CYBERNETICS
and COMPUTER
TECHNOLOGIES

UDC 519.85

DOI:10.34229/2707-451X.20.1.1

V.I. NORKIN

A STOCHASTIC SMOOTHING METHOD FOR NONSMOOTH GLOBAL OPTIMIZATION

The paper presents the results of testing the stochastic smoothing method for global optimization of a multiextremal function in a convex feasible subset of the Euclidean space. Preliminarily, the objective function is extended outside the admissible region so that its global minimum does not change, and it becomes coercive. The smoothing of a function at any point is carried out by averaging the values of the function over some neighborhood of this point. The size of the neighborhood is a smoothing parameter. Smoothing eliminates small local extrema of the original function. With a sufficiently large value of the smoothing parameter, the averaged function can have only one minimum. The smoothing method consists in replacing the original function with a sequence of smoothed approximations with vanishing to zero smoothing parameter and optimization of the latter functions by contemporary stochastic optimization methods. Passing from the minimum of one smoothed function to a close minimum of the next smoothed function, we can gradually come to the region of the global minimum of the original function. The smoothing method is also applicable for the optimization of nonsmooth nonconvex functions. It is shown that the smoothing method steadily solves test global optimization problems of small dimensions from the literature.

Keywords: *global optimization; Steklov smoothing; averaged functions; stochastic optimization; nonsmooth nonconvex optimization*.

Introduction. Function smoothing has long been used in computational mathematics [1]. In the works [2–7] the Steklov local smoothing was used to study and optimize nonsmooth nonconvex functions. In the works [6, 8 – 11] it is proposed to use function smoothing for global optimization purposes.

The smoothing method in global function optimization consists in replacing (approximation) of the original multi-extreme function with a sequence of so-called smoothed (or averaged) functions and optimization the latter with one or another method [6, p. 135–137], [9]. As the smoothing parameter decreases, the averaged functions uniformly converge to the original function, so the global minima of the approximate functions converge to the global minima of the original function. Concerning global optimization, the idea behind the method is that smoothing eliminates small local minima and little changes deep minima.

In the works [10, 11] the behavior of critical points of smoothed functions when changing the smoothing parameter is described by a system of ordinary differential equations. If we take the minimum of a strongly smoothed function as the starting point, then having solved the equations, we can come to the global minimum of the original function. This approach is applied to the global optimization of functions given by analytic expressions, for example, for polynomial functions.

In this paper, a stochastic smoothing method is developed, in which the sequence of smoothed functions is minimized by contemporary stochastic optimization algorithms. The method is applicable for global optimization, in general, non-smooth functions, as well as for solving problems with constraints. The results of testing the method on numerous examples of multi-extremal functions from the literature are presented. This method confidently finds global optimums of small-scale multi-extremal functions. As the dimension of the space increases, the time for solving problems increases significantly, which is due to the need for multiple estimations of multidimensional integrals by the Monte Carlo method.

The problem setting. Suppose that the problem of conditional global optimization is solved:

$$f(x) \to \min_{x \in D \subseteq R^n}, \tag{1}$$

where $f(x)$ is a continuous on a closed set $D \subseteq R^n$ function such that $f(x) \to +\infty$ for $\|x\| \to +\infty$ and $x \in D$; R^n is an n -dimensional arithmetic Euclidean space; $\|\cdot\|$ is some norm in R^n.

There are several ways to reduce the conditional optimization problem (1) to an unconditional optimization problem.

For example, if $D = \{x \mid g_j(x) \le 0, j = 1,..., J; h_k = 0, k = 1,..., K\}$, then in the exact penalty function method the Lipschitz function $f(x)$ is replaced by

$$F(x) := f(x) + M\left(\sum_j \max\{0, g_j(x)\} + \sum_k |h_k(x)|\right)$$

with a sufficiently large penalty parameter M and then consider the problem of unconditional optimization of $F(x)$. In this approach, the problem is in the correct choice of the penalty parameter M.

If some internal point x_0 of a convex set D is known, then the exact penalty function can be constructed as follows. Let $y(x)$ be the point of intersection of the line connecting x_0 and x with the boundary of the set D. Then the penalty function can be taken as

$$F(x) = \begin{cases} f(x), & x \in D, \\ f(y(x)) + \|x - y(x)\|, & x \notin D. \end{cases}$$

Another way to construct a penalty function for a convex set D is as follows. Denote by $\pi_D(x)$ the projection of a point x on the set D, and by $\rho_D(x)$ the distance between x and D, i.e. $\pi_D(x) = \arg\min_{y \in D} \|x - y\|$ and $\rho_D(x) = \min_{y \in D} \|y - x\|$. For a simple set D, the projection search problem is either solved analytically or reduced to the quadratic programming problem. Consider the following exact continuous penalty function

$$F(x) := f(\pi_D(x)) + \rho_D(x).$$

Now, instead of problem (1), let us consider the following equivalent unconditional global optimization problem:

$$F(x) \to \min_{x \in R^n}$$

where $F(x) \to +\infty$ as $\|x\| \to +\infty$.

Averaged functions and their gradients. Let $\mu(x)$ be some probabilistic kernel in R^n, i.e., $\mu(x) \ge 0$ and $\int_{R^n} \mu(x) dx = 1$. Along with function $F(x)$, we consider the so-called smoothed functions

$$F_h(x) = \frac{1}{h^n} \int_{R^n} F(x + y) \mu(y/h) dy = \int_{R^n} F(x + hz) \mu(z) dz = \frac{1}{h^n} \int_{R^n} F(y) \mu\left(\frac{y - x}{h}\right) dy,$$

(it is assumed that the integrals exist), where $h > 0$ is the smoothing parameter. Properties of smoothed functions were studied in [2, 3, 5 – 7]. If function $F(x)$ is locally Lipschitz, then $F_h(x)$ is continuously differentiable and its gradient $\nabla F_h(x)$ can be represented as

$$\nabla F_h(x) = \frac{1}{h^n} \int_{R^n} \partial F(x + y) \mu(y/h) dy = \int_{R^n} \partial F(x + hz) \mu(z) dz \tag{2}$$

(assuming that the integrals exist), where $\partial F(y)$ is the Clark subdifferential of function $F(\cdot)$ at point y. The integral of a multi-valued mapping is understood here as the set of integrals of measurable single-valued selections of this mapping [6, 12].

If the kernel function $\mu(\cdot)$ is differentiable, then gradients of the smoothed function $F_h(x)$ can also be represented as (assuming that the integrals exist),

$$\nabla F_h(x) = -\frac{1}{h^{n+1}}\int_{R^n} F(y)\nabla\mu\left(\frac{y-x}{h}\right)dy = -\frac{1}{h}\int_{R^n} F(x+hz)\nabla\mu(z)dz . \quad (3)$$

Steklov smoothing [1]. Denote $V_h(x)$ some solid h-neighborhood of point x, for example, $V_h(x)=\left\{y\in R^n \,|\, \|y-x\|\le h\right\}$, such that $V_h(x)\to x$ as $h\to 0$. Here $\|\cdot\|$ is some norm in space R^n, for example, $\|x\|=\max_{1\le i\le n}|x_i|$. Let v_h be the volume of the set $V_h(x)$, $S_h(x)$ be the surface of the set $V_h(x)$, $N(y)$ be the external normal to $S_h(x)$ at point $y\in S_h(x)$. Consider the so-called smoothed (or averaged) functions of the following form:

$$F_h(x)=\frac{1}{v_h}\int_{V_h(x)} F(y)dy .$$

The gradient $\nabla F_h(x)$, when $F(x)$ continuous, is calculated by the surface integral

$$\nabla F_h(x)=\int_{S_h(x)} F(y)N(y)dS . \quad (4)$$

If the function $F(\cdot)$ is Lipschitz with a subdifferential $\partial F(\cdot)$, then its gradient can be represented as

$$\nabla F_h(x)=\frac{1}{v_h}\int_{V_h(x)} \partial F(y)dy . \quad (5)$$

If the neighborhood $V_h(x)$ has the form of hypercube,

$$V_h(x)=\left\{y=(y_1,...,y_n)\in R^n \,|\, \max_{1\le i\le n}|y_i-x_i|\le h\right\},$$

then Steklov's function

$$F_h(x)=\frac{1}{h^n}\int_{x_1-h/2}^{x_1+h/2}...\int_{x_n-h/2}^{x_n+h/2} F(y)dy_1...dy_n$$

is differentiable and its partial derivatives have the form [1, 2, 6].

$$\frac{\partial F_h(x)}{\partial x_i}=\frac{1}{h^{n-1}}\int_{x_1-h/2}^{x_1+h/2}...\int_{x_{i-1}-h/2}^{x_{i-1}+h/2}\int_{x_{i+1}-h/2}^{x_{i+1}+h/2}...\int_{x_n-h/2}^{x_n+h/2}\Delta_h(x)dy_1...dy_{i-1}dy_{i+1}...dy_n , \quad (6)$$

where $\Delta_h(y)=h^{-1}\left(F(y_1,...,y_{i-1},x_i+h/2,y_{i+1},...,y_n)-F(y_1,...,y_{i-1},x_i-h/2,y_{i+1},...,y_n)\right)$.

Critical points of the smoothed functions. The critical points of a function $F_h(x)$ are those points x at which the gradient $\nabla F_h(x)$ vanishes. Consider the structure of the set of critical points of function $F_h(x)$ in space $\left\{(x,h)\,|\,x\in R^n,\,h\ge 0\right\}$. The picture is especially transparent in the one-dimensional case [6, p. 135–137], [9].

Let $F(x)$ depends on a scalar variable x and $F(x)\to+\infty$ for $|x|\to+\infty$. Consider smoothed functions $(h\ge 0)$

$$F_h(x) = \frac{1}{h}\int_{x-h/2}^{x+h/2} f(y)dy, \qquad F_0(x) = F(x).$$

Their derivatives in this case have the form $\frac{dF_h(x)}{dx} = \left[F\left(x+\frac{h}{2}\right) - F\left(x-\frac{h}{2}\right)\right]h^{-1},\ h>0$.

Let us construct a picture of stationary points of smoothed functions $F_h(x)$ in the plane (x,h), i.e., construct a set $T = \left\{(x,h) \mid \frac{dF_h(x)}{dx} = 0\right\}$.

Note that if $F(x^1) = F(x^2)$ $\left(x^1 \neq x^2\right)$, then $\left((x^1+x^2)/2, |x^2-x^1|\right) \in T$. Thus, to construct the set T, it suffices to trace all segments parallel to the x axis whose ends lie on the graph of $F(x)$. By continuously moving these segments up or down and monitoring their length, you can easily build T.

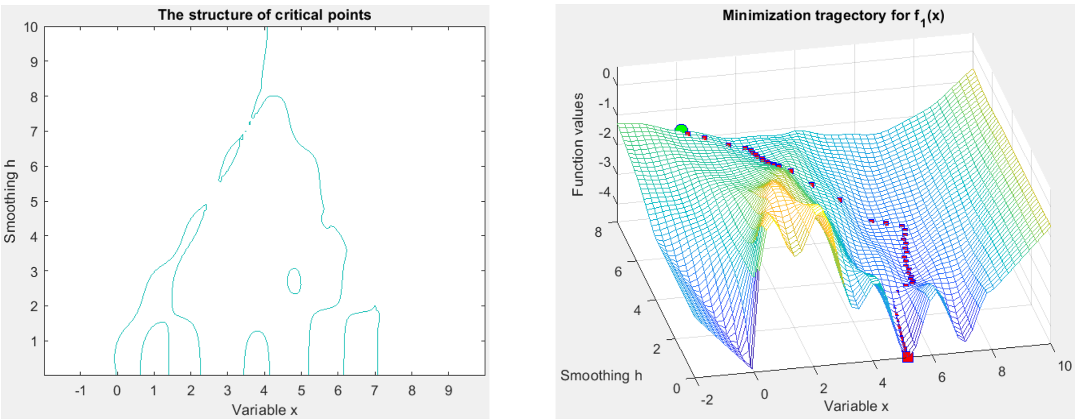
The picture of the set T is most transparent when all extrema of $F(x)$ are different. Then T consists of continuous disjoint lines or bands associated with extrema on the axis x. Moreover, any line starting at a local extremum $\left(x^*,0\right)$ (on the axis x) is bounded and locked to another local extremum $\left(x^{**},0\right)$ (on the axis x). The critical lines do not stick at stationary non-extreme points. And only the line starting at the global minimum $(x_{\min},0)$ goes to infinity in the plane (x,h).

If not all extrema of $F(x)$ are different, then the picture of T is somewhat more complicated (see fig. 1, a, 2, a). However, the set T still splits into connected bounded components, in particular, connecting local extrema, and an unlimited connected component connecting all global minima of $F(x)$ and infinity, as well as the main maxima between neighboring global minima.

Let us illustrate the said with examples.

Example 1. Consider the function of one variable $f_1(x) = \sin x + \sin(10x/3) + \ln x - 0.84x$ and the corresponding penalty function

$$F_1(x) = \begin{cases} f_1(0.01) + 10(0.01 - x), & x < 0.01; \\ f_1(x), & 0.01 \le x \le 8; \\ f_1(8) + 10(x-8), & x > 8. \end{cases}$$



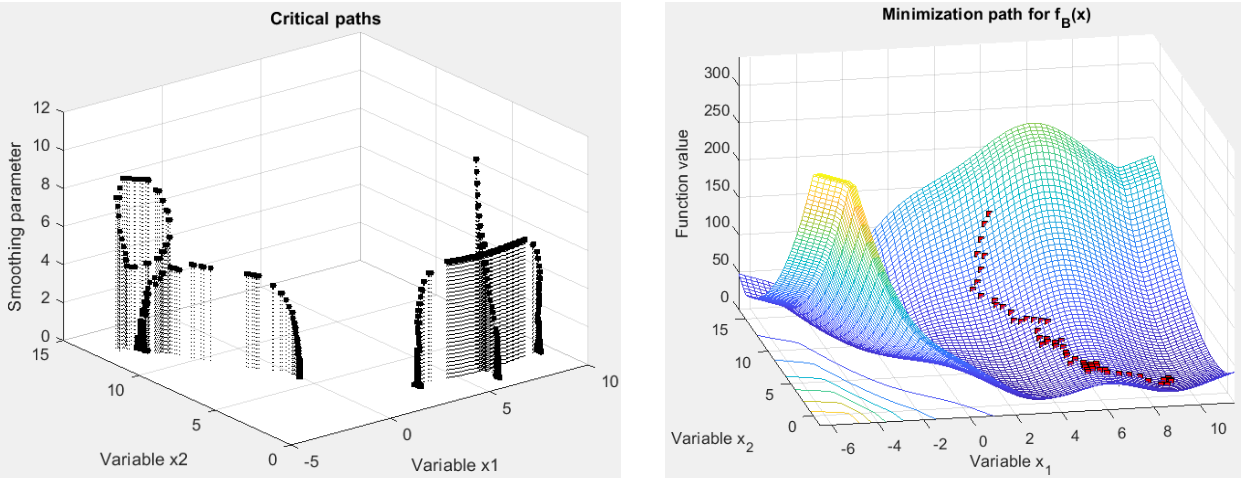
a b

FIG. 1, a – the structure of the critical set $\left\{(x,h) \mid 0 \in \partial_x F_{1h}(x)\right\}$; b – the trajectory of stochastic smoothing method when global minimizing $F_1(x)$

Example 2. Consider the function of two variables

$$f_B(x) = \left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos x_1 + 10,\quad X = \{(x_1, x_2) \mid -5 \le x_1 \le 10, 0 \le x_2 \le 15\},$$

and the corresponding penalty function $F_B(x) = \begin{cases} f_B(x), & x \in X; \\ f_B(\pi_X(x)) + \|x - \pi_X(x)\|, & x \notin X. \end{cases}$



a b

FIG. 2, a – the structure of the critical set $\{(x,h) \mid 0 \in \partial_x F_{Bh}(x)\}$; b – the trajectory of the stochastic smoothing method when minimizing $F_B(x)$

The smoothing method in global optimization. The smoothing method in global optimization consists in replacing (approximating) the original multiextremal function by a sequence of the so-called smoothed (or averaged) functions and minimizing the latter by some optimization method [6, p. 135–137], [8]. When decreasing the smoothing parameter, the average functions uniformly converge to the original function, so the global minima of the approximate functions converge to the global minima of the original function. The smoothing operation eliminates shallow local minima and little changes the global ones.

By minimizing $F_h(x)$ with a sufficiently large smoothing parameter, we obviously fall on a stationary line from T leading to a global minimum. However, it is not easy to go along this line to a global extremum, since it can behave very bizarrely, in particular, it can have kinks. Points on this line can be local minima, local maxima, or just stationary over x for the corresponding smoothed functions $F_h(x)$.

To minimize smoothed functions $F_h(x)$, any smooth optimization methods can be applied if gradients of $F_h(x)$ can be calculated analytically or numerically. This is easy to do for one-dimensional functions and for separable functions $F(x)$ with separable variables, $F(x) = \sum_i f_{i1}(x_1) \times \ldots \times f_{in}(x_n)$.

In a general case, to minimize smoothed functions $F_h(x)$, stochastic gradient methods can be used [2, 6, 13 – 16], that exploit representations of gradients $\nabla F_h(x)$ in the form of multidimensional integrals (2) – (6). The latter, in turn, can be estimated using the Monte Carlo method.

Without any changes, the smoothing method is also applicable to minimize nonsmooth functions under (convex) constraints. Its local convergence was validated in [2, 5, 6, 15].

An experimental study of the stochastic smoothing method in global and non-differentiable optimization.

The algorithm of the stochastic smoothing method.

1. Reduce the conditional optimization problem to the problem of unconstrained optimization of a coercive function.

2. Select a decreasing sequence of smoothing parameters with a sufficiently large initial value of the smoothing parameter.

3. Consequently minimize smoothed functions by any effective deterministic or stochastic method, using the results of minimizing previous smoothed functions to minimize the subsequent smoothed function.

The following multi-extreme test functions are taken from [17], where also the primary sources of these examples are indicated. All calculations were carried out with the same fixed parameters of the algorithm. Calculations always started from the point with the maximum possible coordinates (in a valid hypercube). The only parameter that was changed from an example to an example was the number of smoothing stages.

One-dimensional test functions:

$$f_1(x) = \sin x + \sin\frac{10x}{3} + \ln x - 0.84x, \qquad 2.7 \le x \le 7.5;$$

$$f_2(x) = \sin x + \sin\frac{2x}{3}, \quad 3.1 \le x \le 20.4;$$

$$f_3(x) = -\sum\nolimits_{i=1}^{5} \sin\left((i+1)x + i\right), \quad -10 \le x \le 10;$$

$$f_4(x) = (x + \sin x)e^{-x^2}, \quad -10 \le x \le 10;$$

$$f_5(x) = -\sum_{i=1}^{m} \frac{1}{\left(k_i(x-a_i)\right)^2 + c_i}, \qquad 0 \le x \le 10.$$

Two-dimensional test functions.

$$f_C(x) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4, \quad -5 \le x_i \le 5, \qquad i = 1,2;$$

the global minimum –1.0316285 is attained at (0.08983, –0.7126) and (–0.08983,0.7126).

$$f_B(x) = \left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos x_1 + 10, \quad -5 \le x_1 \le 10, \quad 0 \le x_2 \le 15;$$

the global minimum is approximately 0.398 and it is reached at the three points (–3.142, 12.275), (3.142, 2.275) and (9.425, 2.425).

$$f_G(x) = \left[1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)\right] \times$$
$$\times\left[30 + (2x_1 - 3x_2)^2(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)\right],$$
$$-2 \le x_i \le 2, \qquad i = 1,2;$$

the global minimum is equal to 3 and the minimum point is (0, –1).

$$f_R(x) = x_1^2 + x_2^2 - \cos(18x_1) - \cos(18x_2), \qquad -1 \le x_i \le 1, \qquad i = 1,2;$$

the global minimum is equal to –2 and the minimum point is (0,0).

Multidimensional test functions.

$$f_H(x) = -\sum_{i=1}^{4} c_i \exp\left[-\sum_{j=1}^{n} a_{ij}(x_j - p_{ij})^2\right], \qquad 0 \le x_j \le 1, \qquad j = 1,\dots,n;$$

i		$a_{ij},\ j=1,2,3$		c_i		$p_{ij},\ j=1,2,3$	
1	3	10	30	1	.3689	.1170	.2673
2	.1	10	35	1.2	.4699	.4387	.7470
3	3	10	30	3	.1091	.8732	.5547
4	.1	10	35	3.2	.03815	.5743	.8828

For $n = 3$ the global minimum is equal to –3.86 and it is reached at the point (0.114,0.556,0.852).

$$f_{Q61}(x) = \sum_{i=1}^{6} a_i x_i^4 + x^T Bx + d^T x, \qquad -2 \le x_i \le 2, \qquad i = 1,\dots,6,$$

where $a^{\mathrm{T}} = [9\ 2\ 6\ 4\ 8\ 7]$, $d^{\mathrm{T}} = [2\ 6\ 5\ 0\ 0\ 2]$,

$$B^{\mathrm{T}} = [4\ 4\ 9\ 3\ 4\ 1;\ 4\ 3\ 7\ 9\ 9\ 2;\ 9\ 7\ 4\ 7\ 6\ 6;\ 3\ 9\ 7\ 4\ 2\ 6;\ 4\ 9\ 6\ 2\ 8\ 3;\ 1\ 2\ 6\ 6\ 3\ 5].$$

In this example, the stochastic smoothing method found a deeper minimum, $\min_{\max_i |x_i| \le 2} f_{Q61}(x) = -29.1601$,

$$x_{\min} \approx (-0.559657,\ -1.572471, 0.636230,\ 1.072193,\ 0.759148,\ -0.767645),$$

than the one found in [11, 18].

Nonsmooth test functions:

$$f_{m,n}(x) = \sum_{i=1}^{n} i^m |x_i|, \quad X = \{x \mid -1 \le x_i \le 1, i = 1,\dots,n\}.$$

Test results of the stochastic smoothing method. The performance of the method on functions $f_1(x)$ and $f_B(x)$ is illustrated in fig. 1, b, 2, b.

TABLE 1. Test results of the stochastic smoothing method for global optimization problems

Function	Reference	Reference minimal value	Achieved value	Number of function calculations
$f_1(x)$	[17, p. 177]	-4.6013075	-4.5992	58
$f_2(x)$	[17, p. 177]	-1.9059611	-1.9053	156
$f_3(x)$	[17, p. 177]	-3.3729	-3.3630	830
$f_4(x)$	[17, p. 177]	-0.8242394	$-.8241$	156
$f_5(x)$	[17, p. 177]	-14.5926520	-14.4499	690
$f_B(x_1, x_2)$	[17, p. 184]	0.3980	0.4042	840
$f_C(x_1, x_2)$	[17, p. 183]	-1.0316285	-0.9140	1020
$f_G(x_1, x_2)$	[17, p. 184]	3.0000	3.0988	2400
$f_R(x_1, x_2)$	[17, p. 185]	-2	-1.9021	250
$f_H(x_1, x_2, x_3)$	[17, p. 185]	-3.86	-3.8403	766
$f_H(x_1, \dots, x_6)$	[17, p. 185]	-3.32	-3.3017	952
$f_{Q61}(x_1, \dots, x_6)$	[18]	-28.942817	-29.1601	6384

TABLE 2. Test results of the stochastic smoothing method for nonsmooth convex optimization problems

Function $\sum_{i=1}^{n} i^m \lvert x_i \rvert$	Reference minimal value	Achieved value	Error, $\max_i x_i$	Number of function calculations	nbatch
$m=0, n=1000$	0	1.5120	0.0248	12486304	31
$m=1, n=10$	0	0.0141	0.0032	120000	3
$m=1, n=50$	0	0.1958	0.0077	2100560	7
$m=1, n=100$	0	0.8438	0.0078	5501800	10
$m=1, n=500$	0	18.5758	0.0421	56121560	22
$m=3, n=50$	0	10.4876	0.0043	42000560	7
$m=5, n=10$	0	1.0601	0.0013	2400000	3
$m=10, n=3$	0	0.0309	0.0318	519986	1

Conclusions. The paper describes the stochastic smoothing method for global optimization of nonsmooth nonconvex functions under (convex) constraints. The constrained problem is reduced in one or another way to the problem of unconstrained global optimization of a nonsmooth coercive function. The method consists in optimizing the sequence of smoothed functions with a gradually decreasing smoothing parameter. The initial smoothing parameter (diameter of the smoothing region) should be large enough to erase small local extrema. The justification of the method consists of observing that in the plane “variable – smoothing parameter” among the set of critical points there is a curve connecting the global extremum with infinity. The method is tested on numerous examples of global optimization problems from the literature. It stably finds global minima for low dimensional problems. With increasing the dimension of the problem, the solution time increases significantly due to the need for multiple estimates of multidimensional integrals using the Monte Carlo method.

References

1. Steklov V.A. Sur les expressions asymptotiques de certaines fonctions d´efinies par les equations differentielles du second ordre et leurs applications au probleme du developement dune fonction arbitraire en series procedant suivant les diverses fonctions. *Communications de la Société mathématique de Kharkow*, Série 2. 1907. 10. 97–199. (in French)
2. Gupal A.M. A method for the minimization of almost-differentiable functions. *Cybernetics*. 1977. 13 (1). 115–117. https://doi.org/10.1007/BF01071397
3. Gupal A.M., Norkin, V.I. Algorithm for the minimization of discontinuous functions. *Cybernetics*. 1977. 13 (2). 220–223. https://doi.org/10.1007/BF01073313
4. Norkin V.I. Two random search algorithms for the minimization of non-differentiable functions. In *Matematicheskie metody issledovaniya operatsyi i teorii nadezhnosti*; Ermoliev, Y.M.; Kovalenko, I.N., Eds.; V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine: Kyiv. 1978. 36–40.
5. Gupal A.M. *Stochastic Methods for Solving Nonsmooth Extremal Problems*. Naukova Dumka: Kyiv, 1979. (in Russian)
6. Mikhalevich V.S., Gupal A.M., Norkin V.I. *Methods of Nonconvex Optimization*. Nauka: Moscow, 1987. (in Russian)
7. Ermoliev Y.M., Norkin V.I., Wets R.J.-B. The minimization of semicontinuous functions: mollifier subgradients. *SIAM J. Control Optim*. 1995. 33 (1). 149–167. https://doi.org/10.1137/S0363012992238369
8. Yudin D.B. Quantitative Methods for the Analysis of Complex Systems. Izvestia AN SSSR. *Tehnicheskaia kibernetika*. 1965. 1. 3–13. (in Russian)
9. Norkin V.I. Two notes: on the smoothing method in multi-extreme optimization and on the finite-difference method in non-differentiable optimization. Abstracts of the III All-Union Seminar "*Numerical Methods of Nonlinear Programming*"; Kharkiv State University: Kharkiv. 1979. (in Russian)

10. Arıkan O., Burachik R.S., Kaya C.Y. Steklov Regularization and Trajectory Methods for Univariate Global Optimization. *J. Glob. Optim*. 2020. 76 (1). 91–120. https://doi.org/10.1007/s10898-019-00837-3
11. Burachik R.S., Kaya C.Y. Steklov Convexification and a Trajectory Method for Global Optimization of Multivariate Quartic Polynomials. 2019. https://arxiv.org/abs/1912.00332v1
12. Clarke F.H. Optimization and Nonsmooth Analysis, 2nd ed.; Volume 5 of Classics in Applied Mathematics. SIAM: Philadelphia, PA: 1990. https://doi.org/10.1137/1.9781611971309
13. Ermoliev Y.M. Methods of Stochastic Programming. Moscow: Nauka. 1976. (in Russian)
14. Polyak B.T. Introduction to Optimization. Optimization Software: New York. 1987.
15. Ermoliev Y.M., Norkin V.I. Solution of nonconvex nonsmooth stochastic optimization problems. *Cybern. Syst. Anal*. 2003. 39 (5). 701–715. https://doi.org/10.1023/B:CASA.0000012091.84864.65
16. Bottou L., Curtis F.E., Nocedal J. Optimization Methods for Large-Scale Machine Learning. *SIAM Review*. 2018. 60 (2). 223–311. https://doi.org/10.1137/16m1080173
17. Törn A., Zilinskas A. *Global Optimization*. Springer: Berlin. 1989. https://doi.org/10.1007/3-540-50871-6
18. Qi L., Wan Z., Yang Y.-F. Global Minimization of Normal Quartic Polynomials Based on Global Descent Directions. *SIAM J. Optim*. 2004. 15 (1). 275–302. https://doi.org/10.1137/s1052623403420857

Received 07.02.2020

Vladimir I. Norkin,
Dr. Sci. (Phys. & Math.), Leading research fellow at V.M. Glushkov Institute of Cybernetics, The National Academy of Sciences of Ukraine, Kyiv, Ukraine,
Professor of the Department of Applied Mathematics at the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.
vladimir.norkin@gmail.com

УДК 519.85

В.І. Норкін

СТОХАСТИЧНИЙ МЕТОД ЗГЛАДЖУВАННЯ ДЛЯ НЕГЛАДКОЇ ГЛОБАЛЬНОЇ ОПТИМІЗАЦІЇ

Інститут кібернетики імені В.М. Глушкова НАН України, Київ
Національний технічний університет України «КПІ імені Ігоря Сікорського», Київ
Листування: vladimir.norkin@gmail.com

Вступ. Проблема глобальної оптимізації неопуклих негладких функцій з обмеженнями є актуальною для багатьох інженерних застосувань, зокрема, для навчання неопуклих негладких нейронних мереж. У роботі представлені результати тестування методу згладжування багато екстремальної цільової функції для знаходження її глобального мінімуму в деякій опуклій допустимій області евклідового простору. Попередньо цільова функція довизначається поза опуклої допустимої області так, щоб не змінити її глобального мінімуму, та зробити її коерцитивною. Згладжування функції в будь-якій точці здійснюється шляхом усереднення значень функції по деякому околу цієї точки. Розмір околу є параметром згладжування. Традиційно локальне згладжування використовувалось для гладкої апроксимації недиференційованих або розривних функцій. Більш широке згладжування ліквідує мілкі локальні екстремуми вихідної функції. При великому параметрі згладжування усереднена коерцитивна функція може мати лише один мінімум. Метод згладжування для глобальної оптимізації функцій полягає у заміні вихідної функцій послідовністю згладжених апроксимацій із зменшенням до нуля параметра згладжування та оптимізації останніх сучасними методами стохастичної оптимізації (метод усереднення траєкторії, важкої кульки, яружного кроку). При цьому градієнти згладжених функцій представляються у вигляді багатовимірних інтегралів та оцінюються методом Монте-Карло. Пересуваючись від мінімуму однієї згладженої функції до близького мінімуму другої згладженої функції з меншим параметром згладжування, можна потрапити в область глобального мінімуму вихідної функції. Остаточну дооптимізацію функції можна зробити будь-яким підходящим методом нелінійного програмування. Метод згладжування без яких-небудь змін може бути застосований для оптимізації негладких яружних функцій за опуклих обмежень, а також у комбінації з методом точних негладких штрафів. Показано, що метод згладжування впевнено розв'язує тестові задачі глобальної оптимізації невеликої розмірності з літератури. При збільшенні розмірності задачі час розв'язання значно зростає у зв'язку з необхідністю багатократної оцінки багатовимірних інтегралів методом Монте-Карло.

Ключові слова: глобальна оптимізація, згладжування за Стекловим, усереднені функції, стохастична оптимізація, негладка неопукла оптимізація.

УДК 519.85

В.И. Норкин

СТОХАСТИЧЕСКИЙ МЕТОД СГЛАЖИВАНИЯ ДЛЯ НЕГЛАДКОЙ ГЛОБАЛЬНОЙ ОПТИМИЗАЦИИ

Институт кибернетики имени В.М. Глушкова НАН Украины, Киев
Национальный технический университет Украины «КПИ имени Игоря Сикорского», Киев
Переписка: vladimir.norkin@gmail.com

Введение. Проблема глобальной оптимизации невыпуклых негладких функций при ограничениях актуальна для многих инженерных приложений, в частности, для обучения невыпуклых негладких нейронных сетей. В работе представлены результаты тестирования метода сглаживания многоэкстремальной целевой функции для нахождения ее глобального минимума в некоторой выпуклой допустимой области евклидового пространства. Предварительно целевая функция доопределяется вне допустимой области так, чтобы не изменить ее глобальный минимум, а сделать ее коэрцитивной. Сглаживание функции в какой-либо точке осуществляется путем усреднения значений функции по некоторой окрестности этой точки. Размер окрестности является параметром сглаживания. Традиционно локальное сглаживание использовалось для гладкой аппроксимации недифференцируемых или разрывных функций. Более обширное сглаживание ликвидирует мелкие локальные экстремумы исходной функции. При большом значении параметра сглаживания усредненная функция может иметь только один минимум. Метод сглаживания состоит в замене исходной функции последовательностью сглаженных аппроксимаций с убывающим к нулю параметром сглаживания и оптимизации последних современными методами стохастической оптимизации. При этом градиенты сглаженных функций представляются в виде многомерных интегралов и оцениваются методом Монте-Карло. Переходя от минимума одной сглаженной функции к близкому минимуму другой сглаженной функции, можно прийти в область глобального минимума исходной функции. Окончательную дооптимизацию функции можно сделать любым подходящим методом нелинейного программирования. Метод сглаживания без каких-либо изменений может быть применен для оптимизации негладких овражных функций при выпуклых ограничениях, а также в комбинации с методом точных негладких штрафов. Показано, что метод сглаживания уверенно решает тестовые задачи глобальной оптимизации небольшой размерности из литературы. При увеличении размерности задачи время решения значительно возрастает, учитывая необходимость многократной оценки многомерных интегралов методом Монте-Карло.

Ключевые слова: глобальная оптимизация, сглаживание по Стеклову, усредненные функции, стохастическая оптимизация, негладкая невыпуклая оптимизация.