# СИСТЕМНИЙ АНАЛІЗ І ТЕОРІЯ ПРИЙНЯТТЯ РІШЕНЬ

# SYSTEM ANALYSIS AND DECISION-MAKING THEORY

DOI: 10.20998/2079-0023.2024.02.01
UDC 004:519.24

***A. A. PAVLOV***, Doctor of Technical Sciences, Full Professor, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine, Professor of Informatics and Software Engineering Department; e-mail: pavlov.fiot@gmail.com; ORCID: https://orcid.org/0000-0002-6524-6410
***M. N. HOLOVCHENKO***, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine, Senior Lecturer of Informatics and Software Engineering Department; e-mail: ma4ete25@ukr.net; ORCID: https://orcid.org/0000-0002-9575-8046
***V. V. DROZD***, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Bachelor of Informatics and Software Engineering Department, Kyiv, Ukraine; e-mail: drozdllera@gmail.com; ORCID: https://orcid.org/0000-0003-0418-1139

## MODIFICATION OF THE DECOMPOSITION METHOD OF CONSTRUCTING MULTIVARIATE POLYNOMIAL REGRESSION WHICH IS LINEAR WITH RESPECT TO UNKNOWN COEFFICIENTS

The authors created a universal method of constructing multivariate polynomial regression given by a redundant representation. The method is synthetic, it organically combines a decomposition method and the modified group method of data handling. First, the decomposition method is implemented, it consists in the decomposition of the multivariate problem into a sequence of subproblems of constructing univariate polynomial regressions and the corresponding systems of linear equations, the variables of which are estimates for the nonlinear terms of the multivariate polynomial regression. Partial cases that guarantee the finding of estimates with a predetermined value of their variances were considered. The formal algorithm for constructing coefficient estimates for nonlinear terms of the multivariate polynomial regression stops working on the first coefficient whose estimation with a predetermined accuracy is not achieved under the specified limitations on the number of tests. The estimation of all coefficients that were not found by the decomposition method is done by a heuristic method, which is an efficient modification of the group method of data handling. The increase in the efficiency of the synthetic method is achieved primarily by finding such new theoretically substantiated algorithmic procedures (aggregated operators) of the decomposition method, which significantly, in comparison with its previous version, increases the number of coefficients for nonlinear terms of a multivariate polynomial regression that can be found in advance given accuracy. The authors showed that this effect is achieved due to new theoretical provisions used in the visual analysis of the structure of the multivariate polynomial regression given by the redundant representation by a professional user. The given illustrative example facilitates the use of the presented results when solving practical problems.

**Keywords:** regression analysis, multivariate polynomial regression, redundant representation, decomposition method, individual algorithm, least squares method.

**1. Introduction.** Construction of a multivariate regression based on the results of an active experiment is still theoretically and practically relevant as shown, in particular, in [1–15]. As noted in the abstract, the authors proposed a universal synthetic method for constructing a multivariate polynomial regression (MPR) given by a redundant representation. The originality and efficiency of the method were protected by relevant publications. In this paper, we propose to qualitatively increase its efficiency by expanding the theoretical and practical capabilities of the decomposition method, which is a part of the synthetic method.

***1.1. The problem statement.*** An MPR is given by the following redundant representation [14]:

$$Y(\bar{x}) = \sum_{\forall(i_1,\dots,i_t)\in K,\ \forall(j_1,\dots,j_t)\in K(i_1,\dots,i_t)} b_{i_1\dots i_t}^{j_1\dots j_t} \left(x_{i_1}\right)^{j_1}\cdots\left(x_{i_t}\right)^{j_t} + E, \quad (1)$$

where $\bar{x} = \left(x_1,\dots,x_m\right)^{\mathrm{T}}$ is a deterministic vector of input variables, $x_i \in [c_i, d_i]$, $0 \le c_i < d_i$ (this is the strictest condition for finding estimates with a given accuracy), $E$ is a random variable, its mathematical expectation is $ME = 0$, its variance $\mathrm{Var}(E) = \sigma^2 < \infty$. $\sigma^2$ is known or its efficient upper estimate is known. The values of the coefficients $b_{i_1\dots i_t}^{j_1\dots j_t}$ are unknown ($b_0^0$ is an unknown constant).

We need to estimate the unknown coefficients at nonlinear terms of the MPR according to the results of an active experiment $\left(x_i \to y_i,\ i = \overline{1,n}\right)$.

***1.2. Basic provisions of the decomposition method [1].*** The decomposition method [14] implements a methodology of reducing the finding of estimates for nonlinear terms of the MPR (1) to the sequential construction of univariate polynomial regressions (UPR) and the solution of the corresponding systems of linear equations. Variables in these equations are estimates for nonlinear terms of the MPR (1). The general formal algorithmic procedure for obtaining estimates of all coefficients for nonlinear terms of the MPR



**Research Article:** This article was published by the publishing house of ***NTU "KhPI"*** in the collection "Bulletin of the National Technical University "KhPI" Series: System analysis, management and information technologies." This article is distributed under a Creative Common Creative Common Attribution (CC BY 4.0). **Conflict of Interest:** The author/s declared no conflict of interest.

(1) consists in the sequential execution of the first and then the second subalgorithm. The first subalgorithm is implemented step-by-step for all members of the MPR, each of which contains at least one scalar variable of power greater than or equal to two.

*Remark 1*. The procedure for choosing nonlinear terms and choosing for each of them a single coefficient of the form $\left(x_{i_p}\right)^{j_p}$, $j_p \geq 2$, is substantiated in [1].

At each step of the first subalgorithm, the scalar variable $x_{i_p}$ is replaced by a virtual scalar variable $z$:

$$x_{i_p} = a_{i_p} z + b_{i_p}. \quad (2)$$

The coefficients $a_{i_p}$, $b_{i_p}$ are found according to [1]. The values $x_{i_p,i}$, $i = \overline{1,n}$, of the real input variable $x_{i_p}$ are found according to (2) by the given in advance values of the virtual variable $z$. Other input variables take fixed values. The results of the active experiment on the obtained real univariate regression are transformed as the result of the virtual active experiment on the virtual UPR relative to the virtual scalar input variable $z$.

*Remark 2*. It is substantiated in [1] how many virtual UPRs are built at each step of the first subalgorithm. Their number determines the dimension of the systems of linear equations, which variables are the estimates of the coefficients at the nonlinear terms of the MPR (1).

*Remark 3*. It is shown in [1] that the reduction of real UPRs to virtual ones in the case when a system of linear equations has only one variable allows one to efficiently find the variance of the estimate for the nonlinear term of the MPR (1) before conducting a real experiment and to find the necessary number of repetitions of the main experiment to obtain the variance of the given value. If this quantity is excessive, the corresponding coefficient cannot be found with the specified accuracy by the decomposition method.

The steps of the second subalgorithm [14] relate to finding nonlinear terms of the MPR of the form $b_{i_1 \ldots i_t}^{1 \ldots 1}$. For this case it is proposed in [1] to replace all input variables included in this term by the virtual variable $z$, other input variables take fixed values in the experiments.

It was proved in [1] that the implementation of the first and second subalgorithms formally allow obtaining estimates of the coefficients for all nonlinear members of the MPR (1).

***1.3. The main drawback of the formal algorithmic procedure of the decomposition method.*** The theoretical substantiation of the decomposition method guarantees that the successive use of the first and second subalgorithms at the preliminary stage of the problem analysis leads to the guaranteed obtaining of estimates of all coefficients at nonlinear terms of the MPR. But with the limited number of experiments given in advance, the specified accuracy (by the variance value) is not guaranteed. When finding the estimates of the following coefficients, the estimates of previously found coefficients at nonlinear terms of the MPR are used. Thus, the result obtained may be unsatisfactory in general. Therefore, when solving practical problems, the algorithm of the decomposition method stops working on the first nonlinear term of the MPR that has the estimate of the coefficient significantly different from the given one by variance at the stage of the preliminary analysis.

**2. Methodology of qualitative improvement of the efficiency of the decomposition method.** The methodology of qualitative improvement of the efficiency of the decomposition method is that a qualified user, analyzing the structure of the MPR, the coefficients of the nonlinear terms of which have to be estimated (by the variance value) with a given accuracy, using extended theoretical possibilities of the decomposition method (see Section 3), finds by himself the step by step sequence of the algorithm at the stage of the preliminary analysis of the problem solution (before conducting real experiments). That is, in what sequence will the active experiment be implemented to construct the found number of UPRs for a user-defined sequence of nonlinear terms of the MPR, which guarantees at each step finding the coefficient estimates with a given accuracy. Conducting an active experiment for each UPR in the found order after the preliminary analysis stage is used to build the estimates of the coefficients at the highest degree of each UPR. This sequence of steps can be obtained by a formal procedure only by enumerating all possible options, which is obviously unconstructive.

**3. Expanding the theoretical possibilities of the decomposition method for the efficient construction of an individual algorithm for the preliminary analysis of the problem solution.** Section 3 describes four different theoretically substantiated methods (aggregated operators) of constructing an UPR for estimating the coefficients at nonlinear terms of the MPR at the stage of preliminary analysis of the individual algorithm of the decomposition method. The sequence of their use is set by the user as a result of their analysis of the MPR structure. As a result of the implementation of the previous stage, the individual algorithm (the second stage) becomes fully formalized, namely, the sequence of active experiments to be conducted (with known values of the input variables) is specified, the result of each of the experiments is the estimate of the coefficient at the highest degree of a specially designed UPR, which allows to estimate with a specified accuracy the coefficient at the corresponding nonlinear term of the MPR.

***3.1. The first human-computer algorithmic procedure (the first aggregated operator (AO)).*** The user selects the nonlinear term of the MPR

$$b_{i_1 \ldots i_t}^{j_1 \ldots j_t} \left(x_{i_1}\right)^{j_1} \cdots \left(x_{i_k}\right)^{j_k} \cdots \left(x_{i_t}\right)^{j_t}, \quad (3)$$

that contains his chosen input variable $x_{i_k}$ in the power of $j_k \geq 2$. Algorithmically, the verification of the following condition is carried out: there should not be any nonlinear term of the MPR containing $\left(x_{i_k}\right)^{j_l}$ $\forall j_l \geq j_k$. If such terms are present, the coefficients for them with the specified accuracy will be found as a result of the implementation of an active experiment in the previous steps of the second stage of the individual algorithm. In the main active experiment, the variable $x_{i_k}$ will take the value $x_{i_k,i} = a_{i_k} z_i + b_{i_k}$, $i = \overline{1,n}$, where the coefficients $a_{i_k}$, $b_{i_k}$, and the value $z_i$,

$i=\overline{1,n}$, of the input virtual scalar variable $z$ of the specially constructed virtual UPR [1] are given by formulas (7), (8) [1]. Other input variables $x_i$, $i=\overline{1,m}$, $i\neq i_k$, take fixed values in the active experiment, and the input variables included in (3) take maximally large modulo values (all others take minimally small modulo values). At the same time, according to [1], the statistical data of the results of the virtual active experiment are given in the form

$$\left(z_i \to \overline{y}_i, \overline{y}_i = y_i - \sum b_{l_1\dots l_{l_l}}^{j_{l_1}\cdots j_{l_{l_l}}} \left(x_{l_1,i}\right)^{j_{l_1}} \cdots \left(x_{l_{l_l},i}\right)^{j_{l_{l_l}}}\right), \quad (4)$$

where the coefficients $b_{l_1\dots l_{l_l}}^{j_{l_1}\cdots j_{l_{l_l}}}$ will be found with the specified accuracy in the previous steps of the second stage of the individual algorithm. In this case, according to [1], the variance of the estimate of the coefficient $b_{i_1\dots i_t}^{j_1\cdots j_t}$ (see (3)) has the form

$$Db_{i_1\dots i_t}^{j_1\cdots j_t} = \frac{1}{\left(a_{i_k}\right)^{2j_k}} \cdot \frac{1}{\prod\limits_{l=1,\ l\neq i_k}^{t} \left(x_{i_l,\Phi}\right)^{2j_l}} \cdot D\hat{\gamma}_{j_k}, \quad (5)$$

where, according to [1], $\hat{\gamma}_{j_k}$ is an estimate of the coefficient at $z$ of the maximum power $j_k$ of the virtual UPR (for simplicity, both the random variable and its realization are denoted by $\hat{\gamma}_{j_k}$).

As shown in [1], for the virtual variable $z$, which in the virtual active experiment takes the value $z_1 < z_2 < \dots < z_n$, where $\forall i\ z_i - z_{i-1} = \text{const}$, $n=10$, $z_1=-50$, $z_{10}=50$, the variances of the estimates of the coefficients of the virtual UPR are the following:

$$D\hat{\gamma}_2 = 1.7\cdot 10^{-6}\sigma^2,\ D\hat{\gamma}_3 = 4.7\cdot 10^{-9}\sigma^2,$$

$$D\hat{\gamma}_4 = 2.6\cdot 10^{-13}\sigma^2,\ D\hat{\gamma}_5 = 4.5\cdot 10^{-16}\sigma^2,$$

that is, the variance of each subsequent term decreases by three orders of magnitude. The analysis of expression (5) allows at the previous stage of the individual algorithm to set the values of the input variables in the active experiment (considering possible repetitions of the main active experiment [1]). This allows us to obtain an estimate of the coefficient (3) with the given accuracy or conclude that the coefficient (3) cannot be obtained with the given accuracy by the individual algorithm of the decomposition method.

***3.2. The second human-computer algorithmic procedure (the second AO).*** The user selects the nonlinear term of the MPR

$$b_{i_1\dots i_t}^{j_1\cdots j_t} \left(x_{i_1}\right)^{j_1} \cdots \left(x_{i_k}\right)^{j_k} \cdots \left(x_{i_t}\right)^{j_t}, \quad (6)$$

sets the input variables $x_{i_{l_1}}, x_{i_{l_2}}, \dots, x_{i_{l_l}}$, which in the main experiment will take values

$$\forall l\ x_{i_{l_l},i} = a_{i_{l_l}} z_i + b_{i_{l_l}}, i=\overline{1,n}, \quad (7)$$

where, as in Subsection 3.1, coefficients $a_{i_{l_l}}$, $b_{i_{l_l}}$ are set according to formulas (7), (8) [1]. Other input variables in the active experiment take fixed values, as in Subsection 3.1. Algorithmically, the following condition is checked: there should not be any nonlinear term of the MPR containing input variables $x_{i_{l_1}}, \dots, x_{i_{l_l}}$ or their subset, the total degree of which is greater than or equal to $\sum_{i=1}^{l} j_{t_i}$. If such terms exist, then their coefficients must be found with the specified accuracy in the previous steps of the second stage of the individual algorithm of the decomposition method.

If this condition is fulfilled, the description of the sequence of steps repeats Subsection 3.1 completely. In this case, according to [1], we obtain the variance of the coefficient estimate $b_{i_1\dots i_t}^{j_1\cdots j_t}$ at the stage of preliminary analysis:

$$D\hat{b}_{i_1\dots i_t}^{j_1\cdots j_t} = \frac{1}{\prod\limits_{i=1}^{l}\left(a_{t_i}\right)^{2j_{t_i}}} \cdot \frac{1}{\prod\limits_{\substack{m=1\\ i_m\neq t_1,\dots,t_l}}^{t} \left(x_{i_m,\Phi}\right)^{2j_m}} \cdot D\hat{\gamma}_{\sum\limits_{i=1}^{l} j_{t_i}}. \quad (8)$$

As well as in Subsection 3.1, the analysis of expression (8) allows at the preliminary stage of the individual algorithm of the decomposition method to set the input values of the active experiment (considering possible repetitions of the main experiment [1]). This allows us to obtain an estimate of (8) with the given accuracy or to conclude about the impossibility of implementing such an active experiment.

***3.3. The third human-computer algorithmic procedure (the third AO).*** The user selects the nonlinear term of the MPR

$$b_{i_1\dots i_t}^{1\dots 1} x_{i_1} \cdot x_{i_2} \cdots x_{i_t}. \quad (9)$$

In the main active experiment, the input variables will take the values $x_{i_l,i} = a_{i_l} z_i + b_{i_l}, l=\overline{1,t}, i=\overline{1,n}$; $a_{i_l}$, $b_{i_l}$, $z_i$ are set as in Subsection 3.1. Other input variables in the active experiment take the minimum possible values. Algorithmically, the verification of the following condition is carried out: there should not be any nonlinear term of the MPR containing input variables $x_{i_l}$, $l=\overline{1,t}$, or their subset, the total degree of which is greater than or equal to $t$. If such terms exist, then their coefficients with the specified accuracy must be found in the previous steps of the second stage of the individual algorithm.

If this condition is fulfilled, similarly to Subsections 3.1, 3.2, we obtain the variance of the coefficient estimate of $b_{i_1\dots i_t}^{1\dots 1}$ (8) at the preliminary stage of the individual algorithm:

$$D\hat{b}_{i_1\dots i_t}^{1\dots 1} = \frac{1}{\prod\limits_{l=1}^{t}\left(a_{i_l}\right)^2} \cdot D\hat{\gamma}_t. \quad (10)$$

*Remark 4.* Implementation of Subsection 3.3 is real if the values of all positive coefficients $a_{i_l}$ [1] are close to or exceed one. For example, if in [1] $z_{10} - z_1 = 100$, then the

values of $d_i - c_i$, $i = \overline{1,10}$, should be close to or greater than 100.

***3.4. The fourth human-computer algorithmic procedure (the fourth AO).*** The user selects the nonlinear term of the MPR

$$b_{i_1\dots i_t}^{1\dots 1} x_{i_1} \cdot x_{i_2} \cdots x_{i_t}. \qquad (11)$$

A subset of its input variables $i_{t_1},\dots,i_{t_m}$, $m < t$, $m \geq 2$, in the main active experiment will take the value $x_{i_{t_l},i} = a_{i_{t_l}} z_i + b_{i_{t_l}}$, $l = \overline{1,m}$, $m < t$, $i = \overline{1,n}$, as in the previous Subsections 3.1–3.3, according to [1]. In the main active experiment, the other input terms of (11) take maximally large values, and the rest take minimally possible values. The fulfillment of the following condition is checked: there is no nonlinear term of the MPR containing input variables $x_{i_{t_1}},\dots,x_{i_{t_m}}$ or their subset, the total degree of which is greater than or equal to $m$. If there are such terms, then their estimates will be found with the given accuracy in the previous steps of the second stage of the individual algorithm.

When this condition is met, the variance of the estimate of the coefficient (11) is equal to

$$D\hat{b}_{i_1\dots i_t}^{1\dots 1} = \frac{1}{\prod\limits_{l=1}^{m}\left(a_{i_{t_l}}\right)^2} \cdot \frac{1}{\prod\limits_{\substack{l=1 \\ \forall j_l \notin \{i_{t_1},\dots,i_{t_m}\}}}^{t}\left(x_{j_l,\phi}\right)^2} \cdot D\hat{\gamma}_m. \qquad (12)$$

*Remark 5*. According to [1], the estimate of the coefficient at the corresponding term of the virtual UPR according to the results of the main virtual active experiment is given by the expression

$$\hat{\gamma}_j = \sum_{i=1}^{n} \bar{y}_i Q_j(z_i), \qquad (13)$$

where $Q_j(z)$ are normalized orthogonal polynomials of Forsythe (NOPF) found by the values $z_i$, $i = \overline{1,n}$, of the virtual input variable $z$. Thus, the estimates of coefficients, based on the results of real active experiments, are found using only a single set of NOPFs, the coefficients of which are found in advance with a given accuracy.

*Remark 6*. Estimates of the coefficients for nonlinear terms of the MPR are found by the formulas:

a) for the first human-computer algorithmic procedure (Subsection 3.1):

$$\hat{b}_{i_1\dots i_t}^{j_1\dots j_t} = \frac{1}{\left(a_{i_k}\right)^{j_k} \cdot \prod\limits_{l=1,\, l\neq i_k}^{t}\left(x_{i_l,\phi}\right)^{j_l}} \cdot \hat{\gamma}_{j_k}; \qquad (14)$$

b) for the second human-computer algorithmic procedure (Subsection 3.2):

$$\hat{b}_{i_1\dots i_t}^{j_1\dots j_t} = \frac{1}{\prod\limits_{i=1}^{l}\left(a_{t_i}\right)^{j_{t_i}} \cdot \prod\limits_{\substack{m=1 \\ \forall i_m \neq t_1,\dots,t_l}}^{t}\left(x_{i_m,\phi}\right)^{j_m}} \cdot \hat{\gamma}_{\sum\limits_{i=1}^{l} j_{t_i}}; \qquad (15)$$

c) for the third human-computer algorithmic procedure (Subsection 3.3):

$$\hat{b}_{i_1\dots i_t}^{1\dots 1} = \frac{1}{\prod\limits_{l=1}^{t} a_{i_l}} \cdot \hat{\gamma}_t; \qquad (16)$$

d) for the fourth human-computer algorithmic procedure (Subsection 3.4):

$$\hat{b}_{i_1\dots i_t}^{1\dots 1} = \frac{1}{\prod\limits_{l=1}^{m} a_{i_{t_l}} \cdot \prod\limits_{\substack{l=1 \\ \forall j_l \notin \{i_{t_1},\dots,i_{t_m}\}}}^{t} x_{j_l,\phi}} \cdot \hat{\gamma}_m. \qquad (17)$$

Formulas (14)–(17) are a consequence of the equation expressing the coefficient at the maximum term of the virtual UPR through its corresponding coefficient $b_{i_1\dots i_t}^{j_1\dots j_t}$. The corollaries of these equations are also the formulas for the variances of the coefficients (5), (8), (10), (12).

*Remark 7*. If there are input variables, the range of admissible values of which contains zero, then using the possibility to fix a zero value for an input variable in an active experiment can increase the number of coefficients for nonlinear terms of the MPR, the estimate of which is the solution of a linear equation with a single unknown (see the illustrative example in Section 4).

*Remark 8*. The redundant representation of the MPR may contain input variables of a power greater than the maximum degree of the NOPFs, the coefficients of which are found in advance with a given accuracy. Then, the values of such variables should be fixed in all active experiments.

**4. Illustrative example.** Let us set a redundant representation of the MPR in the form:

$$Y(x_1,x_2,x_3,x_4,x_5,x_6) = b_0 + b_1 x_1^2 x_3^3 x_2 + b_2 x_1^2 x_2^2 x_4 +$$
$$+ b_3 x_2^2 x_1 x_4 + b_4 x_2^2 x_4^2 x_1 + b_5 x_3^3 x_4^2 x_5 + b_6 x_3^3 x_1 x_2 x_6 +$$
$$+ b_7 x_5^6 x_4^3 x_1^2 + b_8 x_4^3 x_6^2 x_1 + b_9 x_1 x_2 x_3 x_4 x_6 + E, \qquad (18)$$

where $ME = 0$, $DE = 4$, $E$ distributed according to the normal law.

The true values of the coefficients are as follows: $b_0 = 1$, $b_1 = 2$, $b_2 = 0$, $b_3 = 1$, $b_4 = 0$, $b_5 = 2$, $b_6 = 1$, $b_7 = 2$, $b_8 = 1$, $b_9 = 2$.

Thus, the true representation of the MPR lacks the terms with coefficients $b_2, b_4$.

Areas in which input variables can take values are as follows: $x_1 \in [1,10]$, $x_2 \in [1,10]$, $x_3 \in [1,10]$, $x_4 \in [1,10]$, $x_5 \in [0,5]$, $x_6 \in [1,5]$.

*Remark 9*. The values of the input variables, which take different values in the simulation of a real active experiment at each step, are set according to formulas (7), (8) [1]:

$$a_j = \frac{d_j - c_j}{z_{10} - z_1};\ b_j = c_j - \frac{d_j - c_j}{z_{10} - z_1} z_1;$$

$$x_{ji} = a_j z_i + b_j,\ i = \overline{1,10}.$$

*Remark 10*. The MPR (18) does not contain linear components because, according to the theoretical provisions of the decomposition method, their estimate is inefficient and is found using the modified group method of data handling [1], and the presence of a linear part in an MPR does not affect the accuracy of coefficient estimates for nonlinear terms of an MPR.

*Remark 11*. The maximum degree of a NOPF is equal to five.

The user visually sets the sequence of steps of the individual algorithm at the preliminary stage of the analysis: the first step for the coefficient $b_1$ (the second AO, Subsection 3.2), the second step for $b_2$ (the second AO), the third step for $b_4$ (the second AO), the fourth step for $b_3$ (the first AO, Subsection 3.1), the fifth step for $b_6$ (the first AO), the sixth step for $b_5$ (the first AO), the seventh step for $b_7$ (the second AO), the eighth step for $b_8$ (the second AO), the ninth step for $b_9$ (the fourth AO, Subsection 3.4). The results of the calculations are given in tables 1–4.

*Remark 12*. Due to the limited size of this article, the values of the input and output variables in the simulation of

Table 1 – Realizations of the random variable $E$ used according to steps 1–9 in the simulation of the main active experiment

| Step 1 | Step 2 | Step 3 |
|---|---|---|
| –1.0659694104455328 | –2.195369811977885 | 0.807950141301234 |
| 3.7710681456850645 | –0.12063419142686603 | 2.6127415396270233 |
| –0.7377368040485024 | –1.7907751802465064 | 1.9923563564870557 |
| –0.0769108594611502 | –2.1026649535817645 | –0.005506315074662242 |
| 1.659240003527758 | –1.5447658191004847 | 1.2290579479034371 |
| 0.6385687090711256 | –0.2002561262551154 | 0.1419919936945139 |
| 1.393884665512836 | 2.6798399955112573 | 0.7288220117347707 |
| –0.9162780984774989 | –2.7977613703718434 | –0.5038414467160576 |
| –4.381571522845034 | 2.4754704037663915 | –0.24207886186565378 |
| –5.414570636039612 | 1.056266088850859 | 1.4330913734546347 |
| **Step 4** | **Step 5** | **Step 6** |
| –0.6893135072283657 | –0.6893135072283657 | 1.540866648948249 |
| –1.1931899114947138 | –1.1931899114947138 | 1.869030207418431 |
| 3.0249102665371534 | 3.0249102665371534 | –2.4896342020289994 |
| 0.5270108534316472 | 0.5270108534316472 | –1.3621035826330083 |
| 1.7601025580636376 | 1.7601025580636376 | –1.3059043929109533 |
| 2.666542829364775 | 2.666542829364775 | 2.3330867000197006 |
| 3.0458456722762604 | 3.0458456722762604 | 0.12292969415601586 |
| 2.130621053328643 | 2.130621053328643 | 1.6969924752545953 |
| –2.6190016516637336 | –2.6190016516637336 | 0.5081198977551179 |
| –0.2229956051703742 | –0.2229956051703742 | 1.0298668767026709 |
| **Step 7** | **Step 8** | **Step 9** |
| –1.0904479093952715 | –1.789867368121218 | –1.789867368121218 |
| 0.9699664610876225 | 1.6833116410256515 | 1.6833116410256515 |
| 3.2737368583418776 | –0.8595534529753378 | –0.8595534529753378 |
| 0.6159984517967693 | –0.7063717798945163 | –0.7063717798945163 |
| –1.4712753850010287 | –0.5767283796110184 | –0.5767283796110184 |
| 1.7289268018974024 | –0.8432436883496297 | –0.8432436883496297 |
| –1.2971084475562134 | –1.2037914210764402 | –1.2037914210764402 |
| –0.7872578872282519 | 1.1739094496264404 | 1.1739094496264404 |
| 0.7516094848925081 | –3.719826795861956 | –3.719826795861956 |
| 0.3787300161358095 | 2.5730133007139497 | 2.5730133007139497 |

Table 2 – Found coefficients of the NOPFs

| | $Q_0$ | $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ | $Q_5$ |
|---|---|---|---|---|---|---|
| $q_{j0}$ | 0.31622776601683794 | 0 | −0.35903046525330308508 | 0 | 0.37601029391503510256 | 0 |
| $q_{j1}$ | | 0.0099087567581727680483 | 0 | −0.023723738974404719567 | 0 | 0.04073337742261610065 |
| $q_{j2}$ | | | 0.00035250853500849863344 | 0 | −0.0012936966199861974775 | 0 |
| $q_{j3}$ | | | | 0.000013116948944919936589 | 0 | −0.000067433562532976568096 |
| $q_{j4}$ | | | | | $5.1116893449436679076 \cdot 10^{-7}$ | 0 |
| $q_{j5}$ | | | | | | $2.1143486041566852428 \cdot 10^{-8}$ |

Table 3 – Calculation results for the illustrative example

| Step number | The maximum degree of virtual UPR | The coefficient variance at the maximum degree of virtual UPR | Estimates of the corresponding coefficients of the virtual UPR |
|---|---|---|---|
| 1 | 5 | $D\hat{\gamma}_5 = 4 \cdot 4.5 \cdot 10^{-16}$ | $\hat{\gamma}_5 = 0.000118172133464$ |
| 2 | 4 | $D\hat{\gamma}_4 = 4 \cdot 2.6 \cdot 10^{-13}$ | $\hat{\gamma}_4 = -0.00000022924202$ |
| 3 | 4 | $D\hat{\gamma}_4 = 4 \cdot 2.6 \cdot 10^{-13}$ | $\hat{\gamma}_4 = -0.0000000981927$ |
| 4 | 2 | $D\hat{\gamma}_2 = 4 \cdot 1.7 \cdot 10^{-6}$ | $\hat{\gamma}_2 = 0.80686953376847$ |
| 5 | 3 | $D\hat{\gamma}_3 = 4 \cdot 4.7 \cdot 10^{-9}$ | $\hat{\gamma}_3 = 0.36436409237099$ |
| 6 | 3 | $D\hat{\gamma}_3 = 4 \cdot 4.7 \cdot 10^{-9}$ | $\hat{\gamma}_3 = 0.72882533637497$ |
| 7 | 5 | $D\hat{\gamma}_5 = 4 \cdot 4.5 \cdot 10^{-16}$ | $\hat{\gamma}_5 = 0.1845281295144$ |
| 8 | 5 | $D\hat{\gamma}_5 = 4 \cdot 4.5 \cdot 10^{-16}$ | $\hat{\gamma}_5 = 0.0000117372678$ |
| 9 | 2 | $D\hat{\gamma}_2 = 4 \cdot 1.7 \cdot 10^{-6}$ | $\hat{\gamma}_2 = 8.09982106030674$ |

Table 4 – Calculation results for the illustrative example (continued)

| Step number, coefficient, input variables which do not take a fixed value in the active experiment | Equation for estimating the coefficient at the nonlinear term of the MPR | Estimate of the coefficient at the nonlinear term | The estimate variance |
|---|---|---|---|
| 1, $b_1$, $x_1$, $x_3$ | $\hat{b}_1 \cdot a_1^2 \cdot a_3^3 \cdot 10 = \hat{\gamma}_5$ | $\hat{b}_1 = 2.001$ | $5.128473 \cdot 10^{-7}$ |
| 2, $b_2$, $x_1$, $x_2$ | $\hat{b}_2 \cdot a_1^2 \cdot a_2^2 \cdot 10 = \hat{\gamma}_4$ | $\hat{b}_2 = -0.0003$ | $2.428001 \cdot 10^{-6}$ |
| 3, $b_4$, $x_2$, $x_4$ | $\hat{b}_4 \cdot a_2^2 \cdot a_4^2 \cdot 10 = \hat{\gamma}_4$ | $\hat{b}_4 = -0.0001$ | $2.428001 \cdot 10^{-6}$ |
| 4, $b_3$, $x_2$ | $\hat{b}_3 \cdot a_2^2 \cdot 10 \cdot 10 = \hat{\gamma}_2$ | $\hat{b}_3 = 0.9961$ | $1.096121 \cdot 10^{-5}$ |
| 5, $b_6$, $x_3$, $x_{5,\phi} = 0$ | $\hat{b}_6 \cdot a_3^3 \cdot 10 \cdot 10 \cdot 5 = \hat{\gamma}_3$ | $\hat{b}_6 = 0.9996$ | $1.420843 \cdot 10^{-7}$ |
| 6, $b_5$, $x_3$ | $\hat{b}_5 \cdot a_3^3 \cdot 10^2 \cdot 5 = \hat{\gamma}_3$ | $\hat{b}_5 = 1.9995$ | $1.420843 \cdot 10^{-7}$ |
| 7, $b_7$, $x_1$, $x_4$ | $\hat{b}_7 \cdot a_1^2 \cdot a_4^3 \cdot 5^6 = \hat{\gamma}_5$ | $\hat{b}_7 = 2.0000$ | $2.100622 \cdot 10^{-13}$ |
| 8, $b_8$, $x_4$, $x_6$ | $\hat{b}_8 \cdot a_6^2 \cdot a_4^3 \cdot 10 = \hat{\gamma}_5$ | $\hat{b}_8 = 1.0062$ | $1.314371 \cdot 10^{-5}$ |
| 9, $b_9$, $x_2$, $x_4$ | $\hat{b}_9 \cdot a_2 \cdot a_4 \cdot 10 \cdot 10 \cdot 5 = \hat{\gamma}_2$ | $\hat{b}_9 = 1.9999$ | $4.384485 \cdot 10^{-7}$ |

a real active experiment at each step are not presented. Below are the tables for NOPFs and intermediate results of calculations necessary to obtain estimates of coefficients $b_1, \ldots, b_9$.

The NOPFs were built for $z_i$, $i = \overline{1,10}$, $z_1 = -50$, $z_{10} = 50$, and their coefficients were found with an accuracy of up to 17 decimal places (Table 2).

Formulas for finding coefficient estimates of the virtual UPR:

$$\hat{\gamma}_j = \sum_{i=1}^{10} \overline{y}_i Q_j(z_i),\ j = \overline{0,5};$$

$$\hat{y}_j = \hat{w}_5 q_{5j} + \ldots + \hat{w}_j q_{jj},\ j = \overline{0,5},$$

where $\overline{y}_i$ is the value of the output variable of the virtual active experiment (given by formula (4)), $y_i$ is the value of the output variable of the corresponding real main experiment.

*Remark 13*. In all experiments, the minimum values of the fixed input variables are equal to one.

*Remark 14*. If the formal procedure of the first subalgorithm [1] were used to find coefficient estimates, then their estimates would be the solution of the corresponding linear systems of algebraic equations of the dimension $2 \times 2$. The coefficients at the variables of these equations depend on the choice of fixed values of the corresponding input variables from the area of their definitions. But, in contrast to equations with a single variable, there is currently no strategy for the efficient selection of their values to minimize the variance of estimates of unknown coefficients.

**Conclusions.**

1. The possibility of using the theoretical provisions of the decomposition method to estimate unknown coefficients at nonlinear terms of an MPR is being investigated in order to build an individual algorithm created by a user based on the analysis of the structure of the MPR given by a redundant representation.

2. Four aggregated operators are proposed, which allow solving the formulated problem. The first two generalized operators are used to estimate the coefficients at nonlinear members of the MPR which contain at least one variable to the power greater than or equal to two; the last two ones are for the case when all variables in the nonlinear term of the MPR are to the power of one.

3. An example is given that illustrates the efficiency of building an individual algorithm for estimating coefficients

at nonlinear terms of an MPR in comparison with the formal algorithmic procedure of the decomposition method.

**References**


1. Pavlov A. A., Holovchenko M. N., Drozd V. V. Efficiency substantiation for a synthetical method of constructing a multivariate polynomial regression given by a redundant representation. *Вісник Нац. техн. ун-ту «ХПІ»: зб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології*. Харків: НТУ «ХПІ», 2023. № 1 (9). С. 3–9. DOI: 10.20998/2079-0023.2023.01.01.
2. Yu L. Using negative binomial regression analysis to predict software faults: a study of Apache Ant. *International Journal of Information Technology and Computer Science (IJITCS)*. 2012. Vol. 4, no. 8. P. 63–70. DOI: 10.5815/ijitcs.2012.08.08.
3. Shahrel M.Z., Mutalib S., Abdul-Rahman S. PriceCop – price monitor and prediction using linear regression and LSVM-ABC methods for e-commerce platform. *International Journal of Information Engineering and Electronic Business (IJIEEB)*. 2021. Vol. 13, no. 1. P. 1–14. DOI: 10.5815/ijieeb.2021.01.01.
4. Satter A., Ibtehaz N. A regression based sensor data prediction technique to analyze data trustworthiness in cyber-physical system. *International Journal of Information Engineering and Electronic Business (IJIEEB).* 2018. Vol. 10, no. 3. P. 15–22. DOI: 10.5815/ ijieeb.2018.03.03.
5. Isabona J., Ojuh D. O. Machine learning based on kernel function controlled gaussian process regression method for in-depth extrapolative analysis of Covid-19 daily cases drift rates. *International Journal of Mathematical Sciences and Computing (IJMSC)*. 2021. Vol. 7, no. 2. P. 14–23. DOI: 10.5815/ ijmsc.2021.02.02.
6. Sinha P. Multivariate polynomial regression in data mining: methodology, problems and solutions. *International Journal of Scientific & Engineering Research*. 2013. Vol. 4, iss. 12. P. 962–965
7. Kalivas J. H. Interrelationships of multivariate regression methods using eigenvector basis sets. *Journal of Chemometrics*. 1999. Vol. 13 (2). P. 111–132. DOI: 10.1002/(SICI)1099-128X(199903/ 04)13:2<111::AID-CEM532>3.0.CO;2-N.
8. Ortiz-Herrero L., Maguregui M. I., Bartolomé L. Multivariate (O)PLS regression methods in forensic dating. *TrAC Trends in Analytical Chemistry*. 2021. Vol. 141. 116278. DOI: 10.1016/ j.trac.2021.116278.
9. Guo G., Niu G., Shi Q., Lin Q., Tian D., Duan Y. Multi-element quantitative analysis of soils by laser induced breakdown spectroscopy (LIBS) coupled with univariate and multivariate regression methods. *Analytical Methods*. 2019. Vol. 11, iss. 23. P. 3006–3013. DOI: 10.1039/C9AY00890J.
10. Настенко Е., Павлов В., Бойко Г., Носовец О. Многокритериальный алгоритм шаговой регрессии. *Біомедична інженерія і технологія*, 2020. № 3. С. 48–53. DOI: 10.20535/2617-8974.2020.3.195661.
11. Babatunde G., Emmanuel A. A., Oluwaseun O. R., Bunmi O. B., Precious A. E. Impact of climatic change on agricultural product yield using *k*-means and multiple linear regressions. *International Journal of Education and Management Engineering (IJEME).* 2019. Vol. 9, no. 3. P. 16–26. DOI: 10.5815/ijeme.2019.03.02.
12. Худсон Д. *Статистика для физиков: Лекции по теории вероятностей и элементарной статистике*. Москва: Мир, 1970. 296 с.
13. Pavlov A. A. Holovchenko M. N., Drozd V. V. Construction of a multivariate polynomial given by a redundant description in stochastic and deterministic formulations using an active experiment. *Вісник Нац. техн. ун-ту «ХПІ»: зб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології*. Харків: НТУ «ХПІ», 2022. № 1 (7). С. 3–8. DOI: 10.20998/2079-0023.2022.01.01.
14. Pavlov A., Holovchenko M., Mukha I., Lishchuk K., Drozd V. A Modified Method and an Architecture of a Software for a Multivariate Polynomial Regression Building Based on the Results of a Conditional Active Experiment. *Lecture Notes on Data Engineering and Communications Technologies.* 2023. Vol. 181. P. 207–222. DOI: 10.1007/978-3-031-36118-0_19
15. Pavlov A. A., Holovchenko M. N. Modified method of constructing a multivariate linear regression given by a redundant description. *Вісник Нац. техн. ун-ту «ХПІ»: зб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології*. Харків: НТУ «ХПІ», 2022. № 2 (8). С. 3–8. DOI: 10.20998/2079-0023.2022.02.01.


**References (transliterated)**


1. Pavlov A. A., Holovchenko M. N., Drozd V. V. Efficiency substantiation for a synthetical method of constructing a multivariate polynomial regression given by a redundant representation. *Visnyk Nats. tekhn. un-tu "KhPI": zb. nauk. pr. Temat. vyp.: Systemnyy analiz, upravlinnya ta informatsiyni tekhnologiyi* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkov, NTU "KhPI" Publ., 2022, no. 1 (9), P. 3–9. DOI: 10.20998/2079-0023.2023.01.01.
2. Yu L. Using negative binomial regression analysis to predict software faults: a study of Apache Ant. *International Journal of Information Technology and Computer Science (IJITCS)*. 2012, vol. 4, no. 8, P. 63–70. DOI: 10.5815/ijitcs.2012.08.08.
3. Shahrel M.Z., Mutalib S., Abdul-Rahman S. PriceCop – price monitor and prediction using linear regression and LSVM-ABC methods for e-commerce platform. *International Journal of Information Engineering and Electronic Business (IJIEEB)*. 2021, vol. 13, no. 1, P. 1–14. DOI: 10.5815/ijieeb.2021.01.01.
4. Satter A., Ibtehaz N. A regression based sensor data prediction technique to analyze data trustworthiness in cyber-physical system. *International Journal of Information Engineering and Electronic Business (IJIEEB).* 2018, vol. 10, no. 3, P. 15–22. DOI: 10.5815/ ijieeb.2018.03.03.
5. Isabona J., Ojuh D. O. Machine learning based on kernel function controlled gaussian process regression method for in-depth extrapolative analysis of Covid-19 daily cases drift rates. *International Journal of Mathematical Sciences and Computing (IJMSC)*. 2021, vol. 7, no. 2, P. 14–23. DOI: 10.5815/ ijmsc.2021.02.02.
6. Sinha P. Multivariate polynomial regression in data mining: methodology, problems and solutions. *International Journal of Scientific & Engineering Research*. 2013. vol. 4, iss. 12, P. 962–965
7. Kalivas J. H. Interrelationships of multivariate regression methods using eigenvector basis sets. *Journal of Chemometrics*. 1999, Vol. 13 (2), P. 111–132. DOI: 10.1002/(SICI)1099-128X(199903/ 04)13:2<111::AID-CEM532>3.0.CO;2-N.
8. Ortiz-Herrero L., Maguregui M. I., Bartolomé L. Multivariate (O)PLS regression methods in forensic dating. *TrAC Trends in Analytical Chemistry*. 2021, vol. 141, 116278. DOI: 10.1016/ j.trac.2021.116278.
9. Guo G., Niu G., Shi Q. et al. Multi-element quantitative analysis of soils by laser induced breakdown spectroscopy (LIBS) coupled with univariate and multivariate regression methods. *Analytical Methods*. 2019, vol. 11, iss. 23, P. 3006–3013. DOI: 10.1039/C9AY00890J.
10. Nastenko E., Pavlov V., Boyko G., Nosovets O. Mnogokriterial'nyj algoritm shagovoj regressii. *Biomedychna inzheneriya i tekhnolohiya* [Biomedical ingeneering and technology]. 2020, no. 3, P. 48–53. DOI: 10.20535/2617-8974.2020.3.195661.
11. Babatunde G., Emmanuel A. A., Oluwaseun O. R., Bunmi O. B., Precious A. E. Impact of climatic change on agricultural product yield using *k*-means and multiple linear regressions. *International Journal of Education and Management Engineering (IJEME).* 2019, vol. 9, no. 3, P. 16–26. DOI: 10.5815/ijeme.2019.03.02.
12. Hudson D. J. Statistics Lectures, Volume 2: Maximum Likelihood and Least Squares Theory. CERN Reports 64(18). Geneva, CERN, 1964. (Russ. ed.: Hudson D. *Statistika dlja fizikov: Lekcii po teorii verojatnostej i jelementarnoj statistike*. Moscow, Mir Publ., 1970. 296 p.). DOI: 10.5170/CERN-1964-018.
13. Pavlov A. A. Holovchenko M. N., Drozd V. V. Construction of a multivariate polynomial given by a redundant description in stochastic and deterministic formulations using an active experiment. *Visnyk Nats. tekhn. un-tu "KhPI": zb. nauk. pr. Temat. vyp.: Systemnyy analiz, upravlinnya ta informatsiyni tekhnologiyi* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkov, NTU "KhPI" Publ., 2022, no. 1 (7), P. 3–8. DOI: 10.20998/2079-0023.2022.01.01.
14. Pavlov A., Holovchenko M., Mukha I. et al. A Modified Method and an Architecture of a Software for a Multivariate Polynomial Regression Building Based on the Results of a Conditional Active Experiment. *Lecture Notes on Data Engineering and*

*Communications Technologies.* 2023, vol. 181, P. 207–222. DOI: 10.1007/978-3-031-36118-0_19

15. Pavlov A. A., Holovchenko M. N. Modified method of constructing a multivariate linear regression given by a redundant description. *Visnyk Nats. tekhn. un-tu "KhPI": zb. nauk. pr. Temat. vyp.: Systemnyy analiz, upravlinnya ta informatsiyni tekhnologiyi* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkov, NTU "KhPI" Publ., 2022, no. 2 (8), P. 3–8. DOI: 10.20998/2079-0023.2022.02.01.

*Received 01.11.2024*

УДК 004:519.24

***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: pavlov.fiot@gmail.com; ORCID: https://orcid.org/0000-0002-6524-6410

***М. М. ГОЛОВЧЕНКО***, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, старший викладач кафедри інформатики та програмної інженерії; e-mail: ma4ete25@ukr.net; ORCID: https://orcid.org/0000-0002-9575-8046

***В. В. ДРОЗД***, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, бакалавр кафедри інформатики та програмної інженерії; e-mail: drozdllera@gmail.com; ORCID: https://orcid.org/0000-0003-0418-1139

## МОДИФІКАЦІЯ ДЕКОМПОЗИЦІЙНОГО МЕТОДУ ПОБУДОВИ БАГАТОВИМІРНОЇ ПОЛІНОМІАЛЬНОЇ РЕГРЕСІЇ, ЛІНІЙНОЇ ВІДНОСНО НЕВІДОМИХ КОЕФІЦІЄНТІВ

Авторами був створений універсальний метод побудови багатовимірної поліноміальної регресії, заданої надлишковим описом. Метод є синтетичним, що органічно поєднує декомпозиційний метод та модифікований метод групового урахування аргументів. Спочатку реалізується декомпозиційний метод, що полягає в декомпозиції багатовимірної задачі на послідовність підзадач побудови одновимірних поліноміальних регресій та відповідних систем лінійних рівнянь, змінними яких є оцінки при нелінійних членах багатовимірної поліноміальної регресії. Розглянуті часткові випадки, що гарантують знаходження оцінок з наперед заданою величиною їх дисперсій. Формальний алгоритм побудови оцінок коефіцієнтів при нелінійних членах багатовимірної поліноміальної регресії припиняє роботу на першому коефіцієнті, оцінка якого з наперед заданою точністю не досягається при заданих обмеженнях на кількість випробувань. Оцінка всіх коефіцієнтів, що не були знайдені декомпозиційним методом, знаходиться евристичним методом, що є ефективною модифікацією метода групового урахування аргументів. Підвищення ефективності синтетичного методу досягається в першу чергу за рахунок знаходження таких нових теоретично обгрунтованих алгоритмічних процедур (агрегованих операторів) декомпозиційного методу, що суттєво, в порівнянні з його попередньою версією, збільшує кількість коефіцієнтів при нелінійних членах багатовимірної поліноміальної регресії, що можуть бути знайдені з наперед заданою точністю. Автори показали, що цей ефект досягається за рахунок нових теоретичних положень, що використовуються при візуальному аналізі професійним користувачем структури багатовимірної поліноміальної регресії, заданої надлишковим описом. Наведений ілюстративний приклад полегшує використання приведених результатів при розв'язанні практичних задач.

**Ключові слова:** регресійний аналіз, багатовимірна поліноміальна регресія, надлишковий опис, декомпозиційний метод, індивідуальний алгоритм, метод найменших квадратів.

*Повні імена авторів / Author's full names*

**Автор 1 / Author 1:** Павлов Олександр Анатолійович, Pavlov Alexander Anatolievich
**Автор 2 / Author 2:** Головченко Максим Миколайович, Holovchenko Maxim Nikolaevich
**Автор 3 / Author 3:** Дрозд Валерія Валеріївна, Drozd Valeriia Valeriivna