;
- 5) $|\phi(x, y, \theta)| \leq M(\theta)$ for all $x, y \in X$, $EM(\theta) < \infty$.

Then $\varphi(x, y) = E\phi(x, y, \theta)$ is a tangent minorant for $F(x) = Ef(x, \theta)$.

Here the situation is similar to the calculation of stochastic gradients of a mathematical expectation function. Finding of deterministic gradients as well as deterministic minorants of mathematical expectation functions can be problematic, however calculation and use of stochastic quasigradients and stochastic minorants is quite possible.

Tangent minorants of a probability function $P(x) = P\{f(x, \theta) \geq 0\}$ are constructed in a similar way, namely, as a tangent at point y minorant of $P(x)$ it is possible to take function $\phi(x, y) = P\{\phi(x, y, \theta) \geq 0\}$, where $\phi(x, y, \theta)$ is a tangent at point y minorant of function $f(x, \theta)$.

Tangent minorants are closely connected to maximum functions. On the one hand, for maximum functions tangent minorants are easily constructed, and on the other hand functions admitting tangent minorants are maximum functions [7].

If $f(x) = \max_{z \in Z} \psi(x, z) = \psi(x, z(x))$ is a maximum function, where $\psi(x, y)$ is continuous in x uniformly in $z \in Z$, then, obviously, function $\varphi(x, y) = \psi(x, z(y))$ is a tangent at point y minorant for $f(x)$ [7].

Let's consider some ways of construction of stochastic tangent minorants.

Tangent cones. If functions $f(x, \theta)$ are Lipschitz (Holder of degree α) with integrable in θ Lipschitz constant $L(\theta)$, then as a tangent at point y minorant for $f(x, \theta)$ it is possible to take function $\phi(x, y, \theta) = f(y, \theta) - L(\theta) \|x - y\|^\alpha$.

Tangent paraboloids. For smooth in x functions $f(x, \theta)$ with Lipschitz gradient (with a constant $L_1(\theta)$) as stochastic tangent minorants it is possible to use tangent to the graph at points y paraboloids.

Minorants of composite functions. Let $f(x) = f_0(f_1(x), \dots, f_m(x))$, $x \in X$, where X is a topological space, $f_0(z)$ is monotonously growing continuous function on the set $Y = \{f_1(x), \dots, f_m(x) \in R^m \mid y \in X\}$. Let functions $f_i(\cdot)$ $i = \overline{1, m}$, have tangent minorants $\varphi_i(x, y)$ $i = \overline{1, m}$. Then functions $\{\varphi(\cdot, y) = f_0(\varphi_1(\cdot, y), \dots, \varphi_m(\cdot, y))\}_{y \in X}$ are tangent minorants for $f(x)$ [7].

Minorants of a difference of convex functions. If function $f(x)$ representable as a difference of two convex on compact $X \in R^n$ functions $f^1(x)$ and $f^2(x)$, i.e. $f(x) = f^1(x) - f^2(x)$, then functions $\{\varphi(x, y, \theta) = f^1(y) + \langle g(y), x - y \rangle - f^2(x)\}_{y \in X}$, where $g(y)$ is a generalized gradient of function $f(\cdot)$ at point y , are concave tangent minorants for $f(x)$ on X [7].

Minorants of a minimum function. Let $f(x) = \inf_{z \in Z} \psi(x, z)$ and functions $\psi(\cdot, z)$ for all $z \in Z$ admit (concave) tangent at points y minorants $\phi(x, y, z)$. Then function $\varphi(x, y) = \inf_{z \in Z} \phi(x, y, z)$ is (concave) tangent at y minorant for $f(x)$.

Minorants of expectation function with measure depending on decisions:

$F(x) = E_x f(x, \theta) = \int_{\Theta} f(x, \theta) p_x(\theta) d\theta$ [17]. Assume that for each (tangent) point y for all X holds:

$$f(x, \theta) \geq f(y, \theta) - L(y, \theta) \|x - y\| \quad \text{and} \quad p_x(\theta) \geq p_y(\theta) - l_y(\theta) \|x - y\|.$$

Then

$$F(x) \geq E_y \left[f(y, \theta) - \left(L(y, \theta) + \frac{l_y(\theta)}{p_y(\theta)} \right) \|x - y\| \right] = E_y \varphi(x, y, \theta_y),$$

i.e. functions $\varphi(x, y, \theta_y)$ constitute stochastic tangent minorants for $F(x)$. If only density depends on x , i.e. $F(x) = E_x f(\theta) = \int_{\Theta} f(\theta) \frac{p_x(\theta)}{p_y(\theta)} dP_y \theta$, and $p_x(\theta)$ admits tangent minorants $\phi(x, y, \theta)$ then functions $\varphi(x, y, \theta) = f(\theta) \phi(x, y, \theta) / p_y(\theta)$ can be taken as stochastic tangent minorants for $F(x)$.

We approximate the original stochastic optimization problem by its empirical approximation:

$$\left[F_k(x) := (1/k) \sum_{i=1}^k f(x, \theta^i) \right] \rightarrow \min_{x \in X},$$

where θ^i are independent observations of random parameter θ . If functions $F_k(x)$ uniformly converge to $F(x) = Ef(x, \theta)$ when $k \rightarrow \infty$, then the initial problem can be solved through the sequence of uniform approximations.

Let functions $\phi(x, y, \theta)$ are tangent minorants for stochastic functions $f(x, \theta)$. Obviously, functions $\varphi_k(x, y) = (1/k) \sum_{i=1}^k \phi(x, y, \theta^i)$, constitute tangent minorants for $F_k(x)$.

Piyavskii's method [18] has been repeatedly rediscovered and is one of popular methods of deterministic global optimization. It has two equivalent forms: for optimization of maximums functions and for functions admitting the so-called tangent minorants [7]. The concept of tangent minorants is the key one for the given method. The basic problem of Piyavskii's method in a multivariate case is how to solve auxiliary approximating multiextremal problems.

The branch and bound method is one of the basic methods of discrete and global deterministic optimization. It is characterized by the way of partitioning of the initial feasible set (for example, on parallelepipeds, simplexes, etc.), a kind of estimations of optimum values of the objective function on subsets (for example, a relaxation of constraints, dual estimation, etc.), strategy of refining of the partitioning. The basic difference of different variants of the method consists in a way of obtaining estimations from below of the optimal value of the objective function on a fragment of partitioning.

A common feature of considered methods applied to a problem of global stochastic optimization is the use of a sequence of uniform approximations of the objective function and tangent minorants of these approximations. Thus, we obtain new modifications of Piyavskii's method and of the branch and bound method for the solution of so-called limit extremal problems in which objective function is optimized through a sequence of approximating functions.

We radically solve a problem of solution of auxiliary approximating problems in a multivariate Piyavskii's method, namely, we solve them not precisely (that is rather difficult) but approximately by partitioning the domain of search into subsimplexes and by searching a subsimplex with the least estimate from below of the approximating function (instead of searching points of its global minimum). Then this variant of the method, in essence, turns into a branch and bound method with minorant estimates of branches.

An important feature of the classical branch and bound method is the opportunity of rejection of unpromising branches. However it cannot be easily done if we use stochastic estimations of branches in stochastic programming problems as there is a probability of loss of a global extremum. In one of modifications of the branch and bound method we do not reject branches (subsets of partitioning) with bad estimates, but aggregate them, i.e. we come back to more rough partitioning of the domain of search, but do it no more, than finite number of times.

References

1. Ermoliev Y.M., Norkin V.I. Methods for solution of nonconvex nonsmooth stochastic optimization problems, *Kibernetika i sistemnyi analiz*, 2003, N 5, 60-81 (in Russian, English translation in *Cybernetics and Systems Analysis*, 2003, Vol. 39, Issue 5, pp. 701-715).
2. Gutjahr W.J., Hellmayr A., Pflug G.C., Optimal stochastic single-machine tardiness scheduling by stochastic branch-and-bound // *European J. of Operational Research*. - 1999. - Vol.117, N 2. - P.396-413.
3. Gutjahr W.J., Strauss C., Wagner E., A Stochastic Branch-and-Bound Approach to Activity Crashing in Project Management // *J. on Computing*. - 2000. - Vol. 12, N 2. - P.125-135.
4. Handbooks in Operations Research and Management Science, 10: Stochastic Programming/Edited by A. Ruszczyński and A. Shapiro, North-Holland: Elsevier, 2003.
5. Hägglöf K. The Implementation of the Stochastic Branch and Bound Method for Applications in River Basin Water Quality Management // Working paper WP-96-89. - Int. Inst. for Appl. Syst. Analysis. Laxenburg, Austria, 1996. - 13 p.
6. Lence B.J., Ruszczyński A. Managing Water Quality under Uncertainty: Application of a New Stochastic Branch and Bound Method // Working paper WP-96-066. - Int. Inst. for Appl. Syst. Analysis. Laxenburg, Austria, June 1996. - 18 p.
7. Norkin V.I. On Piyavski's method for solving general global optimization problem, *Zhurnal Vychislitel'noj Matematiki i Matematicheskoy Fiziki*, 1992, N 7, 992-1006 (in Russian, English translation in *Comp. Maths and Math. Phys.*, 1992, N 7, 873-886).
8. Norkin V.I. Global Stochastic Optimization: Branch and Probabilistic Bound Method, In *Methods of Control and Decision-Making under Risk and Uncertainty*, Ed. Yu. M.Ermoliev, Glushkov Institute of Cybernetics, Kiev, 1993, 3-12 (In Russian).
9. Norkin V.I. Global Optimization of Probabilities by the Stochastic Branch and Bound Method // Stochastic optimization: Numerical methods and technical applications. Proceedings of 3rd GAMM/IFIP Workshop (Neubiberg/Munchen, June 17-20, 1996). - Lecture Notes in Economics and Mathematical Systems 458, Berlin, Springer, 1998. - P.186-201.
10. Norkin V., Ermoliev Yu. M., Ruszczyński A. On optimal allocation of indivisibles under uncertainty // Working paper WP-94-021. - Int. Inst. for Appl. Syst. Analysis. Laxenburg, Austria, 1994.
11. Norkin V., Pflug G.Ch., Ruszczyński A. A branch and bound method for stochastic global optimization // *Math. Progr.* - 1998. - Vol. 83. - P. 425-450.
12. Norkin V., Ermoliev Yu. M., Ruszczyński A. On optimal allocation of indivisibles under uncertainty // *Operations Research*. - 1998. - Vol. 46. - № 3. - P. 381-395.
13. Norkin V.I., Onischenko B.O. On stochastic analogue of Piyavski's global optimization method, *Teoria optimalnyh risheniy (Theory of optimal decisions)*, Ed. N.Z.Shor, Glushkov Institute of Cybernetics, Kiev, 2003, No.2, pp.64-70.
14. Norkin V.I., Onischenko B.O. A branch and bound method with minorant estimates used to solve stochastic global optimization problems, *Komputernaya matematika (Computer mathematics)*, Institute of Cybernetics, Kiev, 2004, N 1, pp. 91-101.
15. Norkin V.I., Onischenko B.O. On the global minimization of minimum functions by the minorant method, *Teoria optimalnyh risheniy (Theory of optimal decisions)*, Ed. N.Z.Shor, Glushkov Institute of Cybernetics, Kiev, 2004, No. 3, pp. 56-63.
16. Norkin V.I., Onischenko B.O. Minorant methods of stochastic global optimization // *Kibernetika i sistemnyi analiz* (translated into English as *Cybernetics and Systems Analysis*) (in print).
17. Norkin V.I., Onischenko B.O. On a branch and approximate bounds method, *Komputernaya matematika (Computer mathematics)*, Institute of Cybernetics, Kiev, 2005, N 1.

18. Pijavski S.A. On an algorithm for searching absolute minimum of a function // Zhurnal vychisl. matem. I matem. fiziki. - 1972. - T. 12, No. 4. - P. 888-896 (in Russian, English translation in *Comp. Maths and Math. Phys.*).