

# GENETIC STOCHASTIC METHOD OF GLOBAL EXTREMUM SEARCH FOR MULTIVARIABLE FUNCTION

**Sergej Ermakov**

St.Petersburg State University  
Russia  
s.m.ermakov@spbu.ru

**Liudmila Vladimirova**

St.Petersburg State University  
Russia  
l.vladimirova@spbu.ru

**Irina Rubtsova**

St.Petersburg State University  
Russia  
i.ribtsova@spbu.ru

**Alexey Rubanik**

St.Petersburg State University  
Russia  
st040092@student.spbu.ru

Article history:

Received 18.02.2022, Accepted 25.05.2022

## Abstract

This article is devoted to the development of stochastic methods of global extremum search. The modification of the annealing simulation algorithm [Ermakov and Semenchikov, 2019] is combined with the covariance matrix adaptation method [Ermakov, Kulikov and Leora, 2017]. In this case, an effective computational approach [Ermakov and Mitioglova, 1977] is used for modeling the multivariate normal distribution. The special algorithms of covariance matrices adaptation are suggested to avoid the obtaining a local extremum instead of a global one. The methods proposed are successfully applied to the problem of nonlinear regression parameters calculation. This problem often arises in physics and mathematics and may be reduced to global extremum search. In particular case considered the extremum of ravine function of 14 variables was found.

## Key words

Genetic stochastic algorithms; Global extremum; Covariance matrix; Normal distribution; Nonlinear regression.

## 1 Introduction

Computational and machine learning methods, among them extremum search algorithms, are widely discussed in the literature [Ermakov, Kulikov and Leora, 2017; Hansen and Ostermeier, 1996; Igel, Hansen and Roth, 2007; Sergeenko, Yakunina and Granichin, 2020; Granichin, Volkov, Petrov and Volkova, 2021]. The problems of finding the extremum of smooth convex functions with a small number of variables are well studied [Zhiglyavsky, 1985; Nesterov, 2010]. Stochastic methods are especially effective in solving problems of large dimensions.

For uniformly distributed trial points, the probability of getting into  $\varepsilon$ -neighborhood of the extremum point has the order  $\varepsilon^{1/s}$ , where  $s$  is the dimension of the search

area. An optimization problem with a large number of variables becomes especially difficult for multi-extreme and ravine functions. Therefore, only adaptive methods such as simulated annealing [Metropolis, Rosenbluth, Teller A. and Teller E., 1953] and related genetic algorithms can count on success. Two examples may be presented.

A. Let us consider an objective function  $f(X): f(X) > 0, |f(X)| \leq M, X \in D \subset E^s$ , where  $M$  is a constant,  $D$  is a bounded domain in Euclidian space  $E^s$ . The sequence of probability densities may be introduced [Vladimirova and Ovsyannikov, 2019]:

$$\bar{p}(T, X) = C(T) \exp\left(-\frac{f(X)}{T}\right), \quad (1)$$

where  $T$  is real parameter,  $T \in E^1$ ,  $C(T) = 1 / \int_D e^{-\frac{f(X)}{T}} dX$  is normalization constant. The sequence (1) converges weakly as  $T \rightarrow \infty$  to the Dirac  $\delta$ -function concentrated at the global minimum point of the function  $f(X)$ . The objective function is supposed to have a unique minimum, otherwise the situation gets more complicated. The extremum search algorithm uses simulation by the Metropolis method [Metropolis, Rosenbluth, Teller A. and Teller E., 1953; Vladimirova and Ovsyannikov, 2019].

B. One of the adaptive methods that have been successfully used in recent years is the covariance matrix method, which was originally proposed in 1977 in [Ermakov and Mitioglova, 1977]. This method was used to find the maximum of target function  $f(X)$ , and another density of the type (1) was proposed:

$$p(X, \nu) = \frac{f^\nu(X)}{\int_D f^\nu(X) dX}. \quad (2)$$

It is assumed that  $f(X) > 0$ , the integral  $\int_D f^\nu(X) dX$  exists and is finite for any finite integer degree  $\nu$ . If the

global maximum of the function  $f(X)$  is unique, then the densities (2) also weakly converge (as  $\nu \rightarrow \infty$ ) to the  $\delta$ -function concentrated at the point of the global maximum  $f(X)$ . The behavior of density (2) in the general case can be complicated. To simplify the simulation of density (2), the authors of the paper [Ermakov and Mitioglova, 1977] replaced it with a similar density

$$\bar{p}(X, k) = C_k f(X) N(X, X_{\max}^{(k)}, B^{(k)}), \quad k = \overline{1, K}. \quad (3)$$

Here  $k$  is iteration number;  $K$  is the total number of search stages;  $C_k$  is normalization constant;  $N(X, X_{\max}^{(k)}, B^{(k)})$  is normal distribution density; the mean  $X_{\max}^{(k)}$  is the estimate of  $f(X)$  extremum point calculated at  $(k-1)$ -th iteration;  $B^{(k)}$  is covariance matrix estimate obtained at  $(k-1)$ -th iteration. At every iteration this matrix reflects the properties of the function  $f(X)$ . The disadvantage of this method is the need to decompose  $B^{(k)}$  into the product of two triangular  $s^2$ -dimensional matrices to simulate the normal density  $N(X, X_{\max}^{(k)}, B^{(k)})$  at each stage of the search. Some additional difficulties can arise when the covariance matrix is close to degenerate.

In [Ermakov and Semenchikov, 2019], the method is proposed that is free from these drawbacks. This method was successfully applied to beam dynamics optimization problem reduced to quality criterion minimization in search area of dimension 84. [Vladimirova, Zhdanova, Rubtsova and Edamenko, 2020]. Now in our paper we suggest essential modifications of the method mentioned to avoid too rapid contracting of the scattering ellipsoid and obtaining local extremum instead of global one.

## 2 Global extremum search method

We will consider the problem of extremum point finding for a function  $f(X)$  defined in a bounded domain  $D : X \in D \subset E^s$ . Suppose we are talking about a global minimum, which is attained at a single point  $X_{\min}$  from the region  $D$ :

$$f(X_{\min}) = \min_{X \in D} f(X). \quad (4)$$

The method for solving problem (4) is to construct a sequence of sample point generations to provide a condensation of points in the neighborhood of the minimum point. So the algorithm consists of a number of stages. Let us present the iterative part of algorithm.

**0-th generation.** If we have no a priori information on the location of global extremum point and the behavior of objective function, we will scan  $D$  using the uniform distribution of  $s$ -dimensional vectors and calculate  $f(X)$  at the points obtained. Let  $N_0$  be the number of trial points composing the zero generation. After that we choose among them  $m$  “best” points  $X_1, X_2, \dots, X_m$  corresponding to the smallest values of the objective function, and determine the minimum point  $X_{\min}^{(0)} = \arg \min_{X_1, \dots, X_{N_0}} f(X)$ :

$$f(X_{\min}^{(0)}) = f(X_1) < f(X_2) < \dots < f(X_m). \quad (5)$$

**$k$ -th generation.** At  $k$ -th stage we calculate  $N_k$  points of the  $k$ -th generation by formulae

$$Z_j = \frac{1}{m} \sum_{i=1}^m \eta_j^{(i)} (X_i - \bar{X}) + \bar{X}, \quad j = \overline{1, N_k}, \quad (6)$$

using the “best” points  $X_1, X_2, \dots, X_m$  of previous generation. In formulae (6)

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i = (\bar{x}^{(1)}, \dots, \bar{x}^{(s)}), \quad (7)$$

$\eta_j^{(i)}, i = \overline{1, m}, j = \overline{1, N_k}$  are independent standard normal variables:

$$\begin{aligned} \eta_j^{(i)} &\sim N(0, 1), \quad M\eta_j^{(i)} = 0, \quad M\left(\eta_j^{(i)}\right)^2 = 1, \\ M\left(\eta_j^{(i)}\eta_j^{(l)}\right) &= 0 \quad (i \neq l), \quad i = \overline{1, m}, \quad j = \overline{1, N_k}. \end{aligned} \quad (8)$$

After that, we perform the selecting of  $m$  “best” points of  $k$ -th generation and determine the minimum point  $Z_{\min}^{(k)} = \arg \min_{Z_1, \dots, Z_{N_k}} f(X)$ :

$$f(Z_{\min}^{(k)}) = f(Z_1) < f(Z_2) < \dots < f(Z_m). \quad (9)$$

After renaming  $Z$  to  $X$ , the next iteration can be performed.

**Covariance matrix estimation.** Let us estimate the parameters of normal distribution at  $k$ -th stage.

$$MZ_j = \frac{1}{m} \sum_{i=1}^m \left( M\eta_j^{(i)} \right) (X_i - \bar{X}) + \bar{X} \stackrel{(6-8)}{=} \bar{X}. \quad (10)$$

$$\begin{aligned} (Cov(Z_j))^{\alpha, \beta} &\stackrel{(6-8, 10)}{=} \\ &= \frac{1}{m^2} \sum_{i=1}^m \sum_{l=1}^m M\left(\eta_j^{(i)}\eta_j^{(l)}\right) \left(x_i^{(\alpha)} - \bar{x}^{(\alpha)}\right) \left(x_l^{(\beta)} - \bar{x}^{(\beta)}\right) = \\ &= \frac{1}{m} \sum_{i=1}^m \left(x_i^{(\alpha)} - \bar{x}^{(\alpha)}\right) \left(x_i^{(\beta)} - \bar{x}^{(\beta)}\right), \quad \alpha, \beta = \overline{1, s}. \end{aligned} \quad (11)$$

The results (10), (11) allow us to draw the following conclusions.

- 1)  $Z_j, j = \overline{1, N_k}$  are normally distributed vectors with covariance matrix (11) and expected value (10).
- 2) The estimates of covariance matrix elements depend only on “best” points  $X_1, X_2, \dots, X_m$  of the previous generation.
- 3) The method presented for random vectors  $Z_j, j = \overline{1, N_k}$  simulation allows not to calculate and store the covariance matrix (which is difficult for large dimension  $s$  of search space). There is no need to decompose the covariance matrix, for example, into triangular ones. (If the matrix is close to degenerate, its decomposition would be difficult.)

The convergence of this algorithm for a unimodal objective function is proved in [Ermakov and Semenchikov, 2019].

**Covariance matrix adaptation.** Consider the formulae (6) for new generation constructing.

*Remark 1.* Instead of the vector  $\bar{X}$ , it is more efficient to use  $X_{\min}^{(k-1)}$  that is “the best” sample point of previous  $(k-1)$ -th generation.

*Remark 2.* With search stage number increasing, the scattering ellipsoid of normally distributed vectors  $Z_j$ ,  $j=\overline{1, N_k}$  quickly contracts to a point and the search practically does not occur. Therefore, the following two options are proposed.

- 1) At each stage we will multiply standard normal random variables  $\eta_j^{(i)}$ ,  $i=\overline{1, m}$ ,  $j=\overline{1, N_k}$  in equations (6) by some constant  $\sigma$ . The optimization experience shows that this constant can depend on the stage number:  $\sigma=\sigma(k)$ . As a result, the scattering ellipsoid of vectors  $Z_j$ ,  $j=\overline{1, N_k}$ , will cover a sufficient part of the domain  $D$  and the search will continue. The value  $\sigma(k)$  is to be determined specially for each objective function  $f(X)$ .
- 2) When modeling, we divide the vectors  $Z_j$ ,  $j=\overline{1, N_k}$  into  $L$  groups, consisting of  $N_k^{(1)}, \dots, N_k^{(L)}$  vectors correspondingly,  $N_k^{(1)} + \dots + N_k^{(L)} = N_k$ . We apply different constants  $\sigma^{(l)}(k)$ ,  $l = \overline{1, L}$  for different groups ( $l$  is the group number). So the formulae for sample points simulation take the form:

$$Z_j = \frac{\sigma^{(l)}(k)}{m} \sum_{i=1}^m \eta_j^{(i)} (X_i - \bar{X}) + \bar{X}, \quad (12)$$

$$j = \overline{1, N_k^{(l)}}, \quad l = \overline{1, L}.$$

Further, we select  $m_l$  “the best” sample points from  $N_k^{(l)}$  points of  $l$ -th group,  $l = \overline{1, L}$ . It is clear that  $m_1 + \dots + m_L = m$ . After renaming  $Z$  to  $X$ , we have the set  $X_1, X_2, \dots, X_m$  of selected points for the next stage.

*Remark 3.* To isolate the global extremum, the objective function is raised to some power  $p$ , and the problem under study takes the form:  $f^p(X) \rightarrow \min$ .

The effectiveness of the modifications proposed is confirmed by numerical results.

### 3 Parameters estimating problem for nonlinear parametric regression

Consider the exponential regression

$$\eta(x, \theta) = \sum_{k=1}^n \beta_k e^{-\lambda_k x} \quad (13)$$

with parameter vector  $\theta = (\beta_1, \lambda_1, \dots, \beta_n, \lambda_n)$  of the dimension  $s=2n$ . Let the vector  $y = (y_1, \dots, y_I)$  be the result of observations of the experiment at the preassigned points  $x_i$ ,  $i = \overline{1, I}$ , where  $I > 0$  is a given number; vector  $x = (x_1, \dots, x_I)$  is regression plan. We will estimate

the regression parameter vector  $\theta$  using the least squares method:

$$\hat{\theta} = \arg \min_{\theta \in P \subset E^s} \sum_{i=0}^I (y_i - \eta(x_i, \theta))^2. \quad (14)$$

Here  $P$  is some compact in Euclidean space  $E^s$ .

The observation values  $y_i$ ,  $i = \overline{1, I}$ , are simulated as follows. Let  $\bar{\theta}$  be some given vector of parameters. We will introduce

$$y_i = \eta(x_i, \bar{\theta}) + \varepsilon_i, \quad (15)$$

where  $\varepsilon_i$ ,  $i = \overline{1, I}$ , are random measurement errors. These values are independent normally distributed and centered:

$$M\varepsilon_i = 0, \quad M\varepsilon_i \varepsilon_j = 0, \quad i \neq j, \quad D\varepsilon_i = v^2, \quad i, j = \overline{1, I}. \quad (16)$$

To find the solution  $\hat{\theta}$  of extremal problem (14), genetic algorithm with covariance matrix adaptation is chosen. For a given plan  $x = (x_1, \dots, x_I)$  and result vector  $y = (y_1, \dots, y_I)$ , the objective function  $f(\theta) = \sum_{i=0}^I (y_i - \eta(x_i, \theta))^2$  depends only on  $s$ -dimensional vector of parameters  $\theta = (\beta_1, \lambda_1, \dots, \beta_n, \lambda_n)$ . Such a problem arises in the field of analysis of the experiment results in various fields of science: mathematics, physics, chemistry, biology, etc.

## 4 Numerical results

**Regression function and experimental data.** The number of terms in the function (13) is  $n=7$ ; consequently, the number of parameters  $s=2n=14$ . The values of independent variable  $x$  for  $I=14$  are as follows:

$$x = (0.07, 0.11, 0.14, 0.18, 0.24, 0.27, 0.38, 0.43, 0.44, 0.50, 0.65, 0.96, 1.28, 1.65). \quad (17)$$

The result of observations is given by vector

$$y = (196.22, 191.39, 187.86, 183.26, 176.61, 173.38, 162.10, 157.25, 156.30, 150.74, 137.78, 114.82, 95.60, 77.83). \quad (18)$$

The resulting vector  $y$  is obtained using simulation by formula (15) for given vector of parameters

$$\bar{\theta} = (5.0, 0.1, 10.0, 0.3, 25.0, 0.35, 30.0, 0.5, 35.0, 0.55, 40.0, 0.7, 60.0, 0.9). \quad (19)$$

The standard deviation of random measurement errors is  $v=0.02$ . Parameter vector values belong to compact  $P$ :

$$\theta \in P = \{5 \leq \beta_i \leq 70, \quad 0 \leq \lambda_i \leq 1, \quad i = \overline{1, 14}\}.$$

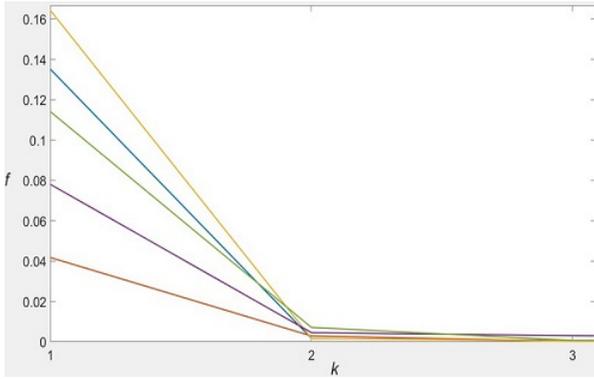


Figure 1. Dependence of the objective function on iteration number for five launches of the program.

**Genetic algorithm parameters.** The numerical experiments were performed for the following values of method parameters. The number of individuals at each generation is  $N=5000$ . The number of “the best” vectors at every stage is  $m=100$ . The number of groups of sample points at each iteration is  $L=2$ . The constant multipliers in formulae (12) are  $\sigma^{(1)}(k)=1$ ,  $\sigma^{(2)}(k)=k\sqrt{k}$ , where  $k$  is generation number. Search stop criterion is  $\varepsilon_f=10^{-5}$ ; if  $f_{k-1}-f_k \leq \varepsilon_f$ , then the search stops.

**The results of the genetic algorithm in figures.** A number of plots are presented below to show the work of the genetic algorithm under consideration.

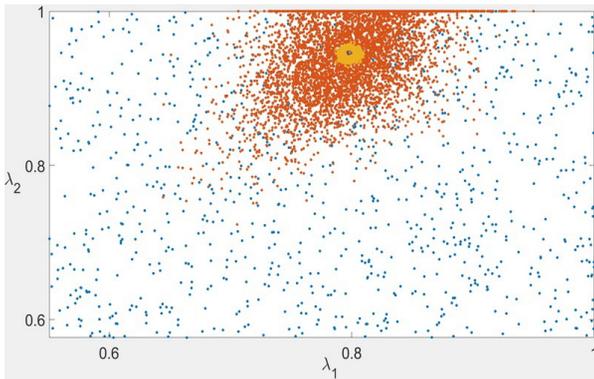


Figure 2. Five generations of sample points obtained by formulae (6).

Fig. 1 shows fast decreasing of target function with iteration number increase. The plot is presented for five launches of the program.

Fig. 2-4 represent the sample points distribution for different variants of algorithm. Multivariate normal distribution is presented graphically for the parameters  $\lambda_1, \lambda_2$ . Each plot shows several generations of trial points, and the images of the generations are overlapping.

Fig. 2 illustrates the results obtained for the whole sample simulation by formulae (6) (i.e.  $N_k^{(1)}=N, N_k^{(2)}=0$  in formulae (12)). In this case the search area contracts to a point.

The next case is as follows. At each iteration, the set of test points was divided in half:  $N_k^{(1)}=N_k^{(2)}$ . For the first part  $\sigma^{(1)}(k)=1$ , and for the second one  $\sigma^{(2)}(k)=k\sqrt{k}$ , where  $k$  is the iteration number. Now scattering ellipsoids do not contract to a point, which can be seen in Fig. 3.

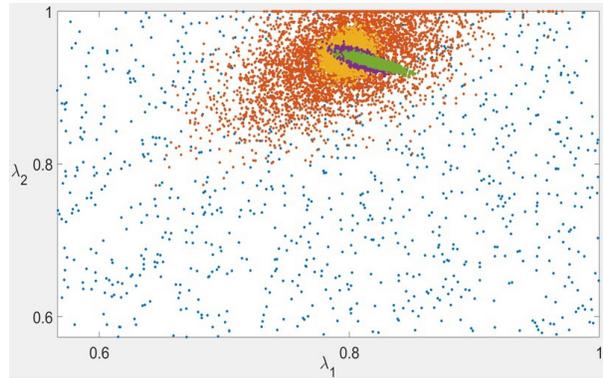


Figure 3. Five generations of trial points obtained using the division of every generation in half.

The next figure (Fig. 4) presents the distribution of trial points simulated by formula (12) with  $N_k^{(1)}=N_k/4, N_k^{(2)}=3N_k/4$ . The objective function is raised to the 5th power. In Fig. 4, one can see that the sample does not contract to a point with increasing the iteration number.

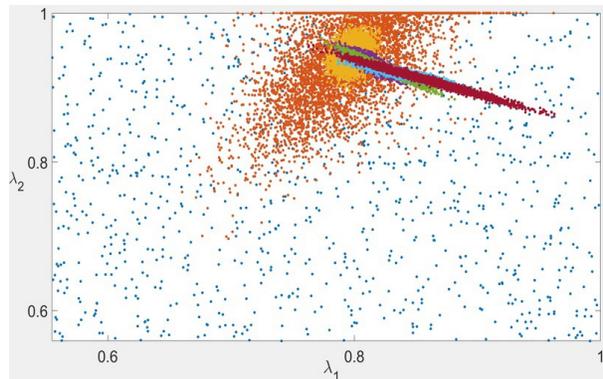


Figure 4. Seven generations of sample points obtained by formulae (12) for two unequal groups.

All results obtained are gathered in the Table. The best result is marked with the asterisk.

Table 1. Optimization results

Two groups of trial points	Objective function value after minimization
$N_k^{(1)}=5000, N_k^{(2)}=0$	$0, 1319 \cdot 10^{-2}$
$N_k^{(1)}=0, N_k^{(2)}=5000$	$0, 5638 \cdot 10^{-4}$
$N_k^{(1)}=2500, N_k^{(2)}=2500$	$0, 2053 \cdot 10^{-4} (*)$
$N_k^{(1)}=3750, N_k^{(2)}=1250$	$0.2098 \cdot 10^{-4}$
$N_k^{(1)}=0, N_k^{(2)}=5000$ Raising to the 5 <sup>th</sup> power	$0, 4819 \cdot 10^{-4}$
$N_k^{(1)}=1250, N_k^{(2)}=3750$ Raising to the 5 <sup>th</sup> power	$0, 3111 \cdot 10^{-3}$

## 5 Conclusion

We present genetic stochastic algorithm using normal distribution with covariance matrix adaptation and introduce some modifications. This method allows one to avoid calculating the covariance matrix, and this is a great advantage. We offer a number of modifications, the effectiveness of which is tested on a practical problem. The method in several variants was applied to parameters estimating problem for nonlinear regression. The worst result was obtained for the method [1] without any modifications. In this case one can observe scattering ellipsoid contracting to the point, and trial points cannot leave the vicinity of the local extremum point. First of modifications is introduction of correcting multiplier in sample points modeling formula. Second is using different multipliers for different groups of sample points. (One can consider the using of multipliers to be mutations of individual properties.) As a result, the trial points cover a sufficient part of the domain  $D$  and avoid the contraction to local extremum point. These modifications led to significant improvement in the result. Third modification is raising the objective function to a sufficiently large degree, which makes it possible to isolate the global extremum point. This is a useful technique that requires further research. When all individuals were divided into two equal parts with using the factors  $\sigma^{(1)}(k)=1, \sigma^{(2)}(k)=k\sqrt{k}$ , where  $k$  is generation number, the best value of the objective function was obtained for a given stop criterion  $\varepsilon_f=0.1 \cdot 10^{-4}$ . The number of iterations in all the cases does not exceed 9.

The objective function near a minimum has a ravine-type shape; this is indicated by the elongated shape of the scattering ellipsoid. The proposed modified method successfully coped with minimization problem. Attempts to use the gradient method in this task were unsuccessful.

The research under discussion was presented by oral contribution at the 10th International Scientific Confer-

ence on Physics and Control (PHYSCON 2021, 4-8 October, 2021, Fudan University, Shanghai, China).

## Acknowledgements

The work is supported by Saint Petersburg State University, project ID: 93024916.

## References

- Ermakov, S.M., Kulikov, D.V. and Leora, S.N. (2017). Towards the analysis of the simulated annealing method in the multiextremal case. *Vestnik St. Petersburg University, Mathematics*, **50**(2), pp. 132–137.
- Ermakov, S.M. and Semenchikov, D.N. (2019). Genetic global optimization algorithms. In *Communications in Statistics, Part B: Simulation and Computation*. <https://doi.org/10.1080/03610918.2019.1672739>.
- Ermakov, S.M. and Mitiglova, L.V. (1977). On extreme search method based on the estimation of the covariance matrix. *Automation and Computer Engineering*, **5**, pp. 38–41 (in Russian).
- Granichin, N., Volkov, G., Petrov, Y. and Volkova, M. (2021). Randomized approach to determine dynamic strength of ice. *Cybernetics and Physics*, **10**(3), pp. 122–126.
- Hansen, N. and Ostermeier, A. (1996). Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. In *Proc. of 1996 IEEE Conference on Evolutionary Computation (ICEC'96)*, Berlin, Germany, May, pp. 312–317.
- Igel, C., Hansen, N. and Roth, S. (2007). Covariance matrix adaptation for multi-objective optimization. *Evolutionary Computation*, **15**(1), pp. 1–28.
- Vladimirova, L., Ovsyannikov, D. (2019). Random search for global extremum of a function using Markov chains simulation. *Journal of Physics: Conference Series*. 1238, **1**, 012073.
- Metropolis, V., Rosenbluth, M., Teller, A. and Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, pp. 1087–1092.
- Nesterov, Y. (2010). *Convex optimization methods*. Moscow: MCCME (in Russian).
- Sergeenko, A., Yakunina, M. and Granichin, O. (2020). Hamiltonian path problem solution using DNA computing. *Cybernetics and Physics*, **9**(1), pp. 69–74.
- Vladimirova, L.V., Zhdanova, A.Y., Rubtsova, I.D. and Edamenko, N.S. (2020). Genetic stochastic algorithm application in beam dynamics optimization problem. Stability and Control Processes. In *Proc. of the 4th International Conference Dedicated to the Memory of Professor Vladimir Zubov, Springer International Publishing (Lecture Notes in Control and Information Sciences - Proceedings)*.
- Zhigljavsky, A.A. (1991). *Theory of Global Random Search* (Mathematics and its Applications, 65), edited by J.D. Pinter. Springer-Verlag, Dordrecht.