UDC 681.513

Sorting Methods in Self-Organization of Models and Clusterizations (Review of New Basic Ideas) Iterative (Multirow) Polynomial GMDH Algorithms

A. G. IVAKHNENKO
The V. M. Glushkov Cybernetics Institute of the Academy of Sciences of the Ukrainian SSR

When solving the parametric identification problem, one must find the estimates of polynomial model coefficients by processing a sample of experimental data.

The readers are, of course, sure that the number of points in the data sample cannot be smaller than the number of members of the polynomial to be estimated. But this is incorrect for iterative (or multi-row) identification algorithms where, instead of the complete polynomial

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_Mx_M$$

one subjects to estimation only partial polynomials of the form

$$y = c_ix_i + c_jx_j \quad (\text{for } x_1 = 1),$$

where x_i, x_j is any pair of regressors taken from the regressor set $x_1, x_2, \ldots, x_M$ indicated in the data sample. The algebraic minimum of data points for estimating the coefficients c_i and c_j in the partial polynomials is two. Following this, we can obtain, by solving simultaneously the system of partial equations the estimates of all the coefficients in the initial complete polynomial that has $M + 1$ members. It is easy to show that this "miracle" is valid under ideal conditions, i.e., when the data are complete and precise. Under actual practical conditions it can also occur, but only with some degree of precision.

Using an iterative GMDH algorithm, one can evaluate a polynomial that is linear in its coefficients and has about 1000 members from a sample that contains on the order of 100 points only. The computation time on a BESM-6 computer (10^6 operations/s) in this case does not exceed 3 hours ("freedom of choice" $F = 3$).

The iterative multi-row algorithm is explained by Fig. 1.

EXTERNAL CRITERIA OF TWO TYPES AND TWO BASIC APPROACHES

The application of external criteria at each iteration step with the aim of selecting only F most effective partial polynomials is a specific feature of GMDH algorithms. In accordance with the well-known D. Gabor paper on sequential decision making, the number F is called "freedom of choice" [1, 2].

External means that it is based on fresh information that has not been used for estimating the coefficients. The data sample is partitioned into two parts, A and B, in order to calculate the so-called external regularity criterion

*From the Editorial Board. The Editorial Board disagrees with many of the statements and evaluations of the author of this paper, and publishes it to stimulate discussion, assuming that the discussed issues are of interest to a broad circle of readers.

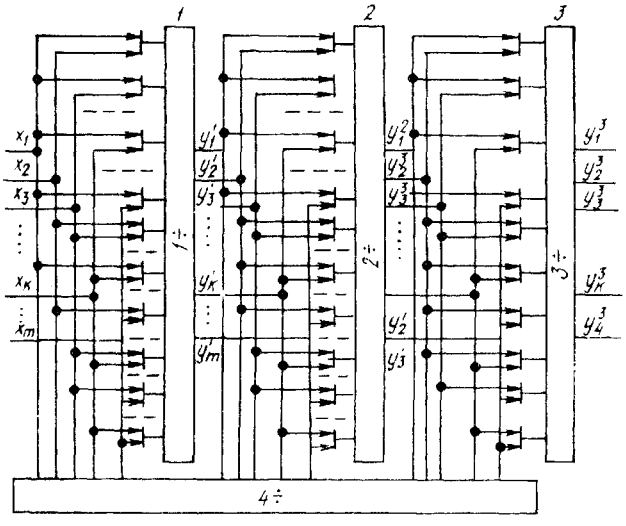

Fig. 1. Iterative (multi-row) GMDH algorithm; 1) first selection; 2) second selection; 3) third selection, etc.; 4) data sample.

$$AB = \sum_{i=1}^{B} (y_B - \hat{y}_B)^2 \to \min$$

or the differential external consistency criterion

$$CN = \sum_{i=1}^{N} (\hat{y}_A - \hat{y}_B)^2 \to \min,$$

where N is the total number of points in the data sample; B is the testing subset of points, $N = A + B$; y_B is the output variable represented in data sample B; y_A, y_B are the outputs of the models with the structure that is subject to evaluation.

Many multi-loop systems with a large number of variables have special invariants: the sum of specific variables is a constant number [3]. For example, in the simplest case, the sum of season averages equals the annual average: $q_1+q_2+q_3+q_4=Q$. Such invariants allow one to arrange a balance of variable criterion

$$b_i = Q_i - (q_1 + q_2 + q_3 + q_4),$$

$$BL = \sum_{i=1}^{N} b_i^2 \to \min.$$

The balance of variables criterion BL and the consistency criterion CN can serve as examples of differential-type criteria that do not include information taken directly from the data sample. Differential-type criteria are incorrect (in the sense of A. N, Tikhonov papers). They can equal zero only for the unique optimal model but also accidentally for some "false" ones. The false "zeros" must be detected and eliminated using a special regularization (or redefinition) procedure in order to find the unique optimal model.

All the criteria (about 50 proposals are known in addition to the informational-type criteria) can be separated into two basic groups.

Accuracy type criteria require the choice of a model that is most accurate on the given data sample. The regularity criterion can serve as an example of an accuracy type criterion. These criteria implement a precision approach in modeling by sorting.

Difference type criteria require the choice of a model that is the same for the two different parts of the sample. The consistency and balance of variables criteria can serve as examples

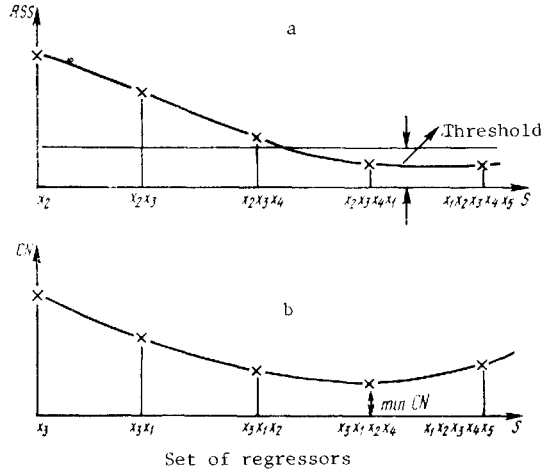

Fig. 2. Regressor ranking for selecting the optimal ensemble: a) subjective method for setting the threshold; b) objective method (by the minimum of an external criterion).

of difference type criteria that implement a new robust approach in modeling by sorting. Robust means here insensitive to the choice of the sample part that is used for modeling.

THE OBJECTIVE NATURE OF THE CHOICE OF THE OPTIMAL MODEL STRUCTURE ACCORDING TO EXTERNAL CRITERIA

External accuracy and differential type criteria have *a* very important feature. When the complexity of the model structure (determined, for example, by the number of members of the polynomial) is increased gradually during the sorting procedure, the external criteria pass through their minimum. The location of the criterion minimum indicates the optimal complexity of the model structure (the self-organization principle for optimal models). The minimum of the external criterion offers us the capability of objective choice of the model.

Internal criteria used in mathematics, for example, the root-mean-square error calculated on all the sample points; RSS (residual sum of error squares), only decrease with increasing model complexity. Here the rule holds: the more complex the model, the more accurate it is. The only way out under such conditions is to set some threshold ("confidence interval") in order to stop the sorting procedure on some complexity level (Fig, 2). One has to be experienced and to have deep knowledge of the modeled object in order to select the threshold value correctly. But there is no other way out if one does not want to use methods of model self-organization according to external criteria.

EXTRAPOLATION OF THE LOCUS OF THE MINIMA IN ORDER TO FIND THE PHYSICAL MODEL MEASURE OF THE NOISE IMMUNITY OF THE CRITERION

A computer chooses (in the presence of interference) under-complicated models as the optimal ones. However, the problem of finding the full physical model in modeling by sorting remains. It can be solved by extrapolating the locus of the minima (LM) of the external criterion, as shown in Fig. 3. The point *0* corresponds to the physical model. It can be determined at the intersection of the LM approximation and the abscissa axis. It is sufficient to add some small portions of noise to the data sample in order to find some points of the LM in order to approximate and extrapolate it. The magnitude of θ_0 is a measure of the noise immunity of the criterion.

Vector, matrix and tensor forms of the model. Algorithms for system analysis (OSA). The same object can be described by a vector, a matrix or a tensor model. The form that provides the deepest minimum of the criterion is the best. The computer chooses the model form.

In this way one can solve the humanly very difficult problems of selecting the limits of the modeling domain. The computer indicates which elements must be included in the model and

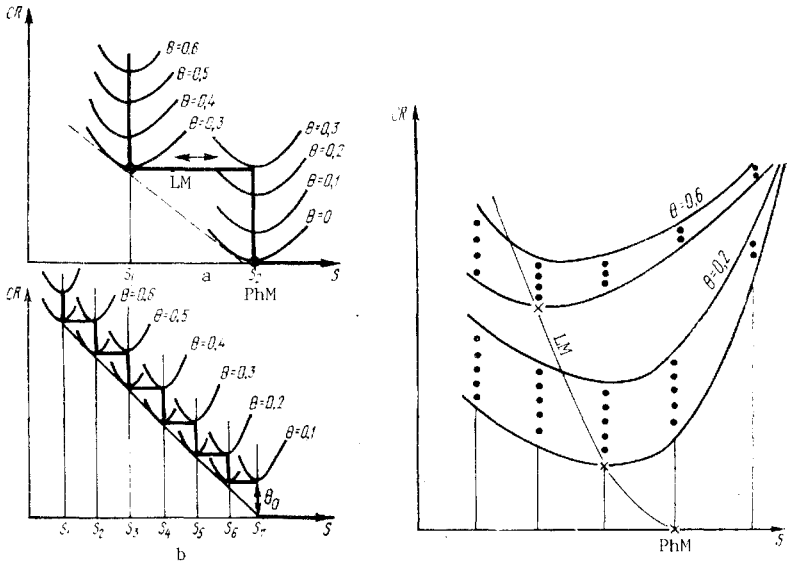

Fig. 3. Extrapolation of the LM for finding the structure of a physical model: a) for two models, S_1, S_2; b) for seven models, S_1, ..., S_7; θ — interference level ($\theta < \sigma^2$).

Fig. 4. Results of a computational experiment of sorting systems of difference equations for interference levels $\theta = 0.2$ and $\theta = 0.6$.

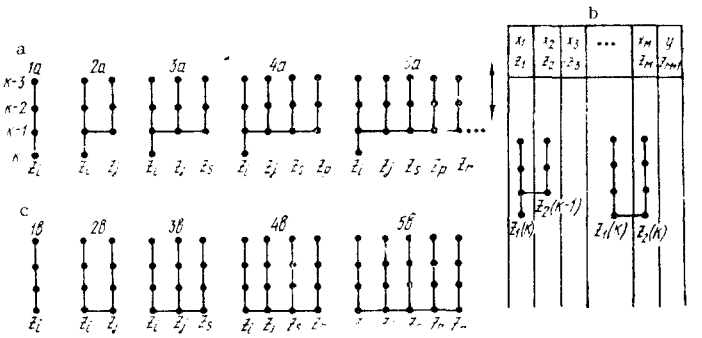

Fig. 5. Ranking of explicit, (a), and implicit, (b), templates S, a method for teaching templates by a data sample.

which must be excluded. A special sorting algorithm which solves this problem is called the Objective System Analysis (OSA) algorithm [4]. Implicit templates (Figs. 4 and 5), i.e., systems of difference equations, are subject to sorting in this algorithm.

WHY IT IS IMPOSSIBLE TO APPLY EXISTING GENERAL IDENTIFICATION THEORY TO THE SOLUTION OF PROBLEMS IN MODELING BY SORTING

The practice of modeling by sorting poses many special questions of theoretical nature. For example, does the iterative procedure of sorting models converge in general? Is the convergence guaranteed only for internal criteria (internal convergence) or also for external criteria (external convergence)? Which method of partitioning the data sample into two sub-samples

A and B is optimal? Under what conditions is unimodality of the external criteria guaranteed and what order of ranking models-candidates is necessary for attaining the unique minimum? How great is the noise immunity of the sorting procedure?

There is a large number of papers on the general identification theory, expecially for the purposes of automatic control [5]. But there are no answers there to the questions stated in the preceding. A special theory of modeling by sorting has to be developed. The situation can be explained as follows: identification theory has been developed in a purely deductive way. All the efforts of this development were aimed at the search for unbiased models, and, in particular, for the simplest unbiased minimal-complexity physical model. The practice of model sorting shows that a computer, controlled by external criteria, chooses under-complicated models that are optimal for approximation as well as the prediction. In presence of interference, the optimal model must have a simpler structure than its physical counterpart.

We can find a similar assertion in communications theory: there, as the signal noise increases, the communications system becomes simpler (Shannon's second theorem for noisy communications channels).

A human cannot invent and propose under-complicated models for prediction, this is excluded psychologically. He or she can consider only the application of physical or over-complicated models, but not of under-complicated ones. Can a human agree to describe pendulum oscillations by only two or one addend of the full three-member equation of oscillations? Of course not! But sorting shows that under noisy conditions, it is useful to exclude some members of the complete equation of the physical model.

TWO METHODS FOR DEVELOPING SORTING METHODS FOR MODELING

The special theory of sorting methods is developed in two ways: mainly, by using computational experiments (that are repeated many times to increase their credibility), or, for some problems, by using ordinary analytic methods.

We will consider the basic results of the analytic investigations.

Convergence of iterative GMDH algorithms. The proof of the convergence of the iterative procedure in GMDH algorithms [6, 7] to a point determined by the minimum of the sorting criterion is an obvious result of analytic research in its general form.

The unimodal nature of the external criterion characteristic. It has been shown analytically that if the data sample is very large, then the "criterion - model complexity" characteristic is unimodal. In stochastic problems, unimodality is guaranteed for the mathematical expectation of the criterion.

Computational experiments show that unimodality is achieved also when the models-candidates are sorted in groups (clusters). At the first step one can use the entire information that is contained in any single column of the sample, at the second step one can use all the information that is contained in any two columns of the data sample, etc. The best models of each step form the unimodal characteristic.

When modeling in the form of difference equations, this rule means that implicit templates of the equations become more complex gradually (Fig. 5).

Computational experiments have demonstrated the advantage of the reverse sorting of templates. First, the most complex template is evaluated, then a simpler one, etc. It has been noted that unimodality is lost very often for the simplest templates ("first-row effect"). Reverse sorting excludes this effect because the sorting procedure stops as soon as the criterion minimum starts to increase. Experiments have shown that unimodality is achieved also for short samples. Theoreticians cannot tell us yet when it appears.

THE PARAMETRIC IDENTIFICATION PROBLEM. NON-PARAMETRIC MODELS

The least squares method (LSM) is applied to the estimation of the coefficients of polynomial models. But it is known that it yields optimal unbiased estimates only for a full number of regressors and when interference acts only on the output variable. In iterative GMDH algorithms the set of variables changes in each step. Under such conditions it is better to use orthogonal polynomials that are optimal when the interference affects all the regressors identically. A few sorting algorithms with orthogonal partial descriptions have been proposed [8].

It has been shown theoretically that when the interference variance can be measured, it is best to apply minimax estimates of the coefficients [9].

However, the possibility of excluding the entire problem exists in modeling by sorting methods. Many algorithms are known in which not polynomials but, for example, Bayes formulas [10], correlation functions [11], or Markov chains [12, 13] are subject to sorting. Very good practical results are shown in papers devoted to non-parametric methods of model sorting. For example, the pollution field in Guinea Bay has been forecast more than a year in advance and proved to be very accurate [14].

Polynomial GMDH algorithms are effective, but non-parametric ones promise to be even more effective. Thus, polynomial models provide accurate forecasting of eleven variables in the model of the development of the economy of the GDR. The remaining 15 variables are forecast using non-parametric programs based on the search for an analog in prehistory [10].

THE PROBLEM OF PARTITIONING THE DATA SAMPLE INTO SUBSAMPLES *A* AND *B*. ASSOCIATION WITH THE THEORY OF INSTRUMENTAL VARIABLES

Partitioning the data sample into two subsamples, *A* and *B*, is one of the methods for calculating external criteria. One sets $A=\frac{2}{3}N$, $B=\frac{1}{3}N$ for the regularity criterion, and $A = B$ for the consistency criterion. The difference between subsamples *A* and *B* can be measured using analysis of variance. It is easy to show that in modeling subsamples *A* and *B* must differ from each other as much as possible. This guarantees a fast increase of the "criterion - model complexity" characteristic and, consequently, the deepest minimum.

Changing the content ("pouring over") of points in subsamples *A* and *B* is one of the methods of obtaining a unique minimum (the regularization method). Adding slight noise to the initial data sample is another method.

It has been shown theoretically that when the data sample is full and precise (no interference), all the differential-type criteria (i.e., the consistency criterion, the balance of variables criterion, etc.) do not work [14]. This means that the number of minima or "false" zeros is too large. Differential-type criteria can work (i.e., provide the unique true minimum) only when (1) there is interference in the data sample or the number of regressors if incomplete; and (2) subsamples *A* and *B* differ in their variances.

The theory of instrumental variables yields a new Idea for modeling by sorting. Instead of partitioning the sample, it is better to apply two different methods for obtaining the instrumental variables and compare the results. The difference between the two instrumental variables is zero for the optimal model. But instrumental variables obtained in this way must be based on different informational bases. Quantization of a data sample into two different numbers of levels (ranks) and comparison of results [10] can serve as an example.

DESCRIPTION OF A MODIFIED OSA ALGORITHM THAT DOES NOT REQUIRE THE DATA SAMPLE TO BE PARTITIONED INTO TWO PARTS

The objective is a choice of a system of difference equations that describes an object the data about which are represented by an $N \times M$-element sample of measurements (N is the number of observation points, *M* is the number of variables or attributes. The output variables are not indicated *a priori* in the sample; they are found by the algorithm as a consequence of choosing the system of equations. The set of system-candidates is sorted according to two criteria. The optimal system of difference equations must correspond to the minimum of the consistency criterion and satisfy the threshold value of the short-term prediction (or variation) criterion $AC=\delta^2\leqslant 1.0$.

Block 1. Construction of the auxiliary samples *A* and *B* of two instrumental variables. To avoid the partitioning of the sample into two parts, one forms two samples *A* and *B* by quantizing the data of the basic sample (except for the points of a small testing sample *C*). Sample *A* is obtained by quantizing the data of the initial sample to *N* levels. Sample *B* is obtained by quantizing it to *N/2* levels (having multiplied the data by the factor of 2).

Many other proposals for forming two instrumental variables for solving the criterion (see, for example, [15], p. 16) are known. The use of various Walsh functions is a promising approach. The consistency criterion (in a combinatorial GMDH algorithm) is determined by the difference in the outputs of two models constructed on the instrumental variables:

$$CN=\frac{1}{N-C}\sum_{i=1}^{C}(\hat{x}_A-\hat{x}_B)_i^2\to\min.$$

To calculate the criterion, one has to determine the models on samples *A* and *B* that are obtained by the indicated method.

Block 2. Determination of the structure of the polynomial difference equations according to the GMDH combinatorial algorithm and filtering out those which carry disinformation (instead of prediction). The following complete polynomials that take into account two delayed arguments (i,j,p,r,s are the indices of the variables) are subject to the sorting.

The first-row polynomial

$$x_{i(k)} = a_0 + a_1 x_{i(k-1)} + a_2 x_{i(k-2)}.$$

The second-row polynomial

$$x_{i(k)} = a_0 + a_1 x_{i(k-1)} + a_2 x_{i(k-2)} + a_3 x_{j(k)} + a_4 x_{j(k-1)} + a_5 x_{j(k-2)}.$$

The third-row polynomial

$$x_{i(k)} = a_0 + a_1 x_{i(k-1)} + a^2 x_{i(k-2)} + a_3 x_{i(k)} + a_4 x_{j(k-1)} + \\ + a_5 x_{j(k-2)} + a_6 x_{p(k)} + a_7 x_{p(k-1)} + a_8 x_{p(k-2)}.$$

The fourth-row polynomial

$$x_{i(k)} = a_0 + a_1 x_{i(k-1)} + a_2 x_{i(k-1)} + a_3 x_{j(k)} + a_4 x_{j(k-1)} + a_5 x_{j(k-2)} + a_6 x_{p(k)} + \\ + a_7 x_{p(k-1)} + a_8 x_{p(k-2)} + a_9 x_{r(k)} + a_{10} x_{r(k-1)} + a_{11} x_{r(k-2)}.$$

The fifth-row polynomial

$$x_{i(k)} = a_0 + a_1 x_{i(k-1)} + a_2 x_{i(k-2)} + a_3 x_{i(k)} + a_4 x_{j(k-1)} + a_5 x_{j(k-2)} + a_6 x_{p(k)} + \\ + a_7 x_{p(k-1)} + a_8 x_{p(k-2)} + a_9 x_{r(k)} + a_{10} x_{r(k-1)} + a_{11} x_{r(k-2)} + a_{12} x_{s(k)} + \\ + a_{13} x_{s(k-1)} + a_{14} x_{s(k-2)}.$$

One uses these complete polynomials to find, according to the combinatorial GMDH algorithm (with the consistency criterion $CN \to \min$ indicated in the preceding), the optimal models of the first, second, etc., rows for each of the variables. These models are checked to the condition $AC = \delta^2 \leqslant 1.0$, where

$$\delta^2 = \sum_{1}^{C} (x_i - \hat{x}_i)^2 / \sum_{1}^{C} (x_i - \bar{x}_i)^2.$$

All the non-optimal models as well as the models that do not satisfy the threshold in the precision of their short-term prediction are excluded from further consideration. Systems of equations are formed further only from polynomials that are optimal for each variable and that have passed the threshold test. This reduces drastically the volume of sorting.

Block 3. Generation of systems of equations and evaluating them according to the system consistency criterion. Systems sorting is reversive, i.e., it starts at the most complex system. The number of systems that are subject to evaluation according to the system criterion $CN_{\text{SYST}} \to \min$ is equal to C^1_M, C^2_M and C^3_M for the first, second and third rows, respectively.

Systems of the fifth row of the selection have the following complete descriptions (only optimal polynomials, chosen according to the combinatorial algorithm, that contain 16 addends participate):

$$x_{i(k)} = f_i (a_0 x_{i(k-1)} x_{i(k-2)} x_{j(k)} x_{j(k-1)} x_{j(k-2)} x_{p(k)} x_{p(k-1)} x_{p(k-2)} \times \\ \times x_{r(k)} x_{r(k-1)} x_{r(k-2)} x_{s(k)} x_{s(k-1)} x_{s(k-2)});$$

$$x_{j(k)} = f_j (a_0 x_{i(k)} x_{i(k-1)} x_{i(k-2)} x_{j(k-1)} x_{j(k-2)} x_{p(k)} x_{p(k-1)} x_{p(k-2)} \times \\ \times x_{r(k)} x_{r(k-1)} x_{r(k-2)} x_{s(k)} x_{s(k-1)} x_{s(k-2)});$$

. . . . . .

$$x_{s(k)} = f_s (a_0 x_{i(k)} x_{i(k-1)} x_{i(k-2)} x_{j(k)} x_{j(k-1)} x_{j(k-2)} x_{p(k)} x_{p(k-1)} x_{p(k-2)} \times \\ \times x_{r(k)} x_{r(k-1)} x_{r(k-2)} x_{s(k-1)} x_{s(k-2)}).$$

All the systems, the number of which is C^5_M, are evaluated according to the consistency system criterion of the form

$$CN_{syst} = \frac{1}{5}(CN_1 + CN_2 + \ldots + CN_5) \to \min,$$

and the fifth-row optimal system is selected.

Next, the systems of the fourth selection row (their number is C^4_M) have a complete description that contains 13 addends. All the systems (that have four equations) are evaluated according to the system criterion of the form

$$CN_{syst} = \frac{1}{4}(CN_1 + CN_2 + \ldots + CN_4) \to \min.$$

The optimal fourth-row system is selected, etc. The characteristic $CN_{SYST} = f(s)$, where s is the number of equations, is constructed, etc.

Reduction of the sorting volume. The properties of many objects allow one to apply the following method for a significant reduction of the sorting volume. In the fourth, third, second and first rows (reverse counting of the rows) of selection one sorts only the polynomials (equations) that turned out to be optimal in the fifth row, i.e., only five equations ($m = 5$ with $m << M$). In the fourth row one sorts C^4_5, in the third, C^3_5, in the second, $C^2_5 = 10$ and in the first, $C^1_5 = 5$ systems of equations.

We satisfy ourselves that the sorting does not require a joint solution of the system of equations.

Experimental confirmation of the adequacy law by multiple application of the OSA algorithm when reducing the variables set. The number of equations in the optimal system found by the OSA algorithm is an index of the degree to which the modeled object is blurred.

Let a data sample that contains $M = 25$ variables be given and let us find, by using the OSA algorithm, the optimal system from a number of equations with the output variables x_7, x_{12}, x_{13}, x_{18}, x_{23}. These variables are the least "blurred" and are predicted well by difference polynomial equations. We eliminate from the sample the found "detail" variables and repeat the OSA algorithm. We find that the optimal system contains only four equations with the output variables x_3, x_8, x_{14}, x_{22}. But the minimum of the system criterion increased (Fig. 6) and moved to the left.

We eliminate now the 9 indicated variables from the sample and repeat the application of the OSA algorithm. We find that the optimal system contains only three equations with the output variables x_5, x_6, x_{11}. The minimum of the criterion moved even farther to the left, etc.

Such a shift of the minimum of the system criterion confirms the adequacy law which asserts that for more blurred systems the optimal description (model) must also be more blurred, i.e., must have a smaller number of equations.

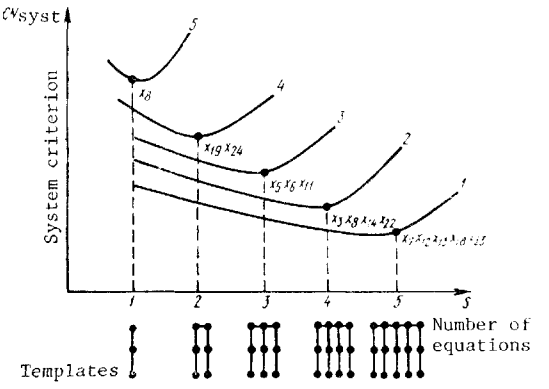

Fig. 6. Results of sorting by the OSA algorithm with subsequent elimination of less "blurred" variables. Processing a sample that contains: 1) 25 variables; 2) 20 variables; 3) 16 variables; 4) 13 variables; 5) 11 variables. The degree to which the object is blurred increases, while the number of equations in the optimal model decreases.

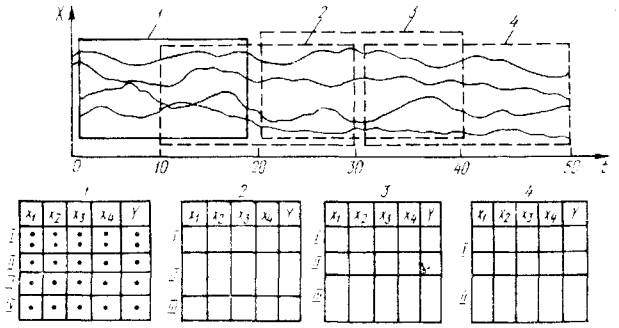

Fig. 7. Four positions of the "sliding window" and the corresponding four clusterizations which demonstrate that the variety (number of clusters) decreases from four to three.

SELECTING THE DEGREE OF BLURRING (FUZZINESS) OF THE DESCRIPTION. INDICATIVE SYSTEMS

The more complex the object, the less detailed must the mathematical language of planning, modeling and control be. Blurred models turned out to be successful for medical diagnostics, for ecology and for economics [10].

Indicative, the term adopted in economics, describes systems in which the degree of blurring of the information signals is optimized, i.e., selected at the optimal level. A combined system of the future will include all the three possible methods for the control of economics: direct planning, indicative planning, and free market. The planning and control signals in indicative systems are blurred. For example, the production of a factory is planned using a "fork" between the minimal and maximal levels. Each local factory will be given some freedom of choice. Indicative planning reduces the volume of information that is processed in the central planning organizations.

In analogy with economics, indicative systems of modeling, clusterization, and control can be defined as non-Godel systems, i.e., systems without set-point elements (without standards) in which the degree of blurring in the information processes language is optimized. Instead of an external specification, one has to specify only the ratio of variables, which is a more blurred information than the specification of concrete values for each of them. In many cases the ratio is known *a priori*; there is no need to consult the experts. This means that the computer becomes an independent arbiter in scientific debates on problems of modeling, clusterization, diagnostics, and pattern recognition [10].

SORTING METHODS OF MODELING AND CLUSTERIZATION OF DATA OBTAINED IN A "SLIDING WINDOW." PROBLEMS OF LONG-TERM PREDICTION

Prof. V. V. Nalimov insists in [14] that predictions for biological, ecological, economic and social systems are possible only in a blurred language. The more blurred the mathematical language of prediction is the longer is its maximum achievable anticipation time.

He proposed a very blurred "pattern-analysis" language (or two-component orthogonal projection of data sample points) as an algorithm for long-term prediction [16].

Clusterizations that are used instead of polynomial equations are also a method of making the mathematical description less detailed or more blurred. Sorting algorithms offer another possibility: to track clusterizations obtained in a "sliding window" that moves along the data sample on the time axis.

For example, the data sample for the ecosystem of Lake Baykal contains measurements over the interval of 50 years (Fig. 7). Moving a 10-year wide sliding window, we can obtain 40 clusterization forms used to track how the ecological system varies in order to predict its further development. The longest anticipation time of a prediction is obtained without using precise differential equations and their difference analogs. The objective clusterization of the sample data into some number of classes is used to calculate the graph of the probability of transition from one class to the other, which makes it possible to find an analog of the current state of the object in prehistory and, consequently, to indicate the long-term prediction. Thus, the following long-term prediction scheme should be considered prospective: (1) clusterization of the data sample for a number of positions of the "sliding window" (Fig. 7); (2) formation of the

graph of canonical coefficients of the pairwise correlation of separate clusterizations; and (3) selection of an analog for the current clusterization and prediction according to the analog or the group analogs method [10].

SOME DISCOVERIES OF MODELING BY SORTING

Sorting methods, being experimental, reveal for us many unexpected facts. It has already been indicated in the preceding that only sorting computational experiments taught us to find optimal models among under-complicated ones.

They also taught us how to build indicative (objective) systems without human-controlled set-point devices.

If a data sample is sufficiently representative, then the design of an experiment becomes a problem of selecting the optimal collection of variables and of eliminating some of the points indicated in the sample. The first experiments in sorting ensembles of attributes have shown that the same minimal value of the criterion corresponds to a number of various ensembles. There is no unique optimal ensemble [10].

Not less important is the lesson learned from the "sliding window" experiments. It turned out that some complex objects must be described by a few different systems of equations that are used sequentially, depending on the initial conditions. This is new in mathematical physics, where each object has a single mathematical description.

For example, it is known that the weather at each point of the atmosphere can be described by a system of four differential equations [17]. Having very long samples of precise measurements, one can find the structure of the matrix model of the local weather. But surprising things begin to occur only when we investigate various positions of the "sliding window" in time [19]. It turns out that the structure of the matrix model varies depending on the window position, sometimes repeating its previous forms. One can organize the clusterization of the weather matrix models and construct a probability graph for its prediction.

Conclusion. The theory of GMDH sorting methods can be presented as a continuous process of theoretical explanation of experimental results of sorting models or clusterizations obtained by a computer.

To speed up the process, both the computational experiments and the theoretical interpretation must be concentrated in the hands of a single researcher or a group of researchers. Lack of unity of theory and practice slows down the development process of the GMDH and causes the appearance of papers that deserve criticism (both from the experimental and the theoretical points of view).

Testing example. Let a model be specified in the form of two differential equations

$$\frac{dx_1}{dt} = -0.5 \text{ and } \frac{dx_2}{dt} = 1.0$$

or their two difference analogs $x_{1(k)} = x_{1(k-1)} - 0.5$ and $x_{2(k)} = x_{2(k-1)} + 1.0$ with the initial conditions $f = 0$, $k = 0$, $x_1 = 10$, $x_2 = 0$. Using the difference equations, one can form the following sample of measurement data:

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_1	10	9.5	20	8.58	8.0	7.5	7.0	6.5	6.0	5.5	5.0	4.5	4.0	3.5	3.0	2.5	2.0	1.5	1.0	0.5	0
x_2	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	10	20

The problem consists of reconstructing the initial difference equations using only the sample data.

Solution. 1. We discretize the variables at $N = 20$ levels and find sub-sample A (first instrumental variable):

$x_1 =$	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
$x_2 =$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20.

2. We discretize the variables at $N/2 = 10$ levels and find sub-sample B (second instrumental variable):

$x_1 =$	20	20	18	18	16	16	14	14	12	12	10	10	8	8	6	6	4	4	2	2
$x_2 =$	2	2	4	4	6	6	8	8	10	10	12	12	14	14	16	16	18	18	20	20.

We look for models of the following full form:

$$x_{1(k)} = a_0 + a_1 x_{1(k-1)} + a_2 x_{2(k)} + a_3 x_{2(k-1)};$$
$$x_{2(k)} = b_0 + b_1 x_{2(k-1)} + b_2 x_{1(k)} + q_2 x_{1(k-1)}.$$

3. Combinatorial algorithm. Fifteen models-candidates in which some of the coefficients are zero and the others are determined by the LSM (coefficients with index A by using sub-sample A, with index B, by using sub-sample 5) are subjected to sorting. We calculate the consistency criterion by the squares of coefficient differences:

$$CN_A = \sqrt{\frac{1}{m}[(a_{0A} - a_{0B})^2 + (a_{1A} - a_{1B})^2 + (a_{2A} - a_{2B})^2 + (a_{3A} - a_{3B})^2]};$$
$$CN_B = \sqrt{\frac{1}{m}[(b_{0A} - b_{0B})^2 + (b_{1A} - b_{1B})^2 + (b_{2A} - b_{2B})^2 + (b_{3A} - b_{3B})^2]}.$$
$$CN_{\text{syst}} = \frac{1}{2}(CN_A + CN_B).$$

The evaluations of the criteria are given in the table.

The system criterion is zero only for the eleventh model. We satisfy ourselves that the problem of model reconstruction by using sorting of two instrumental variables has been solved correctly. The actual model has been reconstructed independently by a computer from a sample of observations without detailed instruction by a human expert. The human only specified the criterion for sorting the models-candidates by the minimum of which the model has been found. No threshold values (nor confidence intervals) were needed. The example demonstrates a direction for creating an independent "artificial intelligence" that does not require cues from a human when making a decision and that often argues with the person.

Model No.	No. of components m	Values of coefficients a_0	a_1	a_2	a_3	Criterion of non-contradictoriness (with respect to coefficients)		
1	4	LSM LSM	LSM LSM	LSM LSM	LSM LSM	$CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$
2	3	0 0	LSM LSM	LSM LSM	LSM LSM	$CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$
3	3	LSM LSM	0 0	LSM LSM	LSM LSM	$CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$
4	3	LSM LSM	LSM LSM	0 0	LSM LSM	$CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$
5	3	LSM LSM	LSM LSM	LSM LSM	0 0	$CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$
6	2	0 0	LSM LSM	0 0	LSM LSM	$CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$
7	2	0 0	LSM LSM	LSM LSM	0 0	$CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$
8	2	LSM LSM	0 0	LSM LSM	0 0	$CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$
9	2	0 0	0 0	LSM 0	LSM LSM	$CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$
10	2	LSM LSM	0 0	0 0	LSM LSM	$CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$
11	2	LSM LSM	LSM LSM	0 0	0 0	$CN_A = 0$	$CN_B = 0$	$CN_{\text{syst}} = 0$
12	1	LSM LSM	0 0	0 0	0 0	Computer choice $CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$
13	1	0 0	LSM LSM	0 0	0 0	$CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$
14	1	0 0	0 0	LSM LSM	0 0	$CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$
15	1	0 0	0 0	0 0	LSM LSM	$CN_A > 0$	$CN_B > 0$	$CN_{\text{syst}} > 0$

REFERENCES

1. Self-organizing Method of Modeling: GMDH-type Algorithms. Stanley Farlow, Marcel Dekker, Inc., New York, Basel, 1986.
2. On Prof. D. Gabor's paper Prospects of Planning, Avtomatika, No. 2, pp. 89-90, 1972.
3. Ivakhnenko, A. G. and L. M. Boychuk. Integral invariants in the problem of modeling multi-dimensional linear dynamic systems, Avtomatika, No. 5, pp. 42-46, 1984.
4. Ivakhnenko, A. G., G. I. Krotov and Yu. V. Kostenko. Self-organization of tensor models (the example of modeling the hydrochemical system of reservoirs), Avtomatika, No. 6, pp. 17-22, 1988.
5. Eykhoff, P. Parameter estimation and structural identification, Avtomatika, No. 6, pp. 21-37, 1987.
6. Koval'chuk, P. I. and E. S. Lakhno. Prediction and Optimization of the Sanitary State of the Environment, Vyssh. shk., Kiev, 1988.
7. Ivakhnenko, A. G. and Yu. P. Yurachkovskiy. Modeling of Complex Systems from Experimental Data. Radio i svyaz', Moscow, 1987.
8. Shelud'ko, 0. I. An GMDH algorithm with orthogonalized complete description for synthesizing models from the results of a planned experiment. Avtomatika, No. 5, pp. 32-42, 1974.
9. Kocherga, Yu. L. and F. Abdullayev. Minimax parameters estimation in linear regression for a J-optimal prediction. Avtomatika, No, 1, pp. 22-28, 1989.
10. Ivakhnenko, A. G., et al. Non-parametric prediction GMDH models, Part 3. Avtomatika, No. 3, pp. 3-10, 1989.
11. Ivakhnenko, A. G., V. V. Osipenko and T. I. Strokova. Self-organization of standard-structure correlative predicting models. Avtomatika, No. 2, pp. 7-13, 1982.
12. Ivakhnenko, A. G. and V. V. Osipenko. An algorithm for inverse transformation of probability characteristics in a deterministic prediction. Avtomatika, No. 2, pp. 7-15, 1982.
13. Ivakhnenko, A. G., V. V. Osipenko and T. I. Strokova. Predicting two-dimensional physical fields using the inverse transform of the transition probabilities matrix. Avtomatika, No. 4, pp. 12-20, 1983.
14. Ikhara, D. About the uniqueness of model choice by the balance of variables criterion. Avtomatika, No. 1, pp. 86-88, 1976.
15. Handbook of Standard Modeling Programs. Tekhnika, Kiev, 1980.
16. Nalimov, V. V. Analysis of the fundamentals of ecological prediction. Man and Nature, No. 8, pp. 31-47, 1983.
17. Marchuk, G. M. Computational Methods of Weather Forecasting. Gidrometlzdat, Leningrad, 1967.
18. Otnes, T. and A. Ipokson. Applied Analysis of Time Series, Mir, Moscow, 1982.