DOI: 10.20998/2079-0023.2026.01.02
UDC 004:519.24:681.3.06

A. A. PAVLOV, Doctor of Technical Sciences, Full Professor, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine, Professor of Computer Science and Software Engineering Department; e-mail: alexanderpavlov1944@gmail.com; ORCID: https://orcid.org/0000-0002-6524-6410
M. Ye. KYSELOV, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Bachelor of Informatics and Software Engineering Department, Kyiv, Ukraine, e-mail: zeusmobilenick@gmail.com, ORCID: https://orcid.org/0009-0005-3686-3419

VERIFICATION OF EMPIRICAL MODELS OF SIMPLEX METHOD COMPLEXITY USING THE MODIFIED GMDH

This paper presents the results of constructing an empirical dependence of the number of arithmetic operations of the standard simplex method as a function of the problem dimensionality defined in canonical form. To address this problem, the principle of model self-organization based on a modified combinatorial algorithm of the Group Method of Data Handling is applied. The modification of Group Method of Data Handling consists in forming a residual representation that is assumed to contain the sought empirical complexity function, as well as in applying a preliminary partitioning algorithm for the basis functions, which are additively included in the residual representation, into two non-overlapping classes: the class of main dominant components and the class of refining residual components. This approach significantly reduces the combinatorial search procedure. Based on theoretical complexity estimates available in the literature, a redundant set of basis functions was constructed. Five fundamental models were considered: Borgwardt’s polynomial estimate, the Adler – Megiddo quadratic bound, a basic polynomial form, the smoothed analysis model of Spielman – Teng, and a general mixed model previously proposed by the authors. To expand the search space, a logarithmic component was added to the basis functions. Thus, the residual representation includes six basis components, along with all their squares and pairwise products. The optimal structure was selected using a symmetric regularity criterion by evaluating more than 134 million alternative models, the number of which is uniquely determined by the cardinality of the set of refining residual components. To ensure the correct application of the least squares method for strongly nonlinear regressors of different scales (whose absolute values differ by several orders of magnitude), the necessity of scaling input data by their standard deviation, with optional centering, was experimentally justified. The efficiency of this approach was confirmed through a specially designed experiment with a known ideal solution, including a constant term (bias) at the order of 100000 and normally distributed noise with an amplitude of 1 % of the mean value of the ideal regression on the values of the input variables of the experiment. The results of simulation modeling on a dataset of 13690 randomly generated linear programming problems demonstrate the clear superiority of the mixed polynomial-logarithmic structure within the extended class of candidate functions.

Keywords: simplex method, model self-organization, modified Group Method of Data Handling, least squares method, redundant representation, regularity criterion, data standardization.

1. Introduction. Publications [1, 2] consider applied problems, the solution of which requires the solution of on average hundreds of linear programming (LP) problems with thousands of variables. The efficient implementation of such algorithms is based on parallelization of calculations, which requires knowledge of efficient estimates of the number of simplex method’s arithmetic operations as a function of the dimensionality of the LP problem given in canonical form [3].

With the aim to verify the objectivity of the empirical complexity formula previously proposed by the authors $(0.63m^{2.96}n^{0.02}\log^{1.62}n+4.04m^{-4.11}n^{2.92}$, where m is the number of equations, n is the number of non-negative variables), which was found using the four common basis functions presented in the abstract, and seven in total, using the modified Group Method of Data Handling (MGMDH) [4]. For this, in this work, the MGMDH is used for the case when the residual representation, according to which the search for the empirical formula of complexity is implemented, is extended to the basis functions, their squares and all pairwise products.

The use of the MGMDH allows us to move away from the a priori defining the model structure and find the optimal dependence by a reduced search of subsets of basis functions, their squares and all pairwise products using the regularity criterion first proposed by Ivakhnenko O. A., as a part of the classical Group Method of Data Handling (GMDH). Finding the coefficients for partial representations (partial models) is done using the least squares method (LSM). The selection of the optimal structure is carried out using the symmetric regularity criterion, which is the sum of two components, each of which is the value of the residual sum of squares on the test sequence for the case when the data from which the partial representations of the sought model are built and the data from which the value of the regularity criterion is found are interchanged. Since the redundant representation contains significantly nonlinear basis functions, the values of which can differ by orders of magnitude, the classical implementation of the residual representation of the MGMDH for linear models requires an appropriate correction.

2. Basic theoretical provisions and data generation. To objectively investigate the computational complexity of the standard simplex method, it is necessary to generate a representative sample of random LP problems in canonical form. To ensure the admissibility of the constraint system, the generation is carried out by the inverse construction method.

That is, the components of the non-negative solution of the vector $\boldsymbol{x}$ and the parameters of the LP problem are generated as realizations of random variables that have a uniform distribution on the intervals $(0.1, f+0.1)$ for the components of the vector $\boldsymbol{x}$, on the interval $(0, f)$ for the parameters of the constraint matrix $\boldsymbol{A}$. The vector of the right-hand side $\boldsymbol{b}$ is generated by the formula $\boldsymbol{b} = \boldsymbol{Ax} + \boldsymbol{\delta}$, where $\boldsymbol{x}$ is the solution of the LP problem, $\boldsymbol{\delta}$ is a vector whose components are realizations of random variables that have a uniform distribution on the interval $(0, 2f)$. The components of the vector $\boldsymbol{c}$ (the optimality criterion of the LP problem) were also generated from a uniform distribution on the interval $(0, f)$. The quantity f acts as a hyperparameter (scaling factor), which allows you to flexibly control the order of quantities in the generated data. This allows you to more correctly find the empirical complexity function of the simplex method.

As mentioned in the introduction, to build an empirical model of complexity on the obtained data, the MGMDH is used [4]. Unlike the classical combinatorial GMDH, which involves a complete enumeration of all possible combinations of basis functions (which is computationally impossible for large bases) and does not specify a residual representation containing the sought solution structure, the modification of the GMDH also consists in a preliminary partitioning of the space of the set of components of the residual representation, linear with respect to unknown coefficients, into two classes. Based on the initial LSM estimate, the algorithm identifies a cluster of main dominant components (set M_1) and assigns the remaining basis functions to a cluster of clarifying residual components (set M_2). The structure of each partial representation contains a mandatory component from the set M_1 and an arbitrary subset of components from the set M_2. It is postulated that the sought empirical formula for the complexity of the simplex method belongs to the set of partial representations formulated above.

3. Data standardization issues. To ensure the correct operation of the LSM within the framework of structural identification, the stage of preliminary data preparation is critically important. In computational complexity estimation problems, the redundant representation contains significantly nonlinear basis functions (see Section 5), for example, such as $\ln m + \ln n$ or $m^3 n^2$. Since their absolute values differ by many orders of magnitude, direct application of the LSM to raw data leads to serious distortions: functions with significant absolute values receive coefficients that are insignificant in absolute value, which makes it impossible to correctly algorithmically select significant basis functions by absolute values of their coefficients. According to the classical theory of regression analysis [5], an efficient solution to this problem is the standardization of predictors that includes centering (subtraction of the mean) and scaling (division by the standard deviation σ of the generated values of the input variables from their mean value) [6]. However, in our case, the input parameters (dimensions of matrices) and the output objective function (the number of operations) have strictly non-negative values. The use of the centering operation inevitably leads to the appearance of artificial negative values, which destroys the physical logic of the problem and complicates the interpretation.

In view of this, this work justifies the use of only the scaling operation without centering. In general, the scaling procedure is as follows.

Let the regression model be of the form:

$$\boldsymbol{y}(x_1,\ldots,x_m) = \omega_0 + \sum_{j=1}^{r} \omega_j f_j(x_1,\ldots,x_m) + \varepsilon, \tag{1}$$

where ε is a random variable, the coefficients ω_j, $j = \overline{0,r}$, are unknown. The test results:

$$\left((x_{1i},\ldots,x_{mi}) \to y_i,\, i = \overline{1,n}\right), \tag{2}$$

we replace model (1) with the following:

$$\hat{\boldsymbol{y}}(x_{1i},\ldots,x_{mi}) = \omega_0 + \sum_{j=1}^{r} \omega_j \frac{f_j(x_{1i},\ldots,x_{mi})}{\sigma_j},\ i = \overline{1,n}, \tag{3}$$

where

$$\sigma_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f_j(x_{1i},\ldots,x_{mi}) - \frac{1}{n}\sum_{i=1}^{n} f_j(x_{1i},\ldots,x_{mi})\right)^2}.$$

According to model (3), based on the results of statistical tests (2), using the LSM we find estimates of unknown coefficients. The final constructed regression has the form:

$$\hat{\boldsymbol{y}}(x_1,\ldots,x_m) = \hat{\omega}_0 + \sum_{j=1}^{r} \frac{\hat{\omega}_j}{\sigma_j} f_j(x_1,\ldots,x_m). \tag{4}$$

Remark. The MGMDH uses the presented scaling procedure for finding estimates of the unknown coefficients for all partial representations.

Such preprocessing is a necessary condition for the LSM to be able to adequately estimate the parameters, and for the clustering algorithm of the GMDH to correctly separate the main representation from the noise one without losing the physical sense.

It is important to note that such an exact algebraic equivalence is a specific property of models that are linear in parameters. In most other machine learning algorithms (e.g., decision trees or multilayer neural networks), such an analytical denormalization of weights is not possible because the scale of arguments is nonlinearly woven into the internal structure of decision making [7].

4. Synthetic test proving the importance of scaling. To experimentally prove the hypothesis of the critical need for data scaling, a verification simulation test was conducted on synthetic data with a known reference structure.

A sample of 500 points was generated, where the independent arguments varied in the range $m, n \in [10, 100]$.

The true target pattern was generated in the form:

$$\boldsymbol{y}(m,n) = 10^5 + m^3 \cdot n^2 + \varepsilon,$$

where ε is a normal noise with a standard deviation of 1 % from the mean of the regression model on the experimental input data. The structural identification algorithm was given a redundant representation of nine candidate basis functions: basis $m^3, n^2, \ln m$, their squares, and all possible pairwise products. The algorithm's task was to automatically find the true pattern.

The results of applying the LSM to unscaled data led to erroneous final results: due to the imbalance of the orders of magnitude of the functions, the algorithm demonstrated instability at the stage of structural identification. When dividing into clusters, the algorithm gave false preference to basis functions with small numerical values, mistakenly selecting as the only main basis function (cluster M_1) a function $\ln m$ with a coefficient of $\sim 1.4126 \cdot 10^9$. At the same time, the algorithm perceived the true dominant term $m^3 \cdot n^2$ as secondary noise and sent it to the residual representation M_2 with a coefficient of 1.0019. Structurally, the model was identified absolutely incorrectly, which makes further analysis of complexity impossible.

Instead, the application of pre-scaling of the input data significantly improved the result. The LSM correctly determined the weights of each function in the total variance, which allowed the algorithm to identify the structure without error. The only true basis function $m^3 \cdot n^2$ was automatically selected for the main representation cluster (M_1). The resulting model took the following form:

$$\boldsymbol{y}(m,n) = 1.2046 \cdot 10^5 + 1.0006 \cdot \left(m^3 \cdot n^2\right). \qquad (5)$$

The terms of the residual representation M_2 and the shift in this case act exclusively as insignificant compensators of the imposed noise. Therefore, the synthetic criterion took the minimum value on the correct structure of the regression model (5). In this example, the following modification of the criterion for selecting the final solution was used: all partial representations were considered, for which the values of the symmetric criterion practically do not differ from the minimum, and from them the partial representation was selected, which contains the minimum number of basis functions. The resulting estimates of the unknown coefficients were found for the entire set of experimental data and practically do not differ from the coefficients of the ideal model.

5. Simulation modeling of input data for finding the empirical complexity function of the simplex method using the MGMDH. Having confirmed the efficiency of the methodology, the MGMDH was applied to objectively verify the empirical formula for the computational complexity of the standard simplex method. For the empirical analysis, random LP problems were generated according to the rules described in Section 2.

To ensure statistical reliability of the results, five series of independent simulation experiments were conducted for each of the two samples (estimating data for each of the models using the LSM and data of the validation subset used to select the best empirical formula from the six candidates found). The values of m (the number of constraints) and n (the number of variables) varied from 200 to 2000 with a step of 50. For each combination of m and n values, five independent LP problems were generated. Thus, the total number of generated and solved problems in one series of tests was $37 \cdot 37 \cdot 5 \cdot 2 = 13690$. For each solved problem, the exact number of arithmetic operations performed by the simplex method was recorded.

To implement the structural identification procedure, a redundant representation was formed based on five fundamental theoretical estimates: the earlier proposed by the authors general mixed model without considering the constant $\left(f_{\text{general}} = 0.63m^{2.96}n^{0.02}\log^{1.62} n + 4.04m^{-4.11}n^{2.92}\right)$, the Adler-Megiddo model $\left(f_{\text{Adler}} = n^4\right)$ [8], the refined Borgwardt model $\left(f_{\text{Borg}} = m^3 n^2\right)$ [9], the basic polynomial model $\left(f_{\text{poly}} = mn\right)$ [10], and the smoothed analysis model $\left(f_{\text{smooth}} = mn^5 \ln n\right)$ [11]. In order to increase the reliability of the verification of the obtained model, the search space was expanded by introducing logarithmic components, namely, supplemented with an additive logarithmic function $f_{\log} = \ln m + \ln n$.

This allowed us to form a set of partial representations from the class of complex polynomial-logarithmic functions, which additively includes six basis functions, their squares, and pairwise products.

Thus, the total number of residual terms is equal to $M = 27$ candidate arguments. Considering the cardinality of the set M_1, which is equal to $M_{11} \bigcap M_{12}$ (each of which is found by its own validation subset), the set of all partial representations is equal to $2^{27} - 1 \approx 1.34 \cdot 10^8$ models, which practically guarantees finding the correct structure of the empirical formula for the complexity of the simplex method within the given residual representation.

The choice of selected basis functions, instead of elementary arguments such as $m, n,$ is due to the multiplicative nature of the interaction of the number of constraints and variables. Since the total complexity is derived from the product of the number of iterations times the complexity of one iteration of the simplex method, any correct estimate of the complexity of the simplex method must consider this physical logic of the algorithm.

To increase the estimation objectivity, as indicated above, a symmetric regularity criterion was used, which requires dividing the test sample into two independent parts: a training subset T and a validation subset V. The algorithm works according to a two-step scheme: first, the estimates of the coefficients of partial representations are found by the sample T and the residual sum of squares of each partial representation is found by the sample V ($T \bigcap V$ is an empty set, $T \bigcup V$ is the set of all test data, the cardinality of the sets T and V is the same); then the procedure is repeated when T performs the functions of the validation subset, and V is used to find the estimates of the coefficients of partial representations (from which the validity of the formula $M_1 = M_{11} \bigcap M_{12}$ follows).

6. Identification results. As a result of using the symmetric regularity criterion, it was found that the above formulated criterion for selecting the correct structure of the partial representation gave the following result: from the set of partial representations, for each of which the values of the symmetric regularity criterion practically do not differ from the minimum, a partial representation with the minimum number of basis functions from the partial representation is selected. Estimates of the unknown coefficients of the selected partial representation are found for the entire set of test data. Thus, the resulting empirical formula for the number of arithmetic operations of the simplex method as a function has the form:

$$\begin{gathered} y(m,n) \approx 3.2705 \cdot 10^8 + \\ +1.0045 \cdot f_{\text{general}} + 1.58 \cdot 10^8 \cdot f_{\log}^2 + 1.97 \cdot 10^{-22} \cdot f_{\text{Borg}}^2 - \\ -1.13 \cdot 10^{-2} \cdot f_{\text{poly}}^2 - 4.49 \cdot 10^{-6} \cdot f_{\text{Adler}} f_{\log} - \\ -1.53 \cdot 10^9 \cdot f_{\log} - 2.35 \cdot 10^{-4} \cdot f_{\text{Borg}} - 2.38 \cdot 10^{-17} \cdot f_{\text{Borg}} + \\ +1.69 \cdot 10^{-5} \cdot f_{\text{Borg}} f_{\log} - 6.17 \cdot 10^{-12} \cdot f_{\text{Borg}} f_{\text{poly}} + \\ +7.10 \cdot 10^{-10} \cdot f_{\text{smooth}} - 2.51 \cdot 10^5 \cdot f_{\text{poly}} - \\ -5.47 \cdot 10^{-9} \cdot f_{\text{poly}} f_{\text{Adler}} + 1.88 \cdot 10^4 \cdot f_{\text{poly}}. \end{gathered} \tag{6}$$

It should be noted separately that there is a free term (shift) in the obtained model, which is about $3.2705 \cdot 10^8$. From the point of view of classical asymptotic analysis of algorithms, a constant term does not affect the complexity class with infinite growth of dimensionality. However, in the context of applied resource estimation problems, this parameter carries information about the fixed overhead of the algorithm [12]. Considering this shift is critically important for accurate prediction of the number of operations on problems of small and medium dimensionality $(m \le 500, n \le 600)$, where asymptotic terms are not yet dominant. It is for these restrictions that it is possible to use a simplified formula:

$$\begin{gathered} y(m,n) \approx 3.2705 \cdot 10^8 + 1.0045 \cdot 0.63 m^{2.96} n^{0.02} \log^{1.62} n + \\ + 4.04 m^{-4.11} n^{2.92}. \end{gathered}$$

Conclusions. 1. We set the problem of finding the correct empirical formula for the number of arithmetic operations of the simplex method for LP problems given in the canonical form of medium complexity (the number of equations does not exceed 500, the number of variables does not exceed 600).

2. We give a justification for a residual representation, to which belongs the structure of the sought empirical formula for the number of arithmetic operations of the simplex method belongs.

3. We chose the MGMDH as a method for finding the empirical complexity formula. The method was adapted for the case when the residual representation is linear with respect to unknown coefficients and nonlinear with respect to given basis functions.

4. As a result of the algorithm, we found an empirical formula for the number of arithmetic operations of the simplex method for LP problems of medium dimensionality (the number of equations does not exceed 500, the number of variables does not exceed 600).

5. We provided a qualitative interpretation of the sense of the constant included in the resulting empirical formula for the number of arithmetic operations of the simplex method.

Declaration on the use of generative AI. The authors did not use any generative artificial intelligence tools in the preparation of this manuscript.

References

1. Pavlov A., Lishchuk K., Melnikov O., Kyselov M., Hu C. Mathematics and software for coordinated planning using aggregated linear volume-time models of discrete manufacturing systems. *International Journal of Information Technology and Computer Science (IJITCS).* 2025. Vol. 17. No. 4. P. 1–15. DOI: https://doi.org/10.5815/ijitcs.2025.04.01.
2. Pavlov A. A., Kushch A. V. Algorithms for constructing a regression linear with respect to unknown coefficients on a limited amount of experimental data. *Вісник Нац. техн. ун-ту «ХПІ»: зб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології. Харків: НТУ «ХПИ». 2025. № 2 (14). С. 8–15. DOI: https://doi.org/10.20998/2079-0023.2025.02.02.*
3. Dadush D., Huiberts S. Smoothed analysis of the simplex method. *Beyond the Worst-Case Analysis of Algorithms* / Ed. by T. Roughgarden, Cambridge University Press. 2021. P. 309–333. DOI: https://doi.org/10.1017/9781108637435.019.
4. Павлов О. А., Головченко М. М., Дрозд В. В., Шаргородський В. С. Підвищення ефективності модифікованого методу урахування аргументів для побудови багатовимірних регресій, заданих надлишковим описом. *Адаптивні системи автоматичного управління.* 2025. Т. 1, № 46. С. 257–266. DOI: https://doi.org/10.20535/1560-8956.46.2025.323833.
5. Marquardt D. W., Snee R. D. Ridge regression in practice. *The American Statistician.* 1975. Vol. 29, No. 1. P. 3–20. DOI: https://doi.org/10.1080/00031305.1975.10479105.
6. Ahsan M. M., Mahmud M. A. P., Saha P. K., Gupta K. D., Siddique Z. Effect of data scaling methods on machine learning algorithms and model performance. *Technologies*. 2021. Vol. 9, No. 3. P. 52. DOI: https://doi.org/10.3390/technologies9030052.
7. Moroz O. H., Stepashko V. S. Comparative features of MIA GMDH and deep feed-forward neural networks. *Cybernetics and Computer Engineering.* 2021. № 4 (206). P. 5–20. DOI: https://doi.org/10.15407/kvt206.04.005.
8. Adler I., Megiddo N. A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension. *Journal of the ACM (JACM).* 1985. Vol. 32. No. 4. P. 871–895. DOI: https://doi.org/10.1145/4221.4222.
9. Borgwardt K.-H. The average number of pivot steps required by the simplex method is polynomial. *Zeitschrift für Operations Research.* 1982. Vol. 26. No. 1. P. 157–177. DOI: https://doi.org/10.1007/bf01917108
10. Klee V., Minty G. J. How good is the simplex algorithm? *Inequalities-III.* New York, Academic Press. 1972. P. 159–175.
11. Spielman D. A., Teng S. H. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM).* 2004. Vol. 51. No. 3. P. 385–463. DOI: https://doi.org/10.1145/990308.990310.
12. Voulgaropoulou S., Samaras N., Ploskas N. Predicting the execution time of the primal and dual simplex algorithms using artificial neural networks. *Mathematics.* 2022. Vol. 10. P. 1038. DOI: https://doi.org/10.3390/math10071038.

References (transliterated)

1. Pavlov A., Lishchuk K., Melnikov O., Kyselov M., Hu C. Mathematics and software for coordinated planning using aggregated linear volume-time models of discrete manufacturing systems. *International Journal of Information Technology and Computer Science (IJITCS).* 2025, vol. 17, no. 4, pp. 1–15. DOI: https://doi.org/10.5815/ijitcs.2025.04.01.
2. Pavlov A. A., Kushch A. V. Algorithms for constructing a regression linear with respect to unknown coefficients on a limited amount of experimental data. *Visnyk Natsionalnoho tekhnichnoho universytetu "KhPI". Seriia: Systemnyi analiz, upravlinnia ta informatsiini tekhnolohii: zb. nauk. prats* [Bulletin of the National Technical

University "KhPI". Series: System analysis, control and information technology: Collection of scientific papers]. Kharkiv, NTU "KhPI" Publ., *2025, no. 2 (14), pp. 8–15. DOI: https://doi.org/10.20998/2079-0023.2025.02.02.*
3. Dadush D., Huiberts S. Smoothed analysis of the simplex method. *Beyond the Worst-Case Analysis of Algorithms* / Ed. by T. Roughgarden, Cambridge University Press Publ., 2021, pp. 309–333. DOI: https://doi.org/10.1017/9781108637435.019.
4. Pavlov A. A., Holovchenko M. N., Drozd V. V., Shargorodskyi V. S. Pidvyshchennia efektyvnosti modyfikovanoho metodu urakhuvannia arhumentiv dlia pobudovy bahatovymirnykh rehresii, zadanykh nadlyshkovym opysom [Improving efficiency of the modified group method of data handling for constructing multivariate regressions given by a redundant representation]. *Adaptyvni systemy avtomatychnoho upravlinnia* [Adaptive systems of automatic control]. 2025, vol. 1, no. 46, pp. 257–266. DOI: https://doi.org/10.20535/1560-8956.46.2025.323833. (In Ukr.).
5. Marquardt D. W., Snee R. D. Ridge regression in practice. *The American Statistician.* 1975, vol. 29, no. 1, pp. 3–20. DOI: https://doi.org/10.1080/00031305.1975.10479105.
6. Ahsan M. M., Mahmud M. A. P., Saha P. K., Gupta K. D., Siddique Z. Effect of data scaling methods on machine learning algorithms and model performance. *Technologies*. 2021, vol. 9, no. 3, p. 52. DOI: https://doi.org/10.3390/technologies9030052.
7. Moroz O. H., Stepashko V. S. Comparative features of MIA GMDH and deep feed-forward neural networks. *Cybernetics and Computer Engineering.* 2021, no. 4 (206). pp. 5–20. DOI: https://doi.org/10.15407/kvt206.04.005.
8. Adler I., Megiddo N. A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension. *Journal of the ACM (JACM).* 1985, vol. 32, no. 4, pp. 871–895. DOI: https://doi.org/10.1145/4221.4222.
9. Borgwardt K.-H. The average number of pivot steps required by the simplex method is polynomial. *Zeitschrift für Operations Research.* 1982, vol. 26, no. 1, pp. 157–177. DOI: https://doi.org/10.1007/bf01917108
10. Klee V., Minty G. J. How good is the simplex algorithm? *Inequalities-III.* New York, Academic Press Publ., 1972, pp. 159–175.
11. Spielman D. A., Teng S. H. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM).* 2004, vol. 51, no. 3, pp. 385–463. DOI: https://doi.org/10.1145/990308.990310.
12. Voulgaropoulou S., Samaras N., Ploskas N. Predicting the execution time of the primal and dual simplex algorithms using artificial neural networks. *Mathematics*. 2022, vol. 10, p. 1038. DOI: https://doi.org/10.3390/math10071038.

Received 25.03.2026
Accepted 17.04.2026
Published 20.05.2026

УДК 004:519.24:681.3.06

О. А. ПАВЛОВ, доктор технічних наук, професор, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, професор кафедри інформатики та програмної інженерії; e-mail: alexanderpavlov1944@gmail.com; ORCID: https://orcid.org/0000-0002-6524-6410
М. Є. КИСЕЛЬОВ, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, бакалавр кафедри інформатики та програмної інженерії; e-mail: zeusmobilenick@gmail.com; ORCID: https://orcid.org/0009-0005-3686-3419

ВЕРИФІКАЦІЯ ЕМПІРИЧНИХ МОДЕЛЕЙ СКЛАДНОСТІ СИМПЛЕКС-МЕТОДУ З ВИКОРИСТАННЯМ МОДИФІКОВАНОГО МГУА

Наведено результати побудови емпіричної залежності кількості арифметичних операцій стандартного симплекс-методу як функції від розмірності задачі, заданої в канонічній формі. Для розв'язання проблеми застосовано принцип самоорганізації моделей на основі модифікованого комбінаторного алгоритму методу групового урахування аргументів. Модифікація методу групового урахування аргументів полягає у формуванні залишкового опису, в якому припускається міститься шукана емпірична функція складності та використання алгоритму попереднього розбиття базисних функцій, з яких адитивно складається залишковий опис, на два класи, що не перетинаються: клас головних домінантних компонент та клас уточнюючих залишкових компонент, що дозволяє суттєво скоротити процедуру комбінаторного перебору. На основі відомих з літератури теоретичних оцінок складності сформовано надлишковий опис базисних функцій. До розгляду включено п’ять фундаментальних моделей: поліноміальну оцінку Боргвардта, квадратичну межу Адлера-Мегіддо, базову поліноміальну форму, модель згладженого аналізу Спілмана-Тенга та загальну змішану модель, раніше запропоновану авторами. З метою розширення простору пошуку, до базисних функцій було додано логарифмічну компоненту. Таким чином, залишковий опис адитивно містить шість базисних компонент та всі їх квадрати та попарні добутки. Відбір оптимальної структури здійснено за симетричним критерієм регулярності шляхом перебору понад 134 мільйонів альтернативних моделей, кількість яких однозначно задається потужністю множини уточнюючих залишкових компонент. Для забезпечення коректного використання методу найменших квадратів для суттєво нелінійних регресорів різного масштабу (абсолютні значення яких відрізняються на багато порядків), експериментально обґрунтовано необхідність масштабування вхідних даних на їхнє стандартне відхилення з необов’язковим застосуванням процедури центрування. Ефективність такого підходу підтверджено на спеціально розробленому експерименті із закладеним ідеальним розв’язком, наявністю вільного члена (зсуву) на рівні 100000 та нормальним шумом з амплітудою 1 % від середнього значення ідеальної регресії на значеннях вхідних змінних експерименту. Результати імітаційного моделювання на вибірці з 13690 випадково згенерованих задач лінійного програмування доводять беззаперечну перевагу змішаної поліноміально-логарифмічної структури у розширеному класі функцій-претендентів.

Ключові слова: симплекс-метод, самоорганізація моделей, модифікований метод групового урахування аргументів, метод найменших квадратів, надлишковий опис, критерій регулярності, стандартизація даних.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Павлов Олександр Анатолійович / Pavlov Alexander Anatolievich
Автор 2 / Author 2: Кисельов Микита Євгенович / Kyselov Mykyta Yevhenovych