Global Optimization Based on a Statistical Model and Simplicial Partitioning

A. ŽILINSKAS
Institute of Mathematics and Informatics
Vilnius, Lithuania
antanaz@klm.mii.lt

J. ŽILINSKAS
Department of Informatics
KUT, Studentu str. 50-214b, Kaunas, Lithuania
jzil@dsplab.ktu.lt

Abstract—A statistical model for global optimization is constructed generalizing some properties of the Wiener process to the multidimensional case. An approach to the construction of global optimization algorithms is developed using the proposed statistical model. The convergence of an algorithm based on the constructed statistical model and simplicial partitioning is proved. Several versions of the algorithm are implemented and investigated. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords—Global optimization, Statistical models, Partitioning, Simplices, Convergence.

1. INTRODUCTION

The first statistical model used for the global optimization was the Wiener process [1]. Several one-dimensional algorithms were developed using Wiener or Wiener related models, e.g., [2–7]. Extensive testing has shown that the implemented algorithms favorably compete with algorithms based on other approaches [7,8]. The sampling functions of Wiener process are not differentiable almost everywhere with probability one. Recently, a statistical one-dimensional model of smooth functions was constructed whose computational complexity is similar to the computational complexity of Wiener model [9]. The nondifferentiability of sampling functions was considered a serious theoretical disadvantage of using Wiener process as a model for global optimization. However, the algorithms based on both mentioned models have similar theoretical properties as shown in [10]. Only the constants of estimates of convergence rate are different: a constant is better if the local features of a model correspond to local features of an objective function.

Standard probabilistic generalization of stochastic processes to the multidimensional case are random fields, i.e., stochastic functions of several variables. A review of global optimization based on multidimensional statistical models may be found in [8,11]. There are some general difficulties in constructing of multidimensional algorithms based on random fields.

First. The current trial point is defined by means of optimization of a merit function (e.g., average improvement at the current step), which is again a multimodal optimization problem.
The computation of the merit function is hard because it requires inversion of correlation matrices of a random field whose dimensionality is \( k \) at \((k+1)\)th minimization step.

A generalization of the Wiener process model to the two-dimensional case is proposed in [12] by further development of the approach in [13,14]. The model, a family of Gaussian variables \( \xi_x \), with linear mean and quadratic variance, is defined over an equilateral triangle: \( x \in S \subset \mathbb{R}^2 \). For such a model, the merit function of the \( P \)-algorithm is unimodal and its optimum can be found analytically. To start the optimization, the feasible region is covered by equilateral triangles, the objective function values are calculated at the vertexes, and the statistical model is extended for the constructed covering. The latter is refined at each iteration by means of selection of a candidate triangle and its cloning, i.e., its partitioning by means of bisection of its sides. The features of the proposed statistical model facilitate easy calculation of the selection criterion. In this way, solving an auxiliary global optimization problem is avoided.

The method proposed in [12] is essentially two-dimensional since cloning method does not exist for \( n \geq 3 \). In the present paper, the statistical model is defined for general simplexes and the cloning is generalized to the partitioning. The generalized model is not restricted to two dimensions. However, many versions of partitioning may be applied instead of cloning. The aim of the present paper is to investigate the basic features of the generalized model, to prove the convergence of a general algorithm, and to investigate experimentally the efficiency of different partitioning methods in two and three dimensions. Such a research seems necessary to approve decisions on rather complicated implementation of the algorithm for higher dimensions.

2. STATISTICAL MODEL

Let us consider the global minimization problem \( \min_{x \in A} f(x), A \subset \mathbb{R}^n \), where \( f(\cdot) \) is a continuous function and \( A \) is a compact set. The very general assumptions on \( f(\cdot) \) imply that the family of random Gaussian variables \( \xi_x \) may be considered as a model of \( f(\cdot) \) [8,13]. The mean value and the variance of \( \xi_x \) depend on assumptions on features of an objective function and on information acquired during the search, i.e., the trial points \( x_i \) and the values \( y_i = f(x_i), i = 1, \ldots, k \). The characteristics of the statistical model will be defined generalizing the properties of the Wiener process. The choice of the latter as a standard is partly justified by the following arguments:

- the functional structure of conditional characteristics (conditional mean and conditional variance) is simplest possible,
- in one-dimensional case the convergence order of \( P \)-algorithm was proved the same for the algorithms based on Wiener and on smooth function models; see [10],
- it is supposed to use the constructed \( P \)-algorithm in combination with a local descent algorithm, i.e., the high precision of local search by the \( P \)-algorithm is not a concern.

The axioms on rationality of extrapolation under uncertainty imply, that the mean value at the point \( x \) is the weighted average of known function values

\[
m_k(x \mid (x_i, y_i), i = 1, \ldots, k) = \sum_{i=1}^{k} w_i(x, x_j, j = 1, \ldots, k) \cdot y_i. \tag{1}
\]

Let \( A \) be covered by simplexes \( S_j, j = 1, \ldots, m \), and let the trial points coincide with the vertexes of \( S_j \); it is assumed that \( k > n \). The Markov property of a one-dimensional stochastic process may be generalized giving the following property of the multidimensional statistical model: for \( x \in S \), all the weights not corresponding to the vertexes of \( S \) are defined equal to zero in (1)

\[
m_k(x \mid (x_i, y_i), i = 1, \ldots, k) = \sum_{i=0}^{n} \nu_i(x, \omega_j, j = 0, \ldots, n) \cdot \varphi_i, \tag{2}
\]

where \( \omega_i, i = 0, \ldots, n \), denote the vertexes of \( S \), and \( \varphi_i \) denote the corresponding function values. However, this is not the only goal of the generalization. The linearity with respect to \( x \) of the
conditional mean of the Wiener process is very important for efficient implementation of a one-dimensional algorithm; therefore, the linearity of \( m_k(x) \) with respect to \( x \) is naturally wanted also in the multidimensional case: (1) is piecewise linear with respect to \( x \) if the weights in (2) are defined equal to baricentric coordinates of \( x \) with respect to \( \omega_i \),

\[
x = \sum_{i=0}^{n} \nu_i(x, \omega_j, j = 0, \ldots, n) \cdot \omega_i.
\]

By similarity to the Wiener process, the variance of \( \xi_x \) for a regular simplex \( x \in S \) may be defined as a quadratic function with zero values at the vertexes and maximum at the point equidistant from the vertexes:

\[
s_k^2(x \mid x_i, i = 1, \ldots, k) = \sigma_0^2 \cdot (D_S^2 - \|x_S - x\|^2),
\]

where \( D_S \) is the distance between \( x_S \) and vertexes, and \( \sigma_0 \) is the only parameter of the model which may be estimated from the data collected during the minimization. Let us start with the case of the unit simplex with vertexes \( \omega_0, \omega_1, \ldots, \omega_n \).

In the case of an arbitrary simplex, the definition of the average (2) remains valid. The variance is obtained by means of the inverse mapping of the considered simplex on the correct simplex with equal perimeter, where the increasingly ordered function values \( \varphi_i \) should correspond to the vertexes \( \omega_i, i = 0, \ldots, n. \) To start the optimization, the feasible region is covered by simplexes.

The family of Gaussian random variables \( \xi_x \) with the characteristics defined above is accepted as a multidimensional statistical model generalizing the Wiener process.

### 3. SELECTION

Let there be given a covering of \( A \) by simplexes. The selection includes the evaluation of simplexes with respect to a rationality criterion and the choice of the best for its further partitioning. Different criteria may be applied, e.g., worst case criteria with respect to a deterministic model of an objective function [15-17].

In the present paper, selection is based on the average case rationality paradigm. The choice of the point for current calculation of \( f(\cdot) \) is justified in [8,14] by assumptions of rationality of a choice with respect to the statistical model \( \xi_x \). It is shown that the rational choice is defined by the maximization of the probability to find a better (up to tolerance \( \varepsilon \)) value than the currently best one, i.e., by the maximization of

\[
P_k(x) = P(\xi_x < z_k),
\]

where

\[
z_k = y_0k - \varepsilon, \quad y_0k = \min(y_1, \ldots, y_k), \quad \varepsilon > 0.
\]

The criterion of maximal probability is adapted for selection of a simplex. Let us consider the maximization of \( P_k(x) \) over the regular simplex \( S \) (4). It is assumed that \( \varphi_0 = 0 \leq \varphi_1 \leq \cdots \leq \varphi_n \). Such an assumption does not reduce generality, but \( z_k \) should be correspondingly normalized for each simplex; after the normalization there holds the inequality \( z_k < 0 \).
THEOREM 1. \( P_k(x) \) has either a single local maximum in the interior of \( S \) at the point \( x_{\text{max}} \) or no local maximum in the interior of \( S \) at all, where

\[
x_{\text{max}} = x_S - \frac{1}{-(n + 1)/n} z_k + \frac{1}{n} \sum \varphi_i \cdot (\varphi - \varphi_* \cdot I),
\]

\[
\varphi = (\varphi_1, \ldots, \varphi_n),
\]

\[
\varphi_* = \frac{\sqrt{n + 1} - 1}{n\sqrt{n + 1}} \sum \varphi_i, \quad I = (1, \ldots, 1).
\]

PROOF. By the definition, there holds the equality

\[
P_k(x) = \prod \left( \frac{z_k - m_k(x | \omega_i, \varphi_i), i = 0, \ldots, n)}{s_k(x | \omega_i, \varphi_i), i = 0, \ldots, n)} \right),
\]

where \( \Pi(z) = (1/\sqrt{2\pi}) \int_0^z \exp(-t^2/2) dt \). Since the function \( \Pi(\cdot) \) is monotonically increasing, then for the constant \( s_k(x | \omega_i, \varphi_i), i = 0, \ldots, n) \), the probability \( P_k(x) \) increases with increasing numerator of (7). Therefore, in the interior of simplex, the maximum of \( P_k(x) \) may be achieved only on the line \( x = x_S - t \cdot \nabla x m_k(x | \omega_i, \varphi_i), i = 0, \ldots, n) \). Since \( m_k(\cdot) \) is a linear function with values \( \varphi_i \) at the points \( \omega_i \), the expression for \( x \) may be rewritten in the following form:

\[
x = x_S - t \cdot (\varphi - \varphi_* \cdot I).
\]

Substituting \( x \) in (8) into (7) reduces the problem of maximization of \( P_k(x) \) to a one-dimensional maximization problem, which is unimodal and similar to the maximization of the probability in the case of construction of the one-dimensional \( P \)-algorithm; see, e.g., [2].

The derivation of formula (6) from the necessary conditions of local maximum is an elementary task but involves operations with rather complicated expressions.

REMARK 1. If the point defined by (6) does not belong to the simplex, then the maximum point of \( P_k(x) \) is at the facet defined by the vertices \( \omega_0, \ldots, \omega_{n-1} \). To find the maximum point taking into account the latter constraint, the proof of the theorem should be repeated for the \( n - 1 \)-dimensional simplex and the projection of the gradient onto the constraint, etc. If not at a previous step, the maximum point, finally, will be found at the edge of the simplex corresponding to the vertices \( \omega_0, \omega_1 \). However, it is very likely that the neighbor simplex should be selected if maximum of \( P_k(x) \) is achieved on a facet of the considered simplex. Therefore, a rough estimate of the maximum point may be sufficient, e.g., \( x_0 \) obtained by means of a step from the point \( x_S \) in the antigradient direction

\[
x_0 = x_S - \frac{\varphi - \varphi_* \cdot I}{\| \varphi - \varphi_* \| \sqrt{n(n + 1)}},
\]

where the step length is equal to the shortest distance from the point \( x_S \) to the facet of the simplex.

REMARK 2. The selection criterion of an arbitrary simplex is defined equal to the selection criterion of the regular simplex whose perimeter is equal to the perimeter of the former simplex. Therefore, \( P_k(x_{\text{max}}) \) may be calculated using formulas (6), (9), (7), (2), and substituting (3) by

\[
s_k^2(x | x_i, i = 1, \ldots, k) = \sigma_0^2 \cdot (D_k^2 - \| x_S - x \|^2) \cdot \frac{d^2}{2},
\]

where \( d = 2p/(n + 1)n \) denotes the average length of the edge of the simplex with perimeter length \( p \).

REMARK 3. If \( \varphi_0 = \varphi_1 = \cdots = \varphi_n \), then for the regular simplex there holds the equality \( x_{\text{max}} - x_S \) and the maximal probability is equal to

\[
P_k(x_{\text{max}}) = \Pi \left( \frac{z_k \sqrt{n + 1}}{\sigma_0 d \sqrt{2n}} \right).
\]
4. SIMPLICIAL PARTITIONING

Let $A_0$ denotes union of simplexes of an initial covering of $A$, $A_0 \supseteq A$. The refinement of the cover includes selection of a simplex and its subdivision into several smaller simplexes, which are, loosely speaking, as regular as possible. In the two-dimensional case, a special case of partitioning, i.e., cloning, may be applied producing descendants similar to the parent. Two examples are presented in Figure 1. No method of partitioning a regular simplex into regular subsimplexes is known for $n \geq 3$. To restrict analysis with regular simplexes cloning might be generalized to covering: instead of partitioning of a parent simplex, it might be covered by similar smaller correct simplexes. The disadvantage of the latter type of cloning is overcovering of $A$. If the requirement of similarity of descendants to the parent is be relaxed, then a variety of partitioning procedures might be proposed. For example, from a regular $n$-dimensional simplex $n + 1$ descendant regular simplexes with half-sized edges may be obtained as a result of cutting of the regular parent simplex by means of $n + 1$ hyperplanes, which are parallel to the facets of the parent simplex and divide the corresponding edges 1:1. The vertexes of descendant simplexes are: one vertex coincides with the vertex of parent simplex and the $n$ other vertexes coincide with centers of edges, whose intersection produces the corresponding vertex of the parent simplex. The rest of the parent simplex should be partitioned taking into account new available vertexes. A three-dimensional version of such a semiregular partitioning is shown in Figures 2a and 3. A simple procedure of bisect the longest is illustrated by Figures 1b and 2b. Similar procedures are successfully applied also in Branch and Bound methods based on deterministic models of the objective function, see, e.g., [15-17]. We will assume that partitioning produces the bounded number of descendants with the bounded ratio of the longest to the shortest edges: $\frac{d_{\text{max}}}{d_{\text{min}}} \leq \Gamma$. 

![Figure 1. Examples of cloning of two-dimensional simplexes.](image1.png)

![Figure 2. Examples of partitioning of three-dimensional simplexes.](image2.png)
To start the optimization, the initial simplexes should be available. We consider the two versions of initial covering presented in Figures 4 and 5.
5. CONVERGENCE

Let \( f_0(x) \) be a continuous function extending \( f(x) \) to \( A_0 \), \( f_0(x) = f(x) \), for \( x \in A \); \( f_0(x) > \min_{x \in A} f(x) \) for \( x \in A_0 \setminus A \). The penalty function \( f_0(x) \) is minimized on \( A_0 \) by means of the algorithm described in previous chapters.

**Theorem 2.** The trial points are everywhere dense in \( A_0 \) implying

\[
\lim_{k \to \infty} y_{0k} = \min_{x \in A} f(x),
\]

where \( y_{0k} = \min\{y_1, \ldots, y_k\} \).

**Proof.** It will be proved that the algorithm generates a sequence of trial points \( x_i \) that are everywhere dense in \( A_0 \); (11) follows immediately from the inequality \( f_0(x) > \min_{x \in A} f(x) \) for \( x \in A_0 \setminus A \), and the continuity of \( f(\cdot) \). Assume that the theorem is false: there exists a point \( x_- \in A_0 \) which is not a limit point of \( x_i \), i.e., there exists an \( \epsilon \) vicinity of \( x_- \) without points \( x_i \). Let us suppose that \( x_- \) belongs to the simplexes of initial cover \( S^1 \), \( l \in L \). By means of bread-first search partitioning of \( S^l \) after a finite number of steps, the descendant simplexes \( S'_l \) with the average edge length no longer than \( d_- \) will be obtained such that \( x_- \) is an inner point of \( S_- = \cup S'_l \), \( S_- \subset Sph(x_-, \epsilon) \). Since \( P_k(x), x \in S^l \), is a monotonic decreasing function of objective function values at the vertexes of \( S'_l \) \[14\], the inequality

\[
\max_{x \in S_-} P_k(x) \geq \Pi \left( \frac{\min_{x \in A} f(x) - \max_{x \in A_0} f_0(x) - \epsilon}{\sigma_0 d_- \sqrt{2n}} \right)^{\frac{n+1}{2}}
\]

follows from Remark 3.

Since the set \( A_0 \) is bounded the sequence \( x_i \) has at least one limit point, e.g., \( x_+ \). The sequence of simplexes, whose average edge lengths \( d^+_1 \to 0 \), should be generated by the algorithm in order to generate the points of \( x_i \) in any vicinity of \( x_+ \). From the above-mentioned monotonicity of \( P_k(x) \) and Remark 3, it follows that for the simplex with edge length \( d_+ \), the maximal probability is not larger than \( \Pi(\epsilon \sqrt{n + 1}/\sigma_0 d_+ \sqrt{2n}) \).

The assumption that \( x_- \) is not a limit point of \( x_i \) implies that the simplexes in \( S_- \) are never selected by the minimization algorithm performing the best first search partitioning. Therefore, the simplexes with average edge lengths

\[
d_+ < \frac{d_- \epsilon}{(\max_{x \in A_0} f_0(x) - \min_{x \in A} f(x) + \epsilon)},
\]

whose maximal probability is smaller than maximal probability of simplexes in \( S_- \), also cannot be selected. Therefore, \( x_1 \) cannot be a limit point of the sequence \( x_i \). The obtained contradiction proves the theorem.

For multidimensional problems, an estimate of the global minimum obtained by means of the global technique in practically acceptable time may be rather rough. Therefore, it may be reasonable to combine this technique with a fast converging local algorithm. Such a combination may be considerably faster than a pure global optimization algorithm \[9\]. Let us suppose that an objective function is smooth enough for the convergence of a descent method with order \( p > 1 \). Suppose that \( N \) iterations of the global technique are alternating with one iteration of local descent method from the best point found.

**Theorem 3.** Let the above assumptions be satisfied and assume that the objective function \( f(x) \) is twice continuously differentiable at the unique global minimum point \( x_0 \). The algorithm including local descent as described above, generates a sequence of estimates converging to the global minimum with convergence order better than one.

**Proof.** Since the global algorithm generates an everywhere dense sequence of points, then, after a finite number of iterations \( m \), the best point found will belong to the vicinity of \( x_0 \). Since the objective function is twice continuously differentiable at \( x_0 \), then

\[
f(x_{0, i}) - f(x_0) = O \left( \|x_{0, i} - x_0\|^2 \right),
\]

where \( x_{0, i} \) is the best point found after \( i > m \) iterations.
The sequence of local descent iterations generates points \( x_{l(t)} \), \( l(t) = t \cdot (N+1) \), \( t > m/(N+1) \), \( t = 1, 2, \ldots \), which converge to \( x_0 \), and the average order of convergence is \( p \) with respect to \( t \)

\[
\limsup \frac{\|x_{0l(t)} - x_0\|^{1/(p^t)}}{t} = 1, \quad p > 1. \tag{13}
\]

Equalities (12), (13) yield

\[
\limsup \frac{\|x_0i - x_0\|^{1/(r^t)}}{t} = 1, \quad \limsup (f(x_0_i) - f(x_0))^{1/2r^t} = 1, \tag{14}
\]

where \( r = p^{1/(N+1)} > 1 \).

6. TESTING EXAMPLES

The test functions from [18] were used to test efficiency of the proposed algorithm. For the criterion of efficiency, we have accepted the number of calls of objective function required to find known global minimum with predefined accuracy.

Table 1. Minimization results with original \( \epsilon \).

(a) | Function | \( \epsilon \) | nfe1 | nfe2 | nfe3 | nfe4 |
--- | --- | --- | --- | --- | --- |
1 | 0.355 | 35 | 29 | 68 | 73 |
2 | 0.0446 | 11 | 11 | 14 | 14 |
3 | 11.9 | 2 | 2 | 2 | 2 |
3.1 | 17.5 | 5 | 5 | 5 | 5 |
3.2 | 13.8 | 2 | 2 | 2 | 2 |
3.3 | 19.6 | 10 | 10 | 12 | 13 |
4 | 0.0141 | 5 | 5 | 5 | 5 |
5 | 0.1 | 42 | 58 | 26 | 33 |
6 | 44.9 | 3 | 3 | 3 | 3 |
7 | 542.0 | 2 | 2 | 2 | 2 |
8 | 3.66 | 3 | 3 | 3 | 3 |
9 | 62900 | 1 | 1 | 1 | 1 |
9.1 | 5.47 \times 10^6 | 1 | 1 | 1 | 1 |
9.2 | 4.93 \times 10^5 | 1 | 1 | 1 | 1 |
9.3 | 3.93 \times 10^3 | 1 | 1 | 1 | 1 |
10 | 0.691 | 10 | 12 | 18 | 23 |
11 | 0.335 | 6 | 6 | 6 | 6 |
12 | 0.804 | 2 | 2 | 2 | 2 |
13 | 0.92 | 1 | 1 | 1 | 1 |

(b) | Function | \( \epsilon \) | nfe1 | nfe2 |
--- | --- | --- | --- |
21 | 0.369 | 166 | 91 |
22 | 101 | 1 | 1 |
23 | 8.33 | 1 | 1 |
24 | 0.672 | 26 | 26 |
25 | 0.0506 | 1781 | 1747 |
26 | 4.51 | 2 | 2 |

In Table 1, the results of minimization by the proposed algorithm with accuracy from Tables XI and XIV in [18] are given. The numbers of the test functions according to [18] are presented in the first column, and the predefined accuracy is presented in the second column. The number of function calls is given in the further columns. The results of the version of the algorithm corresponding to the partitioning of the feasible region (as in Figure 4b for two dimensions and as in Figure 5 for three dimensions) are given under headings nfe1 and nfe2. The columns under headings nfe3 and nfe4 present the results of the version of embedding of a feasible region into a
Global Optimization

Table 2. Minimization results with higher accuracy.

<table>
<thead>
<tr>
<th>Function</th>
<th>ε</th>
<th>nfe1</th>
<th>nfe2</th>
<th>nfe3</th>
<th>nfe4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00355</td>
<td>74</td>
<td>119</td>
<td>1035</td>
<td>1149</td>
</tr>
<tr>
<td>2</td>
<td>0.000446</td>
<td>84</td>
<td>68</td>
<td>35</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>0.119</td>
<td>83</td>
<td>79</td>
<td>199</td>
<td>204</td>
</tr>
<tr>
<td>3.1</td>
<td>0.175</td>
<td>29</td>
<td>44</td>
<td>103</td>
<td>70</td>
</tr>
<tr>
<td>3.2</td>
<td>0.138</td>
<td>46</td>
<td>42</td>
<td>154</td>
<td>164</td>
</tr>
<tr>
<td>3.3</td>
<td>0.196</td>
<td>79</td>
<td>68</td>
<td>41</td>
<td>46</td>
</tr>
<tr>
<td>4</td>
<td>0.000141</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>0.001</td>
<td>38</td>
<td>58</td>
<td>79</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>0.449</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>5.420</td>
<td>25</td>
<td>16</td>
<td>16</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>0.0366</td>
<td>418</td>
<td>309</td>
<td>56</td>
<td>52</td>
</tr>
<tr>
<td>9</td>
<td>629</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>9.1</td>
<td>54700</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>9.2</td>
<td>4030</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>9.3</td>
<td>39.3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>0.00691</td>
<td>365</td>
<td>500</td>
<td>339</td>
<td>532</td>
</tr>
<tr>
<td>11</td>
<td>0.00335</td>
<td>81</td>
<td>75</td>
<td>211</td>
<td>204</td>
</tr>
<tr>
<td>12</td>
<td>0.00804</td>
<td>41</td>
<td>27</td>
<td>71</td>
<td>30</td>
</tr>
<tr>
<td>13</td>
<td>0.0092</td>
<td>230</td>
<td>319</td>
<td>105</td>
<td>76</td>
</tr>
</tbody>
</table>

Table 3. Minimization results with high accuracy.

<table>
<thead>
<tr>
<th>Function</th>
<th>ε</th>
<th>nfe1</th>
<th>nfe2</th>
<th>nfe3</th>
<th>nfe4</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.00449</td>
<td>353</td>
<td>326</td>
<td>95</td>
<td>89</td>
</tr>
<tr>
<td>7</td>
<td>0.0542</td>
<td>2647</td>
<td>3180</td>
<td>996</td>
<td>919</td>
</tr>
<tr>
<td>9</td>
<td>6.29</td>
<td>14</td>
<td>12</td>
<td>17</td>
<td>33</td>
</tr>
<tr>
<td>9.1</td>
<td>547</td>
<td>38</td>
<td>24</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>9.2</td>
<td>49.3</td>
<td>22</td>
<td>14</td>
<td>30</td>
<td>51</td>
</tr>
<tr>
<td>9.3</td>
<td>0.393</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

The results of the superior version (partitioning of the feasible region and “bisect the longest”) of the proposed algorithm is illustrated below for two popular test functions. The functions from [18] are frequently used for testing of global optimization algorithms. They represent specific difficulties, especially for algorithms with guaranteed accuracy. Let us consider two more
test functions which represent a different type of challenge. The function

\[ f(x) = \sum_{i=1}^{2} x_i^2 - \cos(18x_i), \quad -1 \leq x_i \leq 1, \quad i = 1, 2, \]

with global minimum point \( x_0 = (0,0) \) and minimum value \( f(x_0) = 0 \), is known as the Rastrigin test function. It is widely used for testing of global optimization algorithms [8]. The second function is a generalization of a popular one-dimensional test function by Shubert to two dimensions [8,18]

\[ f(x) = \sum_{i=1}^{2} \sum_{j=1}^{5} -j \cdot \sin((j + 1)x_i + j), \quad -10 \leq x_i \leq 10, \quad i = 1, 2, \]

where global minimum of both summands is equal to \(-12.0312\) and it is attained at three points: \(-6.77458, -0.49139, \) and \(5.79179\). Both functions represent oscillating objective functions without subregions of the very steep growth as well as without the subregions of flatness. Such functions were considered as the prototypes of the objective functions for the construction of the statistical models of multimodal functions.

The stopping criterion of Lipshitz algorithm guarantees the estimating of global minimum within tolerance \( \varepsilon \). The stopping criterion of the proposed algorithm is applied meaning the high probability of estimating of global minimum with accuracy \( \varepsilon \). To choose the reasonable \( \varepsilon \), the accuracy guaranteed by the quadratic grid \( 100 \times 100 \) is estimated using the Lipshitz constants \((27.7,96.8)\). They are equal to \(0.392\) and \(13.7\), correspondingly. For the Rastrigin function, the tolerance \(0.1\), \(0.0392\), \(0.01\) was chosen and for the generalized Shubert function, the tolerance \(1.37, 0.1, \) and \(0.01\) was chosen.

Table 4. Minimization results for two oscillating functions by the proposed algorithm.

<table>
<thead>
<tr>
<th>Func.</th>
<th>( \varepsilon )</th>
<th>( N )</th>
<th>( \varepsilon )</th>
<th>( N )</th>
<th>( \varepsilon )</th>
<th>( N )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rastrig.</td>
<td>0.1</td>
<td>668</td>
<td>0.0392</td>
<td>1113</td>
<td>0.01</td>
<td>1140</td>
</tr>
<tr>
<td>Shubert</td>
<td>1.37</td>
<td>366</td>
<td>0.1</td>
<td>2664</td>
<td>0.01</td>
<td>2464</td>
</tr>
</tbody>
</table>

The number of function evaluations by the proposed algorithm in the case of different predefined accuracy are presented in Table 4. In all cases, the value \(-2.0000\) at the point \((0.0000,0.0000)\) was found for the Rastrigin function. The following values were found for the Shubert function: the value \(-23.9665\) at the point \((-0.4688,5.7813)\) in the case \( \varepsilon = 1.37 \), the value \(-24.0534\) at the point \((-0.4883,-6.7676)\) in the case \( \varepsilon = 0.1 \), the value \(-24.0612\) at the point \((-6.7773,5.7910)\) in the case \( \varepsilon = 0.01 \). The rare anomaly of the number of function evaluations is observed for the Shubert function. The algorithm stops with a smaller number of function evaluations for \( \varepsilon = 0.01 \) than for \( \varepsilon = 0.1 \). This anomaly is explained by the influence of \( \varepsilon \) to the search strategy, not only to the stopping condition. For this particular case, the select defined by \( \varepsilon = 0.01 \) was more efficient, and important simplexes were partitioned earlier than in the case \( \varepsilon = 0.1 \).

Table 5. Minimization of two oscillating functions by the algorithm of [19].

<table>
<thead>
<tr>
<th>Func.</th>
<th>( \varepsilon )</th>
<th>( N )</th>
<th>( \varepsilon )</th>
<th>( N )</th>
<th>( \varepsilon )</th>
<th>( N )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rastrign., ( L = 27.7 )</td>
<td>0.1</td>
<td>1015</td>
<td>0.0392</td>
<td>1407</td>
<td>0.01</td>
<td>2057</td>
</tr>
<tr>
<td>Shubert, ( L = 96.8 )</td>
<td>1.37</td>
<td>7776</td>
<td>0.1</td>
<td>11640</td>
<td>0.01</td>
<td>11948</td>
</tr>
</tbody>
</table>

The same test functions were minimized by the Lipshitz algorithm with simplex based covering [19]. The results are presented in Table 5. The number of function evaluations of the proposed algorithm is much smaller than the number of function evaluations by algorithm of [19]. On the other hand, Lipshitz algorithms find the global minimum with guarantee.
7. CONCLUSIONS

A statistical model, widely used for one-dimensional global optimization, is generalized to the multidimensional case. An approach for constructing global optimization algorithms is developed, which is free of some crucial difficulties characteristic to previous implementations based on statistical models of multimodal functions. The proposed algorithm inherits many advantages of the one-dimensional prototype. The results of numerical experiments are promising. New implementations in frame of the proposed approach would be interesting, e.g., combination of partitioning with local descent iterations, and extension to $n > 3$.

REFERENCES