Convergence conditions and numerical comparison of global optimization methods based on dimensionality reduction schemes

Vladimir Grishagin[a,*], Ruslan Israfilov[a], Yaroslav Sergeyev[a,b]

[a]*Department of Software and Supercomputing, Lobachevsky State University, Gagarin Avenue 23, 603950 Nizhni Novgorod, Russia.*
[b]*DIMES, Universita della Calabria, Italy.*

Abstract

This paper is devoted to numerical global optimization algorithms applying several ideas to reduce the problem dimension. Two approaches to the dimensionality reduction are considered. The first one is based on the nested optimization scheme that reduces the multidimensional problem to a family of one-dimensional subproblems connected in a recursive way. The second approach as a reduction scheme uses Peano-type space-filling curves mapping multidimensional domains onto one-dimensional intervals. In the frameworks of both the approaches, several univariate algorithms belonging to the characteristical class of optimization techniques are used for carrying out the one-dimensional optimization. Theoretical part of the paper contains a substantiation of global convergence for the considered methods. The efficiency of the compared global search methods is evaluated experimentally on the well-known GKLS test class generator used broadly for testing global optimization algorithms. Results for representative problem sets of different dimensions demonstrate a convincing advantage of the adaptive nested optimization scheme with respect to other tested methods.

Keywords: Multiextremal functions, global optimization, numerical methods, dimensionality reduction, convergence, comparison of efficiency.
2010 MSC: 65K05, 90C26

*Corresponding author
Email addresses:* `vagris@unn.ru` (Vladimir Grishagin), `ruslan@israfilov.com` (Ruslan Israfilov), `yaro@dimes.unical.it` (Yaroslav Sergeyev)

1. Introduction

In this paper the black-box global optimization problem

$$f^* = f(y^*) = \min_{y \in P} f(y), \tag{1}$$

$$P = \{y \in \mathbb{R}^N : a_i \leq y_i \leq b_i, \ 1 \leq i \leq N\}, \tag{2}$$

is considered as a problem of finding the global minimum value f^* and global minimizers $y^* \in P$ of a real-valued multivariate function $f(y)$ in the hyperparallelepiped (2) of the Euclidean space $\mathbb{R}^N$. The objective function $f(y)$ is supposed to satisfy in the domain P the Lipschitz condition

$$|f(y') - f(y'')| \leq L\|y' - y''\|, \ y', y'' \in P, \tag{3}$$

where $L > 0$ is a finite constant and $\| * \|$ denotes the Euclidean norm in $\mathbb{R}^N$. In general case, these problems are multiextremal and non-smooth.

The global optimization problem under consideration has been drawing attention of many researchers (see, for example, fundamental monographs [1–6]). On the one hand, problems of this kind are very important from the practical point of view because they arise often in scientific and engineering applications (see, for instance, [4, 7–13]). On the other hand, Lipschitzian problems (1)–(2) are very interesting for theoretical study because they have a rich variety of properties and, as a consequence, there is no a "universal" algorithm for solving multiextremal problems. These circumstances generate many fruitful approaches for solving this class of problems. Given approaches are based on ideas of different nature (both stochastic [5, 6, 14–17] and deterministic [18–30]), but, in any case, numerical methods of searching for the global optimum are proposed within the frameworks of approaches as a tool for getting a solution sought.

As it was discussed in [5], the global optimum is an integral characteristic of the problem, i.e., in order to make sure that a point $y^* \in P$ is the global minimizer of the problem (1)–(2) it is required to compare the value $f(y^*)$ with values of the objective function at all points of the domain P, but not in a vicinity of y^* only. As a result, when minimizing an essentially multiextremal function, a numerical global optimization method has to build a grid (random or regular) in the feasible domain and the number of grid nodes increases exponentially when rising the problem dimension. This peculiarity causes the substantial complexity of multiextremal problems and dimension is a crucial factor influencing significantly the efficiency of global optimization algorithms.

In this situation, approaches to elaboration of computational schemes reducing the dimension are widely used in the multiextremal optimization. Here we consider two approaches in which the initial multidimensional problem (1)–(2) is reduced to one or several univariate subproblems solved by efficient one-dimensional algorithms. The first approach applies *the nested optimization scheme* in its classical version [31–37] and in its generalization — *the adaptive nested scheme* [38, 39]. The second approach is based on *Peano space-filling curves* mapping the multidimensional domain (2) onto an interval in one-dimensional space $\mathbb{R}^1$ [5, 40–46].

The rest of the paper is organized in the following way. Section 2 describes the basic structures of the reduction schemes mentioned above and the univariate characteristical methods of global optimization to be applied within the reduction structures for solving the internal one-dimensional subproblems. Section 3 is devoted to a theoretical substantiation of the convergence for multidimensional methods combining the reduction schemes with characteristical methods of global search and contains both the known and new theoretical results. Section 4 presents experimental results of efficiency comparison for the methods described in previous sections on representative sets of multiextremal test functions of various dimensions belonging to the popular test class GKLS [47] with a controllable complexity. Finally, Section 5 contains a brief conclusion.

2. Schemes of dimensionality reduction

The first approach to dimensionality reduction called the scheme of nested optimization is based on the well-known relation (see, e.g., [5, 25, 31])

$$\min_{y \in P} f(y) = \min_{y_1 \in [a_1, b_1]} \cdots \min_{y_N \in [a_N, b_N]} f(y_1, \ldots, y_N). \tag{4}$$

In order to describe the scheme let us define a family of reduced functions as follows:

$$f^N(y) \equiv f(y), \tag{5}$$

$$f^i(y_1, \ldots, y_i) = \min_{y_{i+1} \in [a_{i+1}, b_{i+1}]} f^{i+1}(y_1, \ldots, y_i, y_{i+1}), \ 1 \le i \le N - 1. \tag{6}$$

Then, according to (4), in order to find the solution to the multidimensional problem (1)–(2) it is sufficient to solve the one-dimensional problem

$$f(y^*) = \min_{y_1 \in [a_1, b_1]} f^1(y_1). \tag{7}$$

But in order to evaluate the function f^1 at a fixed point y_1 it is necessary to solve the one-dimensional problem of the second level

$$f^1(y_1) = \min_{y_2 \in [a_2, b_2]} f^2(y_1, y_2), \tag{8}$$

and so on up to the univariate minimization at the N-th level the function $f^N(y) = f(y)$ with fixed coordinates $y_1, \ldots, y_{N-1}$.

Thus, instead of the multidimensional problem (1)–(2) one can solve the set of nested one-dimensional subproblems

$$\min_{y_i \in [a_i, b_i]} f^i(y_1, \ldots, y_{i-1}, y_i), \ 1 \le i \le N, \tag{9}$$

in which the coordinates $y_1, \ldots, y_{i-1}$ have been already determined by the subproblems of preceding levels. The objective functions $f^i(y_1, \ldots, y_i)$ of the subproblems (9) satisfy the Lipschitz condition for the corresponding argument y_i

3

subject to the assumption (3) for the function $f(y)$ (see [48]). Therefore, efficient univariate optimization algorithms can be taken for solving these subproblems. Various combinations of such the algorithms with the nested scheme (4) have been proposed in [5, 9, 25, 33–35].

The recursive character of generating the subproblems (9) during the optimization forms a hierarchical structure of subordination for these subproblems as a tree where the root is the subproblem (7) and the leaves are the subproblems of the N-th level. Fig. A.1 shows such a structure for $N = 3$.

Algorithmic implementation of the nested optimization scheme has some features. In the classical version (see, e.g., [5, 31]) during solving the root problem (7) a new evaluation of the function f^1 at a new point (new search iteration), i.e., solving a new subproblem of minimization of f^2 at the second level is supposed to be carried out only after the completion of the current iteration. It means that only one minimization subproblem at the second level can be active (can be in the course of solving) because the other subproblems of this level either have been already completed or will be initiated later. Analogously, the same situation takes place at all the rest of levels and at every moment of the multidimensional optimization not more than N univariate subproblems are active and they belong to a path from the root to a leaf. In Fig. A.1 one of the possible paths consists of subproblems marked with grey colour. Such organization of computations signifies that the information on the subproblems which have been already completed at the level i is not used during current minimization of a function f^i from (9). For example, in Fig. A.1 all the subproblems located above the active ones (marked with grey) have been already solved and information on the objective function obtained in the course of solving these subproblems is not used in the currently solved subtasks. Such loss of information leads to increasing the number of trials (evaluations of the objective functions) and worsens the efficiency of optimization. Indeed, if during the minimization of a function f^i from (9) we get values higher than ones obtained in a subproblem of the same level solved earlier, then, may be, it is reasonable not to continue solving the current subproblem up to the end and not to spend excessive trials.

This idea is realized *in the adaptive nested optimization scheme* proposed in the paper [38] containing a detailed algorithmic description of the adaptive scheme. The main suggestion is to give up the principle of strict subordination inherent to the classical nested scheme and to use a *simultaneous consideration* of all the univariate subproblems arising in the course of multidimensional optimization. In the adaptive scheme all the generated subproblems are active and a numerical "measure of quality" for each subproblem is introduced (this measure is called the *subproblem characteristic*). An iteration of the multidimensional optimization consists in the choice of the subproblem with maximal characteristic and carrying out a new trial within this subproblem. Such the approach allows choosing the subproblems with lower values of objective univariate functions (9) more often than with the higher ones. The way of defining the subproblem characteristics can be different, but in the case when the univariate characteristical algorithms [49, 50] are used for solving the subproblems (9) (see the general description of these algorithms below) the maximal interval characteristic

4

generating by the algorithm in the course of optimization can be taken as the subproblem characteristic.

The second approach is based on the known fact that a finite interval $[a, b]$ of the real axis and the N-dimensional hyperparallelepiped (2) are equipollent sets and there exist mappings called *Peano space-filling curves* (or *evolvents*) which map the unit interval $[a, b]$ onto the hyperparallelepiped (2) continuously and unambiguously.

Let $y(x)$ be such an evolvent. Let us consider the function

$$\varphi(x) = f(y(x)) \tag{10}$$

and the one-dimensional problem

$$\varphi^* = \varphi(x^*) = \min_{x \in [a,b]} \varphi(x). \tag{11}$$

Owing to (3) the function $f(y)$ is continuous and, because of continuity of $f(y)$ and $y(x)$,

$$\min_{y \in P} f(y) = \min_{x \in [a,b]} \varphi(x), \tag{12}$$

and the point $y^* = y(x^*)$ is the global minimizer of the multidimensional problem (1)–(2).

As a result, solving the univariate problem (11) allows obtaining the solution to the initial problem (1)–(2). A wide spectrum of the methods exploiting this idea is presented, for example, in publications [5, 28, 40, 43–46]. It should be noticed that the objective function $\varphi(x) = f(y(x))$ of the univariate problem (11) satisfies in general case the Hölder condition

$$|\varphi(x') - \varphi(x'')| \leq H \sqrt[N]{|x' - x''|}, \ x', x'' \in [a, b], \tag{13}$$

if the function $f(y)$ is Lipschitzian (3), and the resulting Hölder constant H is finite and depends on the Lipschitz constant L and the dimension N [5, 48] only.

Combinations of the reduction schemes described above with different univariate techniques of global search generate corresponding multidimensional algorithms of solving the multiextremal problem (1)–(2). Since in the nested optimization schemes (both classical and adaptive) all the univariate subproblems (9) satisfy the Lipschitz condition, we considered as one-dimensional methods the known algorithms of Lipschitzian optimization proposed by Piyavskij [25] and Strongin [51] which are theoretically substantiated and quite efficient. In regard to the reduction on the base Peano evolvents, the reduced objective function (10) is Hölderian (13) and for solving the univariate problem (12) it is necessary to apply the methods oriented at this class of optimization problems. As such the methods we took two algorithms which are generalizations [5, 28] of the Piyavskij's and Strongin's methods mentioned above.

All the univariate methods chosen for our research belong to the class of characteristical algorithms introduced by Grishagin [50] (see also [49]). Therefore, at first we describe a general structure of the characteristical methods and then

give in details the specificity of each method in the framework of the general description.

Let us consider the problem of univariate optimization in a standardized form: to find the minimal value

$$\varphi^* = \varphi(x^*) = \min_{x \in [a,b]} \varphi(x) \tag{14}$$

of a function $\varphi(x)$ over a finite interval $[a, b] \subset \mathbb{R}^1$ where the objective function $\varphi(x)$ satisfies either Lipschitz (3) or Hölder (13) condition in the feasible domain $[a, b]$.

Let the term "trial" denote an operation that calculates the value of the function $\varphi(x)$ at a point $x \in [a, b]$. An optimization method is supposed to place sequentially trials at points $x^1, x^2, \ldots, x^k, \ldots$ with corresponding outcomes $z^k = \varphi(x^k)$, $k = 1, 2, \ldots$.

A numerical method of solving the problem (14) is characteristical if its decision rule of planning trials consists of the following actions.

The first two trials are to be carried out at the end-points of the interval $[a, b]$, i.e.,

$$x^1 = a, \ x^2 = b, \tag{15}$$

with the values $z^1 = \varphi(a)$, $z^2 = \varphi(b)$.

The point x^{k+1}, $k > 2$, of any subsequent $(k + 1)$-th trial is chosen in accordance with the following steps:

Step 1. All the points $x^1, x^2, \ldots, x^k$ of preceding trials are renumbered by subscripts in increasing order, i.e.,

$$a = x_1 < x_2 < \cdots < x_k = b. \tag{16}$$

The values $z_i = \varphi(x_i)$, $1 \le i \le k$, are juxtaposed to the points x_i, $1 \le i \le k$, from (16).

Step 2. For each subinterval (x_{i-1}, x_i), $2 \le i \le k$, a numerical value $R(i)$ is calculated (this value is called characteristic of the subinterval).

Step 3. The subinterval (x_{t-1}, x_t) with the maximal characteristic is selected among all the subintervals:

$$R(t) = \max_{2 \le i \le k} R(i). \tag{17}$$

Step 4. The next $(k + 1)$-th trial is executed at a point $x^{k+1} \in (x_{t-1}, x_t)$ and the value $z^{k+1} = \varphi(x^{k+1})$ is calculated.

The general conditions of convergence to global minimum for characteristical algorithms have been proven in [49]. In particular, these results substantiate the termination criterion in the form

$$x_t - x_{t-1} \le \varepsilon, \tag{18}$$

6

where t is the argument of the largest characteristic $R(t)$ from (17) and $\varepsilon > 0$ is a predefined coordinate accuracy of the search, i.e., rules 1–4 are carried out until the length of the interval with maximal characteristic is less than the accuracy ε.

Now in order to describe a characteristical algorithm it is sufficient to define the expressions for its characteristics and the rule of placement of the next trial within the interval with maximal characteristic.

Following this way, we give corresponding descriptions for 4 univariate characteristical methods used in the reduction schemes under consideration.

The first two methods are intended for optimization of univariate Lipschitzian functions and are used in classical and adaptive nested schemes.

The first method was proposed by Piyavskij [25] and it is well-known in the Lipschitzian optimization. Its characteristics $R(i)$ are calculated as

$$R(i) = \frac{m(x_i - x_{i-1})}{2} - \frac{z_{i-i} + z_i}{2}, \qquad (19)$$

and the point of new trial is calculated in accordance with the expression

$$x^{k+1} = \frac{x_{t-1} + x_t}{2} - \frac{z_t - z_{t-1}}{2m}, \qquad (20)$$

where $m > 0$ is the parameter of the method. We will use for short designation of this algorithm the abbreviation PM (*Piyavskij's Method*).

The second algorithm was proposed by Strongin [5, 48, 51]. It has the characteristics

$$R(i) = m(x_i - x_{i-1}) + \frac{(z_i - z_{i-1})^2}{m(x_i - x_{i-1})} - 2(z_{i-1} + z_i), \qquad (21)$$

and its new trial point x^{k+1} is given in accordance with the relation (20). The value $m > 0$ is the parameter of the method. The author of the method called it simply *Global Search Algorithm* and we keep this name in the short form GSA.

If the objective function in the problem (14) satisfies the Lipschitz condition with the Lipchitz constant $L > 0$ then the convergence to all global minima is provided by the methods described when the parameter $m > L$ for PM and $m > 2L$ in GSA. Unfortunately, the value of the Lipchitz constant is unknown, as a rule, and for the parameter m one can use its adaptive estimation [5, 48].

$$m = \begin{cases} rM, & M > 0, \\ 1, & M = 0, \end{cases} \qquad (22)$$

where

$$M = \max_{2 \leq i \leq k} \frac{|z_i - z_{i-1}|}{x_i - x_{i-1}} \qquad (23)$$

and $r > 1$ is the parameter of the method.

The next two algorithms are oriented at solving the problem (12) with a Hölderian objective function and are generalizations of PM and GSA to this

7

case. The third methods called GPM (Generalized Piyavskij's Method) uses the characteristics

$$R(i) = \frac{m \sqrt[N]{x_i - x_{i-1}}}{2} - \frac{z_{i-1} + z_i}{2}, \tag{24}$$

and the new trial point

$$x^{k+1} = \frac{x_{t-1} + x_t}{2} - \mathrm{sign}(z_t - z_{t-1}) \frac{|z_t - z_{t-1}|^N}{2rm^N}. \tag{25}$$

The fourth method — Generalized Global Search Algorithm (GGSA) contains the characteristics

$$R(i) = m \sqrt[N]{x_i - x_{i-1}} + \frac{(z_i - z_{i-1})^2}{m \sqrt[N]{x_i - x_{i-1}}} - 2(z_{i-1} + z_i), \tag{26}$$

and the point of the new trial is calculated in accordance with (25).

In both the methods the parameter m is taken in the form (22), but instead of (23) the relation

$$M = \max_{2 \leq i \leq k} \frac{|z_i - z_{i-1}|}{\sqrt[N]{x_i - x_{i-1}}}, \tag{27}$$

is used.

3. Convergence of multidimensional reduction algorithms

Combinations of the nested dimensionality reduction schemes with the univariate methods PM and GSA generate the multidimensional algorithms for solving the problem (1). Let us call these algorithms in the following manner: PM-C and GSA-C are the classical nested scheme with the methods PM and GSA correspondingly at the univariate levels of reduction (9), and PM-A and GSA-A realize the adaptive nested scheme in which for the univariate optimizations (9) the methods PM and GSA are used respectively.

Convergence to the global solution of the problem (1) was proven for GSA-C in the monograph [48] and for GSA-A in the paper [38]. Moreover, the methods GPM and GGSA provide the property of global convergence as well (see [28, 48]). As for PM-C and PM-A, the problem of their convergence to absolute minimum was open so far, and in the present paper we give a theoretical substantiation of this important property.

First of all, we prove an auxiliary assertion on the sensitivity of PM to the possible errors in calculations of objective function values.

Describing the general structure of the characteristical algorithms we supposed that the results z^k at the trial points coincide with strict values of the objective function (14), i.e., $z^k = \varphi(x^k)$, $k = 0, 1, \dots$. But in the real optimization there are disturbances caused by computational errors. Assume that these errors are bounded by a real number $\xi > 0$. Then [48] the results z^k of trials can be treated as precise values of a function $\omega(x)$ such that

$$|\omega(x) - \varphi(x)| \leq \xi, \ x \in [a, b]. \tag{28}$$

After termination of the search caused by the stopping criterion (18) the value

$$\omega_k^* = \min_{1 \le i \le k} z^i \tag{29}$$

may be considered as an approximation of the global minimum φ^* of the function $\varphi(x)$ from (14). In this case, the sensitivity of the method to the bounded errors can be estimated by a degree of proximity between ω_k^* and φ^*. The next theorem provides such an estimation for the algorithm PM.

Theorem 1. *Assume that after a termination of the search according to the stopping criterion (18) there is a truncated sequence of trials $x^1, x^2 \ldots, x^k$ generated by PM during minimizing the approximation $\omega(x)$ of the objective function $\varphi(x)$, $x \in [a, b]$, satisfying the inequality (28), and the function $\varphi(x)$ is Lipschitzian with constant $L > 0$. Then the following inequality holds:*

$$|\omega_k^* - \varphi^*| \le \frac{L\varepsilon}{2} + \xi, \tag{30}$$

if the PM's parameter $m > L$.

PROOF. Let us order the elements of the truncated sequence $x^1, x^2, \ldots, x^k$ in accordance with the rule (16) and introduce the designations $\alpha_i = (z_{i-1} + z_i)/2$, $\delta_i = x_i - x_{i-1}$, $2 \le i \le k$, where $z_i = \omega(x_i)$, $1 \le i \le k$.

From the Lipschitz condition for the function $\varphi(x)$

$$\varphi(x_i) + \varphi(x_{i-1}) - L\delta_i \le 2\varphi^*, \ 2 \le i \le k, \tag{31}$$

and owing to (28)

$$z_i - \varphi(x_i) \le \xi, \ z_{i-1} - \varphi(x_{i-1}) \le \xi, \ 2 \le i \le k, \tag{32}$$

from where

$$\alpha_i - \frac{L\delta_i}{2} \le \xi + \varphi^*, \ 2 \le i \le k, \tag{33}$$

or

$$\min_{2 \le i \le k} \left(\alpha_i - \frac{L\delta_i}{2} \right) - \xi \le \varphi^*. \tag{34}$$

On the other hand, for the point x_s such that $\omega(x_s) = \omega_k^*$ taking into account (28) and the inequality $\varphi(x_s) \ge \varphi^*$ we have

$$\varphi^* \le \omega_k^* + \xi. \tag{35}$$

Let the number q, $2 \le q \le k$, be such that

$$\alpha_q - \frac{L\delta_q}{2} = \min_{2 \le i \le k} \left(\alpha_i - \frac{L\delta_i}{2} \right). \tag{36}$$

Let us consider two cases. The first one corresponds to the situation $\delta_q \le \varepsilon$ for ε from (18). Since $\omega_k^* \le \alpha_q$ then

$$\omega_k^* - \frac{L\varepsilon}{2} \le \alpha_q - \frac{L\delta_q}{2} \le \alpha_i - \frac{L\delta_i}{2} \tag{37}$$

9

for all i, $2 \le i \le k$.

In the opposite case $\delta_q > \varepsilon$ let us take the interval (x_t, x_{t-1}) with the largest characteristic $R(t)$ from (17) which the inequality (18) is true for. The characteristic $R(q) \le R(t)$ because of (17), i.e.,

$$R(t) - R(q) = \frac{m(\delta_t - \delta_q)}{2} - \alpha_t + \alpha_q \ge 0, \tag{38}$$

from where

$$\alpha_t - \frac{L\delta_t}{2} \le \alpha_q - \frac{L\delta_q}{2} - \frac{(m-L)(\delta_q - \delta_t)}{2} \le \alpha_q - \frac{L\delta_q}{2}, \tag{39}$$

taking into account $m > L$ and $\delta_q > \varepsilon \ge \delta_t$.

But $\alpha_t - L\delta_t/2 \ge \omega_k^* - L\varepsilon/2$, therefore, from (34) it follows that

$$\omega_k^* - \varphi^* \le \frac{L\varepsilon}{2} + \xi. \tag{40}$$

On the other hand, due to (35) we have

$$\varphi^* - \omega_k^* \le \xi \le \xi + \frac{L\varepsilon}{2}. \tag{41}$$

The inequalities (40), (41) prove the main assertion (30) of the theorem. The proof has been completed.

Now let us return to the multidimensional problem (1) and the algorithm PM-C. It should be noted that the univariate subproblems (9) are solved approximately. Indeed, the evaluation of the function $f^{N-1}(y_1, \ldots, y_{N-1})$ for fixed $y_1, \ldots, y_{N-1}$ consists in minimization by means of PM the function $f^N(y_1, \ldots, y_{N-1}, y_N)$ for the coordinate y_N and instead of the precise value $f^{N-1}(y_1, \ldots, y_{N-1})$ we obtain an approximate estimation $\omega^{N-1}(y_1, \ldots, y_{N-1})$. Furthermore, evaluation $f^{N-2}(y_1, \ldots, y_{N-2})$ for fixed $y_1, \ldots, y_{N-2}$ is implemented as an approximate minimization of the function $\omega^{N-1}(y_1, \ldots, y_{N-1})$ and the result of this minimization is an approximate estimation $\omega^{N-2}(y_1, \ldots, y_{N-2})$, etc., up to obtaining an estimation $\omega^* = \omega^1$ in the subproblem $\min\{\omega^1(y_1) : y_1 \in [a_1, b_1]\}$. In this procedure the resulting value ω^* is an estimation of the sought solution f^* from (1). As a consequence, in order to ascertain the convergence of PM-C to global minimum of the problem (1) it is sufficient to prove the convergence the estimation ω^* to f^* when ε_1 (search accuracy in (18) for the subproblem (7)) tends to zero, i.e.,

$$\lim_{\varepsilon_1 \to 0} \omega^* = f^*. \tag{42}$$

If PM is applied for univariate minimization of subproblems (9), its parameters can differ at different levels of recursion, in particular, one can take distinct accuracies in the termination criterion (18) and consider the vector

$$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N), \tag{43}$$

where ε_i is the termination accuracy during solving the subproblems (9) for the coordinate y_i, $1 \le i \le N$.

10

Theorem 2. *If the objective function $f(y)$ in the problem (1) satisfies the Lipschitz condition (3) with constant $L > 0$, all the subproblems (9) are solved by PM with the parameter $m > L$ and accuracies ε_i satisfy the inequalities*

$$\varepsilon_{i+1} < \varepsilon_i, \ 1 \leq i < N, \tag{44}$$

then the assertion (42) is true.

PROOF. Suppose that for a number i, $1 \leq i < N$, the inequality

$$|f^i(y_1, \ldots, y_i) - \omega^i(y_1, \ldots, y_i)| \leq \xi_i, \tag{45}$$

holds. This condition is met for $i = N$ with $\xi_N = 0$ as

$$\omega^N(y_1, \ldots, y_N) = f^N(y_1, \ldots, y_N) = f(y). \tag{46}$$

As the corollary of the Lipschitz condition for the function $f(y)$, all the functions $f^i(y_1, \ldots, y_i)$ are Lipschitzian as well with the same constant L (see [48]). Then, in accordance with the Theorem 1 the inequality (45) involves for the number $i - 1$ the estimation

$$|f^{i-1}(y_1, \ldots, y_{i-1}) - \omega^{i-1}(y_1, \ldots, y_{i-1})| \leq \xi_i + \frac{L\varepsilon_i}{2}, \tag{47}$$

and, as consequence, one can take

$$\xi_{i-1} = \xi_i + \frac{L\varepsilon_i}{2} \tag{48}$$

Since the condition (45) takes place for $i = N$ with $\xi_N = 0$ it is true for all i, $1 \leq i \leq N$, and we get the estimation

$$|\varphi^* - \omega^*| \leq \frac{L}{2} \sum_{i=1}^{N} \varepsilon_i, \tag{49}$$

Then the statement (42) results immediately from (44) and (49). The theorem has been proven.

Finally, for the adaptive nested scheme with the Piyavskij's method inside (PM-A) its convergence follows from the convergence of PM-C by reasons which substantiated in [38] the convergence of GSA-A.

4. Numerical experiments

In this section we consider 7 global search algorithms and compare their efficiency experimentally. 6 methods combine the reduction schemes described above with two Lipschitzian one-dimensional methods and their generalizations to the Hölder case and were considered in previous section. To demonstrate the

efficiency of these algorithms in comparison with a method of another nature we took the well-known method of Lipschitz optimization DIRECT [22, 23] as well.

All the methods were compared on the test class of multiextremal functions GKLS from [47] being an easy-to-use source of functions with tuned complexity. The test sets consisting of 100 GKLS problems were taken for each of dimensions 2, 3, 4. For all the dimensions the hard class of GKLS was chosen with the following parameters:

- 10 local minima;

- radius of the attraction region of the global minimizer equal to 0.12;

- standard function value -1.0 at the global minimizer;

- distance from the global minimizer to the vertex of the paraboloid equal to 0.9.

The efficiency of algorithms was measured in accordance with the following two criteria: the average number K of trials (evaluations of the objective function) executed during optimization of the functions from the test set and the number P of test problems solved successfully. A test problem was supposed to have been solved if the trial with the lowest value obtained in the course of optimization was placed in a δ-neighbourhood of the global minimizer.

In more details, the values K and P were estimated as follows. The algorithm with a given set of parameters solves n optimization problems from the test class. After solving the i-th test problem we have the number K_i of executed trials and the distance

$$P_i = \sqrt{\sum_{j=1}^{N}(y_j^* - y_j^+)^2} \tag{50}$$

between the real global minimizer $y^* = (y_1^*, \ldots, y_N^*)$ and the coordinate vector $y^+ = (y_1^+, \ldots, y_N^+)$ corresponding to the lowest function value found by the algorithm.

After solving all the test problems the criteria K and P are calculated as

$$K = \frac{1}{n}\sum_{i=1}^{n} K_i, \quad P = \frac{1}{n}\sum_{i=1}^{n} p_i, \tag{51}$$

where for given accuracy $\delta > 0$

$$p_i = \begin{cases} 1, & P_i < \delta, \\ 0, & P_i \geq \delta. \end{cases} \tag{52}$$

Taking these criteria into consideration it is possible to apply the *method of operational characteristics* [39, 52] which allows one to present comparison results in a visual form. In accordance with this method an optimization algorithm solves problems of a test set several times using different values of its parameters and obtains a pair (K, P) of the efficiency criteria for each variant of parameters

chosen. The set of these pairs is called *operational characteristic* of the algorithm. Plotting the operational characteristics on the plane (K, P) enables to compare visually the efficiency of different optimization algorithms. Namely, if for the same P the operational characteristic of an algorithm is located to the left of the characteristic of another method, the first algorithm is better because it has spent less number of trials K for successful solving the same number of problems.

In experiments the algorithms on the base of the reduction schemes used in (22) the identical parameter $r = 8$ providing for all 6 methods the sufficient conditions of global convergence and the curves of the operational characteristics for these algorithms and the method DIRECT were obtained by means of variation of parameters in stopping criteria of these methods.

The operational characteristics of the optimization algorithms underwent the examination on the 3-dimensional GKLS set are divided for better visualization into 3 groups and presented in Fig. A.2, A.3, A.4.

Fig. A.2 and Fig. A.3 contain the operational characteristics for the algorithms of PM-type family and for the algorithms of GSA-type family correspondingly. Fig. A.4 compares the operational characteristics of the algorithms based on the adaptive nested scheme. In these figures the number of trials is plotted on the abscissa axis in the logarithmic scale.

As it follows from the results presented in Fig. A.2 and Fig. A.3, among the algorithms combining the methods of the same type with different schemes of dimensionality reduction the adaptive nested scheme is most efficient if it is necessary to solve problems with high reliability P. At the same time, the efficiencies of the reduction by means of Peano-type space-filling curves and the classical nested scheme differ slightly. Moreover, all 6 algorithms realizing the dimensionality reduction outperform significantly the method DIRECT.

With regard to comparison of the algorithms having demonstrated the best results in their families, the method GSA-A is some more efficient than PM-A.

The similar conclusions on the efficiency of the compared algorithms can be drawn after the experiments on the 2- and 4-dimensional GKLS classes. As a confirmation Tables A.2 and A.1 describe the results of testing for these dimensions. The left column contains the number K of trials and the other columns correspond to the studied methods and show the quantity of problems solved successfully depending on the spent trials.

5. Conclusion

Numerical multidimensional global optimization algorithms based on several ideas of a dimensionality reduction have been considered. These algorithms reduce the multidimensional problem to one or several univariate subproblems solved by efficient methods of one-dimensional global optimization. Two approaches to reducing the dimension in multiextremal Lipschitzian optimization problems have been used: the reduction by means of Peano-type space-filling curves and the nested optimization scheme. In the framework of the second approach two versions of the nested reduction have been taken, namely, classical

13

nested scheme and its recent generalization — adaptive nested optimization. For solving the univariate subproblems the known algorithms proposed by Strongin and Piyavskij and the generalizations of these methods to the class of Hölderian functions have been used.

For the considered algorithms the theoretical substantiation of convergence to global optimum has been given including both the known results and the new ones proved in the paper. For evaluation of efficiency a representative experimental testing of the methods has been carried out for different dimensions on sets of multiextremal functions belonging to the well-known GKLS test class being at present a classical tool for testing global optimization algorithms. Additionally, the popular method DIRECT has been tested to compare the reduction methods with an algorithm of the other nature. As a result of experiment, among all competitors the adaptive nested scheme has demonstrated the best efficiency.

The promising way of future researches consists in development of parallel algorithms based on the adaptive nested scheme, theoretical investigation of their properties and evaluation of parallelization efficiency, including comparison with existing parallel optimization methods.

Acknowledgements

References

[1] R. Horst, P. M. Pardalos, Handbook of Global Optimization, Kluwer Academic Publishers, Dordrecht, 1995.

[2] R. Horst, H. Tuy, Global Optimization: Deterministic Approaches, Springer-Verlag, Berlin, 1990.

[3] R. Paulavičius, J. Žilinskas, Simplicial Global Optimization, Springer, NewYork, 2014.

[4] J. D. Pintér, Global Optimization in Action, Kluwer Academic Publishers, Dordrecht, 1996.

[5] R. G. Strongin, Y. D. Sergeyev, Global Optimization with Non-convex Constraints: Sequential and Parallel Algorithms, Kluwer Academic Publishers, Dordrecht, 2000.

[6] A. A. Zhigljavsky, A. Žilinskas, Stochastic Global Optimization, Springer, NewYork, 2008.

[7] D. Famularo, P. Pugliese, Y. D. Sergeyev, A global optimization technique for checking parametric robustness, Automatica 35 (1999) 1605–1611.

[8] V. P. Gergel, M. I. Kuzmin, N. A. Solovyov, V. A. Grishagin, Recognition of surface defects of cold-rolling sheets based on method of localities, International Review of Automatic Control 8 (2015) 51–55.

[9] V. A. Grishagin, R. G. Strongin, Optimization of multiextremal functions subject to monotonically unimodal constraints, Engineering Cybernetics 22 (1984) 117–122.

[10] D. E. Kvasov, D. Menniti, A. Pinnarelli, Y. D. Sergeyev, N. Sorrentino, Tuning fuzzy power-system stabilizers in multi-machine systems by global optimization algorithms based on efficient domain partitions, Electric Power Systems Research 78 (2008) 1217–1229.

[11] D. E. Kvasov, Y. D. Sergeyev, Lipschitz global optimization methods in control problems, Autom. Remote Control 74 (2013) 1435–1448.

[12] V. Y. Modorskii, D. F. Gaynutdinova, V. P. Gergel, K. A. Barkalov, Optimization in design of scientific products for purposes of cavitation problems, AIP Conf. Proc. 1738 (2016) 400013.

[13] Y. D. Sergeyev, P. Daponte, D. Grimaldi, A. Molinaro, Two methods for solving optimization problems arising in electronic measurements and electrical engineering, SIAM J. Optim. 10 (1999) 1–21.

[14] H. J. Kushner, A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise, Transactions of ASME Ser. D. Journal of Basic Engineering 86 (1964) 97–106.

[15] A. A. Zhigljavsky, Theory of Global Random Search, Kluwer Academic Publishers, Dordrecht, 1991.

[16] A. Žilinskas, On similarities between two models of global optimization: statistical models and radial basis functions, J. Global Optim. 48 (2010) 173–182.

[17] A. Žilinskas, J. Žilinskas, P-algorithm based on a simplicial statistical model of multimodal functions, TOP 18 (2010) 396–412.

[18] K. Barkalov, V. Gergel, Parallel global optimization on GPU, J. Global Optim. 66 (2016) 3–20.

[19] Y. G. Evtushenko, M. A. Posypkin, A deterministic approach to global box-constrained optimization, Optim. Lett. 7 (2013) 819–829.

[20] V. P. Gergel, S. V. Sidorov, A two-level parallel global search algorithm for solving computationally intensive multi-extremal optimization problems, Lecture Notes in Comput. Sci. 9251 (2015) 505–515.

[21] E. Gourdin, B. Jaumard, R. Ellaia, Global optimization of Hölder functions, J. Global Optim. 8 (1996) 323–348.

[22] D. R. Jones, C. D. Perttunen, B. E. Stuckman, Lipschitzian optimization without the Lipschitz constant, J. Optim. Theory Appl. 79 (1993) 157–181.

[23] D. R. Jones, The DIRECT global optimization algorithm, in: C. Floudas, P. M. Pardalos (Eds.), Encyclopedia of optimization, Kluwer Academic Publishers, Dordrecht, 2001, pp. 431–440.

[24] D. E. Kvasov, C. Pizzuti, Y. D. Sergeyev, Local tuning and partition strategies for diagonal GO methods, Numer. Math. 94 (2003) 93–106.

[25] S. A. Piyavskij, An algorithm for finding the absolute extremum of a function, Comput. Math. Math. Phys. 12 (1972) 57–67.

[26] R. Paulavicius, Y. D. Sergeyev, D. E. Kvasov, J. Žilinskas, Globally-biased DISIMPL algorithm for expensive global optimization, J. Global Optim. 59 (2014) 545–567.

[27] Y. D. Sergeyev, D. Famularo, P. Pugliese, Index branch-and-bound algorithm for Lipschitz univariate global optimization with multiextremal constraints, J. Global Optim. 21 (2001) 317–341.

[28] Y. D. Sergeyev, V. A. Grishagin, Sequential and parallel algorithms for global optimization, Optim. Methods Softw. 3 (1994) 111–124.

[29] Y. D. Sergeyev, D. E. Kvasov, A deterministic global optimization using smooth diagonal auxiliary functions, Commun. Nonlinear Sci. Numer. Simul. 21 (2015) 99–111.

[30] Y. D. Sergeyev, M. S. Mukhametzhanov, D. E. Kvasov, D. Lera, Derivative-free local tuning and local improvement techniques embedded in the univariate global optimization 171 (2016) 186–208.

[31] C. R. Carr, C. W. Howe, Quantitative Decision Procedures in Management and Economic: Deterministic Theory and Applications, McGraw-Hill, New York, 1964.

[32] E. R. Dam, B. Husslage, D. Hertog, One-dimensional nested maximin designs, J. Global Optim. 46 (2010) 287–306.

[33] V. P. Gergel, V. A. Grishagin, R. A. Israfilov, Local tuning in nested scheme of global optimization, Procedia Computer Science 51 (2015) 865–874.

[34] V. A. Grishagin, R. A. Israfilov, Global search acceleration in the nested optimization scheme, AIP Conf. Proc. 1738 (2016) 400010.

[35] V. A. Grishagin, R. A. Israfilov, Multidimensional constrained global optimization in domains with computable boundaries, CEUR Workshop Proceedings 1513 (2015) 75–84.

[36] L. Shi, S. Ólafsson, Nested partitions method for global optimization, Oper. Res. 48 (2000) 390–407.

[37] Y. D. Sergeyev, V. A. Grishagin, Parallel asynchronous global search and the nested optimization scheme, J. Comput. Anal. Appl. 3 (2001) 123–145.

[38] V. P. Gergel, V. A. Grishagin, A. V. Gergel, Adaptive nested optimization scheme for multidimensional global search, J. Global Optim. 66 (2016) 35–51.

[39] V. A. Grishagin, R. A. Israfilov, Y. D. Sergeyev, Comparative efficiency of dimensionality reduction schemes in global optimization, AIP Conf. Proc. 1776 (2016) 060011.

[40] A. R. Butz, Space-filling curves and mathematical programming, Inform. Control 12 (1968) 314–330.

[41] B. Goertzel, Global optimization with space-filling curves, Appl. Math. Lett. 12 (1999) 133–135.

[42] A. E. Hime, H. A. Oliveira Jr., A. Petraglia, Global optimization using space-filling curves and measure-preserving transformations, Soft Computing in Industrial Applications 96 (2011) 121–130.

[43] D. Lera, Y. D. Sergeyev, Deterministic global optimization using space-filling curves and multiple estimates of Lipschitz and holder constants, Commun. Nonlinear Sci. Numer. Simul. 23 (2015) 328–342.

[44] D. Lera, Y. D. Sergeyev, Global minimization algorithms for Hölder functions, BIT 42 (2002) 119–133.

[45] D. Lera, Y. D. Sergeyev, Lipschitz and Hölder global optimization using space-filling curves, Appl. Numer. Math. 60 (2010) 115–129.

[46] Y. D. Sergeyev, R. G. Strongin, D. Lera, Introduction to Global Optimization Exploiting Space-Filling Curves, Springer, 2013.

[47] M. Gaviano, D. E. Kvasov, D. Lera, Y. D. Sergeyev, Software for generation of classes of test functions with known local and global minima for global optimization, ACM TOMS 29 (2003) 469–480.

[48] R. G. Strongin, Numerical Methods in Multiextremal Problems (Information-Statistical Algorithms), Nauka, Moscow, 1978, in Russian.

[49] V. A. Grishagin, Y. D. Sergeyev, R. G. Strongin, Parallel characteristic algorithms for solving problems of global optimization, J. Global Optim. 10 (1997) 185–206.

[50] V. A. Grishagin, On convergence conditions for a class of global search algorithms, Proceedings of the 3-rd All-Union Seminar Numerical Methods of Nonlinear Programming, 1979, pp. 82–84, in Russian.

[51] R. G. Strongin, The information approach to multiextremal optimization problems, Stoch. Stoch. Rep. 27 (1989) 65–82.

[52] V. A. Grishagin, Operating characteristics of some global search algorithms, Problems of Statistical Optimization 7 (1978) 198–206, in Russian.

Appendix A.

$$\min_{y_1\in[a_1,b_1]} f^1(y_1)$$

$$\min_{y_2\in[a_2,b_2]} f^2(y_1^1,y_2)$$

$$\min_{y_2\in[a_2,b_2]} f^2(y_1^2,y_2)$$

$$\min_{y_2\in[a_2,b_2]} f^2(y_1^{k_1},y_2)$$

$$\min_{y_3\in[a_3,b_3]} f^2(y_1^1,y_2^{1,1},y_3)$$

$$\dots$$

$$\min_{y_3\in[a_3,b_3]} f^2(y_1^1,y_2^{1,k_{21}},y_3)$$

$$\min_{y_3\in[a_3,b_3]} f^2(y_1^2,y_2^{2,1},y_3)$$

$$\dots$$

$$\min_{y_3\in[a_3,b_3]} f^2(y_1^2,y_2^{2,k_{22}},y_3)$$

$$\min_{y_3\in[a_3,b_3]} f^2(y_1^{k_1},y_2^{k_1,1},y_3)$$

$$\dots$$

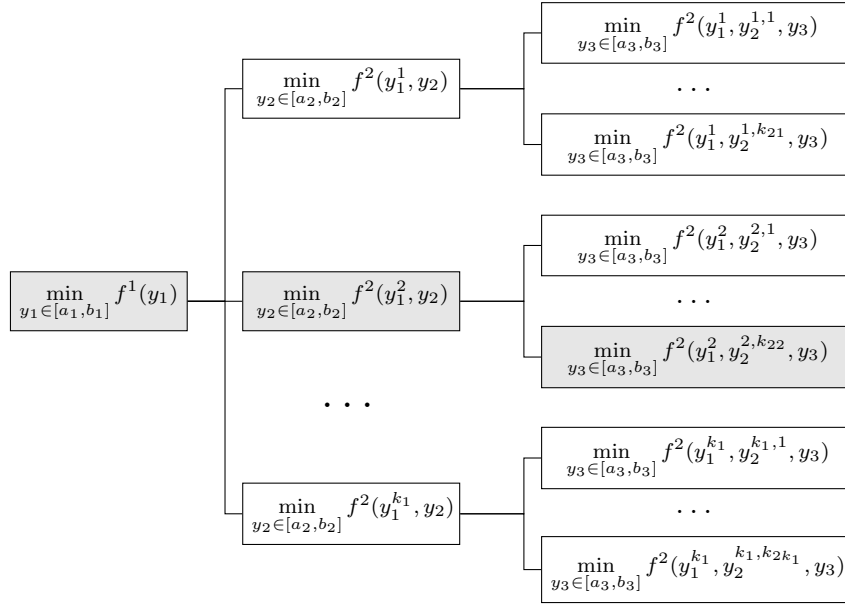$$\min_{y_3\in[a_3,b_3]} f^2(y_1^{k_1},y_2^{k_1,k_{2k_1}},y_3)$$

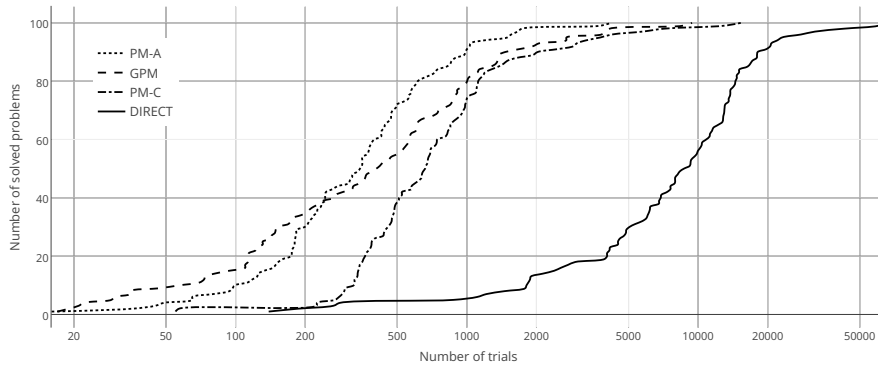Figure A.1: The subordination structure of the univariate subproblems (9)

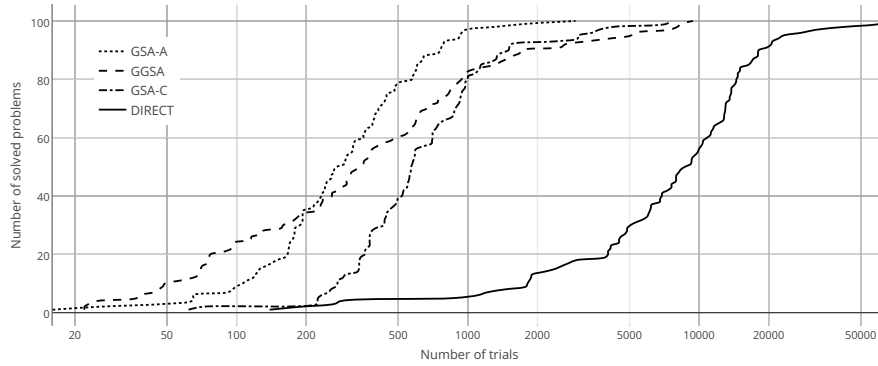Figure A.2: Operational characteristics of PM-type algorithms and DIRECT

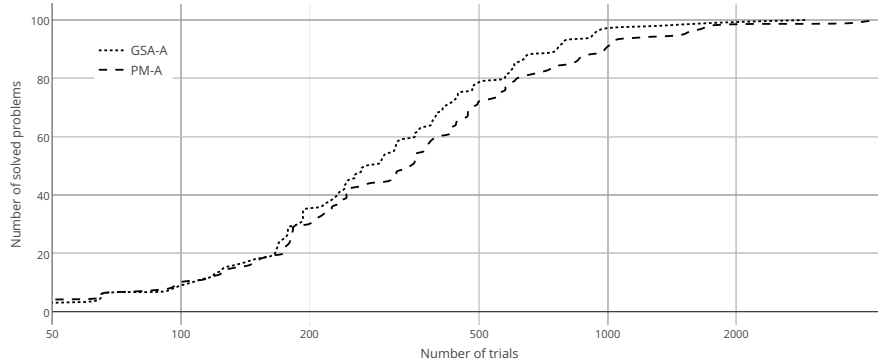Figure A.3: Operational characteristics of GSA-type algorithms and DIRECT

Figure A.4: Operational characteristics of GSA-A and PM-A

Table A.1: Number of problems solved successfully depending on the executed trials on four-dimensional class

K	GSA-A	GSA-C	GGSA	PM-A	PM-C	GPM	DIRECT
500	6	1	1	1	1	1	1
1000	22	1	20	21	1	15	1
2000	45	4	34	46	5	29	1
5000	82	48	53	79	47	57	1
10000	96	78	71	92	72	74	1
16000	100	87	79	100	86	83	1
30000	100	93	89	100	95	90	20
50000	100	100	100	100	100	100	50
70000	100	100	100	100	100	100	71
90000	100	100	100	100	100	100	81
120000	100	100	100	100	100	100	100

19

Table A.2: Number of problems solved successfully depending on the executed trials on two-dimensional class

K	GSA-A	GSA-C	GGSA	PM-A	PM-C	GPM	DIRECT
10	1	1	8	1	1	18	3
20	35	32	53	30	26	52	6
30	82	71	80	68	59	71	9
40	96	94	91	92	86	84	13
50	100	99	97	100	96	92	16
80	100	100	100	100	100	100	24
150	100	100	100	100	100	100	35
250	100	100	100	100	100	100	47
500	100	100	100	100	100	100	80
800	100	100	100	100	100	100	100