Optimization Methods and Software

ISSN: 1055-6788 (Print) 1029-4937 (Online) Journal homepage: https://www.tandfonline.com/loi/goms20

Multidimensional global optimization using numerical estimates of objective function derivatives

Victor Gergel & Alexey Goryachih

OPTIMIZATION METHODS & SOFTWARE
https://doi.org/10.1080/10556788.2019.1630624

Multidimensional global optimization using numerical estimates of objective function derivatives

Victor Gergel and Alexey Goryachih

Department of Software and Supercomputing, Lobachevsky State University, Nizhni Novgorod, Russia

ABSTRACT

This paper proposes a method for solving the computation-costly multidimensional global optimization problems. The method is efficient in the one-dimensional case and its combination with the nested reduction scheme competitive with optimization methods reducing multidimensional problems by using space-filling (Peano) curves. The developed method is based on an approach, in which not only the minimized function values but also the values of derivatives of these functions are used to increase the efficiency of global optimization. The required values of the derivatives are estimated numerically by handling the available search information. The results of the executed experiments confirm the developed approach is promising.

ARTICLE HISTORY
Received 8 July 2018
Accepted 3 June 2019

KEYWORDS
Multiextremal optimization; global search algorithms; Lipschitz condition; numerical estimations of derivative values; dimensionality reduction; numerical experiments

1. Introduction

The problems of global (or multiextremal) optimization [5,6,19,22,25–27,36–39] are among the most complicated optimization problems. The multiextremal function can have several local optima that complicates highly finding the global minimum because of the necessity to explore the whole search domain. The amount of computations when solving the global optimization problems can grow exponentially with increasing number of varied parameters.

The global optimization problem is a problem of finding the minimum value of a real function $\varphi(y)$ and can be defined in the form

$$\varphi(y^*) = \min\{\varphi(y) : y \in D\}, \tag{1}$$

where $y = (y_1, y_2, \ldots, y_N)$ is a vector of varied parameters, N is the dimensionality of the multiextremal optimization problem, and D is the search domain, which is a N-dimensional hyperinterval

$$D = \{y \in \mathbb{R} : a_i \le y_i \le b_i, 1 \le i \le N\}, \tag{2}$$

at given boundary vectors a and b. The minimized function $\varphi(y)$ is assumed to be a multiextremal one, and the evaluation of the function values may require a large amount of computations.

CONTACT Alexey Goryachih a_goryachih@mail.ru

Obtaining the guaranteed global minimum estimate is possible with some *a priori* assumptions on the minimized function behaviour. The fulfilment of the Lipschitz condition, i.e.

$$\left|\varphi(y_2) - \varphi(y_1)\right| \le L \,||\, y_2 - y_1 \,||\, , y_1, y_2 \in D, \tag{3}$$

where $L > 0$ is the Lipschitz constant and $||\cdot||$ denotes the Euclidean norm in $\mathbb{R}^N$ is one of the most often used assumptions.

The Lipschitz condition corresponds to the assumption of a limited variation of the function value at limited variations of its parameters. This condition allows making the estimates of potential behaviour of the function $\varphi(y)$ based on a finite set of its values computed at some points in the search domain D.

Recently, the assumption on the fulfilment of the Lipschitz condition (3) was expanded onto the partial derivatives $\varphi_i'(y)$, $1 \le i \le N$ of the minimized function $\varphi(y)$ as well, i.e.

$$\left|\varphi_i'(y_2) - \varphi_i'(y_1)\right| \le L_i \,||\, y_2 - y_1 \,||\, , y_1, y_2 \in D, 1 \le i \le N, \tag{4}$$

where $L_i > 0$, $1 \le i \le N$, are the corresponding Lipschitz constants for the partial derivatives $\varphi_i'(y)$, $1 \le i \le N$ [1,2,9–11,20,29,30,34].

The fulfilment of the conditions (4) allows making more precisely estimates of the probable values of the minimized function $\varphi(y)$ that provides an opportunity to increase the efficiency of the developed algorithms [9,29,30].

However, in some applied optimization problems the computing of the derivatives may be restricted or even impossible. In this case, the development of the global optimization methods, in which the necessary values of derivatives are computed numerically may be useful [11,15,16,24].

It should be noted that solving multidimensional optimization problems is often based on reduction to a one-dimensional optimization problem. Widely used techniques are the approach based on using Peano (space-filling) curves [28,31], the nested multistep reduction scheme [4,12,13,17,33], the diagonal generalization technique [10,27,30], etc. As a result algorithms of one-dimensional optimization can be applied efficiently also in the multidimensional case [1–4,7,9,11,18,20,21,23,28,29,34,35]. Wherein using the nested multistep reduction scheme can lead for executing some redundant global search iterations – see for example [4,33,38]. This deficiency can be diminished by using the values of derivatives of the minimized function – see the results of numerical experiments given in Section 5. Also some redundancy of the multistep reduction scheme can be diminished by using an adaptive modification of this scheme – see [12,13].

In general, many optimization algorithms were developed based on the approach when algorithms construct minorants of the minimized function using values of the function at optimization iteration points [1,2,9,11,20,28,29,35]. At the same time, developed algorithms can use different information about the minimize functions given *a priori*. For example, algorithms proposed in [7,28,35,38] assume that minimized functions satisfy the Lipschitz condition, while another methods considered in [3,9,11,15] assume that the Lipschitz condition is also fulfilled for the first derivatives of minimized functions.

The proposed approaches differ also that the algorithms considered in [3,28,35] are assumed the value of the Lipschitz constant is known *a priori*, while the algorithms given in [9,11,15,38] are estimated the value of the Lipschitz constant numerically. The algorithms proposed in [9,11,15] use values of the first derivative (the algorithm given in [15] estimates the derivative values numerically) that allows to construct more precisely estimation of the minimizing function values – see the results of numerical experiments in Section 5. So, the main contribution of this paper can be considered as proposing the new multidimensional global optimization algorithm using the multistep dimension reduction scheme and a numerical scheme for computing approximate values of the first derivative of the minimized function. The proposed method generalizes the method of two-dimensional multiextremal optimization given in [15] to the general case of multidimensional global search. The developed algorithm has a high potential for further development by possible use of the perspective adaptive nested reduction scheme [12,13].

In this paper, the development of global optimization algorithms utilizing the numerical derivatives of the minimized function $\varphi(y)$ is conducted. In Section 2, the base one-dimensional algorithm utilizing the numerical derivatives is given. The convergence conditions of this algorithm are substantiated in Section 3. Section 4 describes the nested dimensionality reduction scheme, which allows generalizing the proposed one-dimensional algorithm for solving the multidimensional global optimization problems. In Section 5, the results of the numerical experiments are considered, which confirm the developed approach to be promising.

2. One-dimensional global optimization algorithm utilizing numerical derivatives

The proposed optimization algorithm is based on the Adaptive Global Method using Derivatives (AGMD) [9,10] intended for solving the one-dimensional global optimization problems

$$\varphi(x^*) = \min\{\varphi(x) : x \in [a, b]\}. \tag{5}$$

The Adaptive Global Method using Numerical Derivatives (AGMND) is a modification of AGMD, in which the values of the first derivative of the minimized function are replaced by the numerical estimation of these ones [11,15].

The computational scheme of AGMND is as follows. The first two iterations are performed at the boundary points a and b of the search domain $[a, b]$. Then, let k, $k > 1$ global search iterations were completed, and the values of the minimized function $\varphi(x)$ have been computed at each iteration (hereafter, these computations will be called *trials*). The trial point of the next $(k + 1)$ optimization iteration is determined by the following rules.

Rule 1. Renumber the points of previous trials by subscripts in increasing order

$$a = x_0 < x_1 < \cdots < x_i < \cdots < x_k = b. \tag{6}$$

Rule 2. Compute the numerical estimations of the first derivatives of $\varphi(x)$ at the points of the executed search iterations x_i, $0 \le i \le k$ according to the expressions:

$$\dot{z}_i = \begin{cases} \frac{z_{i+1}-z_i}{x_{i+1}-x_i}, i = 0, \\ \frac{z_i-z_{i-1}}{x_i-x_{i-1}}, 1 \le i \le k, \end{cases} \tag{7}$$

here and the following z_i, $0 \le i \le k$, denotes $\varphi(x_i)$.

Rule 3. Compute the estimation of the Lipschitz constant from (4) for the first derivative of the optimized function

$$m_1 = \begin{cases} rM_1, & M_1 > 0, \\ 1, & M_1 = 0, \end{cases} \tag{8}$$

where

$$M_1 = \max(M_{1i}), 1 \le i \le k, \tag{9}$$

$$M_{1i} = \max \begin{cases} |\dot{z}_i - \dot{z}_{i-1}|/|x_i - x_{i-1}|, \\ -2\left[z_i - z_{i-1} - \dot{z}_{i-1}(x_i - x_{i-1})\right]/(x_i - x_{i-1})^2, \\ 2\left[z_i - z_{i-1} - \dot{z}_i(x_i - x_{i-1})\right]/(x_i - x_{i-1})^2, \end{cases} \tag{10}$$

and $r > 1$ is the reliability parameter of the algorithm.

Rule 4. Compute the characteristic $R(i)$ for each interval (x_{i-1}, x_i), $1 \le i \le k$, according to the following expressions to estimate the minimum possible values of $\varphi(x)$ in the interval (x_{i-1}, x_i)

$$R(i) = \begin{cases} \widehat{\varphi}_i(\widehat{x}_i), & \widehat{x}_i \in [\overline{x}_i, \overline{\overline{x}}_i], \\ \min(\widehat{\varphi}_i(\overline{x}_i), \widehat{\varphi}_i(\overline{\overline{x}}_i)), & \widehat{x}_i \notin [\overline{x}_i, \overline{\overline{x}}_i], \end{cases} \tag{11}$$

where

$$\widehat{x}_i = \frac{-\varphi'(x_{i-1}) + m_1(\overline{x}_i - x_{i-1}) + m_1 x_i}{m_1}, \tag{12}$$

and the auxiliary functions (minorants) $\widehat{\varphi}_i(x)$, $1 \le i \le k$, take the form

$$\widehat{\varphi}_i(x) = \begin{cases} \widehat{\varphi}_{i1}(x) = z_{i-1} + \dot{z}_{i-1}(x - x_{i-1}) - 0.5m_1(x - x_{i-1})^2, & x \in (x_{i-1}, \overline{x}_i), \\ \widehat{\varphi}_{i2}(x) = A_i(x - \overline{x}_i) + 0.5m_1(x - \overline{x}_i)^2 + B_i, & x \in [\overline{x}_i, \overline{\overline{x}}_i], \\ \widehat{\varphi}_{i3}(x) = z_i - \dot{z}_i(x - x_i) - 0.5m_1(x - x_i)^2, & x \in (\overline{\overline{x}}_i, x_i], \end{cases} \tag{13}$$

where

$$\begin{aligned} A_i &= \dot{z}_{i-1} - m_1(\overline{x}_i - x_{i-1}), \\ B_i &= \widehat{\varphi}_{i1}(\overline{x}_i), \\ \overline{x}_i &= \frac{(z_{i-1} - \dot{z}_{i-1}x_{i-1}) - (z_i - \dot{z}_i x_i) + m_1(x_i^2 - x_{i-1}^2)/2 - m_1 d_i^2}{m_1(x_i - x_{i-1}) - (\dot{z}_i - \dot{z}_{i-1})}, \\ \overline{\overline{x}}_i &= \frac{(z_{i-1} - \dot{z}_{i-1}x_{i-1}) - (z_i - \dot{z}_i x_i) + m_1(x_i^2 - x_{i-1}^2)/2 + m_1 d_i^2}{m_1(x_i - x_{i-1}) - (\dot{z}_i - \dot{z}_{i-1})}, \\ d_i &= (x_i - x_{i-1})/2 - (\dot{z}_i - \dot{z}_{i-1})/2m_1. \end{aligned} \tag{14}$$

Each characteristic $R(i)$, $1 \le i \le k$, calculated in this way is an estimation of the minimum possible value of the minorant $\widehat{\varphi}_i(x)$ from (13) in the intervals $[x_{i-1}, x_i]$ and the estimation of the minimum possible values of $\varphi(x)$ in these intervals.

Rule 5. Find the interval (x_{t-1}, x_t) with the minimal characteristic $R(t)$

$$R(t) = \min\{R(i) : 1 \le i \le k\}. \tag{15}$$

In the case when there exist several intervals satisfying (15) the interval with the minimal number t is taken for certainty.

Rule 6. Compute the next point of the next trial x^{k+1} accordingly

$$x^{k+1} = \begin{cases} \widehat{x}_t, & \widehat{x}_t \in [\overline{x}_t, \overline{\overline{x}}_t], \\ \overline{x}_t, & \widehat{\varphi}(\overline{x}_t) \le \widehat{\varphi}(\overline{\overline{x}}_t), \\ \overline{\overline{x}}_t, & \widehat{\varphi}(\overline{x}_t) > \widehat{\varphi}(\overline{\overline{x}}_t). \end{cases} \tag{16}$$

The stopping condition is defined by the following relation:

$$|x_t - x_{t-1}| \le \varepsilon, \tag{17}$$

where ε is the *accuracy*, $\varepsilon > 0$.

The minimum computed value of the minimized function is accepted as the current estimate of the global minimum value, i.e.

$$\varphi^* = \min\{z_i : 0 \le i \le k\}. \tag{18}$$

Notes. 1. The computing of the numerical estimations $\dot{z}_i$, $0 \le i \le k$, of the first derivative of the function $\varphi(x)$ can be performed also using the three-point approximating expressions:

$$\dot{z}_i = \begin{cases} \frac{1}{H_1^2}\left(-(2+\delta_2)z_0 + \frac{(1+\delta_2)^2}{\delta_2}z_1 - \frac{1}{\delta_2}z_2\right), i = 0, \\ \frac{1}{H_i^{i+1}}\left(-\delta_{i+1}z_{i-1} + \frac{(\delta_{i+1}^2-1)}{\delta_{i+1}}z_i + \frac{1}{\delta_{i+1}}z_{i+1}\right), 1 < i < k, \\ \frac{1}{H_{k-1}^k}\left(\delta_k z_{k-2} - \frac{(1+\delta_k)^2}{\delta_k}z_{k-1} + \frac{(2\delta_k+1)}{\delta_k}z_k\right), i = k, \end{cases} \tag{19}$$

where $H_i^{i+1} = h_i + h_{i+1}$, $\delta_{i+1} = \frac{h_{i+1}}{h_i}$ and $h_i = x_i - x_{i-1}$, see [16].

These formula used three points of trials and can be used in **Rule 2** if $k > 2$.

2. For the applicability of the computational scheme described above, the fulfilment of the following inequalities:

$$x_{i-1} < \overline{x}_i < \overline{\overline{x}}_i < x_i \tag{20}$$

for all intervals (x_{i-1}, x_i), $1 \le i \le k$, is necessary. If the estimate of the Lipschitz constant m_1 computed in (8) is insufficient and the condition (20) is violated for some $1 \le i \le k$ the value m_1 should be refined. Thus the maximum root of the equations

$$\begin{cases} -(z_i - z_{i-1}) + 0.5(\dot{z}_i + \dot{z}_{i-1}) + 0.25m_1(x_i - x_{i-1})^2 - \dfrac{(\dot{z}_i - \dot{z}_{i-1})^2}{4m_1} = 0, \\ (z_i - z_{i-1}) - 0.5(\dot{z}_i + \dot{z}_{i-1}) + 0.25m_1(x_i - x_{i-1})^2 - \dfrac{(\dot{z}_i - \dot{z}_{i-1})^2}{4m_1} = 0, \end{cases} \tag{21}$$

was selected as m_1 in this case for AGMD [9].

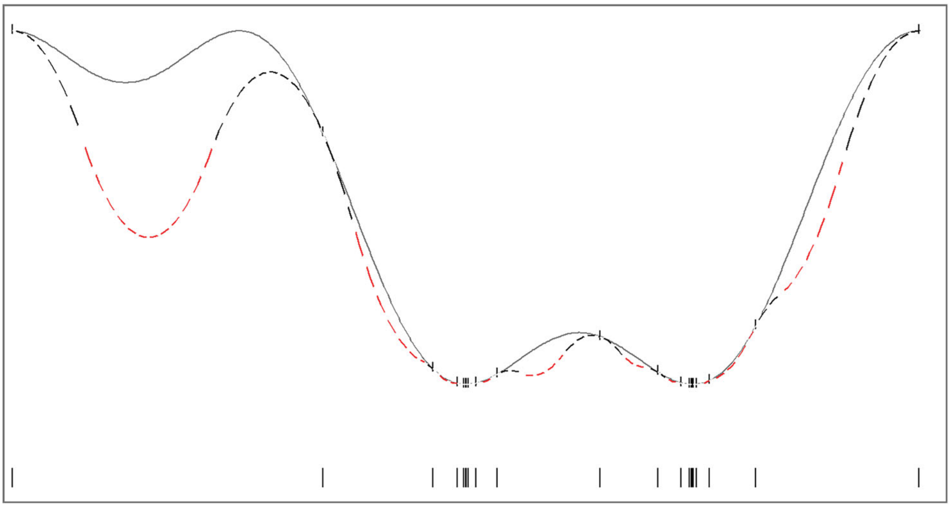

Figure 1. An example of solving a global optimization problem by AGMND. The grey curve is the minimized function graph, the dashed curve is the minorant, the vertical dashes are the points of executed global search iterations.

As an example the first several iterations of AGMD and the minorant of the minimized function are shown in Figure 1.

To clarify the essence of the considered method, one can note the following. If the values $\dot{z}_i$, $0 \le i \le k$, at the points of the executed search iterations x_i, $0 \le i \le k$, are the values of the first derivative $\varphi'(x)$ of the minimized function $\varphi(x)$, the following condition is satisfied:

$$m_1 > L_1, \tag{22}$$

where m_1 from (8) is an adaptive numerical estimation of the Lipschitz constant L_1 of the derivative $\varphi'(x)$, then the auxiliary functions $\widehat{\varphi}(x)$, $1 \le i \le k$, from (13) are the minorants of the minimized function $\varphi(x)$ in the intervals (x_{i-1}, x_i), $1 \le i \le k$, i.e. the condition

$$\forall x \in (x_{i-1}, x_i) \Rightarrow \varphi(x) \ge \widehat{\varphi}_i(x), 1 \le i \le k \tag{23}$$

is satisfied [9,10].

3. Convergence conditions

When using the AGMND algorithm, the values z_i', $0 \le i \le k$, are the numerical estimations of the first derivative $\varphi'(x)$ according to (7). Besides, the adaptive estimate m_1 from (8) can be an underestimate of the Lipschitz constant L_1, and the condition (22) may not be satisfied. In these conditions, the functions $\widehat{\varphi}_i(x)$, $1 \le i \le k$, from (13) may not be the minorants of the minimized function $\varphi(x)$ in the intervals (x_{i-1}, x_i), $1 \le i \le k$.

To investigate the convergence conditions of the AGMND algorithm, first of all, let us ensure the fulfilment of the condition (22) by an appropriate increasing of the reliability parameter r from (8). Taking into account this assumption, let us prove the following theorems.

Theorem 3.1: *Let the point $\widetilde{x}$ is the limit point (the accumulation point) of the sequence $\{x^k\}$ generated by AGMND when minimizing a bounded function $\varphi(x)$, $x \in [a, b]$, the derivative of which satisfies the Lipschitz condition (4), and $\widetilde{x} \neq a$, $\widetilde{x} \neq b$. Then the point $\widetilde{x}$ is a local minimum of the function $\varphi(x)$.*

Proof: Let us suppose the opposite and assume that the point $\widetilde{x}$ is not a local minimum one. Then either the point $\widetilde{x}$ is in the interval of monotonous variation of the function $\varphi(x)$ ($\varphi'(\widetilde{x}) \neq 0$) or the point $\widetilde{x}$ is the point of local maximum or the inflection point. Let us consider sequentially all the cases listed above.

Without a loss in generality, let us assume $\varphi'(\widetilde{x}) < 0$ in the first case. In this case, two following variants are possible.

In the first variant, at the step p the new trial is executed at the point $x^p = \widetilde{x}$. Therefore, at any $k > p$ there exists such index $j = j(k)$ that

$$x_j = \widetilde{x} = x^p. \tag{24}$$

First, let us prove that there is no subsequence $\{x^k\}$ converging to the point $\widetilde{x}$ from the left.

Let us estimate the characteristic of the interval (x_{j-1}, x_j). If the convergence to the point $\widetilde{x}$ from the left takes place, there exists such index $q(k) > p$ that at any $k > q$ for the left boundary point of the interval (x_{j-1}, x_j) the inequality should be fulfilled

$$z_{j-1} > z_j. \tag{25}$$

Therefore, $z'_j < 0$ according to (7). So one can obtain from (13)

$$R(j) \geq z_j + \left|z'_j\right| (x_j - \bar{\bar{x}}_j) - 0.5m_1(x_j - \bar{\bar{x}}_j)^2. \tag{26}$$

From this condition, $R(j) \geq z_j$ when $(x_j - \bar{\bar{x}}_j) < 2|z'_j|/m_1$. The characteristic of the interval $R(j+1)$ would be less than the value of z_j in these conditions. Therefore, starting from some step, new trials cannot be executed in the interval (x_{j-1}, x_j).

Now, let us verify the possibility of the convergence to the point $\widetilde{x}$ from the right to exist. If the convergence takes place, there should exist the trials, which are placed to the right from the point $\widetilde{x}$. The values of $\varphi(x)$ at these trial points are less than z_j and decrease in magnitude. Let us mark one of the intervals formed by the points of such trials as $p = p(k)$, i.e.

$$z_p < z_{p-1} < z_j, \tag{27}$$

since $\varphi'(\widetilde{x}) < 0$. Then the characteristic of the interval (x_p, x_{p+1}) at any optimization iteration is less than the value of z_p (according to (13)). On the other hand, if the convergence to the point $\widetilde{x}$ from the right exists, the following is true:

$$\lim_{k\to\infty} R(j+1) = z_j > z_p. \tag{28}$$

Therefore, the convergence to the point $\widetilde{x}$ from the right is impossible as well.

It remains to consider the variant when $\widetilde{x} \neq x^k$ at any k, i.e. when the trial points do not coincide with the limit point. Let the interval including the point $\widetilde{x}$ is (x_{j-1}, x_j). If the convergence to the point $\widetilde{x}$ exists, it follows from (20) that

$$\lim_{k\to\infty} (x_j - x_{j-1}) = 0, \lim_{k\to\infty} R(j) = z_j = \varphi(\widetilde{x}). \tag{29}$$

On the other hand, as in the case of considering the convergence from the right in the previous variant, one can find the interval p, for which (27) is fulfilled. In this case, the characteristic of the interval (x_p, x_{p+1}) is less than z_p. Therefore, starting from some step, the trials cannot fall into the interval (x_{j-1}, x_j), and the convergence to the point $\widetilde{x}$ does not exist.

Thus if the point $\widetilde{x}$ is not a local minimum, then, at least, the value of derivative at this point should be zero (i. e. $\varphi'(\widetilde{x}) = 0$).

The second case corresponds to the situation when the value of the derivative at the limit point $\widetilde{x}$ equals to zero, but the point $\widetilde{x}$ is the local maximum point of $\varphi(x)$. Let (x_{j-1}, x_j) is the interval including the point $\widetilde{x}$ as before. Let mark the point, for which $z^p < \varphi(\widetilde{x})$ as x^p. For certainty, let us assume that $x^p > \widetilde{x}$ and $\varphi'(x^p) < 0$ (the case of $x^p < \widetilde{x}$ can be considered in the same way). Because of the convergence, this point should exist. However, in this case, when the interval (x_{j-1}, x_j) is small enough, the characteristic $R(j)$ would be greater than the one of the right interval for the point x^p. The convergence to the point $\widetilde{x}$ cannot exist that contradicts to the conditions of the theorem.

The case when the point $\widetilde{x}$ is the inflection point is impossible as well can be proved in the same way.

So far, the truth of statement that $\widetilde{x}$ is a local minimum point is proved. ■

One more convergence property of the AGMND algorithm can be formulated in the form of the following lemma, which is given without proof (the truth of the lemma can be proved using the proving scheme of Theorem 3.1).

Lemma: *At the same assumptions, as in Theorem* 3.1, *the sequence* $\{x^k\}$ *includes two subsequences, one converging to the point* $\widetilde{x}$ *from the left and another one is from the right.*

The statements formulated above allow proving the following theorem.

Theorem 3.2: *Let the point* $\widetilde{x}$ *is the limit point of the sequence* $\{x^k\}$ *generated by AGMND when minimizing a bounded function* $\varphi(x)$, $x \in [a, b]$, *the derivative of which satisfies the Lipschitz condition* (4) *with the constant* L_1, *and* $x \neq a, x \neq b$. *Then the following is true:*

(1) *If along with the point* $\widetilde{x}$ *there exists another limit point* $\widehat{x}$ *of the sequence* $\{x^k\}$, *then* $\varphi(\widetilde{x}) = \varphi(\widehat{x})$,
(2) $z^k = \varphi(x^k) \geq \varphi(\widetilde{x})\ \forall k \geq 0$.

Proof: The first statement of the theorem is true because otherwise under assumption on the existence of a sequence of trials converging to the point $\widetilde{x}$ such that $\varphi(\widetilde{x}) \neq \varphi(\widehat{x})$ one obtains a contradiction to the second statement of the theorem.

Let us demonstrate the truth of the second statement. Let us assume the contrary, i.e. the following results were obtained at some step $q \geq 0$

$$z^q = \varphi(x^q) < \varphi(\widetilde{x}). \tag{30}$$

Let us denote the lower index corresponding to the point x^q at the step k as $j = j(k)$, i.e. $z_j = z^q$, and the index of the interval (x_{t-1}, x_t) including the point $\widetilde{x}$ as $t = t(k)$. Subject to the sign of derivative at the point x^q, the following holds:

$$\begin{aligned} &R(j) < z^q, \varphi'(x^q) > 0, \\ &R(j+1) < z^q, \varphi'(x^q) < 0, \\ &R(j), R(j+1) < z^q, \varphi'(x^q) = 0. \end{aligned} \tag{31}$$

In any case, there exists an interval, the characteristic of which is less than z^q. On the other hand, because of the condition of convergence to the point $\widetilde{x}$, the results of Theorem 3.1 and Lemma 3 the following are true

$$\lim_{k\to\infty} R(t(k)) = \varphi(\widetilde{x}). \tag{32}$$

Therefore, as follows from (32) and taking into account (30), at small enough length of the interval (x_{t-1}, x_t) the characteristic of this interval cannot appear to be the minimum one, and the convergence to the point $\widetilde{x}$ is impossible. This is a contradiction to the conditions of the theorem, and, hence, the assumption made is not true. ■

The main result on the sufficient convergence condition for the AGMND algorithm can be formulated in the form of the following theorem.

Theorem 3.3 (the sufficient conditions of convergence): *If at the same assumptions as in Theorem* 3.2 *the following inequality for m_1 from* (8) *holds at some step*

$$m_1 \geq (1 + \gamma)L_1, \tag{33}$$

where

$$\gamma = 4(b - a)/\varepsilon^2, \tag{34}$$

a, b from (5) *are the boundary points of the search interval $[a, b]$, $\varepsilon > 0$ from* (17) *is the required accuracy of the solution of the global optimization problem, then any global minimum point x^* is a limit point of the sequence $\{x^k\}$ and, besides, any limit point $\widetilde{x}$ of this sequence is the global minimum point of the function $\varphi(x)$.*

Proof: Let us denote the index of the interval including the global minimum point x^* of the function $\varphi(x)$ at the step $k > 1$ as $j = j(k)$. For the proof, it is enough to show that the characteristic $R(j)$ gives the estimate of the probable values of the function $\varphi(x)$ in the interval (x_{j-1}, x_j), which is less than $\varphi(x^*)$, i.e.

$$R(j) \leq \varphi(x^*). \tag{35}$$

For the AGMD algorithm using the precise values of the first derivative of the minimized function $\varphi'(x)$, the requirement (35) is satisfied when the condition (22) is fulfilled.

However for the AGMND algorithm, in which the values z_i', $0 \le i \le k$, are computed numerically, the necessary condition (35) may not be satisfied because of the approximation of the first derivative values. The fulfilment of the condition (35) can be provided by increasing the estimate of m_1 from (8) (even with the overestimate of the precise value of the Lipschitz constant L_1 if necessary).

Let us estimate the characteristics $R(i)$, $1 \le i \le k$, from (11). Omitting the intermediate calculations, one obtains

$$R(j) = z_{j-1} + 2\dot{z}_{j-1}(\overline{x}_j - x_{j-1}) - m_1(\overline{x}_j - x_{j-1})^2. \tag{36}$$

With the exact value of the first derivative $\varphi'(x)$ at the point x_{j-1}, the expression takes the form

$$R_1(j) = z_{j-1} + 2\varphi'(x_{j-1})(\overline{x}_j - x_{j-1}) - m_1(\overline{x}_j - x_{j-1})^2. \tag{37}$$

Substituting the approximate value of the first derivative $\dot{z}_{j-1}$ and increasing the value of estimate $\widehat{m}_1 > m_1$ for (36), one obtains

$$R_2(j) = z_{j-1} + 2\dot{z}_{j-1}(\overline{x}_j - x_{j-1}) - \widehat{m}_1(\overline{x}_j - x_{j-1})^2. \tag{38}$$

To ensure the convergence, the value of characteristic $R_2(j)$ should be less than the one of $R_1(j)$ that corresponds to the condition

$$(\widehat{m}_1 - m_1)(\overline{x}_j - x_{j-1})^2 \ge 2(\dot{z}_{j-1} - \varphi'(x_{j-1}))(\overline{x}_j - x_{j-1}). \tag{39}$$

Taking into account that

$$(\overline{x}_j - x_{j-1}) \ge \varepsilon, (\overline{x}_j - x_{j-1}) \le (b-a), \dot{z}_{j-1} - \varphi'(x_{j-1}) \ge 2L_1, \tag{40}$$

where a,b from (5) are the boundary points of the search interval $[a, b]$, $\varepsilon > 0$ from (17) is the required accuracy of solution of the problem (5) and assuming

$$m_1 = L_1, \widehat{m}_1 = (1+\gamma)L_1, \gamma > 0, \tag{41}$$

one can obtain

$$\gamma L_1 \varepsilon^2 \ge 4L_1(b-a), \tag{42}$$

that corresponds to the initial condition of the theorem (33). ■

No doubt, the condition (33) imposes a requirement to the numerical estimation of the Lipschitz constant m_1 with a significant excess. Thus, in the executed numerical experiments (see Section 5), the magnitude of m_1 was determined according to (8) with the value of the reliability parameter $r = R + d/k$ where k was the number of executed trials, R and d were the parameters given *a priori*.

4. Nested dimensionality reduction scheme

The dimensionality reduction is an efficient approach to the application of the one-dimensional global optimization algorithms for solving the multidimensional optimization problems.

The application of the Peano space-filling curve or evolvent $y(x)$ mapping the interval $[0, 1]$ onto an N-dimensional hypercube D unambiguously is one of the possible approaches to the dimensionality reduction used widely [14,17,32,36–38]. As a result of such reduction, initial multidimensional global optimization problem (1) is reduced to a one-dimensional problem:

$$\varphi(y(x^*)) = \min(\varphi(y(x)) : x \in [0, 1]), \tag{43}$$

where the function $\varphi(y(x))$ satisfies the uniform Hölder condition, i.e.

$$\left|\varphi(y(x_2)) - \varphi(y(x_1))\right| \le H\,|x_2 - x_1|^{1/N}, x_1, x_2 \in [0, 1], \tag{44}$$

where the Hölder constant H is defined by the relation $H = 2L\sqrt{N+3}$, L is the Lipschitz constant from (3) and N is the dimensionality of the optimization problem.

In this paper, it is proposed to apply another method of dimensionality reduction based on the nested dimensionality reduction scheme used widely as well [4,12,13,17,33,36–38]. According to this scheme, the solving of a multidimensional optimization problem (1) can be obtained by solving a series of nested one-dimensional problems:

$$\min\{\varphi(y) : y \in D\} = \min_{[a_1,b_1]} \cdots \min_{[a_N,b_N]} \varphi(y_1, \dots, y_N)\ . \tag{45}$$

The use of this scheme allows us to obtain a general scheme for applying the one-dimensional optimization algorithms to solving the multidimensional optimization problems. According to (45), the solving of a multidimensional optimization problem is reduced to solving a one-dimensional problem:

$$\varphi(y^*) = \min_{y \in D} = \min_{y_1 \in [a_1,b_1]} \widetilde{\varphi}_1(y_1), \tag{46}$$

where

$$\begin{aligned} &\widetilde{\varphi}_i(y_i) = \varphi_i(y_1 \dots y_i) = \min_{y_{i+1} \in [a_{i+1},b_{i+1}]} \varphi_i(y_1 \dots y_i, y_{i+1}), 1 \le i \le N, \\ &\varphi_N(y_1 \dots y_N) = \varphi(y_1 \dots y_N). \end{aligned} \tag{47}$$

The one-dimensional function in (46) is constructed according to a general recursive scheme – in order to compute the values $\widetilde{\varphi}_1(y_1)$ for some given value of the variable $y_1 = \widehat{y}_1$ it is necessary to minimize the function

$$\widetilde{\varphi}_2(y_2) = \varphi_2(\widehat{y}_1, y_2). \tag{48}$$

With respect to y_2 the function $\widetilde{\varphi}_2(y_2)$ is a one-dimensional one as well since the value of the variable y_1 is given and fixed one. Next, in turn, in order to compute the value of $\widetilde{\varphi}_2(y_2)$ at the point $y_2 = \widehat{y}_2$, it is necessary to minimize the function

$$\widetilde{\varphi}_3(y_3) = \varphi_3(\widehat{y}_1, \widehat{y}_2, y_3), \tag{49}$$

and so forth.

Additional information on the nested dimensionality reduction scheme and its applications for solving the multidimensional global optimization problems can be found, for example, in [4,12,13,33,37,38].

5. Results of numerical experiments

Initial estimation of the efficiency of developed approach performed for one-dimensional global optimization problem was presented in [9] (for AGMD) and in [11] (for AGMND). In these experiments, well-known methods were compared: the Galperin method (GM) [7], the Piyavskii–Shubert method (PM) [28,35], the Strongin Algorithm (GSA) [36,38], the Brent Algorithm (BA) [3], AGMD and AGMND. The method performance properties are evaluated by solving the set of 20 one-dimensional test global optimization problems accumulated in [18]. The Brent Algorithm requires estimations of the second derivatives (the Lipschitz constants from (4)) as a parameter and for this series of experiment it used exact upper bound of the second derivatives of the test problems.

The numerical results are given in Table 1 that presents the number of iterations executed by each algorithm to solve the test optimization problems with given accuracy (the last row shows the average number of executed iterations). To analyse the numerical results, one should take into account that at each iteration the value of the minimized function is calculated, except for the AGMD algorithm, for which the value of the first derivative is also calculated at each iteration. The results of experiments demonstrate a significant advantage of the algorithms that rely on fulfilment of the Lipschitz condition (4) for the first derivative of minimized functions. The efficiency of the algorithms AGMD and AGMND is higher than BA. It should be noted that AGMND demonstrates even better results than the one using the exact derivative values (AGMD).

The results of experiments demonstrate a significant advantage of the algorithms that rely on fulfilment of condition (4) for minimizing functions (i.e. limitation of the second derivative possible values). The efficiency of the algorithms AGMD and AGMND is higher than BA, since they use values of the first derivative during constructing minimizing

Table 1. The results of comparison of one-dimensional methods of global optimization.

	GA	PM	GSA	BA	AGMD	AGMND
1	377	149	127	43	27	16
2	308	155	135	24	27	13
3	581	195	224	153	98	50
4	923	413	379	16	27	15
5	326	151	126	45	23	14
6	263	129	112	123	39	22
7	383	153	115	23	25	13
8	530	185	188	148	88	47
9	314	119	125	44	26	12
10	416	203	157	27	25	12
11	779	373	405	47	41	29
12	746	327	271	30	37	23
13	1829	993	472	69	89	59
14	290	145	108	34	30	15
15	1613	629	471	50	47	41
16	992	497	557	109	75	49
17	1412	549	470	124	65	44
18	620	303	243	8	21	10
19	302	131	117	21	21	12
20	1412	493	81	99	32	24
Average number of trials	720.75	314.60	244.15	61.9	26.10	23.95

function minorant. It should be noted that AGMND demonstrates even better results than the one with the exact derivatives (AGMD).

Of course conclusions about the efficiency of the compared algorithms based on solving the 20 test optimization problems are quite preliminary. However, it can be noted that the problem of comparing one-dimensional global optimization algorithms is actively studied – some results of these studies can be found in [20,21,31,34]. These investigation also confirm high efficiency of the algorithms developed in the framework of information-statistical theory of multiextremal optimization (in particularly, the efficiency of the method GSA). Taking into account these results, and also the results of the experiments described above, it can be concluded that the algorithms with the best performance (namely, GSA, BA, AGMD and AGMND) can only be used in further experiments for solving the multidimensional optimization problems with the use of the nested dimensionality reduction scheme.

To evaluate the efficiency of solving the multidimensional problems, the first series of experiments was performed for the two-dimensional global optimization problems. The results of these experiments were presented in [15]. A family of well-known two-dimensional multiextremal test functions [36–38] is defined by the relations:

$$\varphi(y_1,y_2) = -\left\{\left(\sum_{i=1}^{7}\sum_{j=1}^{7}\left[A_{ij}a_{ij}(y_1,y_2)+B_{ij}b_{ij}(y_1,y_2)\right]\right)^2 + \left(\sum_{i=1}^{7}\sum_{j=1}^{7}\left[C_{ij}a_{ij}(y_1,y_2)+D_{ij}b_{ij}(y_1,y_2)\right]\right)^2\right\}^{\frac{1}{2}}, \tag{50}$$

$$a_{ij}(y_1,y_2) = \sin(\pi i y_1)\sin(\pi j y_2),$$
$$b_{ij}(y_1,y_2) = \cos(\pi i y_1)\cos(\pi j y_2),$$

where $0 \le y_1, y_2 \le 1$, were used. The values $-1 \le A_{ij}, Bij, C_{ij}, D_{ij} \le 1$ are the independent random generated parameters distributed uniformly over the interval $[-1, 1]$. The functions of this family are complicated enough and highly multiextremal, see an example in Figure 2.

In this series of experiments, in addition to GSA, BA, AGMD and AGMND, two combined methods (GSA-D and GSA-ND) were evaluated. In the last two methods, GSA was applied to optimize the reduced one-dimensional functions $\widetilde{\varphi}_1(y_1)$ from (47). It should be noted that BA requires estimations of the Lipschitz constants from (4) as method parameters. In case of the executed experiments, these parameters have been set by exact upper bounds of the second derivatives of the test optimization problems. In general, it is not applicable to the nested dimensionality reduction scheme due to non-smoothness of one-dimensional functions $\widetilde{\varphi}_1(y_1)$ from (47). And in this series of experiments BA use estimations of the Lipschitz constants calculated during AGMD execution. In addition, the experiments with the multidimensional variant of the GSA algorithm with the dimensionality reduction using Peano space-filling curves (43) were executed.

In all these methods, an adaptive scheme of computing the reliability parameter r from (8) was applied: $r = 3 + 10/k$ where k is the number of executed trials. The accuracy of solving the optimization problems was $\varepsilon = 0.001$. In order to obtain the reliable

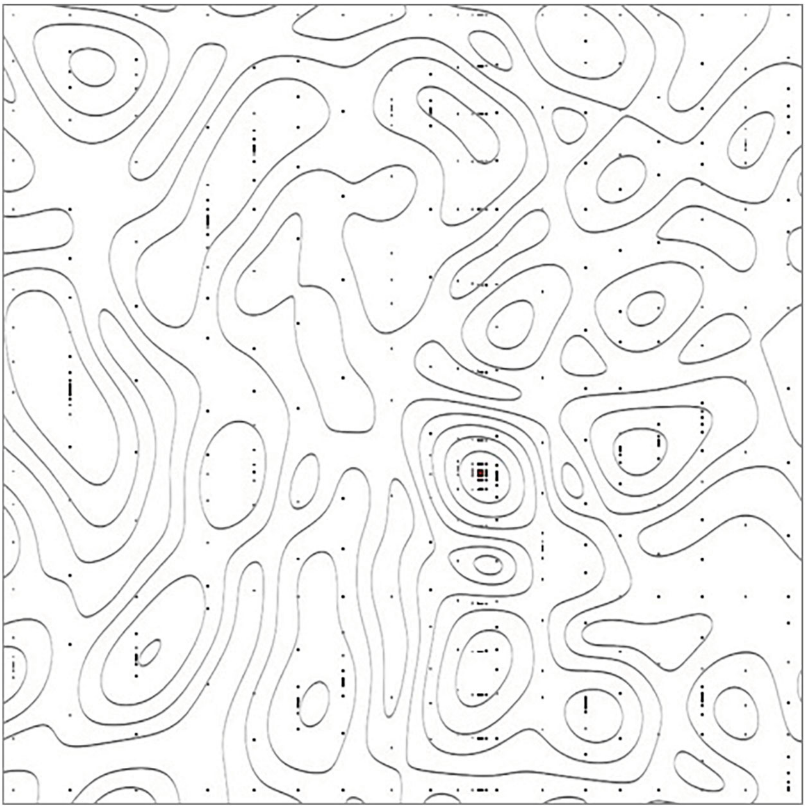

Figure 2. Contour plot of a test function and the trial points generated by AGMND (total 651 global search iterations were executed).

results, 100 multiextremal problems (50) were solved in every series of experiments. From the experimental results, the operational characteristics of the compared methods were constructed. The operational characteristic is a graph of the number of solved problems (the ordinate axis) versus the number of executed trials (the abscissa axis) and is defined by a set of pairs [36–38]:

$$\{< k, p(k) >\} \tag{51}$$

where k is the number of executed iterations and $p(k)$ is the fraction of test problems solved within k iterations. Such characteristics can be obtained from the results of numerical experiments and can be presented as the piecewise broken line graphs. In general, the operational characteristics can be treated as the indicators of probability to find the global minimum with the required accuracy subject to the number of iterations executed by the method.

The average number of iterations executed by each method until the stopping condition is satisfied when solving 100 test problems (50) is presented in Table 2. The operational characteristics of the compared methods are presented in Figure 3. These results demonstrated the AGMND method to show the best performance (the smallest number of trials). The GSA algorithm with the dimensionality reduction using Peano space-filling curves was

Table 2. Average number of executed iterations when solving 100 test problems (50).

	GSA	BA	GSA-D	GSA-ND	AGMD	AGMND	GSA Peano
Average number of trials	1974.75	2626.31	924.86	754.34	824.18	494.74	696.69

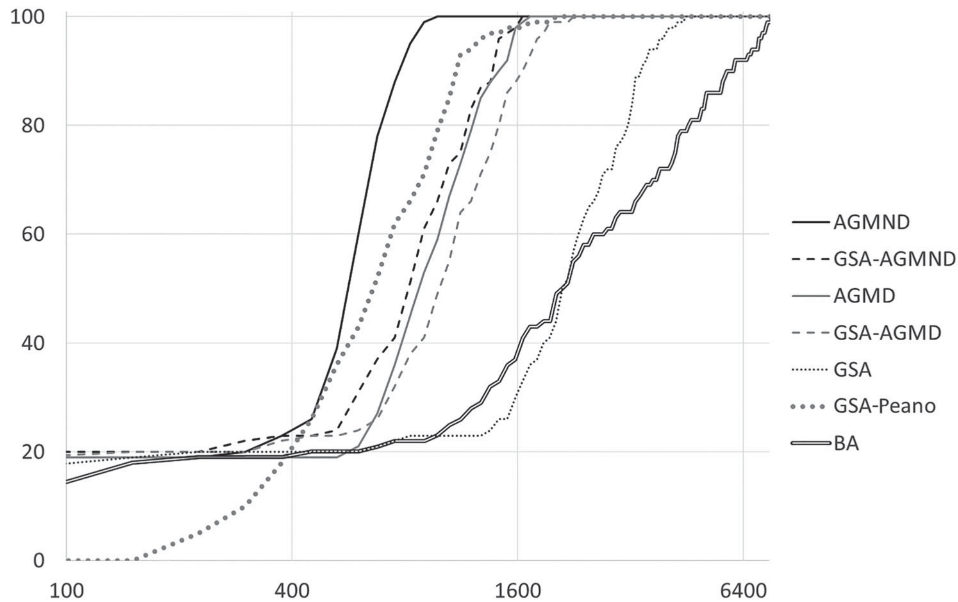

Figure 3. Operational characteristics of the compared optimization methods. The vertical axis is the percentage of problems solved with the required accuracy, the horizontal axis is the number of executed trials in logarithmic scale.

Table 3. Average number of executed iterations when solving 100 test problems generated by GKLS.

	AGMD $N = 3$	AGMND $N = 3$	AGMD $N = 4$	AGMND $N = 4$
Average number of trials	11011.89	10294.68	264175.50	203246.69

the next in efficiency. The combined GSA-ND algorithm and the AGMD method demonstrated close characteristics with a small advantage of GSA-ND. The efficiency indicators of BA were the smallest ones. Again, for the correct analysis of the obtained results, one should keep in mind that in the AGMD method each trial includes computing the derivatives whereas in the GSA, BA and AGMND algorithms the minimized functions are computed only.

In the next series of experiments, the multidimensional problems of higher dimensionality were solved. The test problems were constructed using the GKLS generator [8]. This generator allows us to obtain the global optimization problems with predefined properties: the number of local minima, the size of their attractors, the global minimum point,

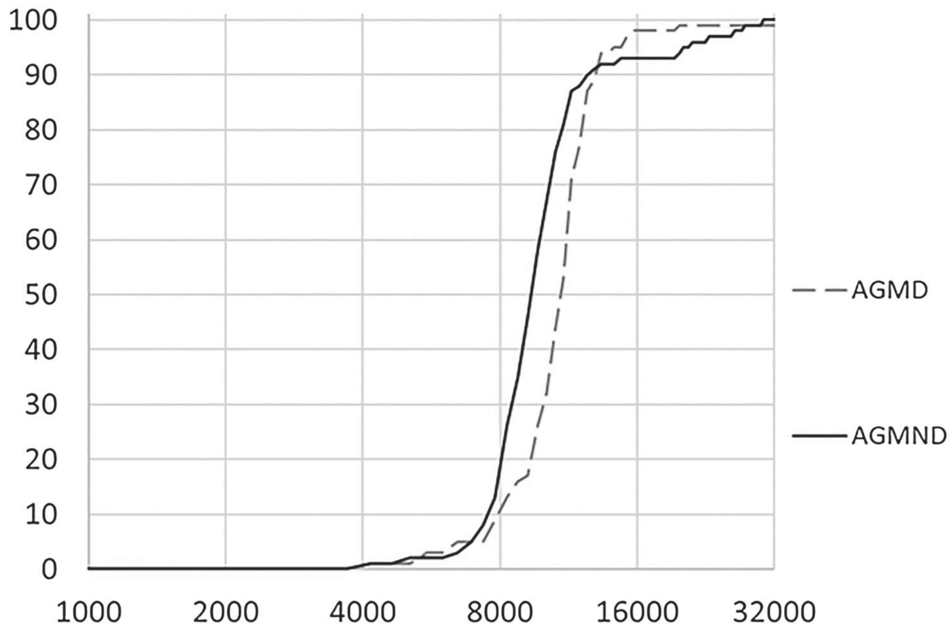

Figure 4. Comparison of the methods for three-dimensional global optimization problems. The vertical axis is a percentage of found solutions, the horizontal axis is a number of trials performed until the global minima is found in logarithmic scale.

the global minimum value, etc. In the current series of experiments, three- and four-dimensional functions were considered, total 100 functions of each dimensionality were generated. The average number of iterations executed by each method until the stopping condition is satisfied when solving 100 test problems is presented in Table 3. The AGMD and AGMND methods utilizing the derivative values were compared. The operational characteristics obtained from the results of solving the three-dimensional problems are presented in Figure 4. The operational characteristics for the four-dimensional problems are shown in Figure 5.

As shown in three-dimensional case, AGMND reached 100% solvability later than AGMD, but AGMND solved 90% of three-dimensional problems and all four-dimensional faster than the method with exact derivatives. And since in AGMD each trial included calculation of derivatives, while in AGMND only values of the functions are computed, AGMND is more efficient.

Analysing the efficiency of the proposed approach, one should keep in mind that the one-dimensional functions $\widetilde{\varphi}_i(y_i)$, $1 \le i < N$, from (47) (except the functions of the last decomposition level $\widetilde{\varphi}_N(y_N)$) in the nested reduction scheme can be non-smooth at some points, i.e. the derivative of these functions can be discontinuous at these points. An example of the function $\widetilde{\varphi}_1(y_1)$ for one of the problems of the family (50) is presented in Figure 6, the points of non-smoothness are marked by the sloping dashes.

The non-smoothness of the functions $\widetilde{\varphi}_i(y_i)$, $1 \le i < N$, raises the question of applicability of the AGMD and AGMND methods utilizing the derivatives in combination with the nested reduction scheme. Strictly speaking, the AGMD and AGMND methods can only be applied for minimizing the one-dimensional functions $\widetilde{\varphi}_N(y_N)$ at the last decomposition

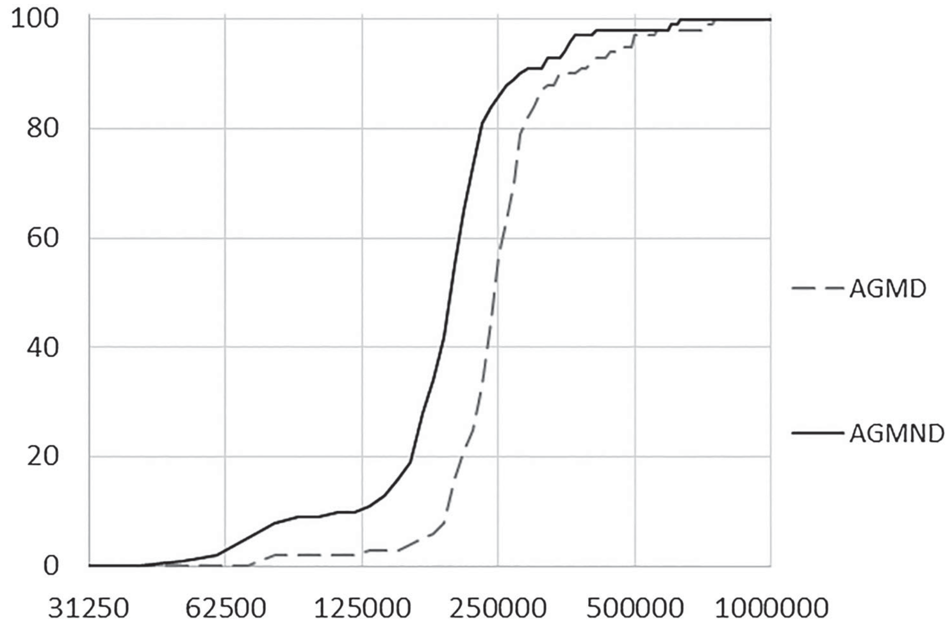

Figure 5. Comparison of the methods for four-dimensional global optimization problems. The vertical axis is a percentage of found solutions, the horizontal axis is a number of trials performed until the global minima is found in logarithmic scale.

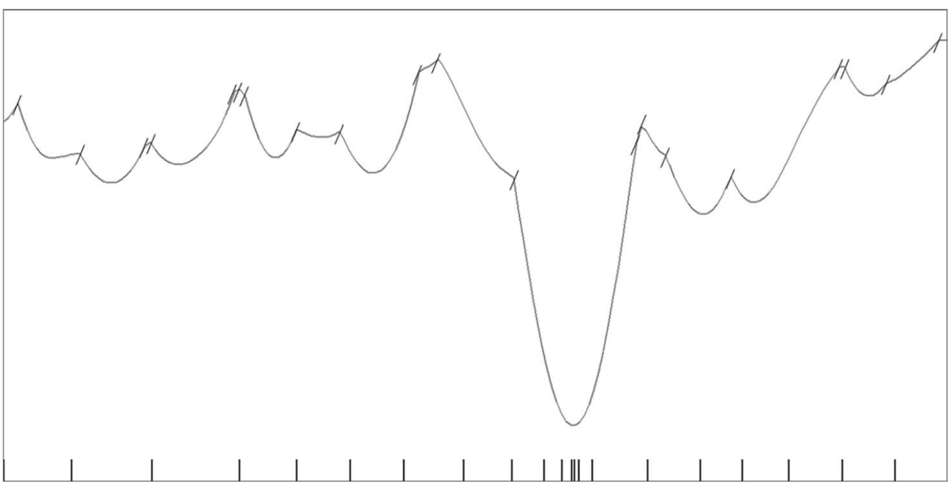

Figure 6. A graph of a reduced one-dimensional function $\widetilde{\varphi}_1(y_1)$ from (47) for one of the problems from the family (50). The points of non-smoothness are marked by the sloping dashes, the vertical dashes under the graph mark the trial points generated by AGMND.

level. This case was demonstrated by the application of the combined GSA-D and GSA-ND algorithms. However, the results of the experiments have shown the AGMD and AGMND methods to provide good efficiency and reliability even in the cases of the non-smooth reduced functions (see Figure 2). To demonstrate this phenomenon, the result of minimizing a reduced one-dimensional function $\widetilde{\varphi}_1(y_1)$ (47) from the test problem family (50) is shown in Figure 6. For more comprehensive evaluation of the effect of the non-smoothness of the reduced functions, an additional analysis of the experimental results was carried out. First of all, the maximum and minimum possible values of the Lipschitz constant estimate m_1 from (8) were computed for the 'true' one-dimensional functions $\widetilde{\varphi}_N(y_N)$ of the last decomposition level and for the reduced one-dimensional functions $\widetilde{\varphi}_1(y_1)$.

The computed estimates are shown in Figure 7. As follows from the presented data, the estimates of the Lipschitz constant when optimizing the reduced functions $\widetilde{\varphi}_1(y_1)$ exceeded the maximum values in 12% of cases only. Thus, for the considered family of test problems (50), the non-smoothness of the reduced functions does not affect greatly the estimates of the Lipschitz constant m_1 from (8).

Similar analysis is carried out for the number of executed trials when optimizing the functions $\widetilde{\varphi}_1(y_1)$ and $\widetilde{\varphi}_2(y_2)$. These data are demonstrated in Figure 8.

The presented data demonstrate the number of executed trials for the reduced one-dimensional functions $\widetilde{\varphi}_1(y_1)$ to exceed insufficiently the maximum number of trials for the 'true' one-dimensional functions $\widetilde{\varphi}_2(y_2)$ in 19% of cases only. The obtained results also confirm the non-smoothness of the reduced functions $\widetilde{\varphi}_1(y_1)$ for the considered test problem family (50) not to affect greatly the number of trials executed by the AGMND method to solve the one-dimensional optimization problems. But with increasing the dimensionality the number of discontinuity points can grow exponentially and increase the number of executed trials. To handle the discontinuity problem the modifications similar to proposed in [23] can be applied for smoothing derivatives in discontinuity points.

In general, the results of the numerical experiments allow us to conclude that the proposed approach is quite promising. No doubt, additional theoretical studies are necessary to investigate the convergence conditions for the global search algorithm utilizing the numerical estimations of derivatives taking into account possible non-smoothness of the reduced functions when using the nested dimensionality reduction scheme.

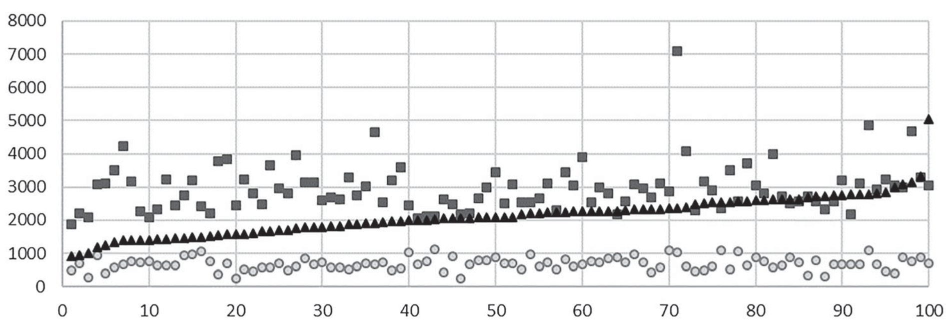

Figure 7. The distribution of the computed estimates of the Lipschitz constant when solving 100 test problems from the family (50). The vertical axis is the estimate of the Lipschitz constant, the horizontal axis is the index of the optimization problem (the presented results are sorted in the order of increasing the Lipschitz constant estimates calculated for the functions $\widetilde{\varphi}_1(y_1)$). The light grey circles mark the lower estimates of the Lipschitz constant for the functions $\widetilde{\varphi}_2(y_2)$, the dark grey squares are the upper estimates for the same functions, the black triangles are the estimates of the Lipschitz constant for the reduced one-dimensional functions $\widetilde{\varphi}_1(y_1)$.

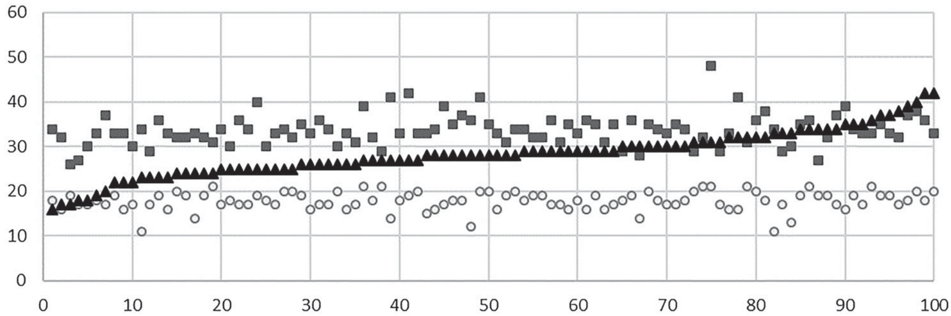

Figure 8. The distribution of the number of trials when solving 100 test problems from the family (50). The vertical axis is the number of executed trials, the horizontal axis is the index of the optimization problem (the presented results were sorted in the order of increasing the numbers of trials required for minimizing the functions $\widetilde{\varphi}_1(y_1)$). The light grey circles mark the minimum number of executed trials for the functions $\widetilde{\varphi}_2(y_2)$, the dark grey squares are the maximum ones for the same functions, the black triangles are the one for the reduced one-dimensional functions $\widetilde{\varphi}_1(y_1)$.

6. Conclusion

In this paper, an efficient method for solving the computation-costly multidimensional global optimization problems is proposed. The developed method is based on the approach where not only the computed values of the minimized functions but also the values of derivatives of the functions are used to increase the efficiency of the global optimization. The necessary values of derivatives are estimated numerically on the base of the available search information.

In order to substantiate the developed approach, the sufficient convergence conditions for the one-dimensional version of the proposed method have been formulated. In the multidimensional case, the nested dimensionality reduction scheme has been applied. Using this scheme, the solving of the multidimensional optimization problems is reduced to solving a sequence of one-dimensional global search problems.

For the experimental evaluation of efficiency of the developed approach, the numerical experiments are executed. In these experiments, 300 global optimization problems were solved. The results of experiments confirm the developed approach to be promising. Its application allows reducing the amount of computations required to solve even the most difficult optimization problems.

In addition, a possible effect of the non-smoothness of the reduced one-dimensional functions when using the nested dimensionality reduction scheme was investigated. The results of experiments demonstrated the non-smoothness (discontinuities of derivatives) not to affect highly the applicability (and the efficiency) of the developed approach. Nevertheless, the situation when the discontinuities of the derivatives are possible requires additional investigations.

Further development of the studies can include carrying out the additional numerical experiments for various classes of test problems of various dimensionality and the application of the parallel algorithms. Also, possible combining the developed approach with the adaptive extension of the nested dimensionality reduction scheme seems to be promising.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The reported study was funded by Russian Foundation for Basic Research (RFBR) according to the research project no. 19-07-00242.

ORCID

Victor Gergel http://orcid.org/0000-0002-4013-2329
Alexey Goryachih http://orcid.org/0000-0002-0066-3678

References

[1] W. Baritompa, *Accelerations for a variety of global optimization methods*, J. Global Optim. 4 (1994), pp. 37–45.
[2] L. Breiman and A. Cutler, *A deterministic algorithm for global optimization*, Math. Program. 58 (1993), pp. 179–199.
[3] R.P. Brent, *Algorithms for Minimization Without Derivatives*, Prentice-Hall, Englewood Cliffs, NJ, 1973.
[4] E.R. Dam, B. Husslage and D. Hertog, *One-dimensional nested maximin designs*, J. Global Optim. 46 (2010), pp. 287–306.
[5] C.A. Floudas and M.P. Pardalos, *State of the Art in Global Optimization: Computational Methods and Applications*, Kluwer Academic Publishers, Dordrecht, 1996.
[6] C.A. Floudas and M.P. Pardalos, *Recent Advances in Global Optimization*, Princeton University Press, Princeton, 2016.
[7] E.A. Galperin, *The cubic algorithm*, J. Math. Anal. Appl. 112 (1985), pp. 635–640.
[8] M. Gaviano, D. Lera, D.E. Kvasov and Y.D. Sergeyev, *Software for generation of classes of test functions with known local and global minima for global optimization*, ACM Trans. Math. Softw. 29 (2003), pp. 469–480.
[9] V.P. Gergel, *A method of using derivatives in the minimization of multiextremum functions*, Comput. Math. Math. Phys. 36 (1996), pp. 729–742. (in Russian)
[10] V.P. Gergel, *A global optimization algorithm for multivariate function with Lipschitzian first derivatives*, J. Global Optim. 10 (1997), pp. 257–281.
[11] V.P. Gergel and A.S. Goryachih, *Global optimization using numerical approximations of derivatives*, Learn. Intell. Optim. LION 2017. Lect. Notes Comput. Sci. 10556 (2017), pp. 320–325.
[12] V.P. Gergel, V.A. Grishagin and A.V. Gergel, *Adaptive nested optimization scheme for multidimensional global search*, J. Global Optim. 66 (2015), pp. 35–51.
[13] V.P. Gergel, V. Grishagin and R. Israfilov, *Local tuning in nested scheme of global optimization*, Procedia. Comput. Sci. 51 (2015), pp. 865–874.
[14] A.S. Goryachih, *A class of smooth modification of space-filling curves for global optimization problems*, Model. Algorithms, Technol. Netw. Anal. 197 (2016), pp. 57–65.
[15] A.S. Goryachih and M.A. Rachinskaya, *Multidimensional global optimization method using numerically calculated derivatives*, Procedia Comput. Sci. 119 (2017), pp. 90–96.
[16] A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, SIAM, Philadelphia, 2008.
[17] V. Grishagin, R. Israfilov and Y.D. Sergeev, *Convergence conditions and numerical comparison of global optimization methods based on dimensionality reduction schemes*, Appl. Math. Comput. 318 (2018), pp. 270–280.
[18] P. Hansen, B. Jaumard and S.H. Lu, *Global optimization of univariate Lipshitz functions: II. New algorithms and computational comparison*, Math. Program. 55 (1992), pp. 273–292.

[19] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches*, Springer-Verlag, Berlin, 1990.
[20] D. Lera and Y.D. Sergeyev, *Acceleration of univariate global optimization algorithms working with Lipschitz functions and Lipschitz first derivatives*, SIAM J. Optim. 23 (2013), pp. 508–529.
[21] M. Locatelli, *Bayesian algorithms for one-dimensional global optimization*, J. Global Optim. 10 (2001), pp. 57–76.
[22] M. Locatelli and F. Schoen, *Global Optimization: Theory, Algorithms, and Applications*, SIAM, Philadelphia, 2013.
[23] I.V. Mayurova and R.G. Strongin, *Minimization of a multi-extremum function with a discontinuity*, USSR Comput. Math. Math. Phys. 24 (1984), pp. 121–126.
[24] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, New York, 2006.
[25] M.P. Pardalos, A.A. Zhigljavsky and J. Žilinskas, *Advances in Stochastic and Deterministic Global Optimization*, Springer, Cham, 2016.
[26] R. Paulavičius and J. Žilinskas, *Simplicial Global Optimization*, Springer Briefs in Optimization. Springer, 2014.
[27] J.D. Pintér, *Global Optimization in Action (Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications)*, Kluwer Academic Publishers, Dordrecht, 1996.
[28] S. Piyavskij, *An algorithm for finding the absolute extremum of a function*, Comput. Math. Math. Phys. 12 (1972), pp. 57–67. (in Russian)
[29] Y.D. Sergeyev, *Global one-dimensional optimization using smooth auxiliary functions*, Math. Program. 81 (1998), pp. 127–146.
[30] Y.D. Sergeyev, *A deterministic global optimization using smooth diagonal auxiliary functions*, Commun. Nonlinear Sci. Numer. Simul. 21 (2015), pp. 99–111.
[31] Y.D. Sergeyev, M.S. Mukhametzhanov and D.E. Kvasov, *Operational zones for comparing metaheuristic and deterministic one-dimensional global optimization algorithms*, Math. Comput. Simul. 141 (2017), pp. 96–109.
[32] Y.D. Sergeyev, R.G. Strongin and D. Lera, *Introduction to Global Optimization Exploiting Space-Filling Curves*, Springer, New York, 2013.
[33] L. Shi and S. Ólafsson, *Nested partitions method for global optimization*, Oper. Res. 48 (2000), pp. 390–407.
[34] A. Shpak, *Global optimization in one-dimensional case using analytically defined derivatives of objective function*, Comput. Sci. J. Moldova 3 (1995), pp. 168–184.
[35] B.O. Shubert, *A sequential method seeking the global maximum of a function*, SIAM J. Numer. Anal. 9 (1972), pp. 379–388.
[36] R.G. Strongin, *Numerical Methods in the Multiextremal Problems (information-statistical algorithms)*, Nauka, Moscow, 1978.(in Russian)
[37] R.G. Strongin, V.P. Gergel, V.A. Grishagin and K.A. Barkalov, *Parallel Computations in the Global Optimization Problems*, MSU Publishing, Moscow, 2013. (in Russian)
[38] R.G. Strongin and Y.D. Sergeyev, *Global Optimization with Non-convex Constraints: Sequential and Parallel Algorithms*, Kluwer Academic Publishers, Dordrecht, 2000. 2nd ed. 2013, 3rd ed. 2014.
[39] A. Zhigljavsky and A. Žilinskas, *Stochastic Global Optimization*, Springer, Berlin, 2008.