СИСТЕМНИЙ АНАЛІЗ І ТЕОРІЯ ПРИЙНЯТТЯ РІШЕНЬ

СИСТЕМНЫЙ АНАЛИЗ И ТЕОРИЯ ПРИНЯТИЯ РЕШЕНИЙ

SYSTEM ANALYSIS AND DECISION-MAKING THEORY

UDC 519.24

DOI: 10.20998/2079-0023.2022.01.01

A. A. PAVLOV, M. N. HOLOVCHENKO, V. V. DROZD

CONSTRUCTION OF A MULTIVARIATE POLYNOMIAL GIVEN BY A REDUNDANT DESCRIPTION IN STOCHASTIC AND DETERMINISTIC FORMULATIONS USING AN ACTIVE EXPERIMENT

We present the methods for constructing a multivariate polynomial given by a redundant representation based on the results of a limited active experiment. We solve the problem in two formulations. The first is the problem of constructing a multivariate polynomial regression given by a redundant representation based on the results of a limited active experiment. The solution method is based on the previous results of Professor A. A. Pavlov and his students showing the fundamental possibility of reducing this problem to the sequential construction of univariate polynomial regressions and solving the corresponding nondegenerate systems of linear equations. There are two modifications of this method. The second modification is based on proving for an arbitrary limited active experiment the possibility of using only one set of normalized orthogonal polynomials of Forsythe. The second formulation refers to the solution of this problem for a particular but sufficient from the practical point of view case when an unknown implementation of a random variable is not added to the initial measurement results during an active experiment. This method is a modification of the solution method for the multivariate polynomial regression problem. Also, we used the main results of the general theory (which reduces the multivariate polynomial regression problem solving to the sequential construction of univariate polynomial regressions and solution of corresponding nondegenerate systems of linear equations) to consider and strictly substantiate fairly wide from the practical point of view particular cases leading to estimating the coefficients at nonlinear terms of the multivariate polynomial regression as a solution of linear equations with a single variable.

Keywords: least squares method, multivariate polynomial regression, normalized orthogonal polynomials of Forsythe, redundant representation, linguistic variable, limited active experiment

О. А. ПАВЛОВ, М. М. ГОЛОВЧЕНКО, В. В. ДРОЗД

ПОБУДОВА БАГАТОВИМІРНОГО ПОЛІНОМА, ЗАДАНОГО НАДЛИШКОВИМ ОПИСОМ В СТОХАСТИЧНІЙ ТА ДЕТЕРМІНОВАНІЙ ПОСТАНОВКАХ, З ВИКОРИСТАННЯМ АКТИВНОГО ЕКСПЕРИМЕНТУ

Наведені методи побудови багатовимірного полінома, заданого надлишковим описом, за результатами обмеженого активного експерименту. Задача розв'язується в двох постановках. Перша – як задача побудови багатовимірної поліноміальної регресії, заданої надлишковим описом, за результатами обмеженого активного експерименту. Наведений метод розв'язання базується на попередніх результатах професора О. А. Павлова та його учнів, в яких показана принципова можливість зведення цієї задачі до послідовної побудови одновимірних поліноміальних регресій та розв'язання відповідних невироджених систем лінійних рівнянь. Приводяться дві модифікації методу. Друга модифікація базується на доведенні для довільного обмеженого активного експерименту можливості використання лише одного набору нормованих ортогональних поліномів Форсайта. Друга постановка – це розв'язання цієї задачі для часткового, але достатнього з точки зору практики випадку, коли на вихідні результати вимірювання при проведенні активного експерименту не додається невідома реалізація випадкової величини. Викладений метод є модифікацією методу розв'язання задачі багатовимірної поліноміальної регресії. Також, використовуючи основні результати загальної теорії, що зводить розв'язання задачі багатовимірної поліноміальної регресії до послідовної побудови одновимірних поліноміальних регресій та розв'язанню відповідних невироджених систем лінійних рівнянь, розглянуті та строго обґрунтовані достатньо широкі з точки зору практики часткові випадки надлишкового опису, що приводять до знаходження коефіцієнтів при нелінійних членах багатовимірної поліноміальної регресії як розв'язку лінійних рівнянь з однією змінною.

Ключові слова: метод найменших квадратів, багатовимірна поліноміальна регресія, нормовані ортогональні поліноми Форсайта, надлишковий опис, лінгвістична змінна, обмежений активний експеримент

Introduction. Modern applied information systems use formal models of informatization objects. Their reliable design is still a non-trivial scientific and applied problem. In particular, the problem of constructing a multivariate polynomial regression (MPR) is still the subject of research by scientists [1–8]. Papers [9, 10] addressed the following problem. We have an MPR given by the redundant representation

$$\overline{y}(\boldsymbol{x}) = \begin{array}{c} \sum b_{i_1,\dots,i_t}^{j_1,\dots,j_t} (x_{i_1})^{j_1} (x_{i_2})^{j_2} \cdots (x_{i_t})^{j_t} + E, \\ \forall (i_1,\dots,i_t) \in K, \forall (j_1,\dots,j_t) \in K(i_1,\dots,i_t) \end{array} \quad (1)$$

where $\overline{\boldsymbol{x}} = (x_1, \dots, x_n)^{\mathrm{T}}$ is a deterministic vector of input variables, E is a random variable with arbitrary distribution, its mathematical expectation $ME = 0$, variance $\mathrm{Var}(E) = \sigma^2 < \infty$. The value or the upper bound σ^2 is known. The work [10] descripts in the most detail the methodology and its implementation algorithms for finding efficient estimates of the values of coefficients at nonlinear terms of the redundant representation (1). The algorithms allow to exclude redundant nonlinear terms from (1) practically exactly. The main idea is that giving an MPR with a redundant representation allows to find with a given accuracy (a limitation on the variance value) the estimates of the coefficient values for nonlinear terms of the univariate polynomial regressions (UPRs) of the form

$$Y(x) = \theta_0 + \theta_1 x + \ldots + \theta_r x^r + E \qquad (2)$$

where x is a scalar deterministic input, with solving the corresponding surely nondegenerate systems of linear equations. The variables of the systems are the estimates of the coefficients at nonlinear terms of the MPR (1). We will give two modifications of this method. The second modification contains the proof of the fact that it is possible to use, without reducing the accuracy, only one set of normalized orthogonal polynomials of Forsythe (NOPFs). We derive formulas for the variance of coefficient estimates for nonlinear terms of UPRs for this case. Paper [9] presents algorithms for estimating the coefficients of linear terms of the MPR (1) and building a linguistic variable, the value of which is a qualitative characteristic of the solution reliability for the problem as a whole. Finally, we give a modification of the proposed algorithms for the case when $\mathrm{Var}(E) = 0$. This problem arises when the researcher can specify, up to the coefficient values, only the redundant representation (1) while performs the output measurements in a limited active experiment ($\forall x \in [c, d]$, $c < d$ are arbitrary real numbers) accurately.

Finding and studying the efficiency of estimates of coefficient values for nonlinear terms of a UPR. Let the UPR look like (2). Paper [11] gives the following result. Let a univariate polynomial have the form

$$y(x) = \theta_0 + \theta_1 x + \ldots + \theta_r x^r, \qquad (3)$$

let $x = az + b$ where a, b are arbitrary constants. Then there is one-to-one correspondence between the coefficients $\gamma_0, \gamma_1, \ldots \gamma_r$ of the polynomial

$$y(z) = \theta_0 + \theta_1(az + b) + \theta_2(az + b)^2 + \cdots + \\ + \theta_r(az + b)^r = \gamma_0 + \gamma_1 z + \cdots + \gamma_r z^r \qquad (4)$$

in the form

$$\boldsymbol{A_r}\begin{pmatrix}\theta_0\\ \theta_1\\ \ldots\\ \theta_r\end{pmatrix} = \begin{pmatrix}\gamma_0\\ \gamma_1\\ \ldots\\ \gamma_r\end{pmatrix} \qquad (5)$$

where the matrix $\boldsymbol{A_r}$ is the upper triangular one:

$$\boldsymbol{A_r} = \begin{pmatrix}1 & & & \\ & a & & a_{ij} \\ & & a^2 & \\ & 0 & & \ddots \\ & & & & a^r\end{pmatrix}. \qquad (6)$$

This result allows to obtain estimates of the coefficients at nonlinear terms of the UPR (2) based on the results of a limited active experiment $x_i \to y_i,\ i = \overline{1,n}\ \forall x_i \in [c, d]$, $c < d$ are arbitrary real numbers, $x_1 = c < x_2 < \cdots < x_n = d$, $x_j - x_{j-1} = \text{const}$, $j = \overline{2,n}$, using only the results of a virtual active experiment for a virtual UPR problem

$$y(z) = \gamma_0 + \gamma_1 z + \cdots + \gamma_r z^r + E \qquad (7)$$

$(z_i \to y_i, i = \overline{1,n})$ where

$$a = \frac{d - c}{z_n - z_1} > 0, \qquad b = x_1 - \frac{d - c}{z_n - z_1} \cdot z_1,$$

setting, for example,

$$z_1 = -|p| < z_2 < \cdots < z_n = p > 1,$$

$$z_j - z_{j-1} = const, j = \overline{2,n}. \qquad (8)$$

The values of the scalar input variable x of the UPR (2) are related to the values of the scalar input variable z of the virtual UPR (7) by the expression

$$x_j = az_j + b, j = \overline{2,n},\ x_1 = c. \qquad (9)$$

This allows for arbitrary $c < d$ to use at the given n only one set of NOPFs $Q_j(z), j = \overline{0,r}$, built for the values of the scalar variable z (8). Indeed [12], let the NOPF be $Q_j(z) = q_{j0} + q_{j1}z + \cdots + q_{jj}z^j,\ j = \overline{0,r}$, where the numbers $q_{ij},\ j = \overline{0,r}$, are obtained by numbers z_j, $j = \overline{1,n}$ (see (8)). Then the estimates $\hat{\gamma}_j,\ j = \overline{0,r}$, obtained by the least squares method (LSM) according to the results of a virtual experiment $z_i \to y_i, i = \overline{1,n}$ ($y_i,\ i = \overline{1,n}$, are the output results of an active experiment for a UPR (2) for $x_i, i = \overline{1,n}$, which are related to the values of $z_i, i = \overline{1,n}$, by the expressions (8), (9)), have the form

$$\hat{\gamma}_j = \widehat{w}_r q_{rj} + \cdots + \widehat{w}_j q_{jj}, j = \overline{0,r}, \qquad (10)$$

$$\widehat{w}_j = \sum_{i=1}^{n} y_i\, Q_j(z_i), D\widehat{w}_j = \sigma^2, \qquad (11)$$

$$M\hat{\gamma}_j = \gamma_j,\ D\hat{\gamma}_j = \sigma^2 \sum_{i=r}^{j} q_{ij}^2. \qquad (12)$$

The values of $\hat{\gamma}_j,\ j = \overline{0,r}$, by virtue of the LSM are

$$\arg \min_{\gamma_0,\gamma_1,\ldots,\gamma_r} \left(y_i - \sum_{j=0}^{r} \gamma_j z_i^j\right)^2. \qquad (13)$$

Let $\hat{\theta}_j,\ j = \overline{0,r}$, be the solution to the system of linear equations

$$\boldsymbol{A_r}\begin{pmatrix}\theta_0\\ \theta_1\\ \ldots\\ \theta_r\end{pmatrix} = \begin{pmatrix}\hat{\gamma}_0\\ \hat{\gamma}_1\\ \ldots\\ \hat{\gamma}_r\end{pmatrix}, \qquad (14)$$

where the variables are $\theta_j,\ j = \overline{0,r}$. Then the following statement is true.

Statement 1. $\hat{\theta}_j,\ j = \overline{0,r}$, are

$$\arg \min_{\theta_0,\theta_1,\ldots,\theta_r} \sum_{i=1}^{n}\left(y_i - \sum_{j=0}^{r} \theta_j x_i^j\right)^2.$$

Proof will be done by contradiction. Suppose

$$\sum_{i=1}^{n}\left(y_i-\sum_{j=0}^{r}\hat{\theta}_j x_i^j\right)^2 >$$

$$> \min_{\theta_0,\theta_1,\dots,\theta_r} \sum_{i=1}^{n}\left(y_i-\sum_{j=0}^{r}\theta_j x_i^j\right)^2$$

and $\theta_j^*,\ j=\overline{0,r}$, are $\arg\min_{\theta_0,\theta_1,\dots,\theta_r} \sum_{j=1}^{n}\left(y_i-\sum_{j=0}^{r}\theta_j x_i^j\right)^2$. Suppose $\gamma_j^*,\ j=\overline{0,r}$, are found from the system of linear equations

$$\begin{pmatrix}\gamma_0^*\\ \gamma_1^*\\ \dots\\ \gamma_r^*\end{pmatrix} = \boldsymbol{A_r}\begin{pmatrix}\theta_0^*\\ \theta_1^*\\ \dots\\ \theta_r^*\end{pmatrix}. \tag{15}$$

Then, due to the one-to-one correspondence of solutions (15) and equations $\sum_{j=0}^{r}\theta_j^* x_i^j=\sum_{j=0}^{r}\gamma_j^* z_i^j$, $i=\overline{1,n}$, the condition (13) is not fulfilled. Statement 1 is proved.

Corollary. When using NOPFs for $r\le r_{\max}$, at a fixed $n\ge r$ for arbitrary $c<d$ ($[c,d]$ is the range of possible values of the scalar deterministic variable in the UPR (2) problem, for $x_1=c<x_2<\dots<x_n=d$, $x_j-x_{j-1}=$ $=\text{const}, j=\overline{2,n}$; the estimates $\hat{\theta}_j, j=\overline{2,r}$, can be found using a single set of NOPFs $Q_j(z)$, $j=\overline{2,r}$, built, for example, for

$$z_1=-|p|<z_2<\dots<z_n=p>1,$$

$$z_j-z_{j-1}=\text{const},\ j=\overline{2,n}. \tag{16}$$

Remark. A set of NOPFs (16) is built for $Q_j(z)$, $j=\overline{0,r_{\max}}$.

Values of the coefficients $q_{ij},\ i,j=\overline{0,r}$, must be found in advance with a predetermined accuracy.

Corollary 1. We suggest choosing the segment $[-|p|,p]$, $p>1$, from empirical considerations related to finding $\forall q_{ij},\ i,j=\overline{0,r}$ (10) with a given accuracy and the fact that for a symmetric segment $[-|p|,p]$, $p>1$, the number of tests n to obtain efficient estimates of the UPR coefficients is the lowest [10].

Corollary 2. Condition (16) increases the efficiency of using an active experiment consisting of a repeating dataset $x_1,\dots,x_n$ for obtaining variances of the estimates $\hat{\theta}_j,\ j=\overline{2,r}$, with a given accuracy.

Corollary 3. The values of variances $\mathrm{Var}(\hat{\theta}_j)$, $j=\overline{0,1}$, do not allow to guarantee their accuracy in advance at a small number of tests [11]. In this case we offer:

- to form a data set $x_1^*,\dots,x_{n_1}^*$ for $n_1\ge n$ such that the matrix $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}$ [12] is nondegenerate;
- check the reliability degree of the found estimates $\hat{\theta}_j,\ j=\overline{0,r}$, that corresponds to the value of the linguistic variable introduced in [9].

To use only one set of NOPFs (16), we need to find the values of variances $\mathrm{Var}(\hat{\theta}_j),\ j=\overline{2,r}$, by the values of $\mathrm{Var}(\hat{\gamma}_j),\ j=\overline{2,r}$, given by the formula (12) [12].

The structure of the coefficient's matrix $\boldsymbol{A_r}$ of the system of linear equations (5) allows us to get a solution for $\hat{\theta}_j,\ j=\overline{2,r}$, in the following form:

$$\hat{\theta}_r=\frac{1}{a^r}\hat{\gamma}_r;\quad \hat{\theta}_{r-1}=\frac{1}{a^{r-1}}\left(\hat{\gamma}_{r-1}-\frac{a_{r-1,r}}{a^r}\hat{\gamma}_r\right)$$

and in general form

$$\hat{\theta}_{r-j}=b_{r-j,1}\hat{\gamma}_{r-j}+b_{r-j,2}\hat{\gamma}_{r-j+1}+\dots+$$

$$+b_{r-j,j+1}\hat{\gamma}_r,\ j=\overline{0,r-2}. \tag{17}$$

Remark. Practical recommendation for choosing a value of $p>1$ is this: the value of p should be such that for almost all values of $c<d$ in any limited active experiment we have $a=\frac{1}{k}$ where $k\ge 1$ is an integer or real number with a limited number of decimal places. Then we have no division operation when finding the coefficients $\forall b_{l,m}$ of the system of equations (17).

By virtue of (10) we have

$$\hat{\theta}_j=b_{r-j,1}\sum_{l=r}^{j}\hat{w}_l q_{lj}+b_{r-j,2}\sum_{l=r}^{j+1}\hat{w}_l q_{lj}+\dots+$$

$$+b_{r-j,j+1}\hat{w}_r q_{rj}=\sum_{l=r}^{j}c_{lj}\hat{w}_l,\ j=\overline{2,r}. \tag{18}$$

By virtue of (11) and the following property of $\hat{w}_j$, $j=\overline{0,r}$ [12]:

$$\forall l\ne m\ cov(\hat{w}_l,\hat{w}_m)=0,$$

if $\hat{w}_l,\hat{w}_m$ considered to be random variables, we obtain:

$$\mathrm{Var}(\hat{\theta}_j)=\sigma^2\left(\sum_{l=r}^{j}c_{lj}\right)^2,\ j=\overline{2,r}. \tag{19}$$

Thus, we have two possible algorithms for finding estimates $\hat{\theta}_j, j=\overline{2,r}$, for the UPR problem (2), $x\in[c,d]$.

The first algorithm. Find such n; $x_1=c<x_2<$ $<\dots<x_n=d$, $x_j-x_{j-1}=\text{const}$, $j=\overline{2,n}$; coefficients with a given accuracy of the NOPFs

$$Q_j(x),\qquad j=\overline{0,r} \tag{20}$$

by the values of $x_1,\dots,x_n$; l as the number of repetitions of the subsequence $x_1,\dots,x_n$ in an active experiment that guarantees compliance with the restrictions on the variances $\mathrm{Var}(\hat{\theta}_j),\ j=\overline{2,r}$ (see the results of Pavlov A. A., Kalashnik V. V., Kovalenko D. A. presented in [10]).

Remark. In this case, we use the set of NOPFs (20), and the active experiment consisting of l repetitions of the subsequence $x_1,\dots,x_n$ allows to reduce the variances $\mathrm{Var}(\hat{\theta}_j),\ j=\overline{2,r}$, in l times compared to an active experiment consisting of a sequence $x_1,\dots,x_n$.

The second algorithm. To find the estimates of $\hat{\theta}_j$, $j=\overline{2,r}$, use for an arbitrary limited active experiment ($x\in[c,d]$) the single set of NOPFs (16), the number l of repetitions of the subsequence $x_1,\dots,x_n$ in a limited active experiment for the UPR problem (2) is set by the analysis of the values of $\mathrm{Var}(\hat{\theta}_j),\ j=\overline{2,r}$, by the formula (19).

Methodology and algorithms for estimating coefficients at nonlinear terms of an MPR given by a redundant representation. As a whole, the methodology and algorithms for estimating coefficients at nonlinear terms of an MPR given by a redundant representation were given in [10]. In this section, based on a brief description of this methodology, we will present partial cases of the general problem which are of practical importance and significantly facilitate the problem solution and increase the accuracy of estimating the coefficients at nonlinear terms of the MPR.

Remark. We suggest having the content of the article [10] for a detailed analysis of the following results.

1. In [10], the sequence of actions is given that leads to the guaranteed possibility of finding the estimates of the coefficients at nonlinear terms of an MPR by estimates of the coefficients at nonlinear terms of UPRs after sequential building of UPRs and solving the appropriate number of systems of nondegenerate linear equations. This sequence of actions is as follows.

1) At first, we select in random order, one by one, nonlinear terms of the MPR, each of which contains at least one variable to a power not lower than two. Let it be a variable x_j. Then we carry out an active experiment at the following values of input variables: $x_{ji} = x_i,\ i = \overline{1,n}$, $\forall x_i \in [c,d], n \geq r_{\max} = \max\limits_{\forall l} j_l$ (see the formula (1)). For the segment $[c,d]$, we have: any input variable can take in a limited active experiment any value from this segment. $x_{li} = x^1_{l\text{fix}},\ i = \overline{1,n},\ \forall l \neq j,\ \ x^1_{l\text{fix}} \in [c,d]$ is the fixed value of the variable x_l when constructing the first UPR in order.

Remark. As calculations have shown, for $n = 10$ and $c < d$ having the same sign, an active experiment must use a sequence consisting of l repetitions of the subsequence $x_1, \ldots, x_n$.

First, we find the estimates of the coefficients at nonlinear terms of the MPR by all the equations containing only one variable. Then, we substitute the values of the estimates into the redundant representation of the MPR (1). Next, for each equation containing more than one indeterminate factor at nonlinear terms of the redundant representation of MPR (1), we execute the appropriate number of active experiments to find the estimates for nonlinear terms of UPRs. In each of the experiments, $x_{ji} = x_i,\ i = \overline{1,n},\ x_{li} = x^t_{l\text{fix}},\ i = \overline{1,n},\ \forall l \neq j,\ \ t = \overline{2,T}$, where T is the number of indeterminate factors for nonlinear terms of the MPR given by the redundant representation which are included in this equation. The values of $\forall x^t_{l\text{fix}}$ are chosen so that the corresponding system of linear equations containing indeterminate factors at nonlinear terms of the MPR given by the redundant representation (1) is nondegenerate. This can always be realized. The estimates found for nonlinear terms of the MPR are substituted into the redundant representation (1). After that, the next equation containing indeterminate factors at nonlinear terms of the MPR is analyzed, etc. Then, we consider the next nonlinear term with an indeterminate factor of the MPR given by the redundant representation (1). The coefficient must contain at least one input variable to a power not lower than two, and the algorithmic procedure described above is repeated for it. As a result, we obtain estimates of the coefficients for all nonlinear terms of the MPR given by the redundant representation which contain at least one input variable to a power not lower than two.

2) The algorithmic procedure described in item 1) is repeated sequentially for all nonlinear terms of the MPR given by the redundant representation (1) which contain input variables to a power not higher than one.

Remark 1. In this case, each active experiment has the form: at least two, but not more than $r_{\max}$ input variables take the values of $x_i,\ i = \overline{1,n},\ \forall x_i \in [c,d]$. Other input variables take fixed values chosen so that the corresponding systems of linear equations are nondegenerate.

Remark 2. The described methodology guarantees finding of all estimates for nonlinear terms of the MPR given by the redundant representation, using the estimates of coefficients at nonlinear terms of the corresponding UPRs. But it should be noted that each specific choice of a sequence of nonlinear terms of the MPR given by the redundant representation (1) differs from each other in its execution complexity.

Remark 3. Paper [10] provides a detailed description of solving a specific example.

Remark 4. If there are additional restrictions on the active experiment, namely, not all input variables can take the same values $(x_{j_1 i} = x_{j_l i} = x_i,\ i = \overline{1,n})$, then the above method solves the problem if for each nonlinear term of the MPR given by the redundant representation of the form $\sum b^{1,\ldots,1}_{i_1,\ldots,i_t} x_{i_1} \cdot x_{i_2} \cdots x_{i_t}$ there are at least two input variables that can take the same values in the active experiment.

Remark 5. The algorithmic procedure for estimating the coefficients at linear terms of the MPR (1) and the reliability degree of the obtained results in general are given in [12].

2. Partial cases of the redundant representation of an MPR and of the possibilities of limited active experiment.

2.1. Suppose that the following situation is fulfilled when using the algorithmic procedure of item 1).

a) in item 1) the first considered nonlinear term of the redundant representation (1) contains the factor $\left(x_{i_l}\right)^{j_l}$, $j_l \geq 2$, and the following holds for the input variable x_{i_l}: there can be only one nonlinear term from the redundant representation (1) that contains $\left(x_{i_l}\right)^{j_p}\ \forall j_p \geq 2$.

b) each subsequent step of the algorithmic procedure of item 1) is executed for the nonlinear term for which the condition a) is fulfilled.

Remark. Nonlinear terms of (1) containing the chosen input variable to a power not lower than two may not correspond to the condition a) but estimates of the coefficients for them were found in the previous steps of the algorithm.

c) the number of factors in each nonlinear term of the redundant representation (1) that contains input variables to a power not higher than one does not exceed $r_{\max}$, while m can be greater than $r_{\max}$, and the steps of the algorithmic procedure of item 2) are executed for nonlinear terms of the redundant representation (1) which contain the maximum number of factors.

Remark. In this case, we estimate at each step only the coefficients given by a linear equality containing a single variable.

The following statement is true.

Statement 2. If the conditions a)–c) are met then the algorithmic procedure of items 1) and 2) surely finds estimates of the coefficients at all nonlinear terms of the redundant representation (1), and an equation with a single variable is solved to estimate each coefficient.

Proof. Statement 2 follows from the construction rule of the linear equations (see [10]).

2.2. Suppose that the following condition is met at each step of the algorithm of item 1) executed for a nonlinear term containing the input variable $\left(x_{i_l}\right)^{j_l}$, $j_l \geq 2$:

d) there are no members containing input variables $\left(x_{i_l}\right)^{j_l}$ in the redundant representation (1) whose coefficients were not estimated in the previous steps of the algorithm.

Remark. A partial case of the condition d): in the redundant representation (1) for $\forall j \geq 2$ any input variable to a power j is included in only one nonlinear term.

Then the following statement is obviously true.

Statement 3. If the conditions d), c) are met, then the algorithmic procedure of items 1), 2) surely finds estimates of all coefficients at nonlinear terms, each of which satisfies a linear equality with a single indeterminate.

2.3. Let the following condition be satisfied.

e) the active experiment allows to set the value of any single input variable to zero; among all nonlinear terms containing $\left(x_{i_l}\right)^{j_p}$, $j_p \geq 2$, whose estimates of the values of their coefficients were not found in the previous steps of the algorithm, we have a single nonlinear term (for which the corresponding step of the algorithm of item 1) is executed) that does not contain an input variable included in all other nonlinear terms with indeterminate factors containing $\left(x_{i_l}\right)^{j_p}$. This variable is set to zero during the active experiment. Then the following obvious statement is true.

Statement 4. If the conditions e) (for each step of the algorithm of item 1)) and c) are met then the algorithmic procedure of items 1), 2) allows to find estimates of coefficients at nonlinear terms of redundant representation (1) by solving only linear equations with a single variable.

Finding the coefficients of a multivariate polynomial given by a redundant representation in a deterministic formulation. Consider the degenerate MPR problem given by a redundant representation (1) for the case when $E \equiv 0$. Then the algorithmic procedure described in items 1), 2), firstly, finds practically exact values of the coefficients (assuming that the corresponding systems of linear equations are solved with a predetermined accuracy), and secondly, allows the following simplifications.

- the total number of tests in the active experiment to build each univariate polynomial does not exceed $r_{\max}$;
- it is quite appropriate to use only one set of NOPFs;
- the coefficient values at the linear terms of a multivariate polynomial given by the redundant representation are found with an algorithmic procedure similar to that of item 2) of the previous section using the NOPFs to the powers zero and one.

Conclusions. 1. We have stated the problem of building a multivariate polynomial given by a redundant representation in stochastic and deterministic formulations, based on the results of an active experiment.

2. A description of the integral algorithmic procedure for solving the problem in both formulations is given. It is based on building univariate polynomial regressions and solving nondegenerate systems of linear equations.

3. We have proved the possibility of using a single set of normalized orthogonal polynomials of Forsythe to estimate the coefficients of a univariate polynomial regression.

4. Quite practical partial cases of redundant representation were considered which simplify the problem solution as much as possible.

5. We have given a modification of the general algorithmic procedure to solve the problem in its deterministic formulation.

References

1. Ивахненко А. Г. *Моделирование сложных систем. Информационный подход*. Киев: Вища школа, 1987. 62 с.
2. 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
3. Draper N. R., Smith H. *Applied Regression Analysis.* 3rd edition. New York: John Wiley & Sons, 1998. 736 p.
4. Большаков А. А., Каримов Р. Н. *Методы обработки многомерных данных и временных рядов: учебн. пособие для вузов*. Москва: Горячая линия–Телеком, 2007. 522 с.
5. 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 (1). P. 1–14. doi: 10.5815/ijieeb.2021.01.01
6. 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 (3). P. 15–22. doi: 10.5815/ijieeb.2018.03.03
7. 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 (2). P. 14–23. doi: 10.5815/ijmsc.2021.02.02
8. 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 (3). P. 16–26. doi: 10.5815/ijeme.2019.03.02
9. Павлов О. А., Головченко М. М., Ревич М. М. Метод оцінки коефіцієнтів при лінійних членах багатовимірної поліноміальної регресії, заданої надлишковим описом. *Адаптивні системи автоматичного управління: міжвідомчий наук.-техн. збірник*. Київ: НТУУ «КПІ», 2022. Том 1, № 40. (у друці)
10. Pavlov A., Holovchenko M., Mukha I., Lishchuk K. Mathematics and software for building nonlinear polynomial regressions using estimates for univariate polynomial regressions coefficients with a given (small) variance. *Lecture Notes on Data Engineering and Communications*

Technologies. 2022. Vol. 134. P. 288–303. doi: 10.1007/978-3-031-04812-8_25

11. Pavlov A. A. Estimating with a given accuracy of the coefficients at nonlinear terms of univariate polynomial regression using a small number of tests in an arbitrary limited active experiment. *Вісник Нац. техн. ун-ту «ХПІ»: зб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології*. Харків: НТУ «ХПІ», 2021. № 2 (6). С. 3–7. doi: 10.20998/2079-0023.2021.02.01
12. Худсон Д. *Статистика для физиков: Лекции по теории вероятностей и элементарной статистике*. 2-е изд. Москва: Мир, 1970. 296 с.

References (transliterated)

1. Ivahnenko A.G. *Modelirovanie slozhnyh sistem. Informacionnyj podhod* [Complex Systems Modeling. Informational Approach]. Kiev, Vyshha shkola Publ., 1987. 62 p.
2. Nastenko E., Pavlov V., Boyko G., Nosovets O. Mnogokriterial'nyj algoritm shagovoj regressii. *Biomedychna inzheneriya i tekhnolohiya* [Biomedical ingeneering and technology]. 2020, no. 3, pp. 48–53. doi: 10.20535/2617-8974.2020.3.195661
3. Draper N. R., Smith H. *Applied Regression Analysis.* 3rd edition. New York, John Wiley & Sons, 1998. 736 p.
4. Bol'shakov A. A., Karimov R. N. *Metody obrabotki mnogomernykh dannykh i vremennykh ryadov: uchebnoe posobie dlya vuzov* [Methods for processing multivariate data and time series: textbook for universities]. Moscow, Goryachaya liniya–Telekom Publ., 2007. 522 p.
5. 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 (1), pp. 1–14. doi: 10.5815/ijieeb.2021.01.01
6. 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 (3), pp. 15–22. doi: 10.5815/ijieeb.2018.03.03
7. 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 (2), pp. 14–23. doi: 10.5815/ijmsc.2021.02.02
8. 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 (3), pp. 16–26. doi: 10.5815/ijeme.2019.03.02
9. Pavlov O. A., Holovchenko M. M., Revych M. M. Metod otsinky koefitsiyentiv pry liniynykh chlenakh bahatovymirnoyi polinomial'noyi rehresiyi, zadanoyi nadlyshkovym opysom. [Method for estimating coefficients for linear terms of multidimensional polynomial regression given by redundant description] *Adaptyvni systemy avtomatychnoho upravlinnya: mizhvidomchyy nauk.-tekhn. zbirnyk* [Adaptive systems of automatic control: interdepartmental scientific and technical. collection]. Kyiv, KPI Publ., 2022, vol. 1, no. 40. (in print)
10. Pavlov A., Holovchenko M., Mukha I., Lishchuk K. Mathematics and software for building nonlinear polynomial regressions using estimates for univariate polynomial regressions coefficients with a given (small) variance. *Lecture Notes on Data Engineering and Communications Technologies*. 2022, vol. 134, pp. 288–303. doi: 10.1007/978-3-031-04812-8_25
11. Pavlov A. A. Estimating with a given accuracy of the coefficients at nonlinear terms of univariate polynomial regression using a small number of tests in an arbitrary limited 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., 2021, no. 2 (6), pp. 3–7. doi: 10.20998/2079-0023.2021.02.01
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

Received 04.05.2022

Відомості про авторів / About the Authors

Павлов Олександр Анатолійович – доктор технічних наук, професор кафедри інформатики та програмної інженерії Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»; м. Київ, Україна; ORCID: https://orcid.org/0000-0002-6524-6410; e-mail: pavlov.fiot@gmail.com

Головченко Максим Миколайович – старший викладач кафедри інформатики та програмної інженерії Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»; м. Київ, Україна; ORCID: https://orcid.org/0000-0002-9575-8046; e-mail: ma4ete25@ukr.net

Дрозд Валерія Валеріївна – бакалавр кафедри інформатики та програмної інженерії Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»; м. Київ, Україна; ORCID: https://orcid.org/ https://orcid.org/0000-0003-0418-1139; e-mail: drozdllera@gmail.com

Pavlov Alexander Anatolievich – Doctor of Technical Sciences, Full Professor of Informatics and Software Engineering Department of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"; Kyiv, Ukraine; ORCID: https://orcid.org/0000-0002-6524-6410; e-mail: pavlov.fiot@gmail.com

Holovchenko Maxim Nikolaevich – Senior Lecturer of Informatics and Software Engineering Department of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"; Kyiv, Ukraine; ORCID: https://orcid.org/0000-0002-6524-6410; e-mail: pavlov.fiot@gmail.com

Drozd Valeria Valerievna – Bachelor of Informatics and Software Engineering Department of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"; Kyiv, Ukraine; ORCID: https://orcid.org/0000-0003-0418-1139; e-mail: drozdllera@gmail.com