

Global optimization method with dual Lipschitz constant estimates for problems with non-convex constraints

Roman Strongin, Konstantin Barkalov, Semen Bevzuk

Lobachevsky State University of Nizhni Novgorod, Nizhni Novgorod, Russia

Abstract

This paper considers the constrained global optimization problems, in which the functions are of the “black box” type and satisfy the Lipschitz condition. The algorithms for solving the problems of this class require the use of adequate estimates of the a priori unknown Lipschitz constants for the problem functions. A novel approach presented in this paper is based on a simultaneous use of two estimates of the Lipschitz constant: an overestimated and an underestimated one. The upper estimate provides the global convergence whereas the lower one reduces the number of trials necessary to find the global optimizer with the required accuracy. In this case, the considered algorithm for solving the constrained problems doesn’t use the ideas of the penalty function method; each constraint of the problem is accounted for separately. The convergence conditions of the proposed algorithm are formulated in the corresponding theorem. The results of the numerical experiments on a series of multiextremal problems with non-convex constraints demonstrating the efficiency of the proposed scheme of dual Lipschitz constant estimates are presented.

Keywords: Global optimization, Multiextremal problems, Non-convex constraints, Lipschitz constant estimates

1. Introduction

In this paper, we consider constrained global optimization problems and the numerical methods for solving problems of this type. Such problems are often

encountered in applications (see, for example, the review [1]). As a rule, problem
 5 functions are defined by a program code, i.e. they are “black-box” functions, for
 which the computing of the values can be a computationally costly operation
 since in applied problems it will require numerical modeling. Any possibility to
 estimate reliably the global optimum in the multiextremal problems with “black-
 box” functions is based fundamentally on the auxiliary assumptions relating the
 10 possible values of the problem functions to their known values at the points of
 the trials already performed.

The assumption that the problem functions satisfy the Lipschitz condition
 is typical for many applied problems (in this case, one can speak about the Lip-
 schitzian optimization problems) because the ratio of the functions’ variations
 15 to corresponding ones of the variables usually cannot exceed some threshold
 determined by the limited energy of variations in the system. This threshold
 can be described using the Lipschitz constant.

The methods for solving the Lipschitzian optimization problems constitute a
 very important area in the development of the global optimization methods and
 20 are the objects of investigations of many researchers. The first algorithms of
 this class were proposed as early as in the 1970s [2, 3, 4, 5]; since that time, this
 area has been continuing to develop extensively (see, for example [6, 7, 8, 9, 10]).
 Note that these global optimization algorithms are, as a rule, oriented towards
 solving either unconstrained optimization problems [? ? ? ?] or problems
 25 with constraints of a simple structure [11, 12]. The solving of the problems
 with complex non-convex constraints is usually performed with the use of the
 penalty functions method [13, 14, 15], which has a number of drawbacks. In
 particular, this method cannot be applied to the problems with the partially
 defined functions (i.e. when any constraint is violated, the values of all the
 30 other functions of the problem remain undefined).

This situation is typical of many optimal design problems, when, in case of
 violation of some conditions of the modeled object’s functioning, other charac-
 teristics of the object appear to be undefined. For example, the modeled object
 is an electronic device, and its response time is minimized. However, the re-

35 sponse time is undefined if at a given combination of parameters the device fails to operate. Such a partial computability of the functions in the constrained optimization problems complicates essentially (in some cases, makes impossible) the application of the penalty functions method.

Within the framework of our research, an original approach to the minimiza-
 40 tion of the multiextremal functions with non-convex constraints called the index method of the accounting for the constraints [?] was used. This approach is based on separate accounting for each constraint of the problem and does not involve the use of the penalty functions. According to the rules of the index method, every iteration called trial at the corresponding point of the search do-
 45 main includes a sequential checking of the feasibility of the problem constraints at this point. When the first violation of any constraint is found, the trial is terminated (the values of the remaining problem functions are not computed at this point) and the transition to the next iteration point is initiated. This allows solving the problems with partially defined functions. The experimental com-
 50 parison of the index method of accounting for the constraints and the penalty function method demonstrated the advantages of the index method in terms of the number of iterations and the number of computations of the problem functions [16, 17].

Note that the values of the Lipschitz constant for the problem functions
 55 are usually not known a priori, which makes their estimation one of the key problems in the construction of the Lipschitzian optimization methods. The methods utilizing the values of the Lipschitz constant predefined a priori (see, for example, [3, 4, 18, 19, 20]) are important in the theoretical aspect, but it is difficult to apply such methods in applied problems (in which the information
 60 of the Lipschitz constant values is absent).

The value of the unknown Lipschitz constant estimate affects the convergence rate of the algorithm essentially. Therefore, the issue of its correct estimate is very important. An underestimate of the true value of the constant may lead to the loss of the algorithm convergence to the global solution. At the same time,
 65 too high a value of the constant estimate not corresponding to real behavior of

the function results in slow convergence of the algorithm to the global minimum point.

Several typical methods for adaptive estimation of the Lipschitz constant according to the results of the performed search trial are known:

- 70 – global estimate of the constant L in the whole search domain D [21, 22, 8].
- local estimates of the constants L_i in different subdomains D_i of the search domain D [23, 24, 25].
- choice of the estimates of the constant L from some set of possible values [26, 27, 10, 28].

75 Each of the approaches listed above has its own advantages and drawbacks. For example, the use of the global estimate only in the entire search domain may slow down the convergence of the algorithm to the global minimum point. On the other hand, the use of the local estimates accelerating the convergence requires an adequate adjustment of the algorithm parameters in order to ensure
80 the global convergence.

 In this paper, we consider an algorithm, in which it is proposed to use two global estimates of the Lipschitz constant, one of them being much greater than the other one. The larger estimate ensures global convergence and the smaller one reduces the total number of trials needed to find the global optimizer. The
85 choice of one of the two estimates to be used in the rules of the algorithm is performed in the adaptive manner, depending on the behavior of the function. This work continues the research reported in the paper [?], in which the preliminary results have been obtained for the unconstrained problems.

 The main part of the paper has the following structure. In Section 2, the
90 index method of accounting for the constraints and the dimensionality reduction scheme for the constrained global optimization problems are described. In Section 3, a description of the computational rules of the index global search algorithm is given and its theoretical properties are formulated. In Section 4, the index algorithm with dual Lipschitz constant estimates is formulated

95 and its convergence is proved. Section 5 presents the results of the numerical experiments carried out on several series of the multiextremal problems with non-convex constraints. Section 6 concludes the paper.

2. Index Method of Accounting for the Constraints and Dimensionality Reduction

100 We consider the global optimization problems of the form

$$\min \{\varphi(y) : y \in D, g_j(y) \leq 0, 1 \leq j \leq m\}, \quad (1)$$

$$D = \{y \in R^N : a_i \leq y_i \leq b_i, 1 \leq i \leq N\}. \quad (2)$$

It is supposed that the objective function $\varphi(y)$ (henceforth denoted by $g_{m+1}(y)$) and the left-hand sides $g_j(y), 1 \leq j \leq m$, of the constraints satisfy the Lipschitz condition

$$|g_j(y') - g_j(y'')| \leq L_j \|y' - y''\|, y', y'' \in D, 1 \leq j \leq m + 1. \quad (3)$$

105 with *a priori* unknown constant $L_j, 1 \leq j \leq m + 1$, and may be multiextremal.

The analytical form of the functions $g_j(y)$ may be unknown, i.e. these functions may be defined algorithmically (“black-box” functions). In this case, it is supposed that even a single search trial may be a computationally costly operation since in applied problems it involves the need to perform numerical
110 modeling.

Along with the exact solution y^* of the problem (1), we will consider also the ε -reserved solution y_ε , for which

$$\varphi(y_\varepsilon) = \min \{\varphi(y) : y \in D, g_j(y) \leq -\varepsilon_j, 1 \leq j \leq m\}, \quad (4)$$

and the set of solutions Y_ε , which are not worse than the ε -reserved solution in the objective function value, i.e.

$$Y_\varepsilon = \{y \in D, g_j(y) \leq 0, 1 \leq j \leq m, \varphi(y) \leq \varphi(y_\varepsilon)\}. \quad (5)$$

115 Fig. 1 illustrates the concepts introduced.

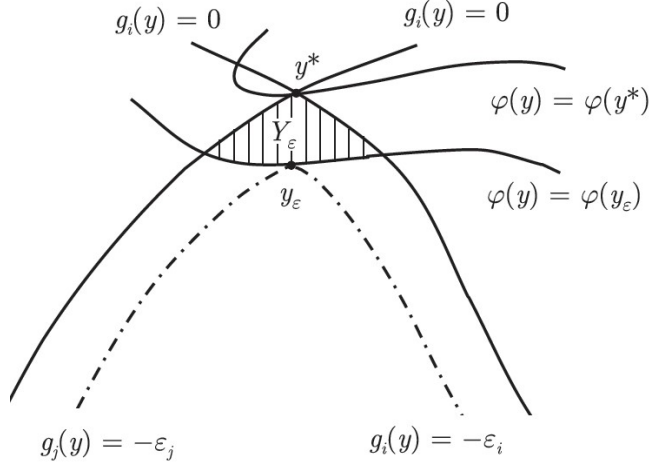


Figure 1: ε -reserved solution

The existence of the ε -reserved solutions of problem (1) can be interpreted as some analog of the regularity condition in classical nonlinear programming problems. The applied role of this condition should be also noted. Even if the exact solution y^* is found, its practical implementation is possible as some
120 approximation of y^* only. Therefore, it is important that there are feasible points from Q_{m+1} close enough to y^* (in coordinates and values). The existence of the ε -reserved solution guarantees the existence of such points. In this case, the domain Y_ε may play the role of a set of suitable approximations.

We will consider the initial problem (1) assuming that the problem functions
125 $g_j(y), 1 \leq j \leq m+1$ may be defined only partly. It means that they are defined and calculable only in the sets

$$D_1 = D, D_{j+1} = \{y \in D_j : g_j(y) \leq 0\}, 1 \leq j \leq m. \quad (6)$$

This situation is typical for the optimal design problems when some conditions for the functioning of the modeled object are not met and its other characteristics appear to be undefined. In this approach, the first violation of any
130 constraint indicating that the point ybeing considered is no longer feasible terminates further computations at the point y .

Many well known approaches to solving the multidimensional global optimization problems are based on either explicit or implicit reduction of multidimensional problems to one-dimensional ones (see, for example, the description
135 of methods using the diagonal partitions [28] or simplicial partitions [29]). In this paper, we will use the approach based on the dimensionality reduction with the use of the space-filling curve $y(x)$ (the Peano curve). Continuous unambiguous mapping $y(x)$ projecting the interval $[0,1]$ onto an N -dimensional hypercube D allows reducing a multidimensional constrained minimization problem in the
140 domain D to a one-dimensional constrained optimization problem in the interval $[0,1]$

$$\varphi(y(x^*)) = \min \{ \varphi(y(x)) : x \in [0, 1], g_j(y(x)) \leq 0, 1 \leq j \leq m \}, \quad (7)$$

The issues of the numerical construction of the Peano space filling curve-type mappings and the corresponding theory were described in detail in [8, 9]. Note that the theoretical Peano curve $y(x)$ is a limit object and in practice is replaced
145 by its approximation (called *evolvent*) with predefined precision 2^{-m} where the number $m > 1$ is a parameter of the construction (the density of the evolvent). In this case, the construction density of the evolvent is determined by the accuracy of the possible practical implementation of the problem solution found. For illustration, Fig. 2 (a), (b) shows the projections of the interval $[0,1]$ are
150 shown for the mapping $y(x)$ with $N = 2$ and $m = 4, 5$, correspondingly.

The dimensionality reduction scheme considered here matches a multidimensional problem with Lipschitzian minimized function and Lipschitzian constraints to a one-dimensional problem, in which the corresponding functions satisfy the uniform Hölder condition (see [8, 9]), i.e.

$$|g_j(y(x')) - g_j(y(x''))| \leq K_j |x' - x''|^{1/N}, \quad x', x'' \in [0, 1], \quad 1 \leq j \leq m + 1,$$

where N is the dimensionality of the initial multidimensional problem and the coefficients K_j are related to the Lipschitz constants L_j of the initial problem by the relations

$$K_j \leq 2L_j \sqrt{N+3}, \quad 1 \leq j \leq m + 1. \quad (8)$$

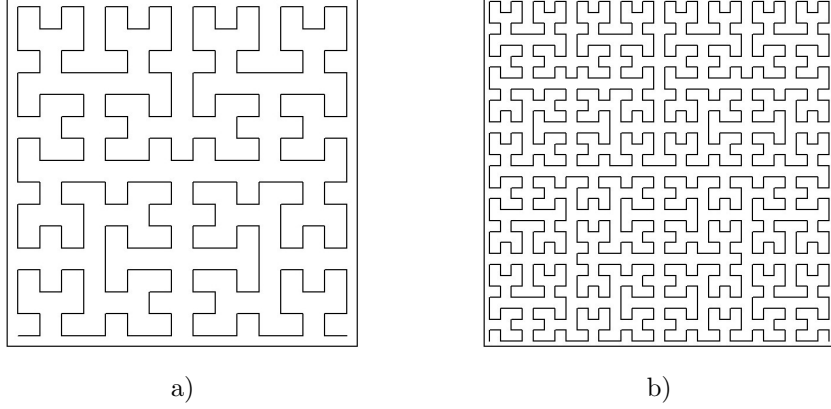


Figure 2: Evolvent for $N = 2$ and $m = 4$ (a), $m = 5$ (b)

According to (6) and (7), the reduced one-dimensional functions $g_j(y(x))$, $1 \leq$
155 $j \leq m + 1$ will also be defined and calculable in the corresponding intervals

$$Q_1 = [0, 1], Q_{j+1} = \{x \in Q_j : g_j(y(x)) \leq 0\}, 1 \leq j \leq m. \quad (9)$$

These relations allow introducing a classification of the points $x \in [0, 1]$ according to the number $\nu = \nu(x)$ of the constraints calculable at the given point. We will call this number $\nu = \nu(x)$ the *index* of the point x . The index can be determined from the conditions

$$g_j(y(x)) \leq 0, 1 \leq j \leq \nu - 1, g_\nu(y(x)) > 0,$$

where the last inequality is insufficient if $\nu = m + 1$.

Thus, a search trial at some point $x^k \in [0, 1]$ performed at the k^{th} iteration of the algorithm will consist of the following sequence of operations:

1. Determine the image $y^k = y(x^k)$ with the use of the evolvent $y(x)$.
- 160 2. Compute the values $g_1(y^k), \dots, g_\nu(y^k)$ where $\nu = \nu(x)$ is the index of the point.
3. Fix the result of the trial as three values

$$x^k, z^k = g_\nu(y(x^k)), \nu = \nu(x^k), \quad (10)$$

According to this scheme, the trial is terminated at the point $y^k = y(x^k)$ when the first violated constraint is found.

165 In the case when the point y^k is a feasible one, i.e. when $y(x^k) \in Q(m+1)$, the trial involves checking all constraints and computing the objective function. Here, the value of the index is accepted to be equal to the quantity $\nu = m+1$.

The main idea of the index scheme of accounting for the constraints consists in the reduction of the constrained problem (7) (the functions of which satisfy
170 the Hölder condition with corresponding coefficients K_j) to an unconstrained problem of the form

$$\psi(x^*) = \min \{ \psi(x) : x \in [0, 1] \}, \quad (11)$$

$$\psi(x) = \frac{g_\nu(y(x))}{K_\nu} - \begin{cases} 0, & \nu \leq m, \\ \frac{\varphi^*}{K_{m+1}}, & \nu = m+1. \end{cases}$$

where $\nu = \nu(x)$ and φ^* is the objective function value at the point of the problem solution (7).

The value $\psi(x)$ at $\nu(x) \leq m$ matches the value of the left-hand part of the first constraint violated at this point $g_\nu(y(x))$ divided by the constant K_ν . In the case when $\nu(x) = m+1$

$$\psi(x) = \frac{\varphi(y(x)) - \varphi^*}{K_{m+1}}$$

175 and, consequently, the set of the points, in which the function $\psi(x)$ turns to zero matches the set of the global minimum points of the problem (7). It should be noted that in any subdomain Q_j the function $\psi(x)$ has the Hölder coefficient equal to 1.

These considerations constitute the basis of the index global search algo-
180 rithm, according to which the unknown constants K_ν and the unknown minimum value φ^* are substituted in the course of computations by their current approximations.

3. Index-Based Global Search Algorithm

The index-based global search algorithm under consideration for solving the
 185 problem (1) (here it is given according to [8]) involves the construction of a
 sequence of points $x^k \in (0, 1)$, at which the search trials are conducted with the
 results of the type (10). The set of triplets $\{x^k, z^k = g_\nu(y(x^k)), \nu = \nu(x^k)\}$, $1 \leq$
 $k \leq n$ makes up the search information accumulated by the method after per-
 forming n steps. The operation of the index-based global search algorithm is
 190 defined by the following rules.

The first trial is executed at an arbitrary internal point $x^1 \in (0, 1)$. The
 selection of the point x^{k+1} , $k \geq 1$, of any next trial is determined by the following
 steps.

Step 1. Renumber the points x^1, \dots, x^k of the preceding trials by the lower
 indices in the increasing order of the coordinate values, i.e.

$$0 = x_0 < x_1 < \dots < x_k < x_{k+1} = 1,$$

and juxtapose to them the values $z_i = g_\nu(y(x_i))$, $\nu = \nu(x_i)$, $1 \leq i \leq k$, computed
 at these points. The points $x_0 = 0$ and $x_{k+1} = 1$ are introduced additionally
 and the values z_0 and z_{k+1} are not defined.

Step 2. Classify the indices i , $1 \leq i \leq k$, of the trial points in accordance
 with the number of the constraints of the problem fulfilled at these points by
 constructing the sets

$$I_\nu = \{i : 1 \leq i \leq k, \nu = \nu(x_i)\}, \quad 1 \leq \nu \leq m + 1,$$

including the numbers of all points x_i , $1 \leq i \leq k$ with the same values of ν . The
 end points $x_0 = 0$ and $x_{k+1} = 1$ are interpreted as the ones having the indices
 equal to zero. An additional set $I_0 = \{0, k + 1\}$ corresponds to these points.

Determine the maximum value of the index

$$M = \max \{\nu = \nu(x_i), \quad 1 \leq i \leq k\}.$$

195 Step 3. Compute the current lower bounds

$$\mu_\nu = \max \left\{ \frac{|z_i - z_j|}{(x_i - x_j)^{1/N}}, i, j \in I_\nu, i > j \right\}, \quad (12)$$

for the unknown Hölder constants K_ν of the functions $g_\nu, 1 \leq \nu \leq m+1$. If the set I_ν includes less than two elements or if μ_ν appears to be equal to zero, then set $\mu_\nu = 1$.

Step 4. For all nonempty sets $I_\nu, 1 \leq \nu \leq m+1$, compute the estimates

$$z_\nu^* = \begin{cases} -\epsilon_\nu, & \nu < M, \\ \min\{z_i : i \in I_\nu\}, & \nu = M, \end{cases} \quad (13)$$

200 where the nonnegative numbers $(\epsilon_1, \dots, \epsilon_m)$ are the algorithm parameters.

Step 5. For each interval $(x_{i-1}, x_i), 1 \leq i \leq k+1$, compute the *characteristics* $R(i)$:

$$R(i) = \Delta_i + \frac{(z_i - z_{i-1})^2}{r_\nu^2 \mu_\nu^2 \Delta_i} - 2 \frac{z_i + z_{i-1} - 2z_\nu^*}{r_\nu \mu_\nu}, \nu = \nu(x_i) = \nu(x_{i-1}), \quad (14)$$

$$R(i) = 2\Delta_i - 4 \frac{z_i - z_\nu^*}{r_\nu \mu_\nu}, \nu = \nu(x_i) > \nu(x_{i-1}), \quad (15)$$

$$R(i) = 2\Delta_i - 4 \frac{z_{i-1} - z_\nu^*}{r_\nu \mu_\nu}, \nu = \nu(x_{i-1}) > \nu(x_i), \quad (16)$$

205 where $\Delta_i = (x_i - x_{i-1})^{1/N}$. The values $r_\nu > 1, 1 \leq \nu \leq m+1$, are the parameters of the algorithm. With an appropriate selection of r_ν one can use the product $r_\nu \mu_\nu$ as an estimate of Hölder constants $K_\nu, 1 \leq \nu \leq m+1$.

Step 6. Find the interval (x_{t-1}, x_t) with the maximum characteristic

$$R(t) = \max \{R(i) : 1 \leq i \leq k+1\}. \quad (17)$$

Step 7. Execute the next trial at the middle point of the interval (x_{t-1}, x_t) if the indices of the points x_{t-1}, x_t are not the same, i.e.,

$$x^{k+1} = \frac{x_t + x_{t-1}}{2}, \nu(x_{t-1}) \neq \nu(x_t).$$

Otherwise, execute a trial at the point

$$x^{k+1} = \frac{x_t + x_{t-1}}{2} + \frac{\text{sign}(z_t - z_{t-1})}{2r_\nu} \left[\frac{|z_t - z_{t-1}|}{\mu_\nu} \right]^N, \nu = \nu(x_{t-1}) = \nu(x_t).$$

The inequality $\Delta_t \leq \epsilon$ can be taken as the termination condition, where t is
 210 from (17) and $\epsilon > 0$ is the predefined accuracy.

The convergence conditions for the index global search algorithm can be formulated in the form of the following theorem from [8].

Theorem. Assume that the following conditions are satisfied:

1. Problem (1) has an ε -reserved solution y_ε from (4).
2. Functions $g_j(y)$, $1 \leq j \leq m+1$ admit Lipschitzian (with the constants L_j) extensions $G_j(y)$ over the whole feasible domain D from (2), i.e.

$$g_j(y(x)) = G_j(y(x)), \quad x \in Q_j, \quad 1 \leq j \leq m+1,$$

215 where the sets Q_j are from (9), and $y(x)$ is the curve from (7).

3. The parameters of the method ϵ_ν , $1 \leq \nu \leq m$ used in the rule (13) are the corresponding components of the reserve vector ϵ_R from (4).
4. For a sufficiently large iteration number q , the values μ_ν from (12) satisfy the inequalities

$$r_\nu \mu_\nu > 2^{3-\frac{1}{N}} L_\nu \sqrt{N+3}, \quad 1 \leq \nu \leq m+1, \quad (18)$$

Then, any limit point \bar{y} of the sequence of the trials $\{y^k\} = \{y(x^k)\}$ generated by the index algorithm for problem (1) belongs to the set Y_ϵ form (5) and satisfies the conditions

$$\varphi(\bar{y}) = \inf \{ \varphi(y^q) : g_j(y^q) \leq 0, 1 \leq j \leq m, q = 1, 2, \dots \} \leq \varphi(y_\epsilon).$$

220 **Note 1.** The computational scheme of the algorithm is described in the form convenient for formulating its theoretical properties. The software implementation of this scheme may be made more frugal. For example, according to Step 7, the point x^{k+1} of the scheduled trial will be an internal point of the interval (x_{t-1}, x_t) , i.e. $x^{k+1} \in (x_{t-1}, x_t)$. Therefore, the set of trial points x^1, \dots, x^k
 225 (and the related trial results (10)) may not be ordered at every step of the algorithm but may be maintained in an ordered state using appropriate data structures.

Note 2. The use of the set of parameters $\epsilon_1, \dots, \epsilon_m$ (the values of which correspond to the coordinates of the reserve vector ϵ_R) in the index algorithm will significantly reduce the density of the trial sequence in the vicinity of the feasible domain boundaries. However, this raises a question about the particular values of the reserve vector components ϵ_R . In [8], it was proposed to choose the components of the vector ϵ_R according to the formula

$$\epsilon_\nu = \mu_\nu \delta, \quad 1 \leq \nu \leq m, \quad (19)$$

where μ_ν are the current estimates of the Hölder constant from (12) and $\delta > 0$ is a parameter. This approach to choosing the reserve vector components allows one to take into account the properties of different problem functions characterized by their Lipschitz constants. This method for choosing the reserve vector was used in the numerical experiments, the results of which are presented below.

Note 3. The index algorithm for solving the constrained problems is based on the information-statistical algorithm of global search (AGS) for solving unconstrained problems. AGS has been compared repeatedly with both deterministic and nature-inspired global optimization algorithms. The results of the experimental comparison (see, for example, [30]) show that AGS is superior to many known methods of similar purpose, including DIRECT and DIRECTL.

At the same time, the index scheme of accounting for the constraints also demonstrated its advantages (in the number of computed values of the problem functions) over the traditional penalty functions method in solving a wide range of model problems (see the results of the experimental comparison in [16, 17].)

4. Index Algorithm with Dual Lipschitz Constant Estimates

The index global search algorithm described in the previous section is intended for solving the multiextremal problems, in which the objective function and the functions of constraints satisfy the Lipschitz condition. In this case, *a priori* estimates of the constant values are not required for the algorithm to work. These estimates are computed in the course of global search, based on the available search information.

According to Theorem 1 from Section 3, the sequence of trial points $\{y^k\}$ will converge to a global minimizer y^* if the condition (18) is satisfied. So far, an appropriate choice of the parameters r_ν from (15) allows using the values

$$\frac{r_\nu \mu_\nu}{2^{3-\frac{1}{N}} L_\nu \sqrt{N+3}}, \quad 1 \leq \nu \leq m+1, \quad (20)$$

as the estimates of the Lipschitz constants for the problem functions $g_j(y)$, $1 \leq j \leq m+1$.

Compliance with condition (20) will be achieved when choosing a high value of the parameters r_ν . However, in this case the constant will be overestimated, and the method will perform a large number of trials until the stopping condition is fulfilled.

The choice of low values of the parameters r_ν (which corresponds to an underestimate of the Lipschitz constant) would reduce the number of trials considerably but could destroy the convergence to the global extremum (due to possible violation of the condition (18)).

An approach, in which two estimates of the Lipschitz constant for each function are used in the rules of the algorithm, is promising. One of these estimates is significantly greater than the other, which corresponds to the use of two sets of parameters r_ν^{glob} and r_ν^{loc} in the algorithm where $r_\nu^{glob} > r_\nu^{loc} > 1$, $1 \leq \nu \leq m+1$. When using the parameter r_ν^{loc} we shall deal with the smaller estimate of the Lipschitz constant, and using the parameter r_ν^{glob} will correspond to the larger one.

We will refer to the characteristics of the i^{th} interval as the global ones ($R_{glob}(i)$) if the parameters $r_\nu = r_\nu^{glob}$ were used when computing the characteristic according to (14), (15), (16) and as the local ones ($R_{loc}(i)$) if the parameters $r_\nu = r_\nu^{loc}$ were used.

The key idea of the novel method is computing two characteristics for each interval and balancing (based on these two) the local and global search. The local search provides a new trial in the vicinity of the current point y_k^* whereas during the global search, a new trial could be performed in a subdomain located far away from y_k^* .

285 Note that Theorem 1 on the convergence of the index method implies the
generation of an infinite sequence of trials $\{y^k\} = \{y(x^k)\}$ by the algorithm.
In this case, at the final stage of the search (i.e. in the case when one of the
points of this trial sequence falls into the vicinity of the ε -reserved solution y^*)
there always exists an interval, at one of the boundaries of which the current
290 minimum value of the objective function z_{m+1}^* from (13) is achieved. It is this
interval that will have the highest characteristic since it is the most promising
one for performing further trials.

Let us establish the following important property for the global and local
characteristics of an interval containing the current optimal point.

295 **Statement.** Let us assume that the current minimum value z_M^* from (13) at
the k^{th} iteration is achieved at the boundary point of some interval (x_{i-1}, x_i) , $i =$
 $i(k)$. Then, $R_{glob}(i) \geq R_{loc}(i)$.

Proof. Let us consider two possible cases of the values of the indices $\nu(x_{i-1})$
and $\nu(x_i)$.

300 1. If $\nu(x_{i-1}) \neq \nu(x_i)$, then according to formulae (15), (16)

$$R(i) = R_{glob}(i) = R_{loc}(i) = 2\Delta_i > 0. \quad (21)$$

2. If $\nu(x_{i-1}) = \nu(x_i) = M$, then, taking into account the rules (12), the
following inequality will hold

$$z_i + z_{i-1} - 2z_M^* = |z_i - z_{i-1}| \leq \mu_M \Delta_i$$

From here, it will follow that for any value $r_M > 1$ the following inequalities
will hold

$$\begin{aligned} R(i) &= \Delta_i + \frac{(z_i - z_{i-1})^2}{r_M^2 \mu_M^2 \Delta_i} - 2 \frac{z_i + z_{i-1} - 2z_M^*}{r_M \mu_M} \geq \Delta_i + \frac{(z_i - z_{i-1})^2}{r_M^2 \mu_M^2 \Delta_i} - 2 \frac{\mu_M \Delta_i}{r_M \mu_M} \\ &\geq \Delta_i + \frac{(z_i - z_{i-1})^2}{r_M^2 \mu_M^2 \Delta_i} - 2 \frac{\Delta_i}{r_M} = \Delta_i + \Delta_i \frac{(z_i - z_{i-1})^2 / \Delta_i^2}{r_M^2 \mu_M^2 \Delta_i^2} - 2 \frac{\Delta_i}{r_M} \\ &\geq \Delta_i + \frac{\Delta_i}{r_M^2} - 2 \frac{\Delta_i}{r_M} = \Delta_i \left(1 - \frac{1}{r_M}\right)^2 > 0 \end{aligned}$$

And because $r_M^{glob} > r_M^{loc} > 1$ then

$$R_{glob}(i) = \Delta_i \left(1 - \frac{1}{r_M^{glob}}\right)^2 > \Delta_i \left(1 - \frac{1}{r_M^{loc}}\right)^2 = R_{loc}(i) \quad (22)$$

The statement is proved.

305 This statement can be interpreted as the incomparability of the local and global characteristics: the global characteristic of the best interval is always larger. At the same time, it follows from the formulae (21), (22) that

$$R_{glob}(i) = \rho R_{loc}(i), \quad (23)$$

where $\rho = \left(\frac{1-1/r_M^{glob}}{1-1/r_M^{loc}}\right)^2$, if $\nu(x_{i-1}) = \nu(x_i) = M$, and $\rho = 1$, if $\nu(x_{i-1}) \neq \nu(x_i)$. In the case of the multiplication by the coefficient ρ , the local and global
310 characteristics of the search intervals become comparable, and one can choose the best (i.e. the largest) one.

Based on the relation (23), one can formulate an *index algorithm with dual estimate of the Lipschitz constants* (hereafter referred to as IA-DL).

The first four steps of the algorithm repeat Step 1 – Step 4 of the initial index
315 algorithm from Section 3. Step 5. For each interval (x_{i-1}, x_i) , $1 \leq i \leq k+1$

- Compute the global characteristics $R_{glob}(i)$ according to formulae (14), (15), (16) using the value $r_\nu = r_\nu^{glob}$
- Compute the local characteristics $R_{loc}(i)$ according to formulae (14), (15), (16) using the value $r_\nu = r_\nu^{loc}$
- 320 – Determine the characteristic $R(i)$ as

$$R(i) = \max \{R_{glob}(i), \rho R_{loc}(i)\} \quad (24)$$

where $\rho = \left(\frac{1-1/r_M^{glob}}{1-1/r_M^{loc}}\right)^2$, if $\nu(x_{i-1}) = \nu(x_i) = \nu$, and $\rho = 1$, if $\nu(x_{i-1}) \neq \nu(x_i)$.

Step 6. Find the interval (x_{t-1}, x_t) with the maximum characteristic

$$R(t) = \max \{R(i) : 1 \leq i \leq k+1\} \quad (25)$$

and fix the value $r_\nu = r_\nu^{loc}$ if $\rho R_{loc}(t) > R_{glob}(t)$, otherwise fix $r_\nu = r_\nu^{glob}$.

Step 7. Execute the next trial at the middle point of the interval (x_{t-1}, x_t) if the indices of the points x_{t-1} , x_t are not the same, i.e.

$$x^{k+1} = \frac{x_t + x_{t-1}}{2}, \quad \nu(x_{t-1}) \neq \nu(x_t).$$

Otherwise, execute a trial at the poin

$$x^{k+1} = \frac{x_t + x_{t-1}}{2} + \frac{\text{sign}(z_t - z_{t-1})}{2r_\nu} \left[\frac{|z_t - z_{t-1}|}{\mu_\nu} \right]^N, \quad \nu = \nu(x_{t-1}) = \nu(x_t).$$

325 When computing the point x^{k+1} , use the value of the parameter r_ν fixed at Step 6. The algorithm terminates if the condition $\Delta_t \leq \epsilon$ is satisfied where t is from (25) and $\epsilon > 0$ is the predefined accuracy.

Let us illustrate the work of the algorithms IA and IA-DL when minimizing the function

$$\begin{aligned} \varphi(y_1, y_2) = & -1.5y_1^2 \exp \{1 - y_1^2 - 20.25(y_1 - y_2)^2\} - \\ & - (0.5(y_1 - 1)(y_2 - 1))^4 \exp \{2 - (0.5(y_1 - 1))^4 - (y_2 - 1)^4\} \end{aligned}$$

330 in the range $0 \leq y_1 \leq 4$, $1 \leq y_2 \leq 3$ subject to the constraints

$$\begin{aligned} g_1(y_1, y_2) &= 0.01 ((y_1 - 2.2)^2 + (y_2 - 1.2)^2 - 2.25) \leq 0, \\ g_2(y_1, y_2) &= 100 (1 - (y_1 - 2)^2/1.44 + (0.5y_2)^2) \leq 0, \\ g_3(y_1, y_2) &= 10 (y_2 - 1.5 - 1.5 \sin(6.283(y_1 - 1.75))) \leq 0 \end{aligned}$$

The feasible domain of this problem is not connected and consists of three non-convex subdomains. The global minimum $\varphi(y^*) = -1.489$ is achieved at the point $y^* = (0.942, 0.944)$ (the values are given with the accuracy of 10^{-3}).

335 In Fig. 3, the search domain (square) is shown, the feasible domain is highlighted by green, and the trial points generated by IA (a) and IA-DL (b) are marked. The best point is marked by red. For both methods, the evolvent with the parameter $m = 10$ was used, the accuracy of the search for the solution $\epsilon = 0.002$; the adaptive reserve parameter $\delta = 0.008$ from (19). For the IA method, the parameters $r_\nu = 2.3, 1 \leq \nu \leq 4$ were used. The same values were

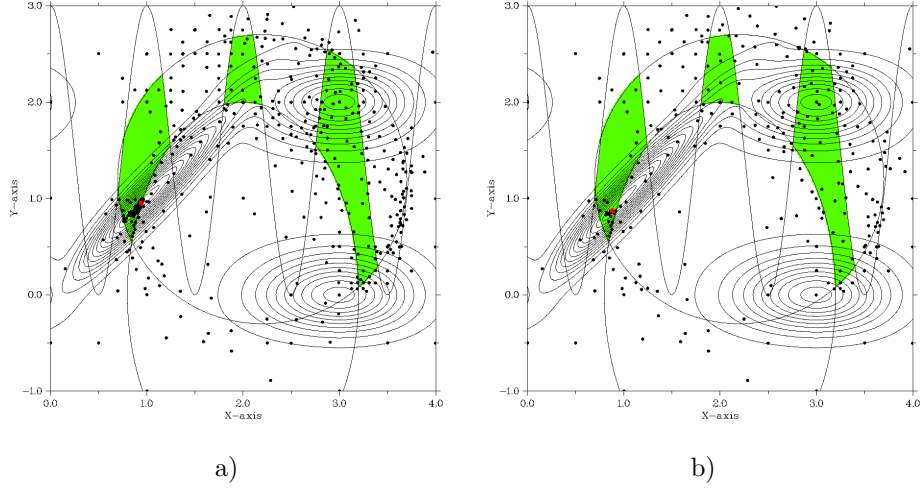


Figure 3: Solving a problem by IA (a) and IA-DL (b) methods

used for the IA-DL method as the parameters $r_\nu^{glob}, 1 \leq \nu \leq 4$. Also, the parameters $r_\nu^{loc} = 1.5, 1 \leq \nu \leq 4$ were used. To solve the problem with a predefined accuracy, 478 trials were required for the IA method compared to only 303 trials for the IA-DL method. The results of larger-scale numerical experiments are presented in Section 4.

Let us formulate and prove the convergence conditions for the new algorithm IA-DL.

Theorem 2. Let us assume the index algorithm (IA) is applied to solve the problem (1) and for the values $r_\nu = r_\nu^{glob}, 1 \leq \nu \leq m + 1$ the conditions of Theorem 1 are fulfilled. Then, the algorithm with dual estimate of the Lipschitz constants (IA-DL) with the parameters $r_\nu^{glob} > r_\nu^{loc} > 1, 1 \leq \nu \leq m + 1$ when solving the same problem also generates a sequence of the trial points $\{y^k\} = \{y(x^k)\}$, any limit point \bar{y} of which belongs to the set Y_ϵ from (9), i.e. satisfies the condition

$$\varphi(\bar{y}) = \inf \{ \varphi(y^k) : g_j(y^k) \leq 0, 1 \leq j \leq m, k = 1, 2, \dots \} \leq \varphi(y_\epsilon). \quad (26)$$

Proof.

Because for the IA with the parameters $r_\nu = r_\nu^{glob}, 1 \leq \nu \leq m + 1$ the

conditions of Theorem 1 are fulfilled, then any limit point $\bar{y} = y(\bar{x})$ of the trial sequence generated by the IA for the problem (1) satisfies the conditions (26).

Let us show that \bar{y} is also the limit point of the sequence of trials $\{y^k\} = \{y(x^k)\}$, generated by the IA-DL, and the sequence $\{y^k\}$ has no new limit points.

360 Let us carry out the proof by contradiction. Assume that \bar{y} is not a limit point of the sequence $\{y^k\}$ and there exists a limit point $y' \neq \bar{y}$.

At every iteration of the IA-DL method there will exist such interval $[x_{j-1}, x_j]$, $j = j(k)$ that $\bar{x} \in [x_{j-1}, x_j]$. According to our assumption that \bar{x} is not the limit point, there will exist such index k , starting from which the trial points will
 365 not fall into the interval $[x_{j-1}, x_j]$. That is, to conduct a new trial at the point $x^{k+1} \in [x_{t-1}, x_t]$ at every iteration such interval $[x_{t-1}, x_t]$, $t = t(k)$ will be selected that $t \neq j$.

Therefore, in the formula (24) for computing the characteristic, the second component takes the highest values, i.e.

$$R_{glob}(t) < \rho R_{loc}(t)$$

Then, there will exist a limit point $x' \neq \bar{x}$, at that either $\nu(x') \leq m$ (i.e. the point $y(x')$ is not a feasible one) or $z_{m+1}^* = g_{m+1}(y(x')) > g_{m+1}(y(\bar{x}))$ (i.e. the
 370 value of the objective function at the limit point $y(x')$ is greater than at the point $y(\bar{x})$).

Let us show that if the conditions of Theorem 1 are satisfied, the inequality $R_{glob}(j) > 0$ will hold. According to the second condition of Theorem 1 and to the condition (8),

$$z_j = g_\nu(y(x_j)) \leq g_\nu(y(\bar{x})) + 2L_\nu \sqrt{N+3}(x_j - \bar{x})^{1/N}, \quad \nu = \nu(x_j)$$

$$z_{j-1} = g_\nu(y(x_{j-1})) \leq g_\nu(y(\bar{x})) + 2L_\nu \sqrt{N+3}(\bar{x} - x_{j-1})^{1/N}, \quad \nu = \nu(x_{j-1})$$

Since the limit point $\bar{u} = y(\bar{x})$ belongs to the set Y_ϵ , then

$$g_\nu(y(\bar{x})) \leq -\epsilon_\nu, \quad 1 \leq \nu \leq m.$$

$$g_{m+1}(y(\bar{x})) \leq z_{m+1}^*$$

Then, the characteristic $R_{glob}(j)$ of the interval $[x_{j-1}, x_j], j = j(k)$ will be strictly positive, i.e. $R_{glob}(j) > 0$. Indeed, at $\nu = \nu(x_j) = \nu(x_{j-1})$ and the value $r_\nu = r_\nu^{glob}$

$$\begin{aligned}
R(j) &= \Delta_i + \frac{(z_j - z_{j-1})^2}{r_\nu^2 \mu_\nu^2 \Delta_i} - 2 \frac{z_j + z_{j-1} - 2z_\nu^*}{r_\nu \mu_\nu} \\
&\geq \Delta_i + 4 \frac{z_\nu^* - \left(g_\nu(y(\bar{x})) + L_\nu \sqrt{N+3} \left((x_j - \bar{x})^{\frac{1}{N}} + (\bar{x} - x_{j-1})^{\frac{1}{N}} \right) \right)}{r_\nu \mu_\nu} \\
&= \Delta_i - 4 \frac{L_\nu \sqrt{N+3} \left((x_j - \bar{x})^{\frac{1}{N}} + (\bar{x} - x_{j-1})^{\frac{1}{N}} \right)}{r_\nu \mu_\nu} + 4 \frac{z_\nu^* - g_\nu(y(\bar{x}))}{r_\nu \mu_\nu} \\
&= \Delta_i - 4 \frac{L_\nu \sqrt{N+3} \left(\alpha^{\frac{1}{N}} + (1-\alpha)^{\frac{1}{N}} \right)}{r_\nu \mu_\nu} + 4 \frac{z_\nu^* - g_\nu(y(\bar{x}))}{r_\nu \mu_\nu} \\
&\geq \Delta_i \left(1 - 4 \frac{L_\nu \sqrt{N+3} \max_{0 \leq \alpha \leq 1} \left(\alpha^{\frac{1}{N}} + (1-\alpha)^{\frac{1}{N}} \right)}{r_\nu \mu_\nu} \right) + 4 \frac{z_\nu^* - g_\nu(y(\bar{x}))}{r_\nu \mu_\nu} \\
&= \Delta_i \left(1 - 4 \frac{2^{3-1/N} L_\nu \sqrt{N+3}}{r_\nu \mu_\nu} \right) + 4 \frac{z_\nu^* - g_\nu(y(\bar{x}))}{r_\nu \mu_\nu} \geq 0
\end{aligned}$$

375 At $\nu = \nu(x_j) > \nu(x_{j-1})$ and the value $r_\nu = r_\nu^{glob}$

$$\begin{aligned}
R(j) &= 2\Delta_i - 4 \frac{z_j - z_\nu^*}{r_\nu \mu_\nu} \geq 2\Delta_i + 4 \frac{z_\nu^* - \left(g_\nu(y(\bar{x})) + 2L_\nu \sqrt{N+3} (x_j - \bar{x})^{\frac{1}{N}} \right)}{r_\nu \mu_\nu} \\
&= 2\Delta_i - 8 \frac{L_\nu \sqrt{N+3} (x_j - \bar{x})^{\frac{1}{N}}}{r_\nu \mu_\nu} + 4 \frac{z_\nu^* - g_\nu(y(\bar{x}))}{r_\nu \mu_\nu} \\
&\geq 2\Delta_i \left(1 - 4 \frac{L_\nu \sqrt{N+3}}{r_\nu \mu_\nu} \right) + 4 \frac{z_\nu^* - g_\nu(y(\bar{x}))}{r_\nu \mu_\nu} > 0
\end{aligned}$$

The case of $\nu = \nu(x_{j-1}) > \nu(x_j)$ is treated similarly and leads to the same inequality $R(j) > 0$.

Since we have assumed that x' is a limit point for the IM-DL and $x' \neq \bar{x}$, then there exists such a sequence of the intervals $[x_{t-1}, x_t], t = t(k)$ that $x' \in$
380 $[x_{t-1}, x_t], R_{loc}(t) > 0$, and $R_{loc}(t) \rightarrow 0$. This property of the characteristics of the intervals shrinking to the limit point has been established when proving the theorem on the convergence of the IA in [9].

Therefore, for large enough, the following inequality will be true

$$R_{glob}(j) > \rho R_{loc}(t) \quad (27)$$

and (according to Rules 5 and 6 of the IA-DL method) the next iteration falls
 385 into the interval $[x_{j-1}, x_j]$, which contradicts our assumption regarding the non-
 limit nature of the point \bar{x} . The theorem is proven.

5. Numerical experiments

6. TEST

[2, 3, 4, 5, 6, 7, 8, 9, 22, 10, 18, 19, 20, 11, 13, 12, 14, 15, 16, 17, 28, 29, 30,
 390 23, 24, 25, 21, 26, 27, 31, 32, 33, 34, 35, 36, 37].

References

- [1] J. Pinter, Global Optimization: Scientific and Engineering Case Studies, Springer, New York, 2006.
- [2] Y. Evtushenko, Numerical methods for finding global extreme (case of a
 395 non-uniform mesh), Comput. Maths. Math. Phys. 11 (6) (1971) 38–54.
- [3] S. Piyavskii, An algorithm for finding the absolute extremum of a function, Comput. Maths. Math. Phys. 12 (4) (1972) 57–67.
- [4] B. Shubert, A sequential method seeking the global maximum of a function, SIAM Journal on numerical analysis 9 (3) (1972) 379–388.
- 400 [5] R. Strongin, On a certain global-minimization algorithm, Radiophysics and Quantum Electronics 13 (4) (1970) 419–424.
- [6] Y. Evtushenko, V. Malkova, A. A. Stanevichyus, Parallel global optimization of functions of several variables, Comput. Math. Math. Phys. 49 (2) (2009) 246–260.
- 405 [7] Y. Evtushenko, M. Posypkin, A deterministic approach to global box-constrained optimization, Optim. Lett. 7 (2013) 819–829.

- [8] Strongin R.G., Sergeyev Ya.D., Global optimization with non-convex constraints. Sequential and parallel algorithms, Kluwer Academic Publishers, Dordrecht, 2000.
- 410 [9] Sergeyev, Y.D., Strongin, R.G., Lera, D., Introduction to global optimization exploiting space-filling curves, Springer Briefs in Optimization, Springer, New York, 2013.
- [10] Jones, D.R., The DIRECT global optimization algorithm, in: The Encyclopedia of Optimization, Springer, Heidelberg, 2009, pp. 725–735.
- 415 [11] Vaz, A. I.F. and Vicente, L.N., PSwarm: a hybrid solver for linearly constrained global derivative-free optimization, Optimization Methods and Software 24 (4-5) (2009) 669–685.
- [12] R. Paulavičius, J. Žilinskas, Advantages of simplicial partitioning for Lipschitz optimization problems with linear constraints, Optim. Lett. 10 (2) 420 (2016) 237–246.
- [13] R. Stripinis, R. Paulavičius, J. Žilinskas, Penalty functions and two-step selection procedure based DIRECT-type algorithm for constrained global optimization, J. Struct Multidisc Optim 59 (6) (2019) 2155.
- [14] G. Di Pillo, S. Lucidi, F. Rinaldi, An approach to constrained global optimization based on exact penalty functions, J Glob Optim 54 (2012) 251. 425
- [15] G. Di Pillo, G. Liuzzi, S. Lucidi, A DIRECT-type approach for derivative-free constrained global optimization, Comput Optim Appl 65 (2016) 361.
- [16] K. Barkalov, I. Lebedev, Comparing two approaches for solving constrained global optimization problems, Lecture Notes in Computer Science 10556 430 (2017) 301–306.
- [17] K. Barkalov, I. Lebedev, Parallel algorithm for solving constrained global optimization problems, Lecture Notes in Computer Science 10421 (2017) 396–404.

- [18] G. Wood, Multidimensional bisection applied to global optimisation, Comput. Math. Appl. 21 (6-7) (1991) 161–172.
- [19] C. Meewella, D. Mayne, An algorithm for global optimization of lipschitz continuous functions, J. Optim. Theory Appl. 57 (2) (1988) 307–322.
- [20] R. Mladineo, An algorithm for finding the global maximum of a multimodal multivariate function, Math. Program. 34 (2) (1986) 188–200.
- [21] R. Horst, H. Tuy, Global Optimization – Deterministic Approaches, Springer-Verlag, Berlin, 1996.
- [22] J. Pinter, Global optimization in action (Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications), Kluwer Academic Publishers, Dordrecht, 1996.
- [23] Kvasov, D.E., Pizzuti, C., Sergeyev, Y.D., Local tuning and partition strategies for diagonal GO methods, Numer. Math. 94 (1) (2003) 93–106.
- [24] Lera, D., Sergeyev, Ya.D., An information global minimization algorithm using the local improvement technique, J. Glob. Optim. 48 (1) (2010) 99–112.
- [25] Sergeyev, Y.D., Mukhametzhanov, M.S., Kvasov, D.E., Lera, D., Derivative-free local tuning and local improvement techniques embedded in the univariate global optimization, J. Optim. Theory Appl. 171 (1) (2016) 186–208.
- [26] Gablonsky, J.M., Kelley, C.T., A locally-biased form of the DIRECT algorithm, J. Glob. Optim. 21 (1) (2001) 27–37.
- [27] D. Jones, C. Perttunen, B. Stuckman, Lipschitzian optimization without the Lipschitz constant, J. Optim. Theory Appl. 79 (1) (1993) 157–181.
- [28] Sergeyev, Y.D., Kvasov, D.E., Global search based on efficient diagonal partitions and a set of Lipschitz constants, SIAM J. Optim. 16 (3) (2006) 910–937.

- [29] J. Žilinskas, Branch and bound with simplicial partitions for global optimization, *Math. Model. Anal.* 13 (1) (2008) 145–159.
- [30] V. Sovrasov, Comparison of several stochastic and deterministic derivative-free global optimization algorithms, *Lecture Notes in Computer Science* 11548 (2019) 70–81.
- [31] Gaviano, M., Kvasov, D.E., Lera, D., Sergeyev, Ya.D., Software for generation of classes of test functions with known local and global minima for global optimization, *ACM Transactions on Mathematical Software* 29 (4) (2003) 469–480.
- [32] K. Barkalov, R. Strongin, Solving a set of global optimization problems by the parallel technique with uniform convergence, *J. Glob. Optim.* 71 (1) (2018) 21–36.
- [33] Paulavičius, R., Sergeyev, Y.D., Kvasov, D.E., Žilinskas, J., Globally-biased DISIMPL algorithm for expensive global optimization, *J. Glob. Optim.* 59 (2-3) (2014) 545–567.
- [34] Sergeyev, Y.D., Kvasov, D.E., A deterministic global optimization using smooth diagonal auxiliary functions, *Commun. Nonlinear. Sci. Numer. Simulat.* 21 (1-3) (2015) 99–111.
- [35] R. Strongin, V. Gergel, K. Barkalov, A. Sysoyev, Generalized parallel computational schemes for time-consuming global optimization, *Lobachevskii Journal of Mathematics* 39 (4) (2018) 576–586.
- [36] V. Gergel, An approach for generating test problems of constrained global optimization, *Lecture Notes in Computer Science* 10556 (2017) 314–319.
- [37] V. Gergel, K. Barkalov, I. Lebedev, M. Rachinskaya, A. Sysoyev, A flexible generator of constrained global optimization test problems, *AIP Conference Proceedings* 2070 (2019) 20009.