A Global Optimization Method Based on the Reduced Simplicial Statistical Model

Antanas Žilinskas and Julius Žilinskas

Institute of Mathematics and Informatics, Vilnius University

Akademijos 4, LT-08663 Vilnius, Lithuania

E-mail(*corresp.*): antanas.zilinskas@mii.vu.lt

E-mail: julius.zilinskas@mii.vu.lt

Abstract. A simplicial statistical model of multimodal functions is used to construct a global optimization algorithm. The search for the global minimum in the multidimensional space is reduced to the search over the edges of simplices covering the feasible region combined with the refinement of the cover. The refinement is performed by subdivision of selected simplices taking into account the point where the objective function value has been computed at the current iteration. For the search over the edges the one-dimensional P-algorithm based on the statistical smooth function model is adapted. Differently from the recently proposed algorithm here the statistical model is used for modelling the behaviour of the objective function not over the whole simplex but only over its edges. Testing results of the proposed algorithm are included.

Keywords: global optimization, statistical models, simplicial partition.

AMS Subject Classification: 90C26.

1 Introduction

Although the majority of global optimization methods are constructed using various heuristics the approaches based on theoretical models of multimodal objective functions remain in the focus of theoretically oriented researchers. To maintain some average optimality of global search statistical models of multimodal objective functions are needed. For the basic theoretical ideas and their implementations we refer to the monographs [5, 6, 9, 10]. A disadvantage of global optimization algorithms aimed at some optimality is time-consuming auxiliary computations which are needed to plan the current iteration. The computational burden depends on the complexity of computing the optimality criterion with respect to a model of objective functions, e.g. the probability to improve the estimate of the global minimum at the current iteration. For the stochastic models and radial basis functions models, those computations

include the inversion of a matrix of the order equal to the number of objective function values computed during the previous iterations [13]. The computational difficulty, caused by the inversion of large matrices, can be avoided using the simplified multidimensional statistical models [14, 15], however auxiliary computations also in this case remain rather intensive. The reduction of auxiliary computations is an urgent and challenging problem of statistical-models-based global optimization.

A sub-class of one-dimensional models, namely the Markovian stochastic functions, has computationally favourable properties that enable (avoiding the inversion of large matrices) the development of efficient one-dimensional global optimization algorithms. An example of that sub-class is the Wiener process which was the first model used for those purposes [10]. However, for smooth functions the P-algorithm based on the smooth function model has better convergence properties as shown in [1]. Our idea is to apply the efficient one-dimensional P-algorithm based on the smooth function model as a subroutine of a multidimensional search algorithm. The search is performed over the edges of an iteratively refined simplicial covering of the feasible region. To justify the use of the P-algorithm for the one-dimensional minimization a reduced statistical model is used modelling the behaviour of the objective function not over whole feasible region but only over the edges of the current simplicial covering.

Some other advantages of simplicial covering of the feasible region are outlined in [17]. Since a simplex is a polyhedron in n-dimensional space with the minimal number of vertices, simplicial partitions are preferable when the values of an objective function at all vertices of partitions are used during optimization. Simplicial partitions may be used to vertex triangulate feasible regions of non rectangular shape defined by linear inequality constraints, that may be used to avoid symmetries in optimization problems [16]. Simplicial statistical models may be used to select a candidate simplex [8] in simplicial branch and bound algorithm with Lipschitz bounds [7] that reduces memory requirements and makes optimization faster compared with the common best-first selection strategy.

2   The Reduced Simplicial Statistical Model

Let us consider the global minimization (GM) 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. As shown in [12] the typical for 'black box' optimization uncertainty about an objective function can be represented by a statistical model of $f(x)$. However, the implementation of multidimensional GM algorithms based on the classical statistical models (random fields) is challenging because of time consuming auxiliary computations. The simplicial statistical model proposed in [15] is computationally simpler than a Gaussian random field however it can be further simplified as it is shown below.

Let us briefly introduce the P-algorithm based on simplicial statistical model [15]. At the $(k + 1)$-th minimization iteration A is covered by simplices S^j, $j = 1, \ldots, m$, and k values of the objective function are computed at the vertices x_i of the simplices. A family of Gaussian variables ξ_x is chosen for

the statistical model of the objective function where the mean of ξ_x, $x \in S^j$ is piecewise linear over S^j and the variance of ξ_x, $x \in S^j$ is piecewise quadratic over S^j. The next value of $f(x)$ is computed at the point of the simplex selected by maximization of the probability $P_k(x) = \mathbf{P}\{\xi_x \leq y_{ok}\}$ where $y_{ok} = \min_{i=1,\dots,k} f(x_i) - \varepsilon$, and ε denotes an improvement threshold. For a discussion about the choice of value of ε we refer to [2, 10]. The maximization of $P_k(x)$ over a particular simplex is facilitated by the properties of $P_k(x)$ considered as the objective function of the auxiliary maximization problem. The selected simplex is partitioned by one of several partitioning methods. The experience in testing a version of the P-algorithm, which is proposed in [15], has shown that the maximum point of $P_k(x)$ is frequently attained at an edge of the considered simplex. In such a case the selected simplex is partitioned into two simplices using the maximum point of $P_k(x)$ as a new vertex of both descendant simplices. Even in the case when the maximum point is an interior point of the considered simplex a method of partition using the maximum point as a pivot point of partition is not necessarily superior to the bisection using e.g. the middle point of the longest edge for the new vertex of the descendant simplices.

Let us construct the GM algorithm based on simplicial covering of the feasible region and the bisection of the selected simplex using a point on an edge for a new vertex of both descendant simplices. Then for the construction of the algorithm only the statistical model defined over edges of simplices is needed. Based on the comparison of various statistical models in [11] and on our past experience, we chose the statistical smooth function model defined in [1] to model the behaviour of the objective function over the edges of the simplicial partition. That model is a stationary Gaussian stochastic process with the correlation function $r(\cdot)$ satisfying the following regularity assumptions

$$r(t) = 1 - \frac{1}{2}\lambda_2 t^2 + \frac{1}{4!}\lambda_4 t^4 + o(t^4),$$

$$\left|\frac{d^4 r(t)}{dt^4} - \lambda_4\right| = \mathcal{O}(|t|), \quad -\frac{d^2 r(t)}{dt^2} = \lambda_2 + \mathcal{O}(|\log^{-a}|t||),$$

for finite λ_2, λ_4, and $a > 1$, as $t \to 0$, and $r(t)\log(t) \to 0$, as $t \to \infty$.

3 The P-algorithm Based on a Reduced Simplicial Statistical Model

The search for the global minimum at the $(k+1)$-th minimization iteration is considered. The feasible region is covered by m simplices, and k function values are computed at the vertices of the simplices. Assume that the next point for computation of $f(x)$ can be chosen on an edge only, and that the behaviour of $f(x)$ over an edge is modeled by the smooth function model. The latter assumption implies that the maximum point of $\max_x P_k^{ij}(x) = \mathbf{P}\{\xi(x) \leq y_{ok}\}$ where x belongs to the segment of line connecting vertices x_i and x_j can be computed by a simple formula [1]. An appropriate monotonically increasing function of $\max_x P_k^{ij}(x)$ can be used to compute the criterion for the selection of simplices. Let γ^{ij} is used, which is monotonically related to $\max_x P_k^{ij}(x)$. Therefore, at

the current search iteration the vertices x_i and x_j are selected according to the maximum criterion γ^{ij}, and the point x_{k+1} for the next computation of the objective function value is defined by the following formulas:

$$x_{k+1} = x_j + \tau^{ij}(x_i - x_j), \tag{3.1}$$

$$\tau^{ij} = \frac{1}{1 + \sqrt{\frac{y_j - y_{ok} + \varepsilon}{y_i - y_{ok} + \varepsilon}}}, \quad y_i = f(x_i), \ \ y_j = f(x_j). \tag{3.2}$$

The point x_{k+1} is not only the site of the current computation of the objective function value but also the pivot of the refinement of the simplicial covering. The refinement means the bisection of all simplices sharing the selected edge: the point x_{k+1} is the common vertex of all new simplices.

The termination condition of a search method is defined according to the general concept of the method. For example, termination conditions of local descent methods normally require fulfilment of necessary minimum conditions with great precision. However such a termination condition would not necessarily be rational to apply for a GM method. Many GM methods are not very efficient for local refinement of the identified local minimizers. Therefore, hybrid methods seem favourable where a GM method is used to find rough approximations of the main minimizers, and the latter are refined by a local minimization method. Although in the present paper we do not consider the hybridization, the termination condition will be defined supposing the refinement of the best or several best candidate solutions by a local method.

During the search the edges incident to the best function values found become shortest because of (3.2). The current point for computation of the objective function value by the P-algorithm is selected on the edge where the improvement is most probable. Therefore, the selection of the shortest edge indicates likelihood that it is located close to the true minimizer. Theoretical investigation of the relation between the length of the shortest edge and the distance between the best point found and the true minimizer is extremely difficult. Nevertheless such a relation can be guessed, and it seems reasonable to stop GM when the length of the selected edge becomes shorter than a predefined Δ guessing that the current best point belongs to the Δ vicinity of a true minimizer not including other local minimizers.

The choice of the parameter ε is motivated by its influence to the strategy of search of the original one-dimensional P-algorithm where the increase of ε makes the search more global, and, contrary, the vanishing of ε implies stronger concentration of trial points in the vicinity of the best found point ensuring better asymptotic convergence of the candidate solution to the minimizer [2, 10]. The proposed multidimensional algorithm consists from single steps of one-dimensional P-algorithm repeated with different input data, and it is supposed for a rough finding of a global minimizer but not for its precise computation. Therefore, ε should be chosen sufficiently large. An acceptable empirical rule of thumb is to choose the value equal to a half of difference between average of computed objective function values and the current minimum.

The search by the constructed algorithm is illustrated by the results of minimization of two well known test functions. The first considered test function

by Rastrigin is defined by the following formula

$$f(x) = x_1^2 + x_2^2 - \cos(18x_1) - \cos(18x_2), \quad -1.5 \le x_1 \le 0.5, \quad -0.5 \le x_2 \le 1.5,$$

it has more than 30 local minima in the feasible region. Contour lines of this function are presented in Figure 1a. The global minimum is equal to -2, and it is attained at the point $(0,0)$.
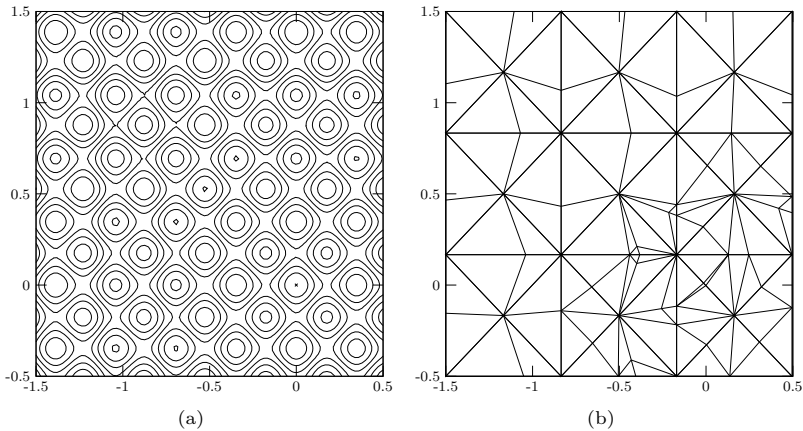

(a)   (b)

Figure 1. The Rastrigin function presented by contour lines and the simplicial covering of the feasible region after termination with $\Delta = 0.1d$

With the parameter defining the very rough termination condition $\Delta = 0.1d$, where d is the length of the diagonal of the feasible region, the algorithm has stopped after 71 computations of the objective function values. The best found value was -1.9986 and the distance between the best found point and the true solution point was $0.001d$. The simplicial covering of the feasible region after termination is shown in Figure 1b. Such a fast location of the global minimizer with a high precision can be considered as occasional. However, the continuation of search (by decreasing Δ) shows rational concentration of points in the attraction regions of the main local minima. In the case of $\Delta = 0.05d$ the algorithm stopped after 182 computations of the objective function values, and the simplicial covering is shown in Figure 2a; the previous approximation was not improved. The distribution of 361 point where the objective function values were computed before stopping with $\Delta = 0.03d$ are shown in Figure 2b. The further concentration of points in the regions of attraction of main local minima seems not to be rational and can be explained by inadequacy of the statistical model to describe the local behaviour of the objective function. For this test function the rational termination condition is defined by $\Delta = 0.05d$.

A similar experiment has been performed with another well known test function by Branin. The function is defined by the formula

$$f(x) = \left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6 \right)^2 + 10\left(1 - \frac{1}{8\pi} \right)\cos(x_1) + 10,$$

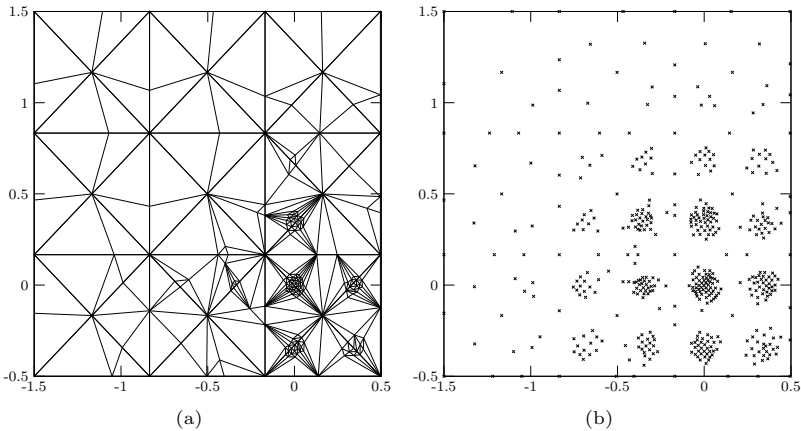$$-5 \le x_1 \le 10, \ 0 \le x_2 \le 15,$$

Figure 2. The illustration of minimization results of the Rastrigin function: simplicial covering of the feasible region after termination with $\Delta = 0.05d$ and points where function values were computed after termination with $\Delta = 0.03d$

and its contour lines are presented in Figures 3a. The global minimum of the Branin function equal to 0.399 is attained at three points $(-3.142, 12.275)$, $(3.142, 2.275)$, and $(9.425, 2.475)$.
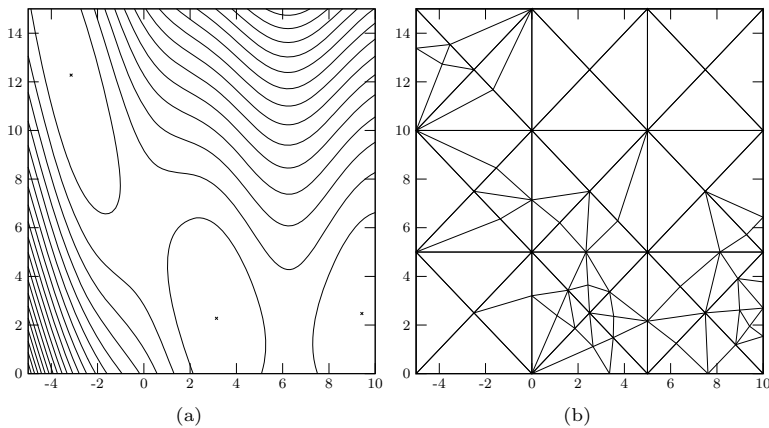

Figure 3. The Branin function presented by contour lines and the simplicial covering of the feasible region after termination with $\Delta = 0.1d$

The minimization was performed similarly as in the previous case. The algorithm with $\Delta = 0.1d$, $\Delta = 0.05d$, and $\Delta = 0.03d$ has stopped after 65, 151, and 293 computations correspondingly. The best function values found were 0.5343, 0.4620 and 0.4056. The distances between the best point found and the global minimizer were $0.0279d$, $0.0096d$, and $0.0117d$. The search results are illustrated by Figures 3b and 4. The results of this experiment support the suggestion of the previous experiment about rationality of choice $\Delta = 0.05d$.
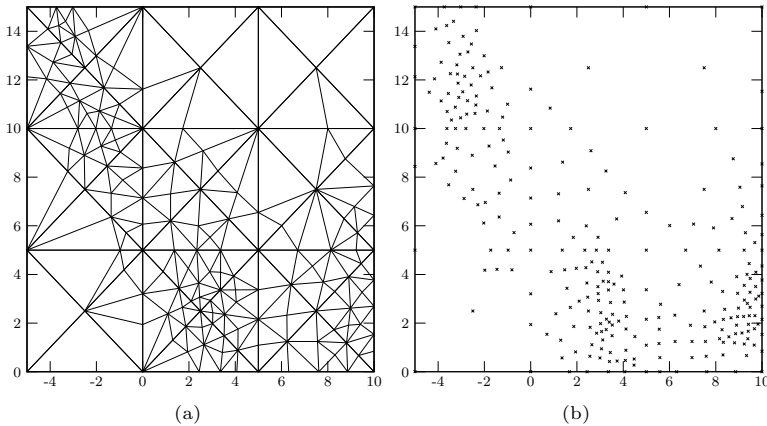
Figure 4. The illustration of minimization results of the Branin function: simplicial covering of the feasible region after termination with $\Delta = 0.05d$ and points where function values were computed after termination with $\Delta = 0.03d$

4 Testing Results

Various test problems ($n \geq 2$) for global optimization from [3, 4] have been used in our experiments. Test problems with ($n = 2$ and $n = 3$) are numbered according to [3]. For ($n \geq 4$) function names from [4] are used. The results are shown in Table 1. The known global minimum is denoted by f^*, the approximation found using the proposed algorithm is indicated as f_{min}, the relative distance between the known minimizer and the best point found is denoted by $\Delta X/d$, the last column shows the number of function evaluations (NFE).

Let us comment on the testing results for $n = 2$ and $n = 3$ first. The tolerance in the termination condition was set equal to $\Delta = 0.05d$ based on the discussion in the previous section. In all cases except 13 and 23 the global minimum point has been found with precision exceeding Δ. The number of function calls is considerably smaller than in the case of a Lipschitz model-based algorithm with rougher precision [3]; this advantage is gained at the price of withdrawal of the guaranteed solution with prescribed accuracy (of any problem with known Lipschitz constant). The failure could occur in specific cases which, we believe, are not very common in practical applications. For example, the best found point for test problem 13 was (0.0762, 0.1829) with function value 0.0599. The true minimum point is (0.0115, 0.0144) with the function value 0.0001. Let us mention that the function value at the point (0.01, 0.01) equal to 1.1304 was computed by the algorithm. Such sharp changes of function values (similar to discontinuity with respect to changes of arguments of size 0.001) are not supposed by the used statistical model. Similar considerations apply to test problem 23. For other test problems both approximation of the global minimum and global minimizer are close to the known ones.

Table 1. Results of experimental investigation

Test problem	n	Δ	f^*	f_{min}	$\Delta X/d$	NFE
1 [3]	2	0.05d	−2.5200	−2.5196	0.0009	111
2 [3]	2	0.05d	−2.8186	−2.8184	0.0068	156
3 [3]	2	0.05d	0.3979	0.4630	0.0089	202
4 [3]	2	0.05d	0.0000	0.0000	0.0000	157
5 [3]	2	0.05d	−0.3967	−0.3658	0.0137	136
6 [3]	2	0.05d	−239.6966	−239.6411	0.0119	171
7 [3]	2	0.05d	0.0000	0.2066	0.0052	349
8 [3]	2	0.05d	0.4500	0.8500	0.0168	128
9 [3]	2	0.05d	3.0001	3.0004	0.0003	407
10 [3]	2	0.05d	−1.9132	−1.9114	0.0051	175
11 [3]	2	0.05d	0.1690	0.1694	0.0111	243
12 [3]	2	0.05d	1.7442	1.7442	0.0056	299
13 [3]	2	0.05d	0.0001	0.0599	0.1289	192
20 [3]	3	0.05d	0.0000	0.0000	0.0021	4031
21 [3]	3	0.05d	−3.8628	−3.8590	0.0416	1002
23 [3]	3	0.05d	0.0155	0.1525	0.1923	1326
24 [3]	3	0.05d	−1.0000	−0.9974	0.0103	127
25 [3]	3	0.05d	−0.5164	−0.5158	0.0047	391
26 [3]	3	0.05d	−36.0000	−35.9919	0.0033	1168
Levy No. 15 [4]	4	0.05d	0.0000	0.8489	0.0129	12834
Schwefel No. 1.2 [4]	4	0.05d	0.0000	0.0000	0.0000	59016
Levy No. 9 [4]	4	0.05d	0.0000	0.0348	0.0168	28178
Levy No. 16 [4]	5	0.05d	0.0000	0.1607	0.0164	> 93573
Levy No. 10 [4]	5	0.05d	0.0000	0.0983	0.0088	> 96551
Schwefel No. 3.7 [4]	5	0.05d	0.0000	0.0000	0.0278	> 78237
Levy No. 15 [4]	4	0.1d	0.0000	2.4479	0.0236	1371
Schwefel No. 1.2 [4]	4	0.1d	0.0000	0.0000	0.0000	4914
Levy No. 9 [4]	4	0.1d	0.0000	0.3474	0.0584	2323
Levy No. 16 [4]	5	0.1d	0.0000	0.9478	0.0319	6328
Levy No. 10 [4]	5	0.1d	0.0000	0.3410	0.0522	13137
Schwefel No. 3.7 [4]	5	0.1d	0.0000	0.0000	0.0278	58278

The number of function calls increases with the dimensionality of problems. It is natural since also the number of simplices (say regular simplices with edge length ϵ) needed to cover a unit cube increases exponentially with n. Assuming that the number of local minimizers of the targeted objective functions remains of the same order as in the case of considered two and three dimensional test functions the tolerance in the termination condition can be increased. The results in Table 1 show that with $\Delta = 0.1d$ the four and five dimensional test problems were solved successfully when the algorithm stopped after a relatively modest number of function evaluations.

5 Conclusions

A simplicial covering of the feasible region is suitable for adapting one-dimensional statistical models of multimodal functions to multidimensional global optimization. The proposed algorithm outperforms the Lipschitz model

based algorithm with respect to the number of function calls; this advantage is gained at the price of possible failures for atypical problems.

Acknowledgments

References

[1] J. Calvin and A. Žilinskas. On the convergence of the P-algorithm for one-dimensional global optimization of smooth functions. *J. Optim. Theory Appl.*, **102**(3):479–495, 1999. Doi:10.1023/A:1022677121193.

[2] J. Calvin and A. Žilinskas. One-dimensional P-algorithm with convergence rate $O(n^{-3+\delta})$ for smooth functions. *J. Optim. Theory Appl.*, **106**(2):297–307, 2000. Doi:10.1023/A:1004699313526.

[3] P. Hansen and B. Jaumard. Lipshitz optimization. In R. Horst and P. M. Pardalos(Eds.), *Handbook of Global Optimization*, volume 1, pp. 407–493. Kluwer Academic Publishers, 1995.

[4] C. Jansson and O. Knuppel. A global minimization method: The multi-dimensional case. Technical report, TU Hamburg-Harburg, 1992.

[5] J. Mockus. *Bayesian Approach to Global Optimization*. Kluwer, 1988.

[6] J. Mockus, W. Eddy and G. Reklaitis. *Bayesian Heuristic Approach to Discrete and Global Optimization*. Kluwer, 1996.

[7] R. Paulavičius and J. Žilinskas. Improved Lipschitz bounds with the first norm for function values over multidimensional simplex. *Math. Model. Anal.*, **13**(4):553–563, 2008. Doi:10.3846/1392-6292.2008.13.553-563.

[8] R. Paulavičius, J. Žilinskas and A. Grothey. Investigation of selection strategies in branch and bound algorithm with simplicial partitions and combination of Lipschitz bounds. *Optim. Lett.*, **4**(2):173–183, 2010. Doi:10.1007/s11590-009-0156-3.

[9] R. Strongin and Y. Sergeyev. *Global Optimization with Non-convex Constraints*. Kluwer, 2000.

[10] A. Törn and A. Žilinskas. Global optimization. *Lecture Notes in Comput. Sci.*, **350**:1–252, 1989. Doi:10.1007/3-540-50871-6.

[11] A. Zhigljavsky and A. Žilinskas. *Stochastic Global Optimization*. Springer, 2008.

[12] A. Žilinskas. Axiomatic approach to statistical models and their use in multimodal optimization theory. *Math. Program.*, **22**(1):104–116, 1982. Doi:10.1007/BF01581029.

[13] A. Žilinskas. On similarities between two models of global optimization: statistical models and radial basis functions. *J. Global Optim.*, **48**(1):173–182, 2010. Doi:10.1007/s10898-009-9517-9.

[14] A. Žilinskas and J. Žilinskas. Global optimization based on a statistical model and simplicial partitioning. *Comput.Math. Appl.*, **44**(7):957–967, 2002. Doi:10.1016/S0898-1221(02)00206-7.

[15] A. Žilinskas and J. Žilinskas. P-algorithm based on a simplicial statistical model of multimodal functions. *TOP*, **18**(2):396–412, 2010.
Doi:10.1007/s11750-010-0153-9.

[16] J. Žilinskas. Reducing of search space of multidimensional scaling problems with data exposing symmetries. *Information Technology and Control*, **36**(4):377–382, 2007.

[17] J. Žilinskas. Branch and bound with simplicial partitions for global optimization. *Math. Model. Anal.*, **13**(1):145–159, 2008.
Doi:10.3846/1392-6292.2008.13.145-159.