SYSTEM ANALYSIS AND DECISION-MAKING THEORY

In recent years, the authors in their publications have developed two different approaches to the construction of a multivariate polynomial (in particular, linear) regressions given by a redundant representation. The first approach allowed us to reduce estimation of coefficients for nonlinear terms of a multivariate polynomial regression to construction of a sequence of univariate polynomial regressions and solution of corresponding nondegenerate systems of linear equations. The second approach was implemented using an example of a multivariate linear regression given by a redundant representation and led to the creation of a method the authors called a modified group method of data handling (GMDH), as it is a modification of the well-known heuristic self-organization method of GMDH (the author of GMDH is an Academician of the National Academy of Sciences of Ukraine O. G. Ivakhnenko). The modification takes into account that giving a multivariate linear regression by redundant representation allows for construction of a set of partial representations, one of which has the structure of the desired regression, to use not a multilevel selection algorithm, but an efficient algorithm for splitting the coefficients of the multivariate linear regression into two classes. As in the classic GMDH, the solution is found using a test sequence of data. This method is easily extended to the case of a multivariate polynomial regression since the unknown coefficients appear in the linear way. Each of the two approaches has its advantages and disadvantages. The obvious next step is to combine both approaches into one. This has led to the creation of a synthetic method that implements the advantages of both approaches, partially compensating for their disadvantages. This paper presents the aggregated algorithmic structure of the synthetic method, the theoretical properties of partial cases and, as a result, the justification of its overall efficiency.

Keywords: univariate polynomial regression, multivariate polynomial regression, redundant representation, least squares method, test sequence, repeated experiment.

Introduction. Multivariate linear and non-linear regressions, constructed based on the results of active or passive experiments, are widely used in modern diagnostic information systems, in particular, medical ones, and in information management systems with a wide range of applications [1–10]. Universal methods for multivariate regressions construction vary from classical statistical methods to heuristics, such as the group method of data handling (GMDH) or genetic algorithms. But none of them, due to the complexity of the problem, dominates the others. Existing methods complement each other. Therefore, scientific research in this field is still relevant.

We give in the abstract, at a qualitative level, the characteristics of a synthetic method of constructing a multivariate polynomial regression (MPR) given by a redundant representation. In this paper, we explain the aggregated algorithmic structure of the synthetic method, substantiate the logic of its construction, its theoretical properties, the properties of partial cases of redundant representations of MPRs that lead to the finding of coefficient estimates for nonlinear members of the MPR with acceptable accuracy.

1. General theoretical provisions that we use. 1.1. Univariate polynomial regression (UPR) [11]. Let a UPR be given in the form:

$$Y(x) = \theta_0 + \theta_1 x + \ldots + \theta_n x^n + E,$$

(1)

where $E$ is a random variable with an arbitrary distribution, its mathematical expectation $ME = 0$, its variance $DE = \sigma^2 < \infty$, the variance or its upper bound is known.

© Pavlov A. A., Holovchenko M. N., Drozd V. V., 2023
According to the results of the experiment \( (x_i \rightarrow y_i, \ i = 1, n) \), where \( r < n \), \( x_i \in [c, d] \),
\[
y_i = \theta_0 + \theta_1 x_1 + \cdots + \theta_k x_k + \delta_i ,
\]
where \( \delta_i \) is an implementation of a random variable \( E \) in the \( i \)-th test, estimates of unknown coefficients can be found using normalized orthogonal polynomials of Forsythe (NOPFs) constructed from the values of the input variable \( x_i, i = 1, n \).

\[ Q_j (x) = q_{j0} + q_{j1} x + \cdots + q_{jk} x^k, \quad j = 0, r . \]

In this case, the UPR (1) has the form
\[
Y(x) = \sum_{j=0}^{r} \omega_j Q_j(x) + E ,
\]
\[
\hat{\omega}_j = \frac{1}{r} \sum_{i=1}^{r} y_i Q_j(x_i), \quad j = 0, r ,
\]
\[
\hat{\theta}_j = \hat{\omega}_j q_{j0} + \cdots + \hat{\omega}_j q_{jk} + \hat{\theta}_j = 0, r ,
\]
\[
M \hat{\omega}_j = \sigma^2, M \hat{\theta}_j = \hat{\theta}_j, D \hat{\theta}_j = \sigma^2 \sum_{j=0}^{r} \hat{\theta}_j^2 .
\]

1.2. A repeated experiment. A. Pavlov and D. Kovalenko proved that the results of a repeated experiment with \( k \) repetitions of the main experiment \( (x_i \rightarrow y_i, i = 1, n) \) are equivalent to the experiment
\[
x_i \rightarrow \frac{1}{k} \sum_{p=0}^{k-1} y_i p^{n+i}, \quad i = 1, n,
\]
and the variances of estimates \( \hat{\theta_j}, \ j = 0, r \), decrease by the factor of \( k \).

1.3. Estimation of coefficients for nonlinear members of a UPR using a single set of NOPFs. It was shown in [12] that the estimates of the coefficients for nonlinear terms of a UPR (1) at arbitrary values of \( c < d \) can be found using only a single set of NOPFs based on the results of the following virtual experiment. We introduce a virtual deterministic scalar variable \( z \) and for its values
\[
z_1 < z_2 < \ldots < z_n
\]
we find with the specified accuracy a set of NOPFs \( Q_j (z) \), \( j = 0, r \). Having put
\[
a = \frac{d - c}{z_n - z_1} > 0, b = c - d - \frac{c - d}{z_n - z_1} z_1 ,
\]
\[
x_i = az_i + b, \quad i = 1, n (x_i = c, x_n = d) ,
\]
substituting \( x = az + b \), we reduce the UPR (1) to a virtual UPR
\[
Y(z) = \theta_0 + \theta_1 (az + b) + \cdots + \theta_k (az + b)^k + E =
\]
\[
= \gamma_0 + \gamma_1 z + \cdots + \gamma_k z^k + E .
\]

The coefficients \( \theta_j, \gamma_j, \ j = 0, r \), are in a mutually unambiguous correspondence
\[
\begin{pmatrix}
1 \\
a \\
a^2 \\
\vdots \\
0
\end{pmatrix}
\begin{pmatrix}
\theta_0 \\
\theta_1 \\
\theta_2 \\
\vdots \\
\theta_r
\end{pmatrix}
= \begin{pmatrix}
y_0 \\
y_1 \\
y_2 \\
\vdots \\
y_r
\end{pmatrix} .
\]

According to results of the main experiment \( (x_i \rightarrow y_i, i = 1, n) \), where \( x_i, i = 1, n \), satisfy (8), we design a virtual experiment \( (z_i \rightarrow y_i, i = 1, n) \). Based on the set of NOPFs constructed for \( z_i, i = 1, n \), we find estimates \( \hat{y}_j, j = 0, r \), and based on the system of equations (10) we find estimates \( \hat{\theta}_j, j = 0, r \). It is shown in [12] that
\[
\sum_{j=0}^{r} \left( y_j - \frac{\hat{\theta}_j z_i}{\sigma} \right)^2 = \min_{\hat{\theta}_j, j = 0, r} \sum_{i=1}^{n} \left( y_j - \hat{\theta}_j z_i \right)^2 .
\]

2. Estimation with acceptable accuracy of the coefficients for nonlinear members of a UPR (1) using a single set of NOPFs. 2.1. Justification of the number of tests \( n \) of the main experiment and the choice of values of \( z_i, i = 1, n \), of the virtual scalar variable \( z \). With the appropriate selection of the values of the virtual scalar variable \( z \), expression (10) significantly simplifies the preliminary analysis and the input data formation for the main active experiment to obtain, with acceptable accuracy, estimates of the coefficients for nonlinear terms of the UPR and, as will be shown later, an MPR. Such a requirement is for \( r_{max} < n \) a compromise between the number of tests \( n \) of the main experiment and the value of variances of the estimates of the coefficients for nonlinear members of the virtual UPR (9). As a result of the analysis of the conducted experiments, the following compromise solution is proposed for \( r_{max} \leq 5 ; n = 10, \ z_1 = -50, \ z_{10} = 50, \Delta z = (z_i - z_{i-1}) = \text{const} . \)

\[
D\hat{\gamma}_2 = 4.26 \cdot 10^{-4} \sigma^2 , \quad D\hat{\gamma}_3 = 7.55 \cdot 10^{-4} \sigma^2 ,
\]
\[
D\hat{\gamma}_4 = 1.4 \cdot 10^{-4} \sigma^2 , \quad D\hat{\gamma}_5 = 1.28 \cdot 10^{-4} \sigma^2 ,
\]
that is, with an increase of \( j \) by one, \( D\hat{\gamma}_j \) is decreased by three orders of magnitude, starting from \( D\hat{\gamma}_2 \).

2.2. Justification of conditions for obtaining estimates for nonlinear members of a UPR (1) with acceptable accuracy. In [12], analytical expressions are given for variances of estimates \( \hat{\theta}_j, \ j \geq 2 \), for a UPR (1). The expressions were obtained from the results of estimation by a virtual UPR (9). This makes it possible to determine, depending on the values of \( c, d, a, b \ (8) \), the number \( k \) of repetitions of the main experiment (variances of the coefficient estimates \( \hat{\theta}_j \) are reduced by the factor of \( k \) that, using the three-
sigma rule, leads to obtaining the estimates \( \hat{\theta}_j, j \geq 2 \), with acceptable accuracy. Thus, when the repeated experiment turned out to be practically feasible, the problem has a solution.

**Remark 1.** The three-sigma rule for an arbitrary random variable \( X \) is formulated as follows. With a probability of 0.89, any realization of \( X \) belongs to the segment \([MX - 3\sigma, MX + 3\sigma]\), where \( \sigma = \sqrt{\text{DX}} \).

**Remark 2.** Due to the roughness of the three-sigma rule, the actual number of repetitions \( k \) is significantly less.

### 3. A decomposition method of estimating the coefficients of an MPR given by a redundant representation.

#### 3.1. The problem formulation.

Let an MPR be given by the following redundant representation [13]:

\[
y(x) = \sum_{\gamma_1, \ldots, \gamma_m} b_{\gamma_1, \ldots, \gamma_m}^i \left( x_1^\gamma_1 \cdots x_m^\gamma_m \right) + E, \quad (12)
\]

where \( x = (x_1, \ldots, x_m) \) is a deterministic vector of input variables,

\[
x_i \in [c_i, d_i], \quad c_i > 0, \quad i = 1, m,
\]

where \( E \) is a random variable, \( ME = 0 \), \( DE = \sigma^2 < \infty \). The values of the coefficients \( b_{\gamma_1, \ldots, \gamma_m}^i \) are unknown (\( b_0 \) is a constant).

**Remark 3.** The condition \( c_i > 0, \quad i = 1, m \), corresponds to the most common practical case. All the results obtained below are trivially extended to arbitrary real values of the numbers \( c_i < d_i, \quad i = 1, m \).

#### 3.2. The decomposition method [13] implements the methodology of reducing the estimation of nonlinear members of an MPR (12) to the sequential construction of UPRs and the solution of the corresponding systems of linear equations, the variables of which are the estimates for the nonlinear members of the MPR (12).

The general algorithmic scheme for obtaining estimates for nonlinear members of the MPR (12) consists of two sub-algorithms [13].

**3.2.1. Aggregated modified algorithmic scheme of the first sub-algorithm [13].** The \( l \)-th step (\( l \leq L_1 \)) of which contains at least one scalar variable raised to a power greater than or equal to two) is implemented for the corresponding coefficient(s) of the first UPR \( (\hat{\theta}_j \neq 0) \) expressed linearly by several coefficients for the nonlinear terms of the MPR (12) not evaluated at the previous steps of the first sub-algorithm.

The right-hand parts of the obtained nondegenerate systems of linear equations are the estimates for nonlinear terms of the UPRs. Their solutions are the estimates of the corresponding coefficients for nonlinear terms of the MPR (12). The number of UPRs constructed at the \( l \)-th step can be more than one if the corresponding coefficient(s) of the first UPR \( (\hat{\theta}_j \neq 0) \) expressed linearly by several coefficients for the nonlinear terms of the MPR (12) not evaluated at the previous steps and has a maximum value of \( t_j \).

Each input variable \( x_{ij}, \quad j = 1, L_1 \), is expressed linearly by a virtual variable \( z : x_{ij} = a_{ij} + b_{ij} \) according to (8) for \( c = c_{ij}, \quad d = d_{ij}, \quad j = 1, L_1 \). In the main experiment, the variables \( x_{ij}, \quad j = 1, L_1 \), vary according to (7), (8). Other scalar input variables take fixed values in each test. The data of the virtual main experiment are determined based on the data of the main experiment

\[
\left( x_{i,j_1}, \ldots, x_{i,j_k} \right) \neq \left( x_{i,j_1}, \ldots, x_{i,j_k} \right) \rightarrow y_i, \quad i = 1, n,
\]

namely:

\[
z_i \rightarrow y_i - \sum_{v \in \{v_k, \ldots, v_1\}(\{j_1, \ldots, j_k\})} \hat{b}_{v_1, \ldots, v_k}^i \left( x_{j_1} \right) \cdots \left( x_{j_k} \right) - \sum_{v \in \{v_k, \ldots, v_1\}(\{j_1, \ldots, j_k\})} \hat{b}_{v_1, \ldots, v_k}^i \prod_{j=1}^{j_k} x_{j, i}, \quad i = 1, n, \quad (15)
\]

where \( K_1 \) is the number of steps of the first sub-algorithm; \( G_n \) is a set of coefficients estimated with sufficient accuracy at the previous steps of the second sub-algorithm.
Remark 6. The data for the repeated virtual experiment is found in a similar way.

Remark 7. The maximum degree of a UPR is \( t_1 \).

3.3. Classes of redundant representations for which the coefficients for nonlinear terms of an MPR (12) are estimated completely or partially with acceptable accuracy.

3.3.1. The redundant representation of (12) satisfies the following three conditions:

1) \( \forall (x_j)^{\hat{y}}, j \geq 2 \) can be included in only one component of (12);

2) for any two nonlinear components of (12) with coefficients \( b_{h,1}^{\hat{y}}, b_{h,2}^{\hat{y}} \), the following is true:

\[
\{i_1, \ldots, i_h\} \neq \{l_1, \ldots, l_h\};
\]

3) \( 0 \in [c_i, d_i], i = 1, m \),

where \( m \) is the number of input variables.

Then the \( l \)-th step of the first sub-algorithm of the decomposition method implemented for the input scalar variable \( (x_j)^{\hat{y}}, j \geq 2 \), of the nonlinear component of (12)

\[
b_{h,1}^{\hat{y}}(x_j)^{\hat{y}} \cdots (x_j)^{\hat{y}}
\]

must satisfy the following additional conditions. In all tests of the active experiment, in each component of (12) \( b_{h,1}^{\hat{y}}(x_j)^{\hat{y}} \cdots (x_j)^{\hat{y}}, b_{h,2}^{\hat{y}} \in \{j_{m+1} \} \) (see (14)), one input scalar variable whose index does not belong to the set \( \{i_1, \ldots, i_h\} \) is equal to zero. In this case, the linear equation for estimating \( b_{h,1}^{\hat{y}} \) (18) has the form

\[
b_{h,1}^{\hat{y}}(a_j)^{\hat{y}} \sum_{i=1}^{l} x_i^p = \hat{\gamma}_{j} = \gamma_{j} \pm \varepsilon_{j},
\]

where \( \hat{\gamma}_{j} \) is the estimate of \( \gamma_{j} \) — the coefficient of the corresponding virtual UPR (depending on its meaning, it is a realization or a random variable), and \( \varepsilon_{j} \) is the realization of a random variable with zero mathematical expectation and the variance equal to \( D \hat{\gamma}_{j} \) (if \( \hat{\gamma}_{j} \) is considered a random variable). Let \( \varepsilon \left(\sum_{i=1}^{l} j_{i}\right) > 0 \) (this is an expertly set upper bound on the value of \( b_{h,1}^{\hat{y}} \)) below which the corresponding component (18) is excluded from the redundant representation (12) as not essential. Then the values of \( x_i^{\hat{y}} \) \( \forall i_{m, i_n} \), \( i_n \neq p \), are set equal to \( x_i^{\hat{y}} = d_{i} \), \( m \neq p \), and thereby the number of repetitions of the main experiment that guarantees the fulfillment of the inequality

\[
|\varepsilon_{j}| \leq 10^{-1} (a_j)^{\hat{y}} \varepsilon \left(\sum_{i=1}^{l} j_{i}\right) \sum_{n=m}^{p} d_{n},
\]

(with the corresponding probability, using the three-sigma rule), is the minimum possible.

Remark 8. To obtain the estimate (20), the three-sigma rule is applied to the random variable \( \hat{\gamma}_{j} \), which realization is used to find the estimate of the coefficient \( b_{h,1}^{\hat{y}} \) (18).

If

\[
|\hat{\gamma}_{j}| < \left( \varepsilon \left(\sum_{i=1}^{l} j_{i}\right) \right) \sum_{n=m}^{p} d_{n},
\]

then the component (18) is excluded from the redundant representation (12). In the opposite case, we have found an estimate of the coefficient for the component (18) with an acceptable accuracy (20).

The \( l \)-th step of the second sub-algorithm of the decomposition method implemented for the component

\[
b_{h,1}^{\hat{y}} \prod_{j=1}^{l} x_j^p, \quad t = \max
\]

must satisfy the following additional conditions. In all tests of the active experiment in each component of (12) of the form \( b_{h,1}^{\hat{y}}(x_j)^{\hat{y}} \cdots (x_j)^{\hat{y}}, b_{h,2}^{\hat{y}} \in \{j_{m+1} \} \cup \{j_{m+1} \} \) (see (15)), one input scalar variable whose index does not belong to the set \( \{i_1, \ldots, i_h\} \) is equal to zero. The linear equation for estimating \( b_{h,1}^{\hat{y}} \) (22) has the form

\[
b_{h,1}^{\hat{y}} \prod_{j=1}^{l} x_j^p = \hat{\gamma}_{j} = \gamma_{j} \pm \varepsilon_{j},
\]

where \( \hat{\gamma}_{j} \) is the estimate of \( \gamma_{j} \) — the coefficient of the corresponding virtual UPR. Let the number of repetitions of the main experiment be feasible to fulfill, based on the three-sigma rule, the limitation

\[
|\varepsilon_{j}| \leq 10^{-1} (a_j)^{\hat{y}} \varepsilon \left(\sum_{i=1}^{l} j_{i}\right) \prod_{n=m}^{p} a_{n},
\]

Then, if

\[
|\hat{\gamma}_{j}| < (a_j)^{\hat{y}} \varepsilon \left(\sum_{i=1}^{l} j_{i}\right) \prod_{n=m}^{p} a_{n},
\]

then the component (22) is excluded from the redundant representation (12). In the opposite case, we have found an estimate of the coefficient \( b_{h,1}^{\hat{y}} \) (22) with an acceptable accuracy (24).

Thus, if all the repeated experiments are feasible, the decomposition method allows you to find all estimates for nonlinear terms of the MPR (12) with acceptable accuracy.

Corollary. If for some steps of the first or second sub-algorithms the required number of repetitions of the main experiment is unfeasible, then the corresponding coefficients are not estimated, and conditions (15), (16) allow such components of the redundant representation (12) to be excluded from other steps of the first and second sub-algorithms.

Remark 9. Each step of the first and second sub-algorithms estimates only one coefficient of the MPR (12).

3.3.2. The redundant representation (12) is given by the following four conditions.
The first condition: all input variables included in the redundant representation (12) of the form \( b^{h_i}_{c_i-1} \times (x_i) \) \( \cdots (x_i) \); \( \sum_{j=1}^{l} j_i \geq 2 \), have a range of acceptable values \([l, d_i]\), \( l = 1, t\).

The second condition: if an arbitrary input variable is included in an arbitrary component of (12) to a power greater than or equal to two, it is no longer included in any other component of (12) of the form \( b^{h_j}_{c_j-1} \times (x_j) \) \( \cdots (x_j) \), \( \sum_{j=1}^{l} j_i \geq t + 1 \).

The third condition: all nonlinear components of the form \( b^{1}_{c_1} \sum_{i=1}^{l} x_i \) do not have common input variables.

The fourth condition: no input variable included in any component of (12) that has \( \exists j_i \geq 2 \), is included in any component of (12) whose coefficient has the form \( b^{1}_{c_1} \), \( t \geq 2 \).

Let the number of repetitions of the main experiment for an arbitrary step of the first and second sub-algorithms be feasible for obtaining, using the three-sigma rule, the estimates:

a) for the first sub-algorithm:
\[
|e_{r_i}| \leq 10^{-4} (a_i)^{1/2} \varepsilon (\sum_{j=1}^{l} j_i); \\
\]

b) for the second sub-algorithm: estimate (24).

Then, if during the implementation of an arbitrary step of the first and second sub-algorithms, the fixed input variables included in the nonlinear terms of the MPR in expressions (14), (15) are set equal to one, then in expressions (14), (15) from the values of \( y_i \) will be subtracted the values of \( \sum_{j=1}^{l} b^{h_j}_{c_j-1} \), which will differ in modulus from the values of \( \sum_{j=1}^{l} b^{h_j}_{c_j-1} \), due to estimates (24), (26) and the roughness of the three-sigma rule, by values that can be practically neglected. Then all estimates for non-linear terms of the MPR (12) are found with acceptable accuracy.

Remark 10. The number of repetitions of the main experiment can be real if \( a_i > \gamma_i, i = 1, m \). This is also true for the second sub-algorithm for the class 3.3.1.

Remark 11. The number of repetitions of the main experiment at some steps of the first sub-algorithm can be significantly smaller, if for this step the estimate for \( |e_{r_i}| \) is of the form (20). That is, some input variables that are not included in (14), (15) can take maximum values.

Remark 12. If we exclude the third and fourth conditions imposed on the redundant representation (12), then the decomposition method implements only the first sub-algorithm.

Remark 13. Each step of the decomposition method for the redundant representation (12) that satisfies the four or the first two conditions of the class 3.3.2, estimates only one coefficient of the redundant representation (12).

3.3.3 (generalization of the class 3.3.2). A redundant representation (12) satisfies the first condition of the class 3.3.1. \( c_i < d_i, c_i > 0, i = 1, m \) are arbitrary numbers. The second condition is set based on the results of the implementation of the following version of the decomposition method. When performing the \( l \)-th (\( l \geq 2 \)) step of the first sub-algorithm for the term \( b^{h_j}_{c_j-1} (x_j) \) \( \cdots (x_j) \), the fixed values of the input variables in (14), which are not part of the set \( \{x_i, \ldots, x_j \} \), are set modulo the minimum possible.

When implementing an arbitrary step of the second sub-algorithm, all fixed variables included in (15) take modulo minimum values.

Let the number of repetitions of the main experiment for each step of the first and second sub-algorithms be feasible to fulfill constraints (20), (24). Then the second condition imposed on the redundant representation (12) is
\[
\forall \left| \hat{b}^{h_j}_{c_j-1} \right| \geq \varepsilon \left( \sum_{j=1}^{l} j_i \right) \cdot 10^p, p \geq 2, \text{ or} \\
\left| \hat{b}^{h_j}_{c_j-1} \right| < \varepsilon \left( \sum_{j=1}^{l} j_i \right).
\]

It is obvious that for any real values of \( c_i < d_i, i = 1, m, r_{max} \) (the maximum degree of the virtual UPR) there is such a small enough value of the natural number \( p \) that the replacement of exact values of \( b^{h_i}_{c_i} \) in expressions (14), (15) with their estimates is statistically guaranteed to have practically no effect on the values of coefficient estimates for nonlinear terms of the MPR (12).

Remark 14. Only one coefficient of the MPR (12) is estimated at each step of the decomposition algorithm.

3.3.4. The redundant representation of the MPR (12) is arbitrary, the random variable \( E = 0 \). Such a case occurs when \( E \) is a measurement error and is neglected having sufficient measurement accuracy. In this case, the general algorithmic procedure of the decomposition method accurately finds the values of all coefficients for the nonlinear terms of (12). The exact values of the coefficients for linear terms, including the constant, are found by the usual interpolation procedure, which consists in solving a system of linear equations with the number of variables \( m+1 \), where \( m \) is the number of input variables.

The advantage of the decomposition method over an arbitrary interpolation method lies is two-fold: the coefficients for nonlinear terms of the redundant representation are estimated using only a single set of NOPFs found with a given accuracy; the needed-to-solve nondegenerate systems of linear equations have a significantly smaller dimension than the number of nonlinear components of (12), and in most cases the dimension is equal to one.

3.3.5. The redundant representation (12) is arbitrary. In general case, the detailed formalization of the first and second sub-algorithms for finding estimates with sufficient accuracy is inefficient. A visual analysis of the specific redundant representation (12) using the above theoretical results allows for each specific case to create an efficient sequence of steps of an individual algorithm of the decomposition method (which can be a step of the first or second
sub-algorithm in any order), which leads to estimation with acceptable accuracy of the maximum number of coefficients for nonlinear terms of the MPR (12).

4. A modified group method of data handling (MGMDH). As a result of the implementation of the decomposition method, we obtained a set \( \{J\} \) of coefficients for nonlinear terms of the MPR (12) estimated with acceptable accuracy.

Remark 15. Insignificant coefficients are excluded from the redundant representation (12) and the set \( \{J\} \).

To obtain the final result, we propose to use the MGMDH described in [14] to construct a multivariate linear regression (MLR) given by a redundant representation. Indeed, the problem of constructing an MPR given by a redundant representation (12) was reduced to the following:

\[
Y(\bar{x}) = \sum_{\forall i \in \{J\}} b_{j_i} x_i^j + f(\bar{x}) + E, \quad (27)
\]

where \( f(\bar{x}) = \sum_{\forall i \in \{J\}} \tilde{b}_{j_i} x_i^j \).

The regression problem without any changes (to an accuracy of the content of the columns of the matrix \( A \) in the expression \( (A^TA)^{-1}A^T \) of the general formula of the least squares method) can be solved by the method [14], since all unknown coefficients are included linearly in expression (27).

Remark 16. As shown in [14], the use of \( k \) repetitions of the main experiment significantly increases the computational efficiency of the MGMDH. Namely, the variance of estimates is reduced by a factor of \( k \), and only the matrices of the main experiment are used in the general formula of the least squares method.

Remark 17. The MGMDH can solve the problem of constructing an MPR given by a redundant representation based also on the results of a passive experiment. In this case, disappear only the advantages of an active experiment: the possibility of using repetitions of the main experiment.

Conclusions. 1. We considered the aggregated algorithmic scheme of the universal synthetic method of constructing an MPR given by a redundant representation. The synthetic method organically combines the decomposition method of estimating coefficients for nonlinear terms of the MPR with acceptable accuracy and the modified group method of data handling.

2. We presented theoretically justified classes of redundant representation of MPR, which allow to estimate fully or partially coefficients for nonlinear terms of an MPR with acceptable accuracy.

3. We substantiated the possibility of extending the modified group method of data handling, created for estimating the coefficients of an MLR given by a redundant representation, to the case of estimating the coefficients of an MPR given by a redundant representation.

References


13. 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. *Advances in Computer Science for Engineering and Education VI (ICCSSEA 2023)*. (у другій)


References (transliterated)

Вісник Національного технічного університету «ХПІ». Серія: Системний аналіз, управління та інформаційні технології, № 1 (9) /2023


Received 15.04.2023