УДК 519.71+510.67

V. Gubarev, V. Romanenko

STAGES AND MAIN TASKS OF THE CENTURY-LONG CONTROL THEORY AND SYSTEM IDENTIFICATION DEVELOPMENT. Part III. THE PROBLEM OF COMPLEX SYSTEMS IDENTIFYING BY INACCURATE DATA

Vyacheslav Gubarev

Space Research Institute of NAS of Ukraine and SSA of Ukraine, Kyiv,
orcid: 0000-0001-6284-1866

v.f.gubarev@gmail.com

Victor Romanenko

National Technical University of Ukraine «Igor Sikorsky Kyiv Politechnic Institute», Institute for Applied Systems Analysis, Kyiv,

ipsa@kpi.ua, romanenko.viktorroman@gmail.com

The article presents the main most noteworthy research results on the identification problem and methods that are widely used in solving identification problems. The most significant here is the problem of modeling especially complex systems, for which it is impossible, even in general, but in constructive form for identification, to write down a class of models in which the system under study has the most accurate description in which there is one of them adequate to the available data obtained from experiments. If the available knowledge on the object under study does not allow us to write down a suitable class of models in mathematical form, then asymptotic modeling should be used, as is done in computational mathematics. Instead of differential, integral and other equations, for example, algebraic equations approximating them are solved, the solutions of which approach the exact ones with increasing dimension and discretization step. Various forms of asymptotic expansions are used in identification, including rational approximation as applied to irrational and infinite-dimensional systems. However, in the presence of errors in the data, the use of asymptotic modeling leads to the fact that, with intentions to increase the accuracy of the resulting model due to its large dimension, in many cases the identification problems becomes incorrectly posed. Therefore, we have to look for a compromise or trade-off between bias and variance, which should determine the best-quality model. In addition to the modeling problem, the article discusses two concepts on the basis of which most existing methods for solving identification problems are based. Both of them give their own interpretation of the errors present in the data, namely, stochastic and nonstochastic. With the stochastic concept, errors are treated as a random process or sequence. The concept of estimation consistency is introduced, which makes it possible to use widely the theory of statistics in justifying methods for solving identification problems. The nonstochastic identification paradigm allows for arbitrary but bounded uncertainties in the da-

ta, i.e. belonging to some bounded sets. The largest number of different methods, as well as algorithms and software implementations, are made within the framework of the stochastic concept. Within the framework of nonstochastic identification, we mainly consider various modifications of the so-called subspace methods. In relation to complex systems, problems were noted that did not allow the development of sufficiently universal and effective methods for solving identification problems. A number of research areas have been pointed out that need further development in order to obtain more accurate models of complex systems.

Keywords: modeling, identification, stochastic identification, nonstochastic paradigm, consistency, error-adjusted regularized solutions.

Introduction

The past century of control theory and practice development is closely related to solution of another important problem: the identification of systems by data obtained in experiments. The point is that most existing control synthesis methods presuppose knowledge in one form or another of a mathematical model that describes the controlled process. At a certain stage, this seemed natural, since the existing knowledge in the field of mechanics, aerodynamics and other branches of physics made it possible based on the laws of conservation, to write equations for the dynamics of controlled processes in a form quite adequate to real processes. First of all, this applied to various aircrafts, electrical power objects, space systems and a number of others. In some cases it is possible to build mathematical models based on system analysis and the parameters of the corresponding equations were found by estimation methods based on observation results. However, as the application scope of control systems expanded or there is the desire to expand it one encountered certain difficulties in constructing mathematical models of controlled processes. That is why a new research direction arose to be called system identification. Moreover, for many of objects, it is impossible to write down a description adequate to the ongoing processes by any means other than identification. For example, these include processes that are implemented in steam-generators of heat power plants, including the processes of combustion, vaporization and heating to high parameters. To regulate effectively and control steam output parameters using controlled fuel supply, a mathematical model is needed to related the input and output variables. It is almost impossible to obtain it using existing knowledge in physics and chemistry because of the great complexity of the technical realization of power units of heat power plants. Therefore, further progress and expansion of the sphere of interests of control science depends on the successful solution of the identification problem.

In this article, we did not set ourselves the goal of writing a review of currently existing approaches, methods and results of research on the identification problem. The main attention is paid to such an extent to which this problem has been solved, and if there are unresolved or unexplored aspects of the problem, then why, for what feason.

Systems modeling

In essence, identification reduces to establishing regularities of behavior of systems under arbitrary permissible external influence on them. For this purpose, data obtained in specially designed experiments (active identification) or in observed and measured processes carried out naturally without human intervention (passive experiments) are used as initial. Data when identifying systems is actually trajectory, i.e. are formed as a pair $\{u_k\ y_k\}_{k=t_0}^{k=T}$, where u_k is the input at the moment $t=k$, and y_k is the measured output. Both input and output can be scalar single-input single-output (SISO-system) or vector multi-input multi-output (MIMO-system). In principle, the continuous case (continuous time) is not excluded, but in most cases it is discrete time. The observation in-

terval can be finite or trends to infinity, i.e. $T \to \infty$. Based on these trajectory data, the regularities of system behaviour is established on the basis of a mathematical model that is assumed to be adequate to them, i.e. in its corresponding model class there is such that the response of this model to any permissible input will be the same or sufficiently close to the response of the system to the same input. It is this circumstance that underlies almost all developed identification methods. Various criteria are formed by which this fact is established. The same models as in control synthesis are taken as model structures or classes. However, in order to link these models to trajectory data, it is necessary to write them in a form that allows one to pose explicitly and solve identification problems. One of these forms, most suitable for trajectory data, for example, is a description in the form of a convolution sum

$$y_k = \sum_{i=0}^{\infty} h_i u_{k-i}, \tag{1}$$

where in the general case y_t and u_t are vectors of dimension m and r, respectively, and h_i are matrices of dimension $m \times r$, i.e. «p, j» — the element of each of these matrices determines the response of the p-th output to the j-th input at the corresponding moment in time. If the data generating system has a finite dimension n,

$$x_{k+1} = Ax_k + Bu_k, \quad y_k = Cx_k + Du_k, \tag{2}$$

where x_k is a vector of dimension n, and the matrices A, B, C, D have the corresponding n, m, r dimensions. The connection between (1) and (2) is determined by the relations [1]

$$h_i = \begin{cases} 0, & t < 0, \\ D, & t = 0, \\ CA^{i-1}B, & t > 0. \end{cases} \tag{3}$$

Sometimes h_i is called Markov parameters of the system. In this case, the dimension n corresponds to the minimum realization that is controllable and observable. Knowledge h_i allows using (3) to find a model in the form (3) for minimal realization. Usually, for this purpose, a Hankel matrix of dimension $(n+2) \times (n+1)$ is formed from elements h_i, which, according to (3), is represented as the product of the observability matrix and the controllability matrix for model (2). Thus, we have a two-stage identification procedure. First, from (1) the sequence $\{h_k\}$ is found, which is then used to find all the parameters of the model (2) using the Hankel matrix. However, the most developed is the so-called 4SID method (Subspace-based state-space system identification), which allows one to solve directly the identification problem using the subspace methods and trajectory data, without going to $\{h_k\}$.

When identifying SISO systems, the most widely used models represented using the rational transfer function of the shift operator, i.e. in the form of fractional-rational functions that connect input and output in various ways. The most common among them are black box or Box-Jenkins models [5]

$$A(q)y_k = \frac{B(q)}{F(q)}u_k + \frac{C(q)}{D(q)}\varepsilon_k, \tag{4}$$

where $A(q)$, $B(q)$, $C(q)$, $F(q)$, $D(q)$ are polynomials with respect to the shift operator q, and ε_k are the existing data errors (assumed to be small). Depending

on which polynomials are preserved in (4) and which are identically equal to one, we obtain various families of models classified as ARX, ARMAX, ARMA, etc.

Models presented in the form of linear regression are very common.

Continuous systems are represented by models based on differential equations or transfer functions (matrices) obtained using the Fourier or Laplace transform [6].

A special class includes models for systems with distributed parameters. Nonlinear systems were also among the interests of the identification problem. As a rule, due to their great generality, they were written in a special form, classified as Wiener–Hammerstein models or their special cases: Wiener models and Hammerstein models [7].

The dimensions of the model or the number of parameters to be estimated when solving an identification problem determine its complexity.

When solving problems of identifying especially SISO systems, methods have been developed that are based on minimizing the prediction error (PE-methods) or the equation error (EE-methods). The criteria are formed from trajectory data over the observation interval [8].

In general, it is difficult to say which class of models mentioned above is most suitable for studying a system. Although many of the models considered in identification are equivalent, i.e. there is a transformation which allows one to pass from one description to another equivalent. However, the appropriate choice of the class of models is sought which contains the most adequate one to the available data this remains an unresolved issue. Here, results on the rational approximation of a wide variety of nuclear-type systems play an important role [9, 10]. These are considered to be those for which there is a Hankel operator with singular values σ_j satisfying the condition $\sum_{j=1}^{\infty} \sigma_j < \infty$. This is not such a strict condition and many physically realizable systems satisfy it. Belonging to nuclear-type systems means that their rational finite-dimensional approximation converges to an exact description according to a number of norms with increasing dimension. Based on this, for many complex systems it is possible to consider in identification the finite-dimensional models as approximating ones, allowing one to approach asymptotically an exact description as their complexity increases. Here we have a complete analogy with numerical methods that allow various approximations, including algebraic ones in the numerical solution of differential or integral equations. Thus, the classes of models described above can be interpreted as asymptotic modeling of complex systems for which it is impossible to write their mathematical model in another form. At the same time, the question remains open which class of asymptotic models is most suitable for the specific system under study, when using not very large dimensions it is possible to construct a high-quality approximate model. In practice, in a number of cases, a big experience working at the object under study and a systematic analysis of observation and measurement results help to do it. Thus, from the accumulated experience in operating power units of heat power plants it follows that the regression model with lag is preferable, i.e. qualitatively close to actual processes.

In recent years, modeling of various environmental, economic, organizational systems, as well as processes in finance, education and many other areas based on cognitive maps (CM) has been intensively developing. Mathematically, CMs are represented as oriented graphs, the vertices of which are coordinates or variables, and the edges establish a cause-and-effect relationship between the vertices using weighting coefficients. External influences are carried out through the vertices. In this case, the measuring system allows one to obtain information about the current values of the coordinates of certain vertices. The propagation of impulses along connected vertices is characterized as an impulse process. As a result, the mathematical model of the CM is reduced

to (2), where the weighting coefficients of the vertices specify the elements of the matrix A, and the matrices B and C for the implementation of the corresponding graph consist of zero and one elements. These features should be taken into account when solving the identification problem. How this should be done is described in [11].

Stochastic identification

The main problems that arise during identification are related to the presence of errors in the data. Accordingly, approaches to solving identification problems developed depending on how the existing uncertainty in the data was interpreted.

The greatest development was achieved by the so-called stochastic identification, when the errors were considered a random process such as white noise. Therefore, statistical methods are widely used in solving identification problems. The concept of assessment consistency was fundamental. The estimation was considered consistent if, with a probability of almost one, the solution to the identification problem converged to an exact solution when the number of data used increased unlimited. For this, the following assumptions were usually carried out.

Assumption 1. The system under study, generating data, is asymptotically stable.

Assumption 2. Uncontrolled disturbances acting at the input of the system and measurement noise at the output are ergodic random processes of the white noise type with zero mean and covariance matrices of the form

$$E\begin{bmatrix} w_t \\ e_t \end{bmatrix}\begin{bmatrix} w_s \\ e_s \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} r_w & r_{we} \\ r_{ew} & r_e \end{bmatrix},$$

where E is the mathematical expectation, w and e are disturbances at the input and measurement errors at the output (t, s are points in time), r_w, r_e, r_{we}, r_{ew} are covariance matrices, and T is the riverse operation. Moreover, w and e are uncorrelated with the input.

Assumption 3. The input action on the system is modeled as an arbitrary quasi-stationary deterministic process. The covariance matrix of the input is represented as

$$r_u = \bar{E}[u_t\, u_t^T],$$

where $\bar{E}$ is the averaging operator defined by the relation

$$\bar{E}[z_t] = \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} E[z_t].$$

Very often, a permanently exciting input is used as an input stimulus in active experiments. Its definition is given in many publications, see, for example, [1, 2, 8].

In order to obtain unbiased estimates (converging to exact ones), it is very important in what form the mathematical models are presented, among which one is searched for that is adequate to the generated data. For example, a prediction model in the form of an updating process is used [8–12]. In certain cases, to ensure the consistency of the assessment, the method of instrumental variables is used [13, 14].

For a long time, stochastic identification used the traditional approach. According to it, a certain parameterized structure of a class of models was considered

$$M = (M(\theta), \theta \in D_M), \tag{5}$$

where θ are the unknown parameters of the model, and D_M defines the set of models on which one is searched fore model that is adequate to the input data. In the classical identification theory, classes of models with different structures were considered: black

box models, state space models with a variable or specified dimension, as well as various polynomial models, ARX, ARMAX, Output Error (OE), Box-Jenkins (BJ) models of the given order polynomials defining a suitable structure, or with varying dimensions. A description of such models classes can be found, for example, in the fundamental monograph [8]. When the dimensions of models class (dimension θ) were established, the problem of their estimation was usually solved using the maximum likelihood method, (predictive error method) (ML/PEM), which reduced to the problem of minimizing the sum of prediction errors, i.e.

$$\min_{\theta \in D_M(\theta)} \sum_{t=1}^{N} \| y(t) - y^{(M)}(t) \|^2, \tag{6}$$

where $y(t)$ is the value of the measured output at point t, and $y^{(M)}(t)$ is the estimate of the model output for different θ. An element $\hat{y}(t)$ with values θ delivering a minimum (6) corresponded to the desired system model.

Great attention in studies of the identification process was paid to the choice of model structure, namely, its dimension or the dimensions of polynomials, i.e. parameters θ. Various procedures have been developed, such as cross validation, as well as a set of criteria known as the Akaike Information Criterion (AIC) or the Baysian order criterion (BIC). Based on these developments, software products were created that became part of Toolboxes, which have found application in solving many identification problems in practice [15].

The original is 4SID identification of the MIMO systems. It offers numerically reliable method to solve identification problem directly from measured data neither nonlinear search is performed, nor a canonical parameterization is used. The computational complexity is modest compared with PEM especially for large dimensional systems. It is based on the results of state-space realization theory, developed still in the 1960's. The primary source of the method can be considered the work of HO and Kalman [16]. Subsequently, a large number of articles and textbooks were published, in which the 4SID-method was developed for various cases and theoretically justified. Let us highlight three main points of the results obtained. The first is associated with the development of the direct 4SID method, when data included in the state-space model is used. To do this, using trajectory data over the observation interval from the standard description of the MIMO system in space, a matrix equation is formed

$$Y = \Gamma_\alpha X + \Phi_\alpha U, \tag{7}$$

where Γ_α is the observability matrix, and Φ_α is the Toeplitz matrix of impulse response, having the form

$$\Gamma_\alpha = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{\alpha-1} \end{bmatrix}, \; \Phi_\alpha = \begin{bmatrix} D & 0 & \dots & 0 \\ CB & D & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{\alpha-2}B & CA^{\alpha-3}B & \dots & D \end{bmatrix},$$

Y and U the blocks Hankel matrices of output and input are defined as

$$Y = [y(1)\, y(2) \dots y(N)], \;\; y(y) = [y_t^T \;\; y_{t+1}^T \dots y_{t+\alpha-1}^T]^{\mathrm{T}},$$

$$U = [u(1)\, u(2) \dots u(N)], \;\; u(y) = [u_t^T \;\; u_{t+1}^T \dots u_{t+\alpha-1}^T]^{\mathrm{T}},$$

X is a state trajectory matrix which has a view

$$X = [x_1\ x_2\ \dots x_N],$$

where x_t is a state vector for time t.

System equations (7) is derived from state-space model

$$x_{t+1} = Ax_t + Bu_t,\quad y_t = Cx_t + Du_t. \tag{8}$$

The second point is determined by separating from (7) the subspace of free motion. Two approaches are used for this. In the first of them (7) was multiplied on the right by the known matrix [1]

$$\Pi_U^{\perp} = I - U^{\mathrm{T}}(UU^{\mathrm{T}})^{-1}U.$$

For a non-singular matrix U, thanks to the condition $U\Pi_U^{\perp} = 0$ after multiplication, we obtain

$$Y\Pi_U^{\perp} = \Gamma_\alpha X\Pi_U^{\perp}. \tag{9}$$

Expression (9) can be obtained from (7) in another way, if we use the RQ decomposition of the composite matrix $\begin{bmatrix} U \\ Y \end{bmatrix}$, namely

$$\begin{bmatrix} U \\ Y \end{bmatrix} = \begin{bmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{bmatrix}\begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix}. \tag{10}$$

Using the orthogonality properties of matrices Q_1, Q_2, from (10) we can arrive at the relation [1]

$$Y\Pi_U^{\perp} = YQ_2Q_2^T = R_{22}Q_2^T,$$

i.e. thus highlight the subspace of free motion of the state-space system.

The third point is related to the original method of finding matrices A and C from (9). According to the theory of realizations, the left side of (9) is invariant under any non-singular transformation. Therefore, under the condition of full rank, any decomposition of it into the product of two matrices should give a matrix Γ_α and a matrix of transformed initial states $X\Pi_U^{\perp}$ for some realization. This is what was used in the HO and Kalman algorithm. For such an expansion, it is proposed to use the SVD decomposition of the left side of (9), i.e.

$$Y\Pi_U^{\perp} = Q\Sigma V^{\mathrm{T}}, \tag{11}$$

where Q and V are orthogonal matrices, and Σ is a matrix of singular numbers with non-zero positive numbers located on the diagonal in non-increasing order. If the system, for some n, has an exact description (8), and the data and all calculations were accurate, then the matrix Σ will have only n non-zero singular values, which allows you to uniquely determine the dimension of the model. With inaccurate data and calculation errors, the matrix Σ will be full-rank with a probability of almost one. However, if these inaccuracies are small, then we should expect that the first n singular values of the matrix Σ will be significant, and all the rest will be insignificant, i.e. corresponding to the errors.

Then you can implement partition

$$Y\Pi_U^{\perp} = Q_s\Sigma_s V_s^T + Q_n\Sigma_n V_n^T, \tag{12}$$

where Q_s contains n the leading left singular vectors, and the diagonal Σ_s has only n non-zero singular values. From (11) it is obvious that $\Sigma_n = 0$ in the absence of errors. However, in practice, the decision to allocate a non-full-rank matrix $Q_s\Sigma_s V_s^T$ can cause difficulties, which reduces the capabilities of this identification method. When this does not cause problems, the expression can be taken as an estimate of the matrix $\hat{\Gamma}_n$

$$\hat{\Gamma}_n = Q_s. \tag{13}$$

From (11) it is easy to obtain an estimate for the matrix C. These will be the first m rows of the matrix Q_s, and find the matrix A using the shift invariance of the matrix Γ_n, namely, from the equation

$$\Gamma_{1:n}A = \Gamma_{2:n+1}, \tag{14}$$

where $\Gamma_{2:n+1}$ is the matrix obtained from Q_s by eliminating the first rows equal to the number of rows of the matrix C, and $\Gamma_{1:n}$ is the last corresponding rows.

After finding the matrices A and C it is not difficult to find the matrix B [1–3] based on (7).

Over the past decade, great attention has been paid to building models that are as close to an accurate description as possible. A systematic analysis of the sources of errors that arise when solving identification problems identifies two main ones, namely: bias and variance. Bias errors occur when the model dimension or its complexity is chosen insufficiently large, such that the exact description does not fit it. Then regardless the amount and quality of the data a correct model cannot be obtained. So the distance between the system and the best model in this case is bias error in the model estimate. It is deterministic in nature and does not depend on disturbances in the available data.

Variance errors are generated by the errors present in the data. Even if a true model is contained in the chosen model class. This error depends on amount and quality of the data. Moreover, it depends on the dynamical properties of the system.

The total discrepancy contains both errors bias and variance. Then the resulting mean square error (MSE) can be written as

$$\text{MSE} = \text{BIAS2} + \text{Variance}$$

The minimum MSE value can be considered the best accurate solution to the identification problem. If such minimum exists it may be considered as BIAS-Variance trade-off in selecting appropriate model flexibility. In certain cases, the result obtained in this way turns out to be acceptable for practice. However, in many cases, for various reasons, the minimum turns out to be poorly expressed or produces a model or models that are unsuitable for practical use, i.e. they are surrogates. This is especially true for complex systems. This is observed when the input is such that a number of modes of the system give a response with a small signal/noise ratio (SNR) at the output, and also when, with increasing model dimension, very quickly grows the conditions number of informative matrices composed of data that have to be inverted when solving identification problems.

This led to the formulation and development of a new approach to solving identification problems about ten years ago [17]. It is based on two key points: input design and

kernel-based regularization. In work Mu and Chen [18] it is proposed two-step procedure for input design. In the first step a quadratic transformation of the input is constructed in such a way that input design problem reduces to optimal convex and their global minima can be found by existing methods. In the second step an expression for the optimal input is derived by solving the inverse image of the quadratic transformation found in the first step.

As was shown in [17], the optimal stabilizer $\Omega(\theta)$ when estimating a vector of parameters θ in the class of linear regression models has the form:

$$\Omega(\theta) = \sigma^2 \theta^{\mathrm{T}} (\theta_0 \theta_0^T)^{-1} \theta, \tag{15}$$

where σ is the dispersion of white noise additively present at the output, and θ_0 is the vector of exact parameters of the data-generating system. Since θ_0 is a priori unknown, it is impossible to use such a stabilizer in practice. Therefore, the main focus of this approach is how to select the stabilizer appropriately. It is proposed to split this procedure into two parts. In the first of them, the structure of the stabilizer Ω is selected, i.e. a parametric representation of the matrix Ω is written. It is usually called the Kernel of the estimation problem. It should be chosen taking into account the features of the vector θ expected when solving the estimation problem. Once the Kernel structure has been selected, the hyperparameters $\eta \in D_\eta$ (their feasible set) should be estimated using the available observations. This problem is of a purely statistical nature and does not depend on the real system. This task is interpreted as a tuning procedure.

Various forms of representing the Kernel structure can be found in [19, 20], and various methods of hyperparameter tuning in a statistical representation are described in [21–23].

This approach is considered in [17] for linear regression models but it can be extended to other classes of models used in identification. In some cases, the equivalence of different descriptions can be used.

The result is that the regularized estimate generally outperforms the standard techniques. It should be noted that when finding a regularized solution, one has to abandon the concept of unbiased estimates. However, in [17] it is shown that the bias obtained when finding it leads, with the correct choice of stabilizer, to significantly smaller errors than those resulting from poor conditioning of the matrices inverted in estimation problems.

Nonstochastic identification

In stochastic identification, fairly strict restrictions are imposed on the errors, which may not be fulfilled in practice. This stimulated the development of another approach, called nonstochastic identification. With this approach, no restrictions are imposed on the a priori unknown errors present in the data other than their belonging to some limited small sets. Then the maximum likelihood principle and all associated statistical methods cannot be used. In this case, other methods are used to find a solution to the identification problem under the specified uncertainty.

One of the first works appeared in the 90s of the last century was so-called Set Membership Identification (SMI) aiming to deliver not a single model of the system to be identified, but a set of models containing true model. Some of the results obtained in this direction are presented in [24–29]. However, the methods developed by SMI have not been widely used in practice because of their complexity and the limited class of models used for this. In addition, set membership can increase significantly with model dimension.

Another approach seems more promising, based on ideas used in the regularization of ill-posed problems. It is proposed to find an approximate solution of identification problems that is consistent in accuracy with errors in the data, and has the property of convergence to an exact solution when the errors in data and calculations tend to zero. As in the stochastic identification of complex systems, firstly, in the class of approximating models under consideration that asymptotically converge to the exact solution, the appropriate complexity is determined, and then the model parameters are estimated using known or modified methods. Most often, this is done iteratively, alternating the above actions. When the estimation problem becomes ill-posed, regularization is introduced and a regularized solution is sought that is consistent in accuracy with the available data [30]. In this case, it is assumed that the exact solution corresponds to the limit obtained as the dimension of the model tends to infinity. The result is always an approximate solution consistent with the errors. As the error in the data, decreases including in calculations, it will approach to the exact solution. The nonstochastic identification paradigm described above is fundamentally new. Some of its fragments are described in [31, 32]. Here we will demonstrate it more comprehensively and describe this approach using the example of solving the problem of non-stochastic identification in the class of state-space models. In recent years, work has been actively carried out on modeling complex systems of various natures based on cognitive maps. It is for them that discrete state-space models are most suitable [11].

Let's consider the modified 4SID method, which allows one to implement fully the stated concept in an active experiment. We will obtain data for identification when excitation intervals alternate with relaxation intervals during which there is no input signal. This can be realized in one experiment or, if permissible, in many experiments for which there are only two intervals: excitation and the following relaxation interval. Such experimental designs make it possible to generate informative data to an acceptable extent for each mode of the system. Each excitation interval uses different signals. In certain cases, inputs in the form of rectangular pulses of different durations and harmonic signals $(\sin\omega t, \cos\omega t)$ with varying frequency and duration can be informative.

At relaxation intervals of the same length l, we form a cascade vector of dimension lM from the output vectors

$$y(t_i, l) = (y^T_{t_i+1}\ y^T_{t_i+2} \dots y^T_{t_i+l})^{\mathrm{T}}, \tag{16}$$

where t_i is the last point of the i-th excitation interval. Then, taking into account that $u_t \equiv 0$ in the relaxation interval, from (7) it is easy to obtain the equation

$$y(t_i, l) = \Gamma_l x_{t_i}, \tag{17}$$

where x_{t_i} is the vector of the initial state at the i-th relaxation interval, and $\Gamma_l = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{l-1} \end{bmatrix}$ is the generalized observability matrix.

Let the total number of relaxation intervals be equal to K and $K >> l$, l exceeds the expected required dimension of the model n, i.e. $l > n$. From equation (17) for $i = 1, 2, \dots, K$ we compose the matrix equation

$$Y_{\text{relax}} = \Gamma_l X_0, \tag{18}$$

where $Y_{\text{relax}} = [y(t_1, l)\ y(t_2, l) \dots y(t_K, l)]$, $X_0 = [x_{t_1}\ x_{t_2} \dots x_{t_K}]$, $x_{t_i} = [x_{1_{t_i}}\ x_{2_{t_i}} \dots x_{n_{t_i}}]^{\mathrm{T}}$.

We will use equation (18) both to determine the dimension and to find the matrices A, C and X_0. This equation is the main for identification method under consideration. This is associated with the incorrectness of the problem as a whole. For the chosen way of the system excitation, finding the remaining matrices B and D, as a rule, does not cause any problems.

So, based on (18), let us consider all the main stages of the method. We represent the informational matrix Y_{relax}, the elements of which are measurement data of the output variable, using its SVD decomposition in the form

$$Y_{\text{relax}} = Q\Sigma V^{\text{T}}, \tag{19}$$

where Q and V are orthogonal matrices, and Σ is a matrix of singular numbers located on the diagonal in non-increasing order. With a probability of almost unit the matrix Σ will be full-rank. A typical picture of the values of singular numbers σ_i, $(i = 1, 2, \ldots, l)$ in the presence of measurement noise (bounded by ε) is shown in Fig. 1. The system generating the data was 19th order. Therefore, the first 19 singular numbers belong to the system, and the rest belong to noise. In asymptotic modelling there will be no such separation.

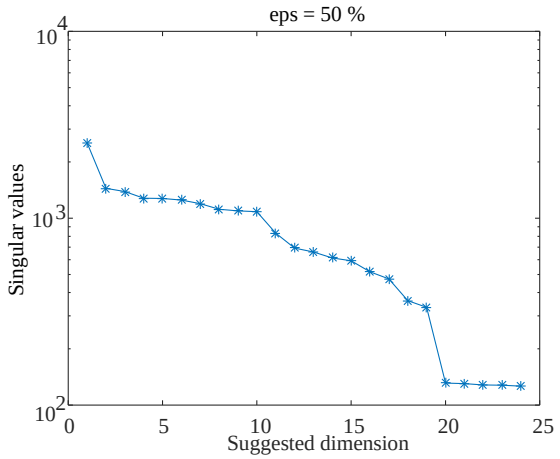

Fig. 1

All singular numbers contain a signal part and noise. At low noise, we consider large values of singular numbers relative to the noise level (set by restrictions) to be significant, since the SNR in the corresponding outputs will be significantly greater than unity. Minor singular numbers have an SNR order or less than one. In some cases, they can be separated by a well noticeable gap in the values of the singular numbers. However, in most cases, instead of a discontinuity, we have a transition zone between them, where the singular numbers attenuation without a noticeable discontinuity. In the first case, with a high probability, the dimension of the model is determined uniquely by the number of the last singular value before the discontinuity. In the second, the dimension will be found iteratively, starting from a certain value from which a rapid decay of singular numbers is visually noticeable or, conversely, the decay significantly slows down at a certain small singular value. Having thus chosen the model dimension, we solve for it the task of estimating its parameters. To do this, we write the matrix Y_{relax} in block form

$$Y_{\text{relax}} = Q_n\Sigma_n V_n^T + Q_e\Sigma_e V_e^T, \tag{20}$$

where Q_n and V_n are rectangular matrices composed of the first main n columns of the matrices Q and V, and Q_e and V_e — from the remaining ones; $\Sigma = \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma_e \end{bmatrix}$.

According to [33], such partition corresponds to extracting from a matrix Y_{relax} a matrix of rank n the closest in norm $\|\cdot\|_2$ to the original one. Next, in accordance with the subspace method, we represent the matrix $Q_n \Sigma_n V_n^T$ as a product of two matrices $Q_n \Sigma_n$ and V_n^T. The first of these two, according to the realizations [34], as well as in accordance with (10), represents the observability matrix $\Gamma_l^{(n)}$, and the second — the matrix of initial conditions X_0 for some realization, i.e.

$$\Gamma_l^{(n)} = Q_n \Sigma_n, \quad X_0^n = V_n^T. \tag{21}$$

Thus, there is a realization in which the initial states and the observability matrix are uniquely determined by expression (21). For this realization, knowing the observability matrix, we find the matrices A and C.

From the representation of the observability matrix it follows that its block containing the first M rows $\Gamma_l^{(n)}$ will be the matrix C for the same implementation, and the matrix A can be found from the following redefined system:

$$\Gamma_{1:l-1} A = \Gamma_{2:l}, \tag{22}$$

where $\Gamma_{1:n-1}$ is obtained from $\Gamma_l^{(n)}$ by crossing out its last M rows, and $\Gamma_{2:n}$ by crossing out $\Gamma_l^{(n)}$ its first M rows. It remains to find the matrix A from (22). To do this, we will use the SVD decomposition of the matrix $\Gamma_{1:l-1}$, presenting it in the form

$$\Gamma_{1:l-1} = Q_1 \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} V_1^T, \tag{23}$$

where Q_1 and V_1 are square orthogonal matrices of dimensions $(l-1)\times(l-1)$ and $n \times n$, respectively, and Σ_1 is a square matrix of dimension $n \times n$. Let's substitute (23) into (22) and get

$$Q_1 \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} V_1^T \cdot a_i = \gamma_i, \quad i = \overline{1, n}, \tag{24}$$

where a_i is the i-th column vector of the matrix A, and γ_i is the i-th column vector of the matrix $\Gamma_{2:l}$. Using the property of orthogonality of matrices Q_1 and V_1, we can reduce (22) to the form

$$\Sigma_1 \cdot \theta_i = \overline{\gamma}_i, \quad \theta_i = V_1^T a_i \text{ or } a_i = V_1 \theta_i, \quad \overline{\gamma}_i = Q_{1n}^T \gamma_i, \tag{25}$$

where Q_{1n} is formed from the first n principal singular vectors of the matrix Q_1.

The ill-posedness of problem (25) is exclusively related to the features of the singular values σ_i of the diagonal matrix Σ_1. If the problem is correct, then solution (25) has the form

$$a_i = V_1 \Sigma_1^{-1} Q_{1n}^T \gamma_i. \tag{26}$$

Otherwise, a regularized solution is found. To find it, you can use the regularization method described in [32] or its modification in the form of stabilizers having a Kernel structure [17], which have been actively developed in recent years in stochastic identification.

However, in relation to the method described above, let us pay attention to the following circumstance. Figure 2 shows the behaviour of singular values depending on the noise level determined by the constraint. The system generating the data was of order nine $(n=9)$. In the absence of noise, all singular numbers, starting from the tenth, practically (to within the computational error) are equal to zero. The presence of noise has made these singular values non-zero and increased, albeit slightly, the significance singular values corresponding to the system. Moreover, there is a high sensitivity of noise singular numbers to measurement errors. Therefore, starting from a certain sufficiently low noise level, the gap between σ_9 and σ_{10}, by which the dimension of the model was determined, decreased significantly, i.e. It becomes impossible to distinguish between essential (signal) and non-essential (noise) values based on the behaviour of singular values. The condition number of the parametric estimation problem has improved for $n=10$ and even for some subsequent values n, i.e. there is no need to use regularization. As a result, if we solve the estimation problem for $n=9$, then $n=10$, etc., we will obtain a poor-quality model due to the fact that the SNR of certain modes is not a large value. Classical regularization [30] will not improve the quality of the model, since it will introduce additional error. Perhaps regularization using the Kernel structure stabilizator [17] will have a positive effect, because in a certain sense it performs the functions of a filter. In these cases, the most appropriate option seems to be a better selection of the input signal at excitation intervals. For good identification of the matrix A, it is necessary to have a sufficient variety of initial conditions at relaxation intervals so that for each mode there is at least one interval where its signal would have a suitable SNR. This should lead to a slower decay of the singular numbers in Fig. 1 and a more noticeable gap between significant and unimportant singular numbers for any noise realized in experiments. In the case of asymptotic modelling, it is necessary that modes with a large SNR propagate to such dimension where, with accurate data, the approximate model has an error consistent with the noise level. How to implement this in each specific task remains an open question.

In multiconnected systems, it is also important to know which of the good or bad ones for identification we are dealing with. As good we mean one in which each simply connected block has its own set of significant modes, different from other blocks, which will determine its model. The complete model will have a dimensional transfer matrix $M \times R$ with different transfer functions as its elements. Then identification can be carried out independently for individual simply connected systems. The most unfavorable case for identification is when all elements of the transfer matrix are the same, i.e. each input excites all modes and at each output all modes are measures. This corresponds to the identity of both the measuring and exciting systems. However, in any case, it is advisable to reduce the identification of a multiply connected systems to solving MR simply connected systems and then combining them into one multiply connected system, taking into account that the simply connected subsystems thus obtained differ in some way from each other.

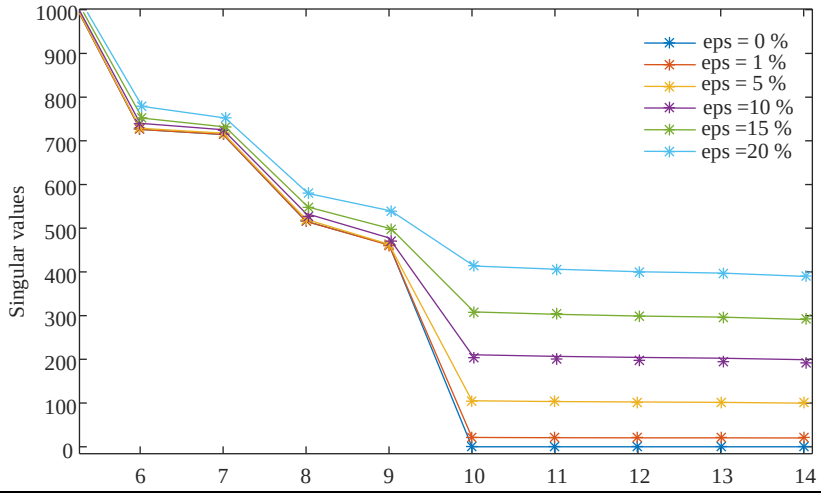

Fig. 2

In this case, the most likely outcome will be different matrix A dimensions. In addition, each SISO block will have different realization. Therefore, to find the matrix B, they should be reduced to one realization. The Jordan form is suitable as such. To go to it, you need to find the eigenvalues of all the resulting matrices A corresponding to simply connected systems. Then, for each m $(m=\overline{1,M})$, select close and different eigenvalues. It is advisable to combine close ones by adjusting both their common eigenvalue and other parameters associated with it.

This is necessary so that the refined model approximates transient processes no worse than the original one with similar roots. Otherwise, you will have to keep close but different eigenvalues. After this, a general model of the multiply connected system is formed. Its matrix will consist of all real values and complex blocks on the diagonal, composed of individual simply connected subsystems. The next step is to reconstruct the matrix C. It is built line by line, i.e. for each exit. All unobservable modes on each of them $(m<n)$ are assigned zero values to the corresponding row vector components c_m.

All other vector values c_m for the Jordan realization can be chosen quite arbitrarily. For matrices A and C a Jordan realization chosen in this way, it is not difficult to find a non-singular transformation P that connects it with the original one, defined by relations (22). To do this, we use the observability matrices for these two realizations, which are related by the relation

$$\Gamma' \cdot P = \Gamma, \quad X_0' = PX_0, \tag{27}$$

where Γ' is the observability matrix, and X_0' is the matrix of initial conditions for the Jordan realization. Accordingly, Γ and X_0 are determined by relations (21).

Let's consider a SISO system corresponding to the *m*-th output and *r*-th input of the original MIMO one. Let each excitation interval begin at the moment t_j, $j=\overline{1,K}$. On each of them we take the interval $[t_j, t_{j+l-1}]$, where $l >> n$, and n is the dimension of the corresponding selected SISO system. Since the lengths of the excitation intervals are different, they can be longer or shorter than n, we proceed as follows. Where they are smaller, we assume $u_r = 0$ at all points exceeding the moment of the end of excita-

ISSN 2786-6491

tion up to l. Where it exceeds l, we take only the first l values u_r. From the scalar outputs u_r formed in this way at each excitation interval we construct a cascade vector $u_r^{(j)}$ of the form

$$u_r^{(j)} = [u_r(t_j)\, u_r(t_j+1) \ldots u_r(t_j+l-1)]^{\mathrm{T}}, \quad j = \overline{1, K}.$$

From these vectors we form a matrix

$$U_r = [u_r^{(2)}\ u_r^{(3)} \ldots u_r^{(K)}]$$

dimensions $l \times K-1$. At the same intervals, by analogy with U_r, a matrix Y_{mr} of the m-th outputs is constructed under the influence of the r-th input

$$Y_{mr} = [y_{mr}^{(2)}\ y_{mr}^{(3)} \ldots y_{mr}^{(K)}].$$

Then the following matrix equations will b $Y_{mr} = \Gamma_l^{mr} \cdot X_0^{mr} + \Phi_l^{mr} \cdot U_r$, e valid at the indicated excitation intervals:

$$Y_{mr} = \Gamma_l^{mr} \cdot X_0^{mr} + \Phi_l^{mr} \cdot U_r, \tag{28}$$

where Γ_l^{mr} is the observability matrix of the m-th output of dimension $l \times n$, X_0^{mr} is the matrix of initial conditions of excitation intervals, i.e. at points t_j, dimension $n \times (K-1)$, Φ_l^{mr} is a lower triangular matrix, as well as a Toeplitz matrix of impulse responses, written for the SISO mr block system as

$$\begin{bmatrix} d_{mr} & 0 & \ldots & 0 \\ c_m^T b_r & d_{mr} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ c_m^T A^{l-2} b_r & c_m^T A^{l-3} b_r & \ldots & d_{mr} \end{bmatrix}.$$

Here d_{mr} is the mr-element of the matrix D, c_m^T is the m-th row vector of the matrix C, and b_r is the r-th column vector of the matrix B.

Now at excitation intervals it is necessary from the output data, i.e. from the matrix Y_{mr}, exclude free motion. To do this, we use estimates of vectors c_m^T and matrices A obtained from (22), i.e. for initial realization. From (21) for the same realization, the initial states at the relaxation intervals are also known. The model parameters obtained in this way make it possible to find an estimate of free movement at all excitation intervals except the first. As a result, we can move on to a matrix equation for determining elements d_{mr} and vectors b_r, which has the form

$$Y_{\text{forced}}^{mr} = \Phi_l^{mr} \cdot U_r, \tag{29}$$

where

$$Y_{\text{forced}}^{mr} = Y_{mr} - \hat{\Gamma}_l^{mr} \cdot \hat{X}_1^{mr},$$

accordingly $\hat{\Gamma}_l^{mr}$, the estimate of the observability matrix obtained on the relaxation interval, and $\hat{X}_1^{mr}$ is the initial value of free motion on the excitation interval.

The identification is completed by restoring the model of the multiply connected system. It will require correction of close eigenvalues obtained for SISO subsystems, and refinement of the components of the vectors c_m and b_r corresponding to the refined eigenvalues.

After this, it is quite easy to construct a model in the cases $R=1$ with arbitrary M, as well as arbitrary R and $M=1$. In other cases, more complex adjustment of the vector components c_m and b_r will be required even for the same eigenvalues. In this case, the quality of the model may deteriorate. This is due to the fact that with the established dimension of the model, the approximation gives a large error that is not consistent with the error in the data. It is necessary to increase the dimensions of SISO models and again carry out identification according to the scheme described above. This leads to an iterative scheme of increasing the complexity of the model until a consistent result is obtained. However, as the results of computational experiments show, starting from a certain dimension, the quality of the simulation does not improve and, moreover, may deteriorate. This means that all modes with a large SNR are already taken into account, and as the model becomes more complex, modes corresponding to the noise and disturbances present are added.

Identification in such cases can only be improved by increasing the SNR of modes with a weak signal, i.e. by increasing the informativity of the input. This can be judged by the behaviour of singular numbers, when a change in excitation leads to an increase in the number of significant singular numbers.

The described approach can also be used to identify systems in the MIMO class of regression models, which are generally written as

$$\theta X = Y, \tag{30}$$

where θ are to be estimated, X is the matrix of repressors, and Y is the matrix of output variables. MIMO system identification (30) can be reduced to MR scalar regression models. After this, using settings and refinement of parameters, we move on to a model of MIMO system.

Conclusion

Today it can be stated that, despite the large amount of research and development carried out, it is premature to talk about a complete solution of the identification problem. It was not possible in relation to complex systems, and for such systems, it is first of all necessary to build their mathematical models based on experimental data, formalize the identification problem and develop universal methods for solving them suitable for engineering practice. First of all, this concerns how it is necessary to excite systems in order to have informative data containing the response of all significant modes of the system with an acceptable SNR. Permanently exiting influence is certainly an effective way in many cases to ensure consistency of assessment. However, with poorly chosen excitation (control) and measurement systems of the object under study, many identification problems become ill-posed, which leads to low-quality (surrogate) models. Moreover, poor measuring and exciting systems can make the identification task completely meaningless.

Another important consideration is how to choose the appropriate class of models. It is unlikely that the selected class will contain an accurate description. Therefore, it is based on asymptotic modelling, when the description error tends to zero as the model

dimension increases. However, as the model becomes more complex due to the error in the data, the error in the approximate solution increases. As a result, a trade-off between these errors can occur with a sufficiently poor approximation, which naturally leads to a poor quality compromise solution.

Thus, identifying or establishing laws in the behaviour of dynamic systems is a poorly formalized process. In each specific task, you should develop your own original approach, taking into account the characteristics of the object under study, existing developments in theory and recommendations for their use, including standard software located in the appropriate Toolboxes.

В.Ф. Губарев, В.Д. Романенко

ЕТАПИ ТА ОСНОВНІ ЗАВДАННЯ СТОЛІТНЬОЇ ТЕОРІЇ КОНТРОЛЮ І РОЗРОБКА СИСТЕМИ ІДЕНТИФІКАЦІЇ. Частина 3. ПРОБЛЕМА ІДЕНТИФІКАЦІЇ СКЛАДНИХ СИСТЕМ ЗА НЕТОЧНИМИ ДАНИМИ

Губарев Вячеслав Федорович

Інститут космічних досліджень НАН України та ДКА України, м. Київ,

v.f.gubarev@gmail.com

Романенко Віктор Демидович

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Навчально-науковий інститут прикладного системного аналізу, м. Київ,

ipsa@kpi.ua, romanenko.viktorroman@gmail.com

Подано основні результати досліджень проблеми ідентифікації та методи, що широко використовуються при розв'язуванні задач ідентифікації і найбільше заслуговують на увагу. Значущою тут є проблема моделювання особливо складних систем, для яких не можна навіть у загальному, але конструктивному для ідентифікації вигляді записати клас моделей, в якому система, що досліджується, має найбільш точний опис, і в ньому існує одна з них, адекватна наявним даним, отриманим з експериментів. Якщо наявні знання про об'єкт, що досліджується, не дають можливості записати в математичному вигляді відповідний клас моделей, то слід використовувати асимптотичне моделювання, як це робиться в обчислювальній математиці. Замість диференціальних, інтегральних та інших рівнянь розв'язуються, наприклад, алгебраїчні рівняння, що їх апроксимують, розв'язки яких наближаються до точних зі збільшенням розмірності та кроку дискретизації. В ідентифікації використовуються різні форми асимптотичних розкладів, зокрема раціональна апроксимація стосовно ірраціональних і нескінченновимірних систем. Однак за наявності похибок в даних використання асимптотичного моделювання призводить до того, що при намаганні підвищити точність одержуваної моделі за рахунок її великої розмірності в багатьох випадках задачі ідентифікації стають некоректно поставленими. Тому доводиться шукати компроміс (trade-off) між bias та variance, що має визначати найбільш якісну модель. Крім проблеми моделювання, у статті розглянуто дві концепції, на основі яких побудовано більшість існуючих методів розв'язання задач ідентифікації. Обидві дають свою інтерпретацію присутніх у даних похибок, а саме, стохастичну та нестохастичну. При стохастичній концепції похибка трактується як випадковий процес чи послідовність. Вводиться поняття спроможності оцінювання, що дає змогу широко вико-

ристовувати теорію статистики при обґрунтуванні методів розв'язання задач ідентифікації. Парадигма нестохастичної ідентифікації допускає довільні, але обмежені невизначеності в даних, тобто такі, що належать деяким обмеженим множинам. Найбільше різних методів, а також алгоритмів і програмних реалізацій створено у рамках стохастичної концепції. У межах нестохастичної ідентифікації розглянуто переважно різні модифікації так званих subspace methods. Щодо складних систем відмічено проблеми, які не дали змоги розробити досить універсальні та ефективні методи розв'язання задач ідентифікації. Виділено низку напрямів, які підлягають розвитку з метою отримання точніших моделей складних систем.

Ключові слова: моделювання, ідентифікація, стохастична ідентифікація, нестохастична парадигма, спроможність, регуляризовані розв'язки, узгоджені з похибкою.

REFERENCES

1. Viberg M. Subspace-based methods for the identification of linear time-invariant systems. *Automatica*. 1995. Vol. 31, N 12. P. 1835–1851. DOI: https://doi.org/10.1016/0005-1098(95)00107-5
2. Verhaegen M., Dewilde P. Subspace model identification. Part 1. The output-error state space model identification class of algorithms. *International Journal of Control*. 1992. Vol. 56, N 5. P. 1187–1210. DOI: https://doi.org/10.1080/00207179208934363
3. Verhaegen M., Dewilde P. Subspace model identification. Part 2. Analysis of the elementary output-error state space model identification algorithm. *International Journal of Control.* 1992. Vol. 56, N 5. P. 1211–1241. DOI: https://doi.org/10.1080/00207179208934364
4. Verhaegen M., Dewilde P. Subspace model identification. Part 3. Analysis of the ordinary output-error state-space model identification class of algorithm. *International Journal of Control.* 1993. Vol. 58, N 3. P. 555–586. DOI: https://doi.org/10.1080/00207179308923017
5. Box G.E.P., Jenkins G.W. Time series analysis, forecasting and control (2nd ed.). Holden Day, San Francisco, USA, 1970.
6. Unbehauen H., Rao G.P. Continuous-time approaches to system identification — a survey. *Automatica*. 1990. Vol. 26, N 1. P. 23–35. DOI: https://doi.org/10.1016/0005-1098(90)90155-B
7. Grzegorz Mzyk. Combined parametric-nonparametric identification of bloc-oriented systems. Springer International Publishing Switzerland, 2014. 238 p. DOI: 10.1007/978-3-319-03596-3
8. Ljung L. System identification — theory for the user (2nd ed.). Upper Saddle River, NJ : Prentice-Hall. 1999.
9. Glover K., Curtain R.F., Partington J.R. Realization and approximation of linear infinite-dimensional systems with error bounds. *SIAM Journal of Control and Optimization.* 1988. Vol. 26, N 4. P. 863–898. DOI: https://doi.org/10.1137/0326049
10. Curtain R.F., Precopa A., Selection J. Straznicky B. Balanced realisations for infinite-dimensional discrete-time systems. *System Modelling and Optimization*. 2006. Vol. 84. P. 181–187. DOI: 10.1007/BFb0043837
11. Gubarev V.F., Miliavsky Yu.L. Features of cognitive maps modeling and identification under uncertainty. *Cybern. Syst. Anal*. 2023. Vol. 59, N 4. P. 546–560. DOI*:* https://doi.org/10.1007/s10559-023-00590-2
12. Verhaegen M. Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. *Automatica. Special Issue on Statistical Signal Processing and Control.* 1994. Vol. 30, N 1. P. 61–74. DOI: https://doi.org/10.1016/0005-1098(94)90229-1
13. Newey W.K., Powell J.L. Instrumental variable estimation of nonparametric models. *Econometrica*. 2003. Vol. 71, N 5. P. 1565–1578. DOI:10.1111/1468-0262.00459
14. Söderström T., Stoica P. Instrumental variable methods for System identification. New York, NY : Springer-Verlag, 1983. 252 p.
15. Ljung L. System identification toolbox for use with MATHLAB. Version 7. The MathWorks, Inc., Natick, MA, 2007.
16. Ho B., Kalman R.E. Efficient construction of linear state variable models from input-output functions. *Regelungstechnik.* 1966. Vol. 14. P. 545–548.
17. Ljung L., Chen T., Mu B. A shift in paradigm for system identification. *International Journal of Control*. 2020. Vol. 93, N 2. P. 173–180. DOI: https://doi.org/10.1080/00207179.2019.1578407
18. Mu B., Chen T. On the input design for regularized LTI system identification: power constrained input. *Automatica.* 2018. Vol. 97. P. 327–338. DOI: https://doi.org/10.1016/j.automatica.2018.08.010

19. Pilonetto G., De Nicolao G. A new kernel-based approach for linear system identification. *Automatica*. 2010. Vol. 46, N 1. P. 81–93. DOI: https://doi.org/10.1016/j.automatica.2009.10.031
20. Chen T. On kernel design for regularized LTI system identification. *Automatica.* 2018. Vol. 90, N 1. P. 109–122. DOI: https://doi.org/10.1016/j.automatica.2017.12.039
21. Chen T., Ljung L. Implementation of algorithms for tuning parameters in regularized least squares problems in system identification. *Automatica*. 2013. Vol. 49, N 7. P. 2213–2220. DOI: https://doi.org/10.1016/j.automatica.2013.03.030
22. Mu B., Chen T., Ljung L. On asymptotic properties of hyperparameter estimations for kernel-based regularization methods. *Automatica*. 2018. Vol. 94. P. 381–395. DOI: https://doi.org/10.1016/j.automatica.2018.04.035
23. Pilonetto G., Dinuzzo F., Chen T., De Nicolao G., Ljung L. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica.* 2014. Vol. 50, N 3. P. 657–682. DOI: https://doi.org/10.1016/j.automatica.2014.01.001
24. Milanese M., Norton J., Piet-Lahanier H., Walter E., eds. Bounding approaches to system identification. New York : Plenum Press, 1996.
25. Giarre L., Milanese M. Model quality evaluation in H2 Identification. *IEEE Transactions on Automatic Control*. 1997. Vol. 42, N 5. P. 691–698, May,1997. DOI: 10.1109/9.580876
26. Chisci L., Garulli A., Vicino A., Zappa G. Block recursive parallelotopic bounding in set membership identification. *Automatica*. 1998. Vol. 34, N 1. P. 15–22. DOI: https://doi.org/10.1016/S0005-1098(97)00160-X
27. Goodwin G.C., Gevers M., Ninness B. Quantifying the error in estimated transfer functions with application to model order selection. *IEEE Transactions on Automatic Control*. 1992. Vol. AC-37, N 7. P. 913–928. DOI: 10.1109/9.148344.
28. Milanese M. Learning models from data: the set membership approach. *Proceedings of the 1998 American Control Conference. ACC (IEEE Cat. No.98CH36207)*. Philadelphia, PA, USA, 1998. Vol. 1. P. 178–182. DOI: 10.1109/ACC.1998.694653
29. Giarre L., Kacewicz B.Z., Milanese M. Model quality evaluation in set membership identification. *Automatica*. 1997. Vol. 33, N 6. P. 1133–1139. DOI: https://doi.org/10.1016/S0005-1098(97)00007-1
30. Tikhonov A.N., Arsenin V.Yu. Solutions of ill-posed problems. Washington, DC: Winston/ Wiley, 1977.
31. Gubarev V.F. Modeling and identification of complex systems. Kyiv : Naukova Dumka, 2019. 248 p. (in Ukrainian).
32. Gubarev V., Melnychuk S., Salnikov N. Paradigm of nonstochastic approach to system identification. *Міжнародний науково-технічний журнал Проблеми керування та інформатики*. 2023. N 1. P. 42–58. DOI: https://doi.org/10.34229/1028-0979-2023-1-4
33. Golub G.H., Van Loan Ch.F. Matrix computations (2nd edition). The John Hopkins University Press, 1989. 642 p.
34. Kailath T. Linear systems. Prentice-Hall. Englewood Cliffs, N.J., 1980. 680 p.

Submitted 12.12.2023