

ON MINORANTS METHODS OF STOCHASTIC GLOBAL OPTIMIZATION

Onishchenko B.O.

Glushkov Institute of Cybernetics NAS of Ukraine

boris@cdtu.edu.ua

Outline

- Problem of stochastic global optimization.
- Stochastic tangents minorants.
- Ways of construction of stochastic tangents minorants.
- Method of successive empirical approximations.
- Modifications of Pijavskii's method.
- Stochastic branch and bound method.
- Numerical experiments.

References:

- Norkin V.I., Onishchenko B.O. On stochastic analogue of a Pijavskii's method of global optimization // The Theory of Optimum Decisions. – Kiev: Ins-t of Cybernetics NAS of Ukraine, 2003.
- Norkin V.I., Onishchenko B.O. The branch and bound method with minorants estimations for the decision of problems of stochastic global optimization // Computer mathematics. – Kiev: Ins-t of Cybernetics NAS of Ukraine, 2003.
- Pijavskii (1967); Danilin, Pijavskii (1967); Norkin (1992); Horst, Tuy (1996).

Problem of stochastic global optimization.

Let's consider a problem **of stochastic global optimization**:

$$\min_{x \in X} [F(x) = Ef(x, \theta)], \quad (1)$$

where θ – casual parameter, E – a symbol of a mathematical expectation on θ , $f(x, \theta)$ – some continuous on x and integrated on θ function, $\theta \in \Theta$, (Θ, Σ, P) – probability space of a problem, X – continuous or discrete set.

Definition 1. Let X – topological space, functions $F(x)$, $x \in X$, also $\varphi(x, y)$, $x \in X$, $y \in X$, 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) function $\varphi(x, y)$ is continuous on (x, y) .

Then functions $\{\varphi(\cdot, y), y \in X\}$ refer to as **tangents** (in points y) **minorants** for $F(x)$.

Stochastic tangents minorants.

Definition 2. Functions $\{\varphi(\cdot, y, \theta), y \in X, \theta \in \Theta\}$ where Θ – the carrier of the some probability space (Θ, Σ, P) , refer to **as stochastic tangents minorants** for $F(x)$ if functions $\varphi(x, y, \theta)$ are measurable on θ , and mathematical expectation $\varphi(x, y) = E\varphi(x, y, \theta)$ are final and for everyone $y \in X$ are tangents minorants for $F(x)$.

Lemma 1. We shall assume, that functions $f(\cdot, \theta)$ supposes *tangents minorants* $\varphi(x, y, \theta)$ in points $y \in X$, i.e. almost for all θ it is executed:

- 1) $f(x, \theta) \geq \varphi(x, y, \theta)$, for all $x \in X, y \in X$;
- 2) $f(y, \theta) = \varphi(y, y, \theta)$, for all $y \in X$;
- 3) function $\varphi(x, y, \theta)$ is continuous on (x, y) almost for all θ ;
- 4) $\varphi(x, y, \theta)$ – it is measurable on θ for anyone $x, y \in X$;
- 5) $|\varphi(x, y, \theta)| \leq M(\theta)$, for all $x, y \in X$ with some integrated function $M(\theta)$.

Then functions $\varphi(x, y) = E\varphi(x, y, \theta)$ are **tangents minorants** for function of a mathematical expectation $F(x) = Ef(x, \theta)$.

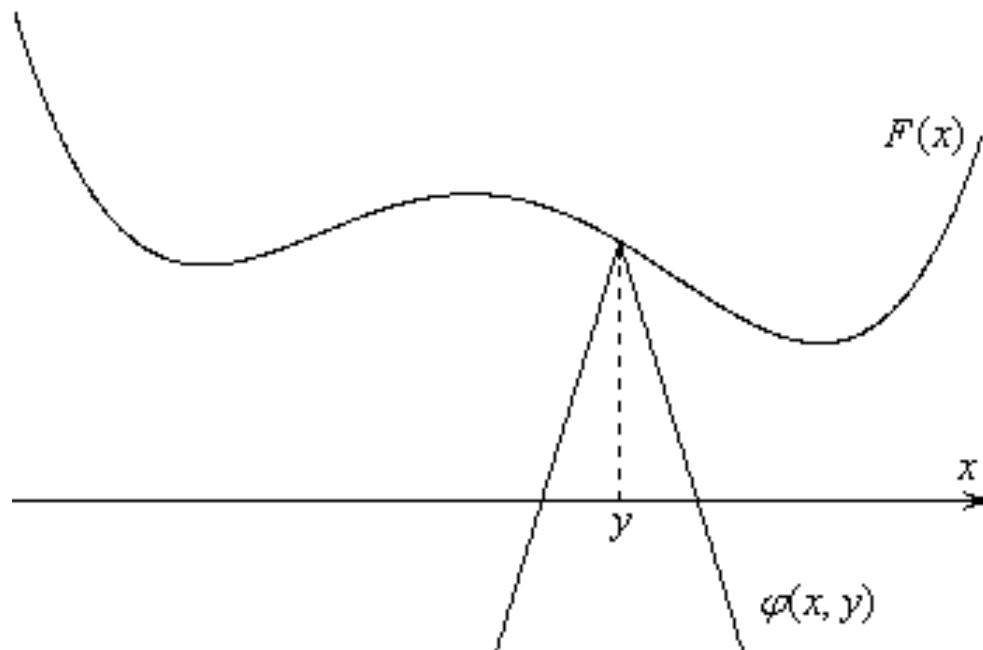
Ways of construction of stochastic tangents minorants (tangents cones).

1. If functions $f(x, \theta)$ Lipschitz (Gelder's) with integrated on θ constant of Lipschitz $L(\theta)$ and a parameter α :

$$|f(x^1, \theta) - f(x^2, \theta)| \leq L(\theta) \|x^1 - x^2\|^\alpha \quad \forall x^1, x^2 \in X \quad 0 < \alpha \leq 1,$$

then as a **stochastic tangent in a point y minorant** for $f(x, \theta)$ is possible to take function

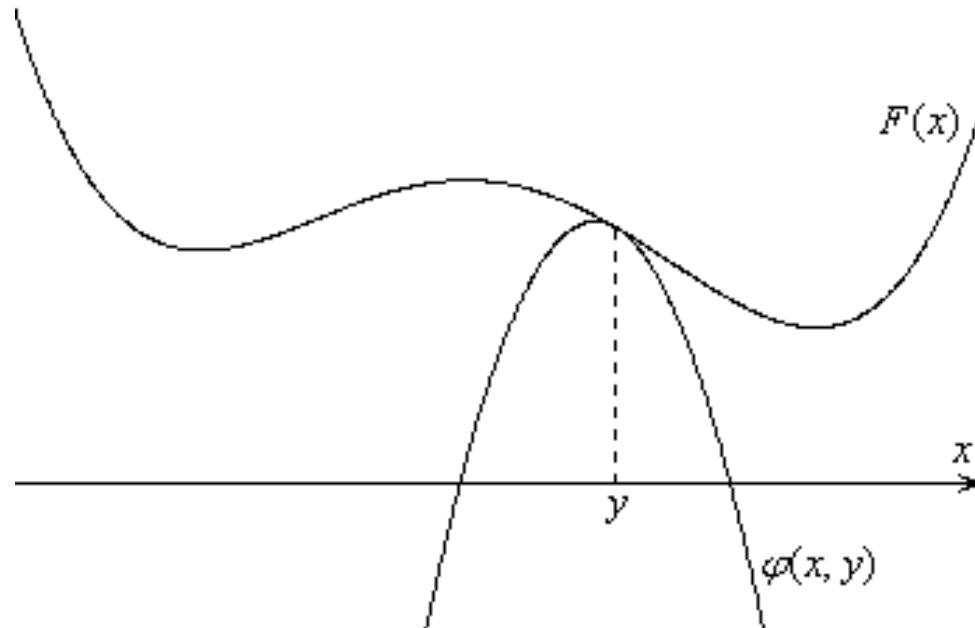
$$\varphi(x, y, \theta) = f(y, \theta) - L(\theta) \|x - y\|^\alpha.$$



Ways of construction of stochastic tangents minorants (tangents paraboloids).

2. For smooth on x functions $f(x, \theta)$ with Lipschitz' gradient (with a constant $L(\theta)$) as **stochastic tangents minotants** it is possible to use tangents to $f(\cdot, \theta)$ in points y paraboloids:

$$\varphi(x, y, \theta) = f(y, \theta) + \frac{1}{2L(\theta)} \|\nabla f(y, \theta)\|^2 - \frac{L(\theta)}{2} \left\| x - y - \frac{1}{L(\theta)} \nabla f(y, \theta) \right\|^2.$$



In these examples various norms are possible, for example, for n -dimensional vector z it is possible to use norm $\|z\| = \left(\sum_{i=1}^n z_i^2 \right)^{1/2}$ or $\|z\| = \max_{1 \leq i \leq n} |z_i|$.

Method of successive empirical approximations.

We approximate a problem (1) using empirical average:

$$\min_{x \in X} [F_N(x) = (1/N) \sum_{k=1}^N f(x, \theta^k)], \quad (2)$$

where θ^k - independent supervision of casual parameter θ . If functions $F_N(x)$ in regular intervals converge to $F(x) = Ef(x, \theta)$, instead of an initial problem (1) it is possible to solve the approached problem (2). The following statements prove such approach.

Theorem 1 (LeCam (1953)). Let function $f(x, \theta)$ is continuous on x for almost all $\theta \in \Theta$ and is measurable on θ for everything $x \in X$, X – compact in R^n . We shall assume, that the family $\{f(x, \cdot), x \in X\}$ are integrable in regular intervals. Then with probability of 1 function $F_N(x)$ in regular intervals on X converge to $F(x) = Ef(x, \theta)$.

Functions $\varphi_N(x, y) = (1/N) \sum_{k=1}^N \varphi(x, y, \theta^k)$, obviously, are **tangents minorants** for $F_N(x)$.

It is important to notice, that in the given method it is possible to use as **the fixed values** N , and gradually **to increase** its value (i.e. $N \rightarrow \infty$).

Modifications of Pijavskii's method.

Let we have a problem without the general restrictions:

$$F(x) \rightarrow \min_{x \in X}. \quad (3)$$

The point $y^0 \in X$ is any; $\varphi_0(x, t) = \varphi(x, y^0)$. Let points $y^1, \dots, y^k \in X$ are already constructed.

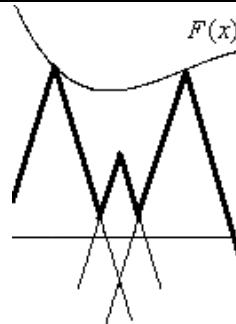
1) Classical Pijavskii's method:

$$\varphi_k(x) = \max \{ \varphi_{k-1}(x), \varphi(x, y^k) \}; \quad (4)$$

2) Modification 1:

$$\varphi_k(x, t) = \max \{ \varphi_{k-1}(x), \varphi(x, y^k) - t(F(y^k) - \varphi_{k-1}(y^k)) \}, \quad (5)$$

where $0 \leq t \leq \bar{t} < 1$. The form of components minorants here does not vary;



3) Modification 2:

$$\begin{aligned} \phi^k(x) &= \max_{0 \leq i \leq k} \varphi(x, y^i), \\ \varphi_k(x, t) &= (1-t)\varphi_{k-1}(x) + t\phi^k(x), \end{aligned} \quad (6)$$

where $0 < \underline{t} \leq t \leq 1$. Here used minorants are deformed.

Point y^{k+1} , $k \geq 0$, we shall find as the decision of a special multiextreme problem:

$$\varphi_k(x) := \varphi_k(x, t) \rightarrow \min_{x \in X}. \quad (7)$$

It is obvious, that both modifications of a method generally use **not tangent minorants**.

All modifications, as well as classical variant of a method, is possible to use for the decision **of stochastic problems of global optimization**, using **a method of successive empirical approximations**.

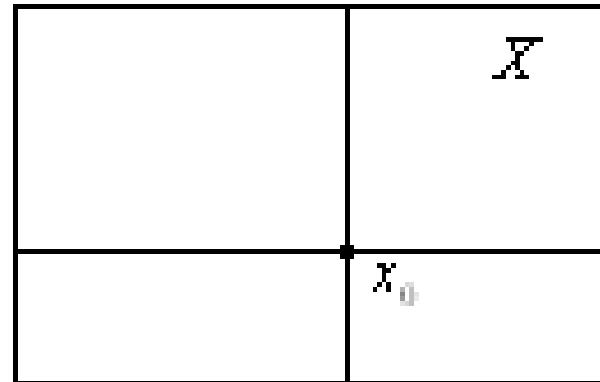
Stochastic branch and bound method.

Let we have a problem without the general restrictions:

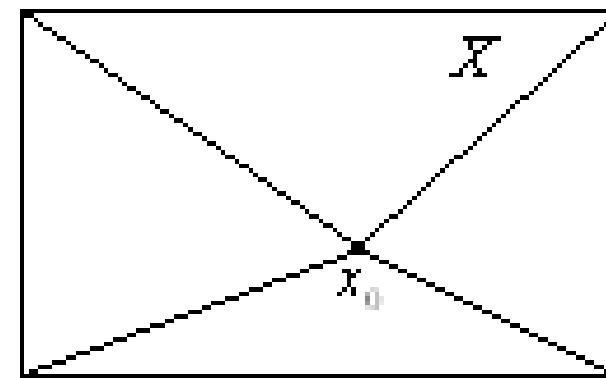
$$F(x) \rightarrow \min_{x \in X}, \quad X = [a, b]^N \quad (8)$$

Ways of initial splitting of set X :

1) Parallelepipeds;



2) Simplexs.

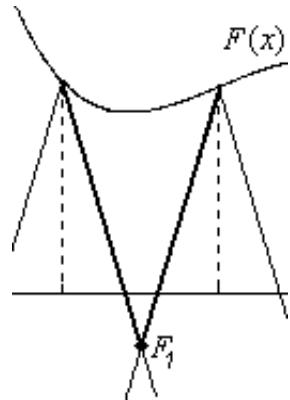


Let $\{\varphi(x, y)\}_{y \in X}$ – the family of tangents in points $y \in X$ minorants for $F(x)$,
 $\{y \in Z \subseteq X\}$ – final or infinite set of points from X .

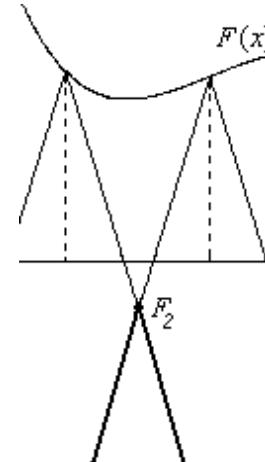
Stochastic branch and bound method.

Minorants bounds of optimum values of criterion function on a fragment of splitting.

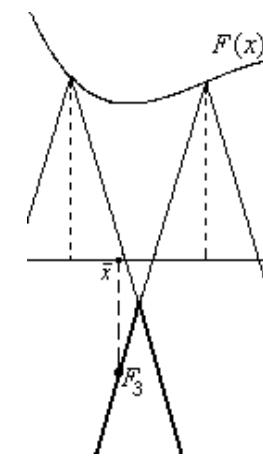
Let's consider some estimations from below for function $F(x)$:



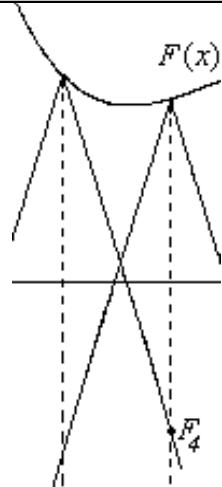
$$F_1 = \min_{x \in X} \max_{y \in Z} \varphi(x, y)$$



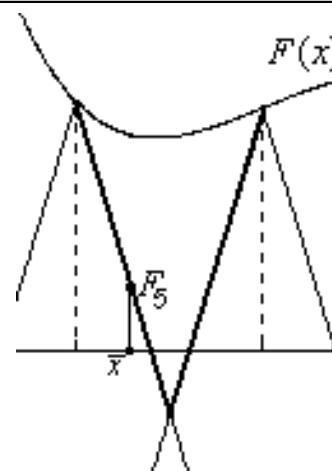
$$F_2 = \max_{x \in X} \min_{y \in Z} \varphi(x, y)$$



$$F_3 = \min_{y \in Z} \varphi(\bar{x}, y)$$



$$F_4 = \max_{y \in Z} \min_{x \in X} \varphi(x, y)$$



$$F_5 = \max_{y \in Z} \varphi(\bar{x}, y)$$



$$F_6 = \min_{x \in Z} \{\varphi(x, \bar{y})\}$$

Stochastic branch and bound method.

Specific minorants bounds of optimum values for function of a mathematical expectation.

If to take into account, that $F(x) = Ef(x, \theta)$ $\varphi(x, y) = E\varphi(x, y, \theta)$ it is possible to construct specific estimations from below for function $F(x)$:

- $\Phi_1 = \min_{x \in X} E \max_{y \in X} \varphi(x, y, \theta);$
- $\Phi_2 = E \min_{x \in X} \max_{y \in X} \varphi(x, y, \theta);$
- $\Phi_3 = E \max_{y \in X} \min_{x \in X} \varphi(x, y, \theta);$
- $\Phi_4 = E \max_{x \in X} \min_{y \in X} \varphi(x, y, \theta);$
- $\Phi_5 = E \min_{y \in X} \varphi(\bar{x}, y, \theta), \bar{x} \in X.$

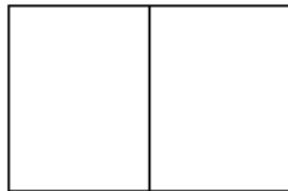
Apparently from the submitted estimations if criterion function is **function of a mathematical expectation** of some function entering operations of a capture of a maximum or a minimum **under a sign on a mathematical expectation** is possible.

Stochastic branch and bound method.

Ways of making small a fragment of the splitting having the least estimation from below.

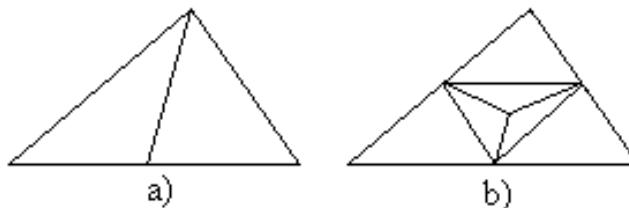
At making small a "record" fragment of splitting it is necessary to aspire its diameter to zero. For performance of the given condition is possible to use the next ways of making small:

- 1) initial splitting by parallelepipeds – edges of the greatest length of a parallelepiped share in half, and is broken into two new parallelepipeds;



- 2) initial splitting by simplexes:

- a) edge of the greatest length of a simplex share in half, and is broken into two new simplexes;
- b) all edges of a simplex share in half, also there is a centre of gravity of middle edges, and is broken on new simplexes with participation of all received points.



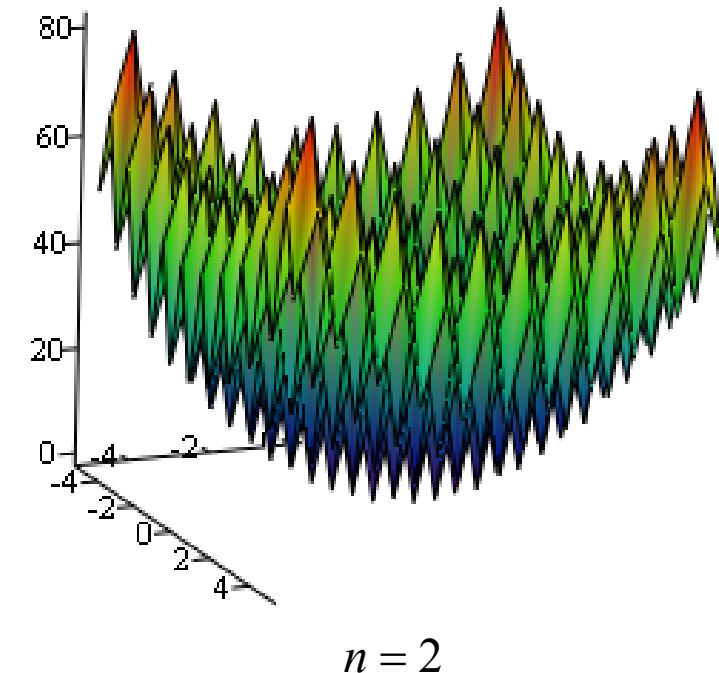
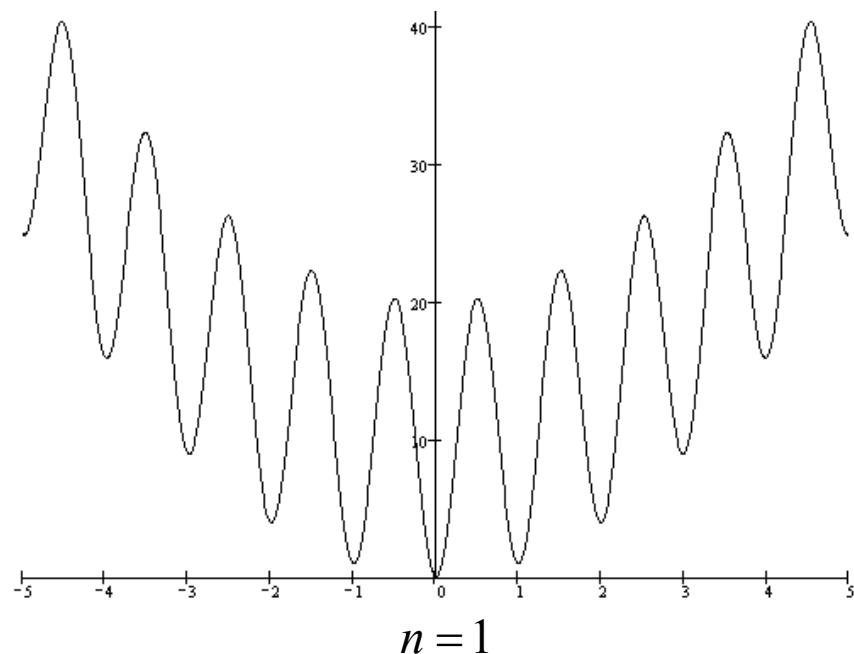
Numerical experiments.

Test function.

For experiments the following test function was used:

$$f(x) = 10n + \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i)),$$
 where n – dimension of space $x \in [-5; 5]$, with

estimations of constants of Lipschits for function $l = 72$ and its gradient $L = 397$; $x^* = 0$, $f(x^*) = 0$.



Numerical experiments.

Modifications of Pijavskii's method.

TABLE 1. Results of work of classical Pijavskii's method (4).

Accuracy	10-4	10-3	10-2	10-1
Paraboloids	47	43	39	37
Cones	1390	451	198	92

Table 2 shows, that the using of tangents of paraboloids conducts to essential reduction of quantity of the iterations, which are necessary to achieve set accuracy.

TABLE 2. Results of work of modified Pijavskii's algorithm (5).

Number of iteration	Quantity of minorants	Accuracy	Number of iteration	Quantity of minorants	Accuracy
10	8	543,44	70	38	28,14
20	12	122,01	80	43	7,49
30	19	59,53	90	48	0,26
40	23	23,34	100	52	0,0008
50	27	13,26	105	56	0,00006
60	33	37,51			

Table 2 shows, that the using of not concerning minorants results in rejection (full majorization) some components of minorant, that in turn simplifies search of a minimum resulting minorant.

Numerical experiments.

Branch and bound method.

TABLE 3. Results of work of algorithm of a branch and bound method.

Dimension of space	Way of splitting	Kind of minorants	Characteristics of algorithm	Kind of an estimation from below				
				F_3	F_4	F_5	F_6	
1	1, 2a, 2b	Paraboloid	Iterations	47	67	25	59	
			Left simplexs	3	2	3	3	
		Cone	Iterations	197	293	93	163	
			Left simplexs	71	141	58	75	
2	1	Paraboloid	Iterations	1245	3184	618	1612	
	2a		Left simplexs	18	41	7	21	
	2b		Iterations	2555	4531	517	2847	
			Left simplexs	27	55	8	49	
			Iterations	2271	3707	269	2661	
			Left simplexs	22	67	17	69	

Table 3 shows, that the most effectively method works using the heuristic estimation from below $F_5 = \max_{y \in Z} \varphi(\bar{x}, y)$.

Numerical experiments.

Stochastic global optimization.

For testing stochastic algorithms function was used:

$$f(x, \theta) = (x - 1 - \theta)(x - 3 - \theta)(x - 7 - \theta)(x - 11 - \theta),$$

where θ – random variable in regular intervals distributed on interval $[0;1]$, $x \in [1;11]$.

It is easy to calculate, that

$$F(x) = Ef(x, \theta) = x^4 - 24x^3 + 187x^2 - 537x + \frac{14051}{30}.$$

The determined estimations of constants of Lipschitz $l = 537$, $L = 374$ of functions $f(x, \theta)$ and its gradient. $x^* = 9.977$; $F(x^*) = -201.667$.

We approximate $F(x)$ using the **empirical function** $F_N(x) = (1/N) \sum_{i=1}^N f(x, \theta_i)$, where θ_i – the independent random variables in regular intervals distributed on interval $[0;1]$.

TABLE 4. Results of work of algorithm of a branch and bound method at $N = 300$.

Kind of minorants	Characteristic	Kind of an estimation from below					
		F_3	F_4	Φ_3	Φ_5	F_5	F_6
Paraboloid	Iterations	20	32	31	20	19	26
Cone	Iterations	1991	3935	3886	2041	1961	1973

Table 4 shows, that estimations operations of a capture of a maximum using entering or a minimum under a sign on a mathematical expectation works not less effectively, than estimations of the function of a mathematical expectation.

Numerical experiments.
Stochastic global optimization.

TABLE 5. An assessment of works of algorithm, at N proportional to number of iterations.

Iterations	F_3		F_4		Φ_3	
	Algorithmic accuracy	Accuracy	Algorithmic accuracy	Accuracy	Algorithmic accuracy	Accuracy
10	22.645	11.656	321.518	186.017	321.518	186.017
20	0.095	2.091	4.221	0.842	3.423	0.842
50	0.049	0.715	0.056	0.709	0.018	0.678
100	0.001	0.452	0.000	0.452	0.006	0.445
200	0.005	0.412	0.069	0.359	0.055	0.343
300	0.009	0.335	0.009	0.335	0.011	0.300
	Φ_5		F_5		F_6	
10	22.645	11.656	6.426	1.920	35.065	11.656
20	0.095	2.091	0.117	2.093	0.057	2.007
50	0.049	0.715	0.049	0.715	0.054	0.709
100	0.001	0.452	0.001	0.452	0.001	0.452
200	0.073	0.356	0.073	0.356	0.068	0.359
300	0.009	0.335	0.009	0.335	0.009	0.335

Table 5 shows, that the use of successive empirical approximations in a stochastic branch and bound method is probably to use not fixed number of supervision of a random variable, and gradually to increase it with growth of number of iterations.