



Simple statistical tests selection based parallel computing method ensures the guaranteed global extremum identification

Viacheslav Kovtun^{a,*}, Torki Altameem^b, Mohammed Al-Maitah^b, Wojciech Kempa^c

^a Department of Computer Control Systems, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Khmelnytske Shose str., 95, Vinnytsia 21000, Ukraine

^b Computer Science Department, Community College, King Saud University, 11451 Riyadh, Saudi Arabia

^c Department of Mathematics Applications and Methods for Artificial Intelligence, Faculty of Applied Mathematics, Silesian University of Technology, ul. Akademicka 2A, 44-100 Gliwice, Poland

ARTICLE INFO

Keywords:

Parallel computing
Applied mathematics
Optimization problem
Global extremum
Statistical tests method
Equilibrium

ABSTRACT

The article proposes a parallel computing oriented method for solving the global minimum finding problems, in which continuous objective functions satisfy the Hölder condition, and the control parameters domain limited by continuous functions is characterized by a positive Lebesgue measure. A typical example of such a task is the discrepancy minimizing problem between the left and right parts of some large system of equilibrium equations (this is a usual situation when describing real process using Markov chain). The method is based on simple statistical tests, thanks to which, at each iteration, growing sets of potential global minima and sets of decrements necessary for estimating the values of the Hölder constants are formed. The article theoretically substantiates and empirically proves the guaranteed convergence of the authors' method to the real global minimum, which occurs at an exponential rate. For the continuous iterations number, analytical upper estimates of the spacing between the potential global minima and real global minima are formalized, as well as an estimate of the probability of overcoming this spacing is formalized. The decrements sequence approximation, estimation of a priori unknown Hölder constants, estimation of the average number of iterations of the method and probabilistic characteristics of the final solution are analytically justified. In addition to the theoretical proof, the adequacy of the authors' method has been confirmed empirically. It turned out that both the quality characteristics of the initial results calculated by the authors' method and the time to obtain them are practically independent of the size of the search area. This expected result is a significant advantage of the authors' method over analogues.

1. Introduction

Optimization is the process of finding the point of the extreme value of a certain (target) function (Stracquadanio and Pardalos, 2019; McNaughton, 2023). It is one of the biggest cornerstones of applied mathematics, physics, engineering, economics, and industry. The scope of its application is vast and includes, for example, minimization of physical quantities at micro and macro levels, maximization of profit or efficiency of logistics chains, etc. Machine learning is also focused on optimization (Huang et al., 2022; Alexiadis, 2023): various regressions and neural networks try to minimize the discrepancy between predicted and real data.

Optimization is an active and relevant sector of scientific research. This sector includes a huge number of thematic fields and their extreme

variations. The number of optimization problem statements and their solution methods has been growing like mushrooms after the rain for decades. For example, there is a huge field of Mixed-Integer Programming (Alfant et al., 2023; Zhang et al., 2023), which deals with discrete scenarios. Nondeterministic optimization (Zhang et al., 2022; Yang et al., 2019), based on stochastic principles, is known. There is robust optimization (Fransen and Langelaar, 2023; Castelli et al., 2023), in which fixed parameters are taken into account. The optimization of dynamic systems evolving in time (Yang et al., 2023; Xie et al., 2022) is known. There are various meta-heuristic methods (Marulanda-Durango and Zuluaga-Ríos, 2023; Hirsching et al., 2022): simulated annealing method, genetic algorithms, and swarm evolution methods. There are well-known approaches to optimization using fuzzy logic (Pei et al., 2023).

Let's focus on the global optimization problem of continuously

* Corresponding author.

E-mail addresses: kovtun_v_v@vntu.edu.ua (V. Kovtun), altameem@ksu.edu.sa (T. Altameem), malmaitah@ksu.edu.sa (M. Al-Maitah), Wojciech.Kempa@polsl.pl (W. Kempa).

<https://doi.org/10.1016/j.jksus.2024.103165>

Received 18 July 2023; Received in revised form 6 February 2024; Accepted 13 March 2024

Available online 16 March 2024

1018-3647/© 2024 The Authors. Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Nomenclature

q	is the dimension of the vector z
N_i	is the volume of the set of stochastic vectors at the i -th iteration ($N_i = M_i$)
$Z_i = \{z^{(1)}, \dots, z^{(N_i)}\}$	is a set of stochastic, uniformly distributed Z_+^n , vectors z
$Y_i = \left\{y^{(1)}, \dots, y^{(\tilde{N}_i)}\right\}$	is a set of admissible stochastic vectors belonging to the admissible set L (see (4)), $\tilde{N}_i \leq N_i$
\tilde{N}_i	is the volume of the set Y_i
$F_i^* = \min_{y \in Y_i} F(y)$	is the minimum of the objective function F on the admissible set Y_i
$F^{(i)} = \{F_0^*, F_1^*, \dots, F_i^*\}$	is the cumulative set of potentially global minima
$\delta_k = F_k^* - F_{k-1}^* $	is the decrement of the set $F^{(i)}$, $k = \overline{1, i}$
$\Delta^{(i)} = \{\delta_1, \dots, \delta_i\}$	is the cumulative set of decrements; $w_k = \log \delta_k$, $k = \overline{1, i}$
$W^{(i)} = \{w_1, \dots, w_i\}$	is the accumulated set of logarithms of decrements
$\tilde{W}^{(i)}(\rho) = \{\tilde{w}_{i-\rho}, \tilde{w}_{i-\rho+1}, \dots, \tilde{w}_i\}$	is the accumulated set of logarithms of decrements of smaller $\log \gamma$; $n_k = \log \tilde{N}_k$, $k = \overline{0, i}$
$N^{(i)} = \{n_0, n_1, \dots, n_i\}$	is the accumulated set of logarithms of the volumes of sets
γ	is the maximum acceptable error for calculating the values of decrements
ρ	is the applied number of elements of the accumulated sequence $\tilde{W}^{(i)}(\rho)$
$\hat{C}^{(i)}, \hat{s}^{(i)}$	are estimates of the values of the Hölder constants (see expression (24))
ξ_j	is a situation when $w_j \leq \log \gamma$
r_ϕ^+	is the upper estimate of the spacing from the current potentially optimal minimum F_i^* , obtained on the $i = \phi$ -th iteration of the method, to the real minimum F^* (see expression (33))
$P_{r_\phi}^*$	is the lower probability estimate (34)

differentiable functions (this will allow us to narrow the area of interest at least a little). The analysis involves evaluating the applicants in a certain quality criteria metric. In this context, first of all, we will mention such criteria as:

- global convergence (global convergence in optimization refers to the ability of an optimization algorithm to guarantee convergence to the global minimum or maximum of the objective function, regardless of the initial conditions. It ensures that the algorithm approximates the optimal solution across the entire search space, not just a local minimum or maximum. This property is crucial to avoid getting stuck in suboptimal solutions, ensuring the discovery of the best solution in the general case of the optimization problem);

- speed of local convergence (the speed of local convergence in optimization refers to the rate at which an optimization algorithm approaches the local minimum or maximum of the objective function near the convergence point. This characteristic determines how quickly the algorithm converges to the optimal solution in the vicinity of the current point, assessing the rate of decrease in the function value. A high speed of local convergence implies rapid approximation to the local optimum, which can be crucial for efficiently solving optimization problems);

- the dimension of the optimization problem, the solution of which is aimed at the method;

- the need to store matrices in the computer memory during the solution process (yes or no);

- use of the Hesse's matrices in the solution process (yes or no);

- the need for scaling (yes or no) (scaling in optimization refers to the process of transforming variables or parameters in an optimization problem in such a way that they have a similar scale. This is important because different variables may have different orders of magnitude, which can hinder the convergence of optimization algorithms. Scaling helps balance the contribution of each variable, thereby facilitating the optimization process and enhancing the efficiency of converging to the optimal solution).

The first two criteria are analytical, the others characterize the aspects of the directly applied implementation of a particular method.

Let us pay closer attention to the first two of the mentioned criteria. An optimization method is said to have the property of global convergence if its iterations converge to a local minimizer regardless of the initial position. The speed of local convergence shows how quickly the method will find the minimizer after getting into the vicinity of the latter. Note that these two criteria are potentially competing. The so-called hybrid methods (Xu et al., 2023; Wu et al., 2023) are an actual answer to the conflict that manifests itself in an attempt to achieve simultaneous compliance with both the first and second criteria.

Now, having mentioned the main optimization method's qualitative criteria, we can go directly to the methods as such (at least, to the most used of them). Let's recall the most used methods of smooth unconditional optimization (Ali et al., 2021; Smith and Nair, 2005): the lines method, the fastest descent method, the Newton method, quasi-Newton methods (incl. DFP, SR1, BFGS (Broyden, Fletcher, Goldfarb, Shanno) and its modifications: muffled, with limited memory, etc.), the nonlinear conjugate gradients method, the truncated Newton method. Note that from this list, only the fastest descent method and the truncated Newton method can claim (with limitations) to pass the first criterion (global convergence).

Of course, there are methods for which compliance with the first criterion was the dominant requirement at the creation time. We will distinguish three classes of such methods. The methods we refer to in the first class are focused on the configuration of the objective function and the admissible solutions domain. The most striking representative of this class is the DC functions minimizing method (Fang et al., 2012; Montano et al., 2022). In the method, both the objective function and the constraint functions for the admissible solutions domain are represented as the differences between two convex functions. A complete view of the evolution of this class of global optimization methods can be made by reading the works (Fang et al., 2012; Montano et al., 2022; Shilaja et al., 2022). Methods of the second class are focused on solving global optimization problems with simple configurations of admissible solutions regions (which have, for example, the shape of a parallelepiped) and objective functions characterized a priori by known Hölder constants

(Chen and Zheng, 2021; Sovrasov, 2016). The same class can also include methods focused on reducing a multidimensional optimization problem to a one-dimensional one using Peano curves (Sovrasov, 2016). However, if the Hölder constants are unknown, the application of second-class methods is accompanied by considerable uncertainty. Finally, methods belonging to the third class focus on random search and the formalization of intellectualized heuristics to interpret its results (Gorawski et al., 2021).

We would like to separately mention Sub-gradient methods for non-smooth optimization (Tymchenko et al., 2019; Zaiats et al., 2019). Sub-gradient methods for non-smooth optimization are optimization techniques specifically designed to address problems where the objective function is not everywhere differentiable. Unlike traditional gradient methods applicable to smooth functions, sub-gradient methods use sub-gradients, a generalization of gradients, to navigate through optimization landscapes containing non-differentiable components. These methods are particularly useful in convex optimization scenarios involving functions with nonsmooth elements, such as those encountered in support vector machines and lasso regression. Sub-gradient methods iteratively adjust the solution in the direction of the sub-gradient, aiming to converge to a point where the sub-gradient approaches zero, indicating a potential solution to the non-smooth optimization problem. While they may exhibit slower convergence rates compared to methods for smooth optimization, sub-gradient methods are essential for effectively tackling problems with inherent nonsmoothness.

One of the essential problems accompanying the use of stochastic tests in the context of solving optimization problems is the generation of uniformly distributed random vectors in a certain region of the search space. Scientists are looking for ways to effectively apply this problem (Bisikalo and Kharchenko, 2023; Izonin et al., 2022). In this context, we note Hit-and-Run-like methods, Markov chains using methods, and methods that take into account relative entropy. However, a universal method has not yet been found.

Taking into account the strengths and weaknesses of the mentioned analogue methods, we will formulate the necessary attributes of scientific research.

Research object is a process of solving the global minimum finding problems, which continuous objective functions satisfy the Hölder condition, and the control parameters domain limited by continuous functions is characterized by a positive Lebesgue measure and is limited by multidimensional parallelepipeds.

Research subject includes functional analysis theory, probability theory and mathematical statistics, experiment planning theory, and computational methods.

Research aim is to formalize the process of guaranteed solutions to the optimization problems outlined by the research object in the form of a parallel computing oriented method.

Research objectives are:

- formalize the parallel computing oriented method of the guaranteed finding of the global solution of the optimization problem with the objective function and restrictions imposed by algorithmically calculated continuous functions that satisfy the Hölder condition (with a priori unknown values of the corresponding constants);
- formalize the process of generating sets of potentially global extrema as the results of simple statistical tests (as a basic element of each iteration of the method);
- formalize the probabilistic characteristics of the approximate solution, which is the result of the implementation of a finite number of iterations of the authors' method;
- justify the adequacy of the proposed mathematical apparatus and demonstrate its functionality with an example.

Main contribution. The article proposes a parallel computing oriented method for solving the global minimum finding problems, in which continuous objective functions satisfy the Hölder condition, and the control parameters domain limited by continuous functions is

characterized by a positive Lebesgue measure and is limited by multidimensional parallelepipeds. A typical example of such a task is the discrepancy minimizing problem between the left and right parts of some system of equilibrium equations. The method is based on simple statistical tests, thanks to which, at each iteration, growing sets of potential global minima and sets of decrements necessary for estimating the values of the Hölder constants are formed. The article theoretically substantiates and empirically proves the guaranteed convergence of the authors' method to the real global minimum, which occurs at an exponential rate. For the continuous iterations number, analytical upper estimates of the spacing between the potential global minima and real global minima are formalized, as well as an estimate of the probability of overcoming this spacing is formalized. The decrements sequence approximation, estimation of a priori unknown Hölder constants, estimation of the average number of iterations of the method and probabilistic characteristics of the final solution are analytically justified.

- Statement of the research, which introduces general definitions and clarifies the purpose and objective of the research;
- Formalization of the parallel computing oriented method of the guaranteed finding of global extremum with simple statistical tests selection;
- Specific aspects of the applied application of the presented method;
- Sections devoted to demonstration and analysis of the results of the intended use of the proposed method.

2. Materials and methods

2.1. Statement of the research

The problem of finding the extremum of a function is to determine its largest (maximum) or smallest (minimum) value in a certain range of values of its arguments. The task of the boundaries of this area (as well as the rest of the additional conditions) is implemented in the form of a system of equations and (or) inequalities. In this case, we talk about the conditional problem of finding the extremum of a function or the problem of finding a local extremum. If the range of admissible values of the arguments is not limited, then we are dealing with the problem of finding the global extremum of the function or the optimization problem. Next, we will investigate the problem of finding the global extremum of the function $f(x)$:

$$globminf(x) \tag{1}$$

for $x \in H_x, H_x = \{x : h(x) \leq 0, x \in R^q\}, h \in R^r, r < q$, functions $h(x)$ are continuous. Function (1) satisfies the Hölder condition with constant $G > 0, v > 0$, i.e. $v(t) = \max_{(\bar{w}, \bar{y}) \in R^q: \|\bar{w} - \bar{y}\| \leq t} |f(\bar{w}) - f(\bar{y})| \leq Gt^v$.

We formalize the canonical form of the optimization problem (1):

$$globminF(z) \tag{2}$$

for $z \in L_z, L_z = H_z \cap Z_+^q, Z_+^q = \{z : 0 \leq z \leq 1\} \subset R^q, H_z = \{z : v(z) \leq 0, z \in Z_+^q\}, v \in R^r, r < q$.

The functional relationship between the optimization problems (1) and (2) can be expressed through the continuously differentiable and mutually unique transformation

$$z = T(x)x = 1/T(z) \tag{3}$$

A functional instance of (3) is, for example, the transformation $z = 1/(1 + \overrightarrow{\exp}(-c \otimes x))$, $x = (-1/c) \otimes \overrightarrow{\ln}((1-z)/z)$, where $\overrightarrow{\exp}(-\bullet \otimes \bullet)$ and $\overrightarrow{\ln}(\bullet)$ - n -dimensional vectors are formed by the components $\exp(-\bullet \otimes \bullet)$ and $\ln(-\bullet)$, respectively.

In the problems of applied design, the area of determination of controlled parameters x is not the entire space R^q , but its area is limited by the q -dimensional parallelepiped P_x . Let's redefine the optimization problem (1) taking into account this restriction:

$$\min_x f(x) \tag{4}$$

for $x \in P_x \cap H_x$: $P_x = \{x : c \leq x \leq d\}$, $c = \{c_i\}$, $d = \{d_i\}$, $i = \overline{1, n}$, $x \in R^q$; $H_x = \{x : h(x) = \{h_j(x), j = \overline{1, r}\} \leq 0, x \in R^q\}$, $h \in R^r$, $r < q$, where the generalized functions $h(x)$ are continuous. In the context of the optimization problem (4), the transformation (3) is significantly simplified: $z = (x - c) \otimes (1/(d - c))$.

Let the objective function $F(z)$ of the optimization problem (4) acquire a global minimum at the values of the argument $z^* \in Z^*$: $F(z^*) = F^*$. Taking into account the continuity of the transformation (3), we can state that the function $F(z)$ is bounded from below by L_z , that is, there is a constant a for which the inequality $F^* \geq a = \text{const}$ holds.

In the general case, a is unknown, but $a \rightarrow 0$ in the context of specific optimization problems (for example, where the function $F(z)$ characterizes the discrepancy between the left and right parts of some system of equilibrium equations). At the same time, the function $F(z)$ also satisfies the Hölder condition, but with constants K and s , that is:

$$v(g) = \max_{(w,y) \in L_s: \|w-y\| \leq g} |F(w) - F(y)| \leq Kg^s \tag{5}$$

Therefore, the aim of our investigation is to create a parallel computing oriented method that will allow us to getting the global extreme $z^* \in Z^*$ of the objective function $F(z)$ on the set L_z . The subject of our research will be the simple statistical tests method.

2.2. Formalization of the parallel computing oriented method of the guaranteed finding of global extremum with simple statistical tests selection

The basic element of the authors' method is the elementary positive cube Z_{i+}^n mentioned in the formulation of the optimization problem (4). Each i -th iteration ($i = 0, 1, 2, \dots$) of the method begins with the generation of a set Z_i , which contains N_i uniformly distributed and independent Z_{i+}^n stochastic vectors $\{z^{(1)}, \dots, z^{(N_i)}\}$. Next, among the calculated for the i -th set of the function $F(z^{(j)})$ values, $j = \overline{1, N_i}$, the minimum value F_i^* and the corresponding value of the argument z_i^* are determined and fixed. Considering that the generation of the set Z_i is carried out by the statistical tests method, the determined optimal value F_i^* is stochastic. The transition to the $(i+1)$ -th iteration is accompanied by a regular increase in the number of stochastic vectors in the set Z_{i+1} : $N_{i+1} > N_i$. The result of execution of the $(i+1)$ -th iteration is the set $\{F_{i+1}^*, z_{i+1}^*\}$. The course of the iterative process is accompanied by the growth of accumulated sets $F^* = \{F_i^*\}$ and $\Delta = \{\delta_{i+1}, \delta_{i+1} = F_i^* - F_{i+1}^*\}$, $i = 0, 1, 2, \dots$. The iterative process is completed if during ρ iterations the dynamics of changes in the values in the set Δ does not exceed the certain threshold γ :

$$\phi = \min\{i : \max_{i-\rho \leq s \leq i} \delta_s \leq \gamma\} \tag{6}$$

The output result of the method described above is the set $S = \{F_\phi^*, z_\phi^*, r_\phi^+, P(\xi_{r_\phi^+})\}$, where F_ϕ^* is the found approximate value of the global extremum of the objective function (4), z_ϕ^* is the found approximate value of the argument of the objective function (4) for F_ϕ^* , r_ϕ^+ is an estimate of the spacing between the approximate and actual values of the global extremum of the objective function (4), $P(\xi_{r_\phi^+})$ is an estimate of the event $\xi_{r_\phi^+} = \{F_\phi^* - F^* \leq r_\phi^+\}$ probability.

The set Z_i is formed from a matrix of independent stochastic values of dimension $(q \times M_i)$, the elements of which are uniformly distributed over the segment $[0, 1]$. The source of this matrix for the i -th iteration is a random number generator. This matrix is the basis for the formation of N_i independent and uniformly distributed positive elementary parallelepiped stochastic vectors $Z_i = \{z^{(j)}\}$, $j = \overline{1, N_i}$. Let's examine the probabilistic characteristics of the set Z_i .

Let's form the desired set Z_i from the existing stochastic matrix of dimension $(q \times M_i)$ in the following way: $N_i = M_i^q$ or $N_i = M_i$, where the parameter N_i characterizes the number of stochastic vectors in the set Z_i , and the parameter M_i characterizes the coordinates of $z_j^{(i)}$ the vector $z^{(i)}$. Taking this $N_i = M_i$, we formulate the rule for the formation of the set Z_{i+1} (taking into account the declared iterative increase in the volume of the latter: $N_{i+1} > N_i$):

$$N_{i+1} = \alpha N_i M_{i+1} = \alpha M_i \alpha > 1 \quad i = 0, 1, 2, \dots \tag{7}$$

Let us divide the q -dimensional parallelepiped Z_{i+}^q mentioned in (4) into unit cubes in each of the dimensions using a lattice with a uniform step

$$\mu_i = M_i^{-q} 0 < \varphi \leq o \leq \beta < 1 \tag{8}$$

where the parameter o characterizes the volume of the unit cube and the values of the boundary parameters φ, β should be chosen taking into account the results of the analysis of probabilistic characteristics of the set of global minima F^* and the set of their corresponding arguments Z^* .

We characterize as $P(N_i, q, o)$ the probability of such a situation that there is an cube that does not contain any of the N_i vectors generated on the i -th iteration:

$$P(N_i, q, o) = \left((1 + N_i^o)(1 - N_i^{-o})^{N_i} \right)^q = Q(i, q, o) \tag{9}$$

For $i \rightarrow \infty$, expression (9) will take the form

$$Q(i, q, o) \sim Q(i, q, o) = N_i^{qo} \exp(-qN_i^{1-o}) \tag{10}$$

Let's justify analytically the correctness of expressions (9), (10) in the context of the optimization problem (4). Let us characterize by the probability $P(A_k)$ the situation A_k when none of the M_i generated independent and uniformly distributed $[0, 1]$ stochastic values falls into the interval μ_i for the k -th dimension:

$$P(A_k) \leq R(i, q, o) = (1 + M_i^{-o})(1 - M_i^{-o})^{M_i}$$

where $(1 + M_i^{-o})$ is the upper limit of the number of segments formed as a result of dividing the interval $[0, 1]$ by a grid with a step M_k^{-q} .

Considering that $\lim_{x \rightarrow \infty} \frac{(1-x^{-o})^x}{x^{-o} \exp(-x^{1-o})} = 1$, the expression (10) can be transformed into the form $P(A_k) \leq R(i, q, o) \sim_{i \rightarrow \infty} M_i^{qo} \exp(-M_i^{1-o})$.

Let us characterize B the situation when an o -dimensional elementary parallelepiped is found that does not contain any of the N_i generated vectors, i.e.: $B = \bigcap_{k=1}^o A_k$. Since an independent set of stochastic values M_i is generated for each i -th dimension, then for $i \rightarrow \infty$ we get

$$P(B) = P(N_i, q, o) = (P(A_i))^{o \leq} \left((1 + N_i^o)(1 - N_i^{-o}) \right)^o \tag{11}$$

Expression (11) is identical to the left side of expression (9). Therefore, with a sufficiently large number of iterations i , at least one of the stochastic vectors N_i from the set Z_i generated on the i -th iteration of the method falls into each unit cube with a side M_i^{-o} with a probability not less than $(1 - Q(i, q, o))$.

We established that for optimization problem (4), both the set of global minima v and the set of $Z^* = \{z_0^*, \dots, z_i^*, \dots\}$ (their corresponding arguments) are not empty. Let us focus attention on the set of global minima F^* and investigate its probabilistic characteristics.

Suppose, Z^* be the set of arguments of the global extrema points of the function F on the unit cube. Let's define the spacing l from a sufficiently compact set Z^* to an arbitrary point z as

$$l(z, Z^*) = \min_{y \in Z^*} \|z - y\| \tag{12}$$

Suppose that the point $\hat{z} \in Z_i$ is the closest to the Z^* in the context of the spacing (12). As we proved earlier, with probability $(1 - Q(i, q, o))$ at least one of the points of the set Z_i falls into the unit cube with side μ_i ,

Taking into account this fact, we redefine the expression (12) for the point \tilde{z} : $\|z^* - \tilde{z}\| \leq \mu_i \sqrt{q}/2 = g_i$. This expression characterizes the situation when the point z^* is located in the center of the unit cube, and the point \tilde{z} is located at one of its vertices. Accordingly: $|F^* - F(\tilde{z})| \leq v(g_i)$, where $v(g_i)$ is the modulus of continuity of the function $F(z)$ characterized by expression (5). On the other hand: $F^* = F(z^*) = \min_{z \in Z_i^*} F(z) \leq \min_{1 \leq s \leq i} F^s < F(\tilde{z})$. Let's generalize the last two expressions and get $-v(g_i) \leq F^* - F(\tilde{z}) \leq \min_{z \in Z_i^*} F(z) - \min_{1 \leq s \leq i} F^s \leq 0$ or

$$\left| \min_{z \in Z_i^*} F(z) - \min_{1 \leq s \leq i} F^s \right| = |F^* - F_i^*| \leq v(g_i) \quad (13)$$

Inequality (13) is fulfilled with a probability not less than $(1 - Q(i, q, o))$. Relying on inequality (13), it can be asserted that for the function of the form (5) there are such positive constants K, s that the estimate

$$r_i(o) = K(\sqrt{q}/2)^p N_i^{-op} \geq |F^* - F_i^*| \quad (14)$$

is reliable with a probability not less than $(1 - Q(i, q, o))$. The estimate (14) depends on the values of the parameters o and φ, β entered in (8).

Operating with probabilistic characteristics (10), (14) it is possible to substantiate the convergence of the parallel computing oriented method proposed at the beginning of Section 2.2. Let's take a sufficiently small positive number ξ and find the value of the parameter $L(\xi) = \min\{i : r_i \leq \xi\}$ (the value of the estimate r_i is determined according to (14)). For $i < L(\xi)$, the situation $\{|F^* - F_i^*| > \xi\} \subset \{|F^* - F_i^*| > r_i\}$ is relevant. The probability of realization of this situation in the context of expressions (9), (14) is characterized by the expression

$$P\{|F^* - F_i^*| > \xi\} \leq Q(i, q, o) \quad (15)$$

Based on expression (7), $\alpha > 1, i = 1, 2, \dots$ we write

$$M_i = \alpha^i M_0 \quad (16)$$

We present the estimate of the probability characteristic (9) as

$$Q(i, n, o) = (\alpha^{qoi} M_0^{qo}) \exp(- (q/2) \alpha^{(1-o)i} M_0^{1-o}) \exp(- (i(1-o)/2) M_0^{1-o}) \quad (17)$$

Considering (16), we note that in expression (17) the third multiplier is greater than the second in terms of value. Taking this into account, as well as $x = \alpha^i M_0$, we present the expression (17) in a compact form: $Q(i, q, o) \leq \tilde{Q}(- (i(1-o)/2) M_0^{1-o})$, where $\tilde{Q} = \max_{x \geq 0} x^{qo} \exp(- qx^{1-o}/2)$.

The result allows for $i > L(\xi)$ to estimate the probability (15) as

$$P\{|F^* - F_i^*| > \xi\} \leq \tilde{Q} \exp(- Ai) \quad (18)$$

where $A = (1 - o)M^{1-o}/2$. Analyzing expression (18) $i > L(\xi/2)$, we write

$$P\left\{ \sup_{s \geq i} |F_i^* - F_{s+i}^*| > \xi \right\} \leq 2P\left\{ \sup_{s \geq i} |F^* - F_s^*| > \frac{\xi}{2} \right\} \leq 2Q \sum_{s=i}^{\infty} \exp(- As) \quad (19)$$

Examining the limiting values of (19) for sufficiently small positives ξ , we obtain

$$\lim_{i \rightarrow \infty} P\left\{ \sup_{s \geq 0} |F_i^* - F_{i+s}^*| \geq \xi \right\} = 0 \quad (20)$$

Expression (20) can be interpreted as a necessary and sufficient condition for the convergence of the set $F^* = \{F_i^*\}, i = 0, 1, 2, \dots$. Expression (19) allows us to state that the speed of this convergence is exponential.

Based on the proven conclusion regarding the convergence of the F^* to the global extrema of the optimization problem (4) value, it can be

predicted that the set of arguments z_i^* is guaranteed to converge to the set Z^* in the spacing metric (12). Suppose that there is a subset $\{z_k^*\}$ that converges to the $\tilde{z} \notin Z^*$ point with non-zero probability. This (considering the continuity of the F function) means that the $\lim_{i \rightarrow \infty} F(z_i^*) = F(\tilde{z}) > F^*$ inequality holds with nonzero probability. But, relying on (20), we can write: $\lim_{i \rightarrow \infty} F(z_i^*) = \lim_{i \rightarrow \infty} F_i^* = F^* = F(\tilde{z})$. In the resulting contradiction, it is the second construction that is analytically justified (see the sequence of obtaining expression (20)). Therefore, the hypothesis that the set of arguments z_i^* is guaranteed to coincide with the set Z^* in the spacing metric (12) can be considered proven. Analyzing expression (20), we can conclude that the set of decrements Δ introduced at the beginning of Section 2.2 converges to zero.

2.3. Specific points of applied use of the authors' method

The iterative process of solving the optimization problem (4) by the authors' method is accompanied by both the calculation of the elements of the set of decrements Δ and the determination of the constants K and s (to comply with the Hölder condition (5)). Let us approximate the sequence of decrements Δ and formalize the process of estimating the values of the constants K, s .

We adapt condition (14) for functions that satisfy the Hölder condition. For the i -th iteration, we get

$$|F_i^* - F^*| \approx K(\sqrt{q}/2N_i^o)^s = r_i(o) \quad (21)$$

and for the $i+1$ -th iteration, we get

$$|F_{i+1}^* - F^*| \approx K(\sqrt{q}/2N_{i+1}^o)^s \quad (22)$$

Subtract expression (22) from expression (21) taking into account rule (7) and that $\alpha^{-os} \ll 1: \delta_i \approx K(\sqrt{q}/2N_i^o)(1 - \alpha^{-op}) \sim K(\sqrt{q}/2N_i^o) = \tilde{\delta}_i \partial$.

After differentiating the obtained expression, we write

$$w_i = \log \tilde{\delta}_i = \log K - s(o \log N_i - \log(\sqrt{q}/2)) \quad (23)$$

Denoting $C^{(i)} = \log K, n_i = \log N_i, \tilde{n}_i(o) = on_i - \log(\sqrt{q}/2)$, we present the expression (23) as

$$w_i = C^{(i)} - s^{(i)} \tilde{n}_i(o) + \psi_i \quad (24)$$

where ψ_i is a set of stochastic values with zero mathematical expectation and variance V , and n_i is considered sufficiently large.

The constants K, s are related to the parameters $C^{(i)}, s^{(i)}$ using expression (24). Estimates $\hat{C}^{(i)}, \hat{s}^{(i)}$ can be determined by the "Least Squares" method based on the current values of the elements of sets $\{w_i\}, \{n_i\}$. At the same time, we take into account $\hat{K}^{(i)}(o) = c^{C^{(i)}}$, where c is the base of the logarithm. Considering the determined estimates of the constants K, s , we present expression (21) as

$$r_i(o) = \hat{K}^{(i)}(o) (\sqrt{q}/2N_i^o)^{\hat{s}^{(i)}(o)} \quad (25)$$

Expression (25) allows us to estimate the spacing r_i to the exact global extrema value, taking into account the unit cube volume o , characteristic of the i -th iteration of the parallel computing oriented method.

Now let's analytically estimate the average number of iterations until the realization of the situation ξ when condition (6) is fulfilled. Suppose that this situation is realized on the ϕ -th iteration. Let's introduce a positive integer variable T for which $\xi = \{\phi > T\}$. For the situation ξ , we can write $\xi \subset \bigcup_{s=T-\rho}^T \{w_s > \gamma\}$, based on which we will get

$$P(\xi) \leq \sum_{s=T-\rho}^T P\{w_s > \log \gamma\} \leq (\rho + 1) \max_{T-\rho \leq s \leq T} P\{w_s > \log \gamma\} \quad (26)$$

Let us present expression (24) in the context of the Bienaime inequality:

$$P\{w_s > \log\gamma\} = P\{\psi_s > \log\gamma - \widehat{C}^{(T)} + \widehat{s}^{(T)}_s\} \leq \widehat{V}^{(T)} / (\log\gamma - \widehat{C}^{(T)} + \widehat{s}^{(T)}_s)^2 \tag{27}$$

Substitute expression (27) into expression (26) and get

$$P(\xi) = (\rho + 1) \max_{T-\rho \leq s \leq T} \widehat{V}^{(T)} / (\log\gamma - \widehat{C}^{(T)} + \widehat{s}^{(T)}_s)^2 \leq (\rho + 1) \widehat{V}^{(T)} / (\log\gamma - \widehat{C}^{(T)} + \widehat{s}^{(T)}(T - \rho))^2 = I(1/T^2), \tag{28}$$

where $\widehat{V}^{(T)}$ is an estimate of the variance V of the stochastic set ψ_i .

Based on the terminology used, we express the average number of iterations of the parallel computing oriented method before the event ξ occurs as

$$E\{\phi\} = \sum_{m=1}^{\infty} mP\{\phi = m\} \tag{29}$$

Let's open the expression (29) for $T > 2$:

$$\begin{aligned} \sum_{m=1}^T mP\{\phi = m\} &= \sum_{m=2}^T m(P\{\phi > m-1\} - P\{\phi > m\}) = \\ &= \sum_{m=2}^T (m-1)P\{\phi > m-1\} - \sum_{m=2}^T mP\{\phi > m\} + \\ &+ \sum_{m=2}^T P\{\phi > m-1\} = \sum_{m=1}^{T-1} P\{\phi > m\} - TP\{\phi > T\}. \end{aligned} \tag{30}$$

Taking expression (28) into account, we write:

$$\lim_{T \rightarrow \infty} TP\{\phi > T\} \leq \lim_{T \rightarrow \infty} TI(1/T^2) = 0 \tag{31}$$

Consider inequality (31) when passing to the limiting form of expression (30) for $T \rightarrow \infty$:

$$E\{\phi\} = \sum_{T=1}^{\infty} P\{\phi > T\} \leq I \left(\sum_{T=1}^{\infty} 1/T^2 \right) \tag{32}$$

Expression (32) is an estimate of the average number of iterations of the method before it stops at the ϕ -th iteration. Assume that condition (32) is fulfilled. Based on expressions (14), (18), we can conclude that with a sufficiently large i , probability that at least one of the values of $F_{\phi-\rho}^*$, $F_{\phi-\rho+1}^*$, ..., F_{ϕ}^* is not in the vicinity of the exact global minimum of the optimization problem (4) tends to zero when the situation ξ arises.

When the situation ξ occurs on the ϕ -th iteration, it is possible to calculate the estimates $\widehat{K}^{(\phi)}(o)$, $s^{(\phi)}(o)$ using expression (25). In turn, these estimates can be used to estimate the spacing of the found global minimum F_{ϕ}^* from the exact value F^* . Taking into account expressions (14), (25), we write:

$$\left| F_{\phi}^* - F^* \right| \leq r_{\phi}^+ \tag{33}$$

$$r_{\phi}^+ = \max_{0 < \varphi \leq \sigma \leq \beta < 1} r_{\phi}(\sigma)$$

Taking this into account, we characterize the lower limit of the probability of occurrence of the situation $\xi_{r_{\phi}^+}$ from expression (6) as

$$P_{r_{\phi}^+}^* = \min_{0 < \varphi \leq \sigma \leq \beta < 1} \left(1 - N_{\phi}^{\varphi\sigma} \exp\left(-qN_{\phi}^{1-\sigma}\right) \right) \tag{34}$$

3. Results

The authors' method is focused on finding the global minimum of the optimization problem (4), which is an adaptation of the canonical form of the optimization problem (2). We algorithmically implement the method as a composition of two functions. The function *MinGen()* implements the generation of a set of local minima Z_i^* and a set of

decrements δ_i , as well as checking the fulfilment of the stop condition (6). The function *MinEst()* implements the quality check of the found extrema using the probabilistic characteristic (34), which allows for determining the global minimum. The search for the extremum is implemented iteratively, and at each i -th iteration, s computational operations are performed. For the convenience of readers, all variables used below are presented in the Nomenclature at the end of the article.

At each i -th iteration of the application of the function *MinGen()*, the following actions are performed:

1. Using the stochastic tests method a pool of stochastic vectors $Z_i = \{z^{(1)}, \dots, z^{(N_i)}\}$, $N_i = \alpha N_{i-1}$ is formed;
2. From the pool Z_i , stochastic vectors belonging to the admissible set L are selected. As a result, a set Y_i with capacity \widetilde{N}_i is formed;
3. The values of the objective function (4) on the set of admissible points $y^{(i)} \in Y$ are calculated: $F(y^{(k)})$, $k = \overline{1, \widetilde{N}_i}$;
4. From the values $F(y^{(k)})$ calculated at stage 3, the smallest one is chosen: $\min_{1 \leq k \leq \widetilde{N}_i} F(y^{(k)})$;
5. The decrement $\delta_i = |F_i^* - F_{i-1}^*|$ and derived parameters $w_i = \log \delta_i$, $n_i = \log N_i$ are calculated.

Stages 1-5 are repeated for ρ iterations (as long as the condition $w_i < \gamma$ is fulfilled). After performing ρ iterations, the function *MinGen()* completes its work. The output parameter of the function is a tuple of storage sets (F, Δ, W, N) of capacity ρ each. Next, the function *MinEst()* comes into action, which for the input tuple (F, Δ, W, N) implements the following structured calculation procedure:

1. Estimates $\widehat{C}^{(i)}$, $\widehat{s}^{(i)}$ are calculated for sets N, W by the least squares method (see expression (24));
2. Such quality indicators of the found solution as the spacing r_{ϕ}^+ (see expression (33)) and the probability characteristic $P_{r_{\phi}^+}^*$ (see expression (34)) are calculated.

The algorithmic constructions described above were implemented in the Python programming environment with a focus on supporting *GPU Programming with Python and CUDA* parallel computing technologies, the description of which is freely available at the link <https://github.com/PacktPublishing/Hands-On-GPU-Programming-with-Python-and-CUDA>. The created software was based on a software platform consisting of Windows, Python 2.7, Anaconda 5, CUDA 10 (incl. cuBLAS, cuSolver), PyCUDA, Scikit-CUDA. The hardware configuration of the test bench included a 64-bit Intel CPU, 32 Gigabytes of RAM, NVIDIA GPU RTX 3070.

The plan of experiments is designed to empirically confirm the adequacy of the presented method, as well as to obtain data for evaluating its effectiveness.

Empirical confirmation of the adequacy of the authors' method was carried out based on the content of the *Optimization Test Problems* section of the specialized resource *Virtual Library of Simulation Experiments: Test Functions and Datasets*, freely available at the link <https://www.sfu.ca/~ssurjano/optimization.html>.

Among the instances available in the *Optimization Test Problems* section, we selected ten optimization problems, the of admissible solutions domains of which are limited by P_x -shaped parallelograms. Here is a list of selected problems with the values of the corresponding controlled parameters and the output results x^*, f^* in the context of the terminology used in the article:

$$T1: f(x) = \sum_{i=1}^k 1 / \left(a_i + \sum_{j=1}^q (x_j - c_{ji})^2 \right). \text{ For } k = 5, q = 4, A = \begin{pmatrix} 4 & 1 & 6 & 8 & 3 \\ 4 & 1 & 6 & 8 & 7 \\ 4 & 1 & 6 & 8 & 3 \\ 4 & 1 & 6 & 8 & 7 \end{pmatrix} \text{ we have: } x^* = 4,000,$$

$$f^* = -10,153;$$

$$T2: f(x) = \left(\sum_{i=1}^q x_i^4 - 16x_i^2 + 5x_i \right) / 2. \text{ For } q = 4 \text{ we have: } x^* = 2,904, f^* = -156,664;$$

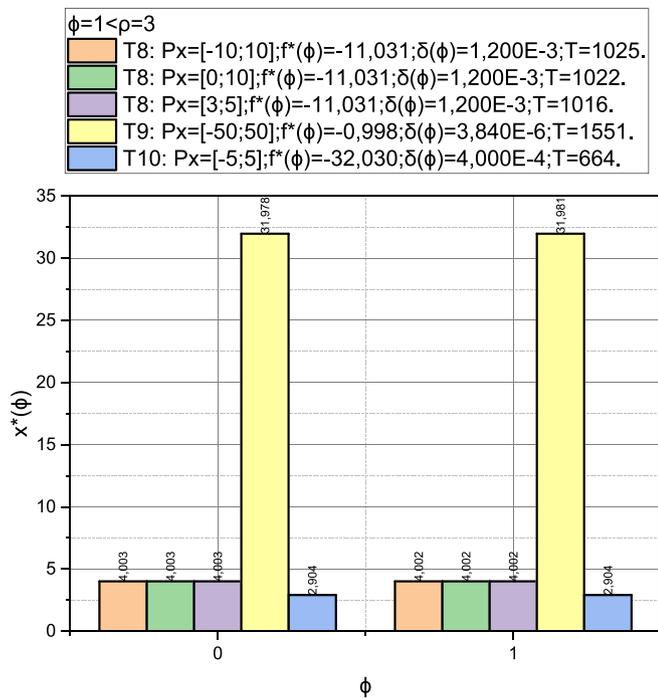


Fig. 4. Results of solving test problems T8, T9, T10 by the authors' method.

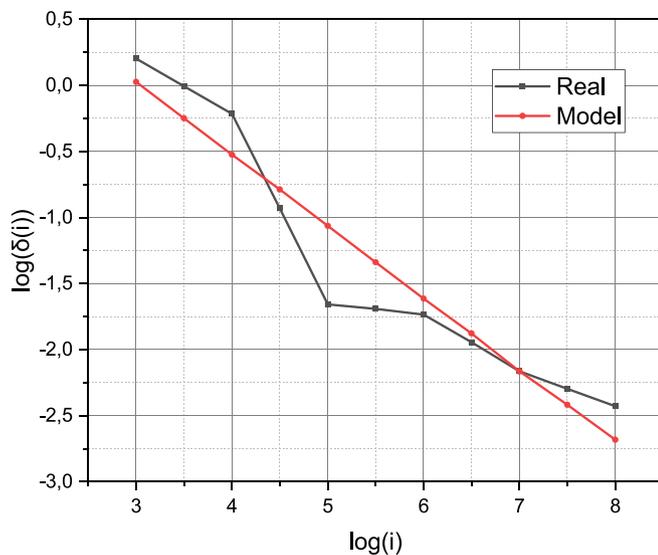


Fig. 5. Real and simulated functional dependence of $\log \delta_i = f(\log(i))$ at $F^* = 0$.

4. Discussion

Let's start the discussion by stating the empirically proven fact of the adequacy of the proposed method for solving the global optimization problems, in which continuous objective functions satisfy the Hölder condition, and the control parameters domain limited by continuous functions is characterized by a positive Lebesgue measure and is limited by multidimensional parallelepipeds. This fact is confirmed by the results of comparing the solutions obtained using the authors' method with etalon solutions of optimization problems selected from the *Optimization Test Problems* section of the specialized resource *Virtual Library of Simulation Experiments: Test Functions and Datasets*. Figs. 1-4 shows that for all optimization test problems T1-T10, approximate values of the global minimum with an error in the range of $[10^{-9}, 10^{-2}]$ were found.

Moreover, for test problems, T1-T7, adequate values of the upper estimates of the spacing to the exact global minimum (metric r_ϕ) with a probability close to one were obtained. Separately, we note that both the quality characteristics of the initial results calculated according to the authors' method and the time to obtain them turned out to be practically independent of the size of the search area. This expected result (provided by a selection of simple statistical tests performed on iterations of the method) is a significant advantage of the authors' method over analogues.

We will also note several theoretically grounded remarks concerning the authors' method. In particular, the values of estimate (14) depend on the value of the established parameter ρ , which is associated with the selected value of the unit cube (see (8)). The fact that the set F_i^* is guaranteed to converge with an exponential rate is theoretically confirmed by the guaranteed convergence of the strictly monotonic set $F^* = \{F_0^* > F_1^* > \dots\}$ to the value of the exact global extrema with an exponential rate. Recall that the cumulative set F^* grows with each transition from the i -th to the $i+1$ -th iteration of the authors' method (see the beginning of Section 2.2). Finally, if the admissible set is part of a unit cube (4), then when evaluating the probabilistic characteristics mentioned in Section 2.3, only the number of unit cubes in the partition of the admissible set will change. Accordingly, the estimates of the probabilistic characteristics from Section 2.3 (as well as the statements regarding the convergence of the authors' method) are valid under the condition of compactness (4), which is characterized by the positivity of the Lebesgue measure. With this in mind, when implementing the authors' method, at each iteration, you should check whether each stochastic point belongs to an admissible set.

5. Conclusion and future directions

The article proposes a parallel computing oriented method for solving the global minimum finding problems, in which continuous objective functions satisfy the Hölder condition, and the control parameters domain limited by continuous functions is characterized by a positive Lebesgue measure and is limited by multidimensional parallelepipeds. A typical example of such a task is the discrepancy minimizing problem between the left and right parts of some system of equilibrium equations. The method is based on simple statistical tests, thanks to which, at each iteration, growing sets of potential global minima and sets of decrements necessary for estimating the values of the Hölder constants are formed. The article theoretically substantiates and empirically proves the guaranteed convergence of the authors' method to the real global minimum, which occurs at an exponential rate. For the continuous iterations number, analytical upper estimates of the spacing between the potential global minima and real global minima are formalized, as well as an estimate of the probability of overcoming this spacing is formalized. The decrements sequence approximation, estimation of a priori unknown Hölder constants, estimation of the average number of iterations of the method and probabilistic characteristics of the final solution are analytically justified. In addition to the theoretical proof, the adequacy of the authors' method has been confirmed empirically. It turned out that both the quality characteristics of the initial results calculated by the authors' method and the time to obtain them are practically independent of the size of the search area. This expected result is a significant advantage of the authors' method over analogues.

Further research is planned to be devoted to the applied use of the proposed method, in particular, in the field of analysis of small stochastic data. The theoretical basis for these studies are articles such as (Kovtun et al., 2024; Kovtun et al., 2023).

Funding

This work was funded by Researchers Supporting Project number (RSP2024R503), King Saud University, Saudi Arabia.

Acknowledgments

The authors are grateful to all colleagues and institutions that contributed to the research and made it possible to publish its results.

References

- Alexiadis, A., 2023. A minimalistic approach to physics-informed machine learning using neighbour lists as physics-optimized convolutions for inverse problems involving particle systems. *J. Comput. Phys.* 473, 111750 <https://doi.org/10.1016/j.jcp.2022.111750>.
- Alfant, R.M., Ajayi, T., Schaefer, A.J., 2023. Evaluating mixed-integer programming models over multiple right-hand sides. *Oper. Res. Lett.* <https://doi.org/10.1016/j.orl.2023.05.004>.
- Ali, H., Tariq, U. U., Hardy, J., 2021. Bensaali, F.; Amira, A.; Fatema, K.; Antonopoulos, N. A Survey on System Level Energy Optimisation for MPSoCs in IoT and Consumer Electronics. *Comput. Sci. Rev.*, 41, 100416. [10.1016/j.cosrev.2021.100416](https://doi.org/10.1016/j.cosrev.2021.100416).
- Bisikalo, O., Kharchenko, V., et al., 2023. Parameterization of the stochastic model for evaluating variable small data in the Shannon entropy basis. *Entropy* 25, 184. <https://doi.org/10.3390/e25020184>.
- Castelli, A.F., Moretti, L., Manzolini, G., et al., 2023. Robust optimization of seasonal, day-ahead and real time operation of aggregated energy systems. *Int. J. Electr. Power Energy Syst.* 152, 109190 <https://doi.org/10.1016/j.ijepes.2023.109190>.
- Chen, W., Zheng, M., 2021. Multi-objective optimization for pavement maintenance and rehabilitation decision-making: a critical review and future directions. *Autom. Constr.* 130, 103840 <https://doi.org/10.1016/j.autcon.2021.103840>.
- Fang, D.H., Li, C., Yang, X.Q., 2012. Asymptotic closure condition and Fenchel duality for DC optimization problems in locally convex spaces. *Nonlinear Anal. Theory Methods Appl.* 75, 3672–3681. <https://doi.org/10.1016/j.na.2012.01.023>.
- Fransen, M. P., Langelaar, M., 2023. Schott, D. L. Deterministic vs. Robust Design Optimization Using DEM-Based Metamodels. *Powder Technology*, 425, 118526. [10.1016/j.powtec.2023.118526](https://doi.org/10.1016/j.powtec.2023.118526).
- Gorawski, M., Grochla, K., Marjasz, R., et al., 2021. Energy minimization algorithm for estimation of clock skew and reception window selection in wireless networks. *Sensors* 2021 (21), 1768. <https://doi.org/10.3390/s21051768>.
- Hirsching, C., de Jongh, S., Eser, D., et al., 2022. Meta-heuristic optimization of control structure and design for MMC-HVdc applications. *Electr. Pow. Syst. Res.* 213, 108371 <https://doi.org/10.1016/j.epsr.2022.108371>.
- Huang, M., Du, Z., Liu, C., et al., 2022. Problem-independent machine learning (PIML)-based topology optimization—A universal approach. *Extreme Mech. Lett.* 56, 101887 <https://doi.org/10.1016/j.eml.2022.101887>.
- Izonin, I., Tkachenko, R., Shakhovska, N., et al., 2022. Two-step data normalization approach for improving classification accuracy in the medical diagnosis domain. *Mathematics* 10, 1942. <https://doi.org/10.3390/math10111942>.
- Kovtun, V., Grochla, K., Kharchenko, V., et al., 2023. Stochastic forecasting of variable small data as a basis for analyzing an early stage of a cyber epidemic. *Sci. Rep.* 13 <https://doi.org/10.1038/s41598-023-49007-2>.
- Kovtun, V., Altameem, T., Al-Maitah, M., Kempa, W., 2024. Entropy-metric estimation of the small data models with stochastic parameters. *Heliyon* 10, e24708.
- Marulanda-Durango, J.J., Zuluaga-Ríos, C.D., 2023. A meta-heuristic optimization-based method for parameter estimation of an electric arc furnace model. *Results Eng.* 17, 100850 <https://doi.org/10.1016/j.rineng.2022.100850>.
- McNaughton, S., 2023. Optimizing Learning: An Overview. *International Encyclopedia of Education (Fourth Edition)*, 560–567. [10.1016/b978-0-12-818630-5.14065-5](https://doi.org/10.1016/b978-0-12-818630-5.14065-5).
- Montano, J., Garzón, O.D., Rosales Muñoz, A.A., et al., 2022. Application of the arithmetic optimization algorithm to solve the optimal power flow problem in direct current networks. *Results Eng.* 16, 100654 <https://doi.org/10.1016/j.rineng.2022.100654>.
- Pei, P., Quek, S.T., Peng, Y., 2023. Enriched global-local multi-objective optimization scheme for fuzzy logic controller-driven magnetorheological damper-based structural system. *Mech. Syst. Sig. Process.* 193, 110267 <https://doi.org/10.1016/j.ymsp.2023.110267>.
- Shilaja, C., Kiran, S.R., Murali, M., et al., 2022. Design and analysis of global optimization methods for proton exchange membrane fuel cell powered electric vehicle system with single switch DC-DC converter. *Mater. Today: Proc.* 52, 2057–2064. <https://doi.org/10.1016/j.matpr.2021.12.204>.
- Smith, J.E., Nair, R., 2005. Dynamic binary optimization. *Virtual Mach.* 147–219 <https://doi.org/10.1016/b978-155860910-5/50005-7>.
- Sovrasov, V.V., 2016. Local tuning in peano curves-based global optimization scheme. *Procedia Comput. Sci.* 101, 27–34. <https://doi.org/10.1016/j.procs.2016.11.005>.
- Stracquadanio, G., Pardalos, P.M., 2019. Stochastic methods for global optimization and problem solving. *Encyclopedia Bioinform. Computat. Biol.* 321–327 <https://doi.org/10.1016/b978-0-12-809633-8.20329-4>.
- Tymchenko, O., Havrysh, B., Khamula, O., et al., 2019. Methods of converting weight sequences in digital subtraction filtration. In: 2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT). <https://doi.org/10.1109/stc-csit.2019.8929750>.
- Wu, H., Niu, W., Zhang, Y., et al., 2023. A hybrid polynomial-based optimization method for underwater gliders with parameter uncertainty. *Appl. Ocean Res.* 133, 103486 <https://doi.org/10.1016/j.apor.2023.103486>.
- Xie, W., Tang, W., Kuang, Y., 2022. A new hybrid optimizer for stochastic optimization acceleration of deep neural networks: dynamical system perspective. *Neurocomputing* 514, 341–350. <https://doi.org/10.1016/j.neucom.2022.09.147>.
- Xu, A., Li, S., Fu, J., et al., 2023. A hybrid method for optimization of frame structures with good constructability. *Eng. Struct.* 276, 115338 <https://doi.org/10.1016/j.engstruct.2022.115338>.
- Yang, Z., Kang, R., Luo, X., et al., 2019. Rigorous modelling and deterministic multi-objective optimization of a super-critical CO2 power system based on equation of state and non-linear programming. *Energ. Convers. Manage.* 198, 111798 <https://doi.org/10.1016/j.enconman.2019.111798>.
- Yang, C., Zhu, Y., Zhou, J., et al., 2023. Dynamic flexibility optimization of integrated energy system based on two-timescale model predictive control. *Energy* 276, 127501. <https://doi.org/10.1016/j.energy.2023.127501>.
- Zaiats, V.M., Rybyska, O.M., Zaiats, M.M., 2019. An approach to assessment of the value and quantity of information in queueing systems based on pattern recognition and fuzzy sets theories. *Cybern. Syst. Anal.* 638–648 <https://doi.org/10.1007/s10559-019-00172-1>.
- Zhang, X., Huang, G., Xie, Y., et al., 2022. A coupled non-deterministic optimization and mixed-level factorial analysis model for power generation expansion planning – a case study of Jing-Jin-Ji metropolitan region, China. *Appl. Energy* 311, 118621. <https://doi.org/10.1016/j.apenergy.2022.118621>.
- Zhang, J., Liu, C., Li, X., et al., 2023. A survey for solving mixed integer programming via machine learning. *Neurocomputing* 519, 205–217. <https://doi.org/10.1016/j.neucom.2022.11.024>.