

# The Review of Problems Solvable by Algorithms of the Group Method of Data Handling (GMDH)

*Published in Pattern Recognition and Image Analysis, Vol. 5, No. 4, 1995, pp.527-535*

A.G.Ivakhnenko and G.A.Ivakhnenko

Glushkov Institute of Cybernetics, National Academy of Sciences of Ukraina,  
pr. Akad. Glushkova 40, Kyiv -127, Ukraina [Gai@gmdh.kiev.ua](mailto:Gai@gmdh.kiev.ua)

**Abstract:** *This paper describes the use of algorithms of the Group Method of Data Handling (GMDH) in solving various problems of experimental data processing. A spectrum of parametric (polynomial) algorithms and of nonparametric algorithms using clusterings or analogues was developed. The choice of an algorithm for practical use depends on the type of the problem, the level of noise variance, sufficiency of sampling, and on whether the sample contains only continuous variables. Basic problems solvable by the GMDH are listed, including:*

- *identification of physical laws,*
- *approximation of multidimensional processes,*
- *short-term stepwise forecasting of processes and events,*
- *long-term stepwise forecasting,*
- *extrapolation of physical fields,*
- *clustering of data samples and search for a physical clustering that corresponds to the physical model of an object,*
- *pattern recognition in the case of continuous or discrete variables,*
- *diagnostics and recognition using probabilistic sorting algorithms,*
- *self-organization of multilayered neural nets with active neurons,*
- *normative vector forecasting of processes,*
- *process forecasting without models using analogues complexing.*

*The main result of the GMDH theory is that, for inaccurate noisy data and short samples, the criterion minimum indicates a nonphysical model (decision rule) whose accuracy is higher and structure is simpler than those of a complete physical model. Sorting of a set of candidate models by external criteria is required only for non-physical models. For low-variance noise, we recommend deductive algorithms that use ordinary internal sorting criteria. As the noise variance increases, we have to turn to nonparametric algorithms that use clustering and search for analogues in previous history and to adopt evolution modeling for process forecasting.*

*The neural net with active neurons is regarded as a means of improving the accuracy even further, beyond limits, and of increasing the lead time of long-term forecasts due to expansion of the regression domain (the procedure of expansion and narrowing of a complete set of variables in the layers of a neural net).*

## 1. Introduction

Solving practical problems and developing theoretical questions of the Group Method of Data Handling (GMDH) produced a broad spectrum of computing algorithms, each designed for a specific application [1-6]. The choice of an algorithm depends both on the accuracy and completeness of information presented in the sample of experimental data and on the type of the

problem to be solved. The purpose of this review is to demonstrate GMDH algorithms for different applications.

We will first describe basic features of the considered algorithms.

### 1.1. Polynomial Support Function

This method involves sorting, that is successive testing of models selected out of a set of candidate models according to a specified criterion. Nearly all known GMDH algorithms use polynomial support functions. General connection between input and output variables can be found in the form of a functional Volterra series, whose discrete analogue is known as the Kolmogorov-Gabor polynomial [6],

$$y = a_0 + \sum_{i=1}^M a_i x_i + \sum_{i=1}^M \sum_{j=1}^M a_{ij} x_i x_j + \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M a_{ijk} x_i x_j x_k$$

where  $X(x_1, x_2, \dots, x_M)$  is the vector of input variables and  $A(a_1, a_2, \dots, a_M)$  is the vector of the summand coefficients.

Other support functions are also used, e.g., harmonic or logistic ones:

$$y = a_0 + \frac{\sum_{i=1}^M a_i}{1 + \exp(-x_i)}$$

The complexity of the model structure is estimated by the number of polynomial terms used. The sorting procedure consists in computing the criterion as the structure of the model gradually changes.

### 1.2. Sorting of Models by Groups of Equal Structure

Being an iterative method, GMDH is close to the best-regression method but differs from it by efficient organization of the search for the optimum structure of the model and, in addition, by the utilization of both internal and special external sorting criteria. Models are sorted by groups or series of equal structural complexity and the best model is found for each series according to the criterion. It was proved theoretically that, for noisy data and a short sample, the minimum of mathematical expectation of the external criterion is unique [1].

The uniqueness of the minimum also remains for sufficiently large groups of models, which is used for the choice of the optimum model. At the same time, if we start estimating models one by one, it becomes virtually impossible to find an optimum one.

### 1.3. External and Internal Criteria

Recall that a criterion is called internal if it is computed for the entire data sample. The external criterion is computed from new information that was not used for estimating the coefficients of the model. For example, to compute the criterion of regularity, we rank points (rows) of a sample according to the variance and every third point is included into a test subsample used to estimate the structure of the model. The other points of the sample are used for estimating the coefficients of the models.

#### 1.4. Physical and Nonphysical Models

The GMDH algorithms are close in structure to self-training algorithms of multilayered pattern recognition systems, such as perceptrons or neural nets [6]. An important difference is that polynomial GMDH algorithms deal with continuous variables. The discrete nature of the output variable of a perceptron, which indicates the membership of the given image in a particular pattern, makes a finer allowance for the inaccuracy of recognition to select the structure of a perceptron impossible. No nonphysical models can be obtained after discretization or clustering. It is only continuous variables that allow one to find the minimum of the external criterion, which determines the optimum structure of non-physical models.

The physical model corresponds to the notion of mathematical description adopted in mathematical physics. Approximation of a physical model by means of polynomials or in the clustering language is also called sometimes the physical model of an object. The physical model is unique for each object and for language used to describe this object.

The main result of the GMDH theory is that, for inaccurate noisy data and short samples, the criterion minimum indicates the so-called nonphysical model whose accuracy is higher and structure is simpler than those of a complete physical model [1]. Nonphysical models can only be produced by the GMDH.

The structure of the nonphysical model becomes simpler as the variance of noise increases. An increase in the length of sample is equivalent to a decrease of noise. The structure of the nonphysical model becomes closer to the physical model as the sample grows. Thus, many nonphysical models may exist for a given object, which depends on the variance of noise and the length of the sample. Nonphysical models may be obtained not only by eliminating some members of a physical model but also at random so that a deeper minimum of the external criterion can be obtained [7].

#### 1.5. Deductive and Inductive GMDH Algorithms

Self-organization of models can be defined as their construction with the required *a priori* information being reduced as much as possible. In particular, the number of instructions of the author of modeling is minimized. In deductive algorithms called algorithms of the GMDH type [5], internal accuracy criteria are applied for model sorting, the result of the computation being used just once for selecting the best model of each iterative series. The number of the series is specified by an expert or by the author of modeling. Still the volume of the required *a priori* information is relatively small and so we can speak about self-organization of models by both inductive and deductive GMDH algorithms.

In contrast to that, models are sorted in inductive GMDH algorithms according to external accuracy criteria, the criterion computation results being used twice, for selecting best models in each series and for an objective choice of the number of iterative series. The most accurate optimum nonphysical model corresponds to the minimum of the external criterion.

Laws described by differential equations are identified in the form of difference analogues, i.e., in the form of algebraic polynomials containing retarded arguments.

1.5.1. The Combinatorial GMDH Algorithm. The main Combinatorial GMDH algorithm has a multilayered iterative structure. Its specific feature is that the iteration rule (particular description) does not remain constant but expands with each new series. In the first series, all the models of the simplest structure in the form

$$y = a_0 + a_1x_i \quad i = 1, 2, \dots, M$$

are subject to sorting and a number  $F$  of the models that are the best by the criterion are selected.

In the second series, models of a more complex structure are sorted. These models were constructed for output variables of the best models of the first series:

$$y = a_0 + a_1x_i + a_2x_j \quad i = 1, 2, \dots, F ; j = 1, 2, \dots, M ; F \leq M$$

In the third series, the sorting involves even more complex structures of the form

$$y = a_0 + a_1x_i + a_2x_j + a_3x_k \quad i = 1, 2, \dots, F ; j = 1, 2, \dots, F ; k = 1, 2, \dots, M ,$$

and so on. The build-up of the series continues as long as the criterion minimum decreases. For large values of the freedom of choice  $F = M$ , the algorithm ensures a complete sorting of all the models in the polynomial form.

1.5.2. The Iterative Multilayered GMDH Algorithm. In the Multilayered algorithm, the iteration rule (particular description) remains the same for all series. For example, a particular description in

the form

$$y = a_0 + a_1x_i + a_2x_j + a_3x_ix_j$$

is used in the first series, a particular description

$$z = b_0 + b_1y_i + b_2y_j + b_3y_iy_j$$

in the second series, a particular description

$$w = c_0 + c_1z_i + c_2z_j + c_3z_iz_j$$

is used in the third series, and so on. In other words, the output values of a previous series serve as arguments in the next series. With this method of iteration, some models may be missed, which may result in the so-called multilayeredness error. There is a need to investigate the convergence of the model self-organization procedure to the results produced by regressive analysis with the same sorting criterion.

1.5.3. The Algorithm of Objective Systems Analysis (OSA). The algorithm involves sorting of sets of equations rather than individual equations. The sets were obtained by means of implicit difference templates, such as

$$\begin{aligned} M=1: & \quad x_{i(k)} = f_1(x_{i(k-1)}, x_{i(k-2)}); \\ M=2: & \quad \begin{cases} x_{i(k)} = f_1(x_{i(k-1)}, x_{i(k-2)}, x_{j(k)}, x_{j(k-1)}, x_{j(k-2)}) \\ x_{j(k)} = f_2(x_{j(k-1)}, x_{j(k-2)}, x_{i(k)}, x_{i(k-1)}, x_{i(k-2)}) \end{cases} \end{aligned}$$

and so on. Each step of the template movement along a sample calls for a solution of a set of linear equations. The result is estimated according to the convolution of criteria calculated for individual equations.

## 2. The Problem of Identifying Physical Laws

A polynomial linear in coefficients and approximating the dependence of the output value on several input argument variables is to be found so as to obtain the minimum of the defined accuracy criterion. This polynomial may be the sum of simple nonlinear functions. The original information is specified in the sample of data produced by observing the operation of an object.

The traditional approach to the problem consists in sorting sets of candidate models to select one model that is the best by the criterion. As noted above, it is appropriate to sort models by groups of equal structure, which makes the minimum of the criterion unique in many cases.

Where the minimum is not clearly expressed, an auxiliary procedure of extending the definition of the minimum can be used. The lower part of the sorting characteristic is approximated by the second-order parabolic equation and the coordinate of the minimum of the parabola is defined.

The model corresponding to the minimum is the desired nonphysical model, whose accuracy is estimated by the criterion of the forecasting error variation [8] as follows

$$\delta_i^2 = \frac{\sum_1^N (y_i - \hat{y}_i)^2}{\sum_1^N (y_i - \bar{y})^2} \rightarrow \min,$$

where:  $y_i$  - is the variable value in the table;  $\hat{y}_i$  - is the value calculated according to the model and  $\bar{y}$  is the mean value.

With a low variance of noise,  $0 < \delta^2 < 0.1$ , and with long samples, we can apply a deductive method, i.e., we can select best physical models by an internal criterion in each group and entrust the stopping of the iteration to an expert. When the noise is greater, special methods of searching for the physical model should be adopted. For really great noise typical of fuzzy objects, self-organization of a physical model should be replaced by an algorithm of searching for physical clustering of a data sample. These recommendations cover the solution of the problem of finding and identifying of laws.

## 2.1. Special Methods of Searching for a Physical Model to Identify Laws for Noisy Data and Short Samples

To find out how an object functions, we need to know its physical model. The algorithms described below, which are based on the criterion of balance of variables, are designed to find such a model.

The criterion requires selecting the structure of the model such that it remains optimum when new observations of the object become available. To compute the criterion of balance, we divide the sample into two equal parts: subsample A and subsample B. The Combinatorial GMDH algorithm yields a series of models that gradually become more complicated in each subsample. The physical model can be found by the criterion of similarity or balance, e.g., in the form

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

where  $\hat{y}_A$  is the output value of the model found in subsample A and  $\hat{y}_B$  is the same in subsample B.

With a low noise variance, the criterion of balance is multivalued, i.e., many models give its zero value. Special methods were developed to achieve uniqueness (regularization) of the choice, including:

- (1) model sorting that uses a combined criterion equal to the sum of the criterion of balance and the criterion of regularity taken with a small weight,
- (2) model sorting by the criterion of balance of variables to select an optimum model with another regularization method, e.g., by superimposing normal noise on the original data sample or by reducing the number of the sample rows used,
- (3) discretization of the variables defined in a sample into a small number of levels using internal criteria, the discretization noise filters, whereupon the criterion minimum indicates only a physical model,
- (4) extrapolation of the locus of minima (LM) to its intersection with the abscissa axis of the plane where the sorting of the criterion versus the complexity of the model structure takes place, as the noise variance changes and the length of the sample (the number of points being considered) varies, the minimum of point criteria moves along the LM, which is used to extrapolate the LM and to define the structure of the physical model.

## 3. The Problem of Approximation of Multidimensional Processes

It is required to find the multidimensional polynomial function of time that approximates the function represented in the data sample by examination of variants.

With a low noise variance and a long sample, the deductive method of self-organization of a physical model can be used to solve the problem. Where the variance is considerable, the inductive method with the search for a nonphysical model is called for. For a greater variance, we

must abandon polynomial models and turn from parametric polynomial models to nonparametric clusterings and the search for analogues in previous history [9]. These recommendations cover solutions of approximation and process forecasting problems.

#### **4. The Problem of Short-Term Stepwise Process Forecasting**

The original information is the same as in the process approximation problem: a data sample is defined and the accuracy criterion is specified. The latter may be internal (for low noise) or external (for considerable noise). The noise variance can be estimated from the criterion of error variation in the forecast of the output variable. The error can be calculated only after the forecasting:  $\xi^2 \approx \delta_{opt}^2$ .

The main difference between the solutions of approximation and forecasting problems lies in the criterion computing method. To solve the approximation problem, we compute the error at each current moment. To select a forecasting model, we compute the error on the forecast with a lead time of one step ahead, i.e., for short-term forecasting. The Combinatorial GMDH algorithm with allowance for retarded arguments is recommended for the selection of the set of effective regressors (arguments) and for self-organization of the difference forecasting model. A full polynomial containing all arguments and their retarded values measured with a delay of one or two steps is set.

The regression domain should be extended in order to improve the accuracy of forecasts: current and retarded values of other variables correlated with the one to be predicted may be introduced into the original full polynomial. It is also advisable that the computer be offered pair products (covariances) of arguments for selection. The algorithm will leave in the nonphysical forecasting model only the summands that ensure the deepest criterion minimum.

Where the data sample is too great, it is recommended to count time by several scales. For example, to forecast the water level in a river, the level was measured by averaging over months and years. A separate model was obtained for each month. Then, the models were used in turn for stepwise forecasting one year ahead [10]. The use of several time scales is an efficient method of processing large data arrays.

#### **5. The Problem of Long-Term Stepwise Process Forecasting**

A forecast is referred to as long-term if its lead time is equal to, or exceeds 10 variable-measuring intervals (steps) in data samples. The forecasting error variation criterion should be less than one over the entire forecast. The long-term forecast is produced by repeating of one-step short-term forecast many times. However, the accuracy diminishes with each step. Thus, the problem arises of increasing the lead time of forecasting.

In addition, there is a danger of self-excited oscillations starting from a certain step of the forecast for many structures of difference forecasting models. Stability of difference schemes (called

templates) was studied in detail in discrete mathematics. We mention here only two ways of increasing the stability, including:

- (1) use of two time-reading scales with transfer to implicit templates and
- (2) introduction of feedback that excludes a certain row in the sample after each forecasting step [11].

To describe the first method, we use again the problem of forecasting the river flow that was measured on average by months and years [10]. The sample fills a rectangular that contains  $12 \times N$  ( $M = 12$ ) digits. The Combinatorial GMDH algorithm yields at once 12 monthly difference models. The original fall equations correspond to L-shaped implicit templates whose short arms are directed inside the rectangular.

For example, the full equation for the January model is

$$x_{(k)(s)} = f_1(x_{(k)(s-\Delta t)}, x_{(k)(s-2\Delta t)}, x_{(k+\Delta T)(s)}, x_{(k+2\Delta T)(s)});$$

where  $k$  and  $s$  are the discrete values of the number of years and months, respectively, and  $\Delta t$  is the lead time of a month; and  $\Delta T$  is that of a year. For the December model we have:

$$x_{(k)(s)} = f_2(x_{(k)(s-\Delta t)}, x_{(k)(s-2\Delta t)}, x_{(k-\Delta T)(s)}, x_{(k-2\Delta T)(s)}).$$

The best practice is to include only the elements of the rectangular in the orthogonal axes.

The second method that may be adopted to ensure the stability is applicable only to long-term forecasts of steady-state processes, which may include, in many cases, a trend of deviations of the predicted value from the median [2]. This method uses the concept of maintaining the distribution of discrete values of the variable to be forecasted. According to the concept, all discrete values should repeat themselves in the same number of cases over the forecasting interval equal in length to the original data sample. The stepwise forecasting is carried out as usual but the discrete value once selected for a forecast is skipped in the next steps. Then, the discrete value that is next is selected by the criterion. The algorithm was used to forecast the average annual flow of the Volga river when its level was discretized in 10 values [11]. In forecasting multidimensional processes, values of several variables are given in each node of the template to increase the stability of the long-term stepwise forecast.

An alternative to the use of difference equations for long-term forecasting is a Harmonic GMDH algorithm [6]. Its applications are associated with unidimensional oscillatory processes. The oscillation of the processes can be estimated by computing the integral of the harmonic intensity [12].

## 6. The Problem of Extrapolation of Physical Fields

The physical field is represented as a rectangular net of measurements where some nodes contain values of one or more output variables correlated with each other. The problem is to find the values of these variables in the nodes where they are unknown.



The problem of near extrapolation to one cell of the net is similar to that of short-term forecasting one time step ahead. The problem can be solved by the same algorithm, a nonphysical model yielding the most accurate extrapolation for inaccurate measurements. Distant extrapolation is similar to long-term forecasting. Self-excited oscillations are also possible depending on the choice of the difference scheme, i.e., the template. In [13], we described a program for selecting a cross-like template for the extrapolation of a two-dimensional productivity field of an oil-bearing horizon.

## 7. The Problem of Clustering of a Data Sample and of Search for Physical Clustering to Correspond to the Physical Model of an Object

Clustering of a data sample means its division into clusters whose number and the points of the sample they contain are selected so that each cluster includes a compact group of closely situated sample points. Each point is present in only one cluster. Clustering can be regarded as a discrete description of an object. We can find the discrete analogue of the physical model of the object its physical clustering among all possible clusterings [14-16].

Clustering variants are sorted by an algorithm of Objective Computer Clusterization (OCC) [6]. Construction of a hierarchical tree of clusterings orders and reduces the sorting while the clustering optimum according to a criterion is not lost [17]. Physical clustering can be found by the criterion of balance of clusterings. To compute the criterion, the sample is divided into two equal parts. A clustering tree is constructed on each subsample and the criterion of balance is computed at each step with the same number of clusters. The criterion requires finding the clustering where both the number and coordinates of the centers (middle points) of corresponding clusters coincide:

$$BL = \frac{1}{MK} \sum_{j=1}^M \sum_{i=1}^K (x_{oA} - x_{oB})^2 \rightarrow \min$$

where  $K$  is the number of clusters at a given step of the tree construction,  $M$  is the number of coordinates,  $x_{oA}$  are the coordinates of the centers of clusters constructed within part  $A$ , and  $x_{oB}$  are the coordinates of the centers of clusters within part  $B$ .

The physical clustering can be recommended for process forecasting where noise is very high for  $\delta^2 > 0.8$  and the physical and non-physical polynomial models are not accurate enough. Clustering makes it possible to replace the problem of forecasting multidimensional processes by that of one-dimensional forecasting of cluster numbers following one another. Algorithms of evolutionary modeling are quite suitable for this forecasting [18,19].

## 8. The Pattern Recognition Problem

The GMDH algorithms can be directly used to solve a problem of pattern recognition if we turn from the discrete formulation of the problem to the continuous function of the membership measure of each image in a particular pattern. The advantages of nonphysical models and of decision rules can be utilized only for continuous variables.

Two approaches to the problem of pattern recognition are well known: a statistical approach and a deterministic approach [20]. The GMDH sorting methods carry out both of them. The statistical approach makes use of the Multilayer Theory of Statistical Decisions (MTSD) for the distribution of the probability of pairwise random events [16].

The deterministic approach involves the hypothesis of compactness. The criteria of the minimum of empirical risk are computed for a learning sample (the internal criterion) or for a separate verification sample (the external criterion). The problem of recognition of new images is the most urgent for the recognition theory.

The statistical approach to pattern recognition solves the problem of selecting the decision rule under the condition of error minimization both for the learning sample and for the sample of new images, while it is *a priori* assumed that the learning sample is sufficient and the noise variance is low. Thus, to minimize the error on new data we should know something about them. The deterministic approach also allows one to minimize the error on new data but it requires different *a priori* information. Availability of the information determines the choice of an approach [21].

### **8.1. Pattern Recognition for Continuous Variables**

In this case, *a priori* information about the noise variance can be taken into account. Polynomial GMDH algorithms can be used to find the most accurate simplified decision rule corresponding to the nonphysical model for each level of variance as well as for large variation of the length of the image learning sample. The rule will be optimum for each level of noise variance, i.e., for each state of the recognition object. The main advantage of nonphysical decision rules is that, when obtained for a sample with a certain noise variance, they also remain optimum for the next sample if the noise level does not change.

Continuous variables carry information about noise, which is lost in discretization of variables into a small number of levels. Removing thresholds to return to continuous measurement of variables and considering continuous functions of the membership measure of each image in a particular pattern allow one to utilize this advantage.

### **8.2. Pattern Recognition for Discrete Variables**

For this problem *a priori* information about the availability of physical clustering both for the learning sample and for the new data sample can be considered. The output variable indicating the class is usually discretized and, thus, an accurate nonphysical decision rule cannot be found by GMDH algorithms. We have to use a less accurate physical decision rule (model) that is not simplified as noise variance increases. This rule can be found by many methods, e.g., by sorting variants according to the criterion of balance of clusterings [21].

This method is convenient when a perceptron or a neural-net recognition system is used to solve the problem of image classification. The flow of images arriving at the input of the perceptron can be divided into a number of samples containing equal numbers of images. A hierarchical tree of clusterings is constructed for each sample. By comparing step by step clusterings of two

neighboring trees during the construction, according to the criterion of balance of clusterings, we can find the general or physical clustering. Classification of images is possible only if the physical clustering is available for all image samples. For self-organization of the perceptron designed for the classification of images, it is preferable that the number of internal  $A$ -elements be selected equal to the number of clusters and the coefficients of connections with  $S$ -elements be assumed to be equal to the coordinates of the middle points (centers) of the clusters. This adjustment minimizes the recognition error for all samples, including the sample of new data.

### **8.3. Recognition and Diagnostics on the Basis of the Multilayered Theory of Statistical Decisions**

As noted above, the construction of sorting GMDH algorithms may use different support functions, including polynomials, harmonic, or logistic functions. The Bayes formulas or the Wald theories of statistical decisions are also applied.<sup>1</sup>

As the algorithm was developed, we found that computations of *a posteriori* probability can be reduced dramatically if we assume the achievement of two defined levels by two variables instead of the achievement of one level by one variable to be a single accidental event [16]. The probabilistic sorting GMDH algorithms with dual accidental events are used for medical diagnosis and to evaluate the treatment method prescribed by a doctor.

### **8.4. Neural Nets with Passive Binary or Logistic Neurons (Perceptrons)**

The process of optimizing a set of input variables is not included in passive neurons. It is established in the complex process of self-organization of the entire system of many neurons as a whole. The polynomial decision rule is replaced by a piecewise linear rule as a result.

Perceptrons do not involve structural optimization according to the external criterion and optimum non-physical models. Only parametric optimization by back propagation is carried out. Perceptrons are almost as good as GMDH algorithms in accuracy if the learning sample is long enough, noise variance is low, and the sample contains variables discretized into a small number of levels, i.e., in the cases where the physical model is optimum. The GMDH algorithms gain the advantage in accuracy for short samples of continuous noisy variables, i.e., when the nonphysical model is optimum. The advantages of perceptrons and of polynomial GMDH algorithms are combined in neural nets with active neurons [22].

## **9. The Problem of Self-Organization of Multilayered Neural Nets With Active Neurons**

The GMDH algorithms described above produce a high accuracy of solving the problems in the specified regression domain. The domain can be extended by computing additional regressors. The extension of a set of regressors by computing pair products (covariances) or inverse values of variables is a well-known practice. A problem arises here of selecting simple nonlinear

---

<sup>1</sup> Recall the difference: the former involve multiplication of specific probabilities of events, whereas the latter perform summation of these probabilities. Formulas are used in each sorting series according to the MTSD

transformations of input variables to get new, generalized features. Computations have shown that the extension of a set of variables by introducing output variables found by GMDH algorithms can be very effective [22,23].

To implement this idea, we form a neural net, which is a collection of active neurons, each following a GMDH algorithm. The number of neurons in each layer of the net equals the number of variables defined in the sample. Each neuron forecasts one of the variables. Output variables in each layer are added to the set of variables in the next layer. The build-up of the layers continues until it ceases to improve the accuracy of approximation or that of the forecasting of the output variable we are interested in.

One variable may require several layers while another may need no net at all. Considering the limited capabilities of computers, we extend the set of variables in each layer of the net and narrow it at the same time by discarding the variables that proved to be ineffective in the previous layers. Ordinary GMDH algorithms select a complete set only once. In the neural net, the selection takes place in each layer by introducing the extension-narrowing procedure for the set of variables.

The neural net is designed to solve the same problem that is solved by its neurons: if GMDH algorithms are used, the neural net solves the problem of approximation or forecasting. If perceptrons or other recognition systems are used as active neurons, the neural net as a whole solves the same problem of pattern recognition but does it more accurately than the neurons it consists of. The forecasting neural net is a means for increasing the lead time of the stepwise forecasts. This time is equal to the number of steps at which the variation criterion meets the condition  $\delta^2 < 1.0$  [8].

Neural nets with active neurons are twice-multilayered since multilayered neurons are connected into a multilayered net. The construction of a neural net at each forecasting step increases the accuracy of each step and thus extends the lead time. Roughly, we have

$$\tau_y = \Delta t / \delta_{i_{opt}}^2,$$

where  $\tau_y$  is the lead time,  $\Delta t$  is the variable computing step, and  $\delta_{i_{opt}}^2$  is the value of the forecast variation criterion.

The advantage of the neural net with active neurons over an ordinary neural net with binary neurons is that self-organization of the net is simplified as each neuron finds the required connections in the course of its self-organization by itself. Thus, one of the difficult problems in the theory of designing an artificial intelligence is solved.

## **10. The Problem of Normative Process Forecasting by Simplified Linear Programming And Specifying GMDH Algorithm Limitations**

Specifying limitations of the optimization domain is the main difficulty in the practical application of linear programming. As described in [24], this problem can be resolved by GMDH algorithms. The result is a set of tools that can be recommended as a basic one for normative forecasting (“If-

Then”) of processes and for computer control of open-loop objects [24,25]. Particularly promising is the algorithm of simplified linear programming where the number of controlled input variables is selected equal to the number of optimized output variables. There is no need to sort vertices of the optimization domain and to use the simplex method since the optimum mode is unique.

### **10.1. Normative Vector Forecasting of Results of Physical Experiments**

The problem is to find values of a set of variables after an experiment when values of the set of variables measured before the experiment are known. The multidimensional function required to solve this problem is defined as a sample of observations of a small number of experiments [23]. The sample is processed following the algorithm of simplified linear programming with specification of limitations by the GMDH.

### **10.2. Normative Vector Forecasting for a Macroeconomic System [25]**

The variables that characterize the system are divided into three sets as follows:

- (a) controllable variables that may be set (taxes, discount rates, deposit rates, issue of money, etc.),
- (b) variables that should be optimized (national income, consumption of resources, etc.), and
- (c) external disturbances (world oil prices, etc.).

According to the law of necessary variety, the number of input variables (a) should be no less than the number of output variables (b). The law is satisfied when the simplified linear programming algorithm is applied [23,25]. The technique involves a number of attempts or scenarios. When we set too favorable results of a forecast, we obtain impracticable relations for output variables. The best relation of controlling variables is taken from previous history as an optimum example.

### **10.3. The Problem of Supporting Decisions in Systems of Automatic Control of Open-Loop Objects**

The optimization interval is defined as a length of time wherein values of controlled output variables (resources) are defined and for which optimum values of controlling actions should be found. For macroeconomic systems, the optimization interval may be a quarter of a year or a year. If the period of natural oscillations of a system is many times greater than the interval assigned for optimization, it is useful to consider only the problem of the object behavior within this interval, i.e., to regard the object as an open-loop one or, to be more exact, a quasi-open-loop one. The problem of control is reduced to computing the vector of controlling variables one or more steps ahead.

To control open-loop objects, we can apply algorithms of mathematical programming using procedures of stepwise forecasting. Computerized automatic control of open-loop objects includes two stages. The first stage calls for a short-term forecast of external disturbances one time step ahead while the second stage involves computation of optimum values of controlling variables by the simplified algorithm of linear programming using the GMDH to specify the limitations. The

control should be multidimensional or vector-type and should satisfy the law of necessary variety, i.e., the number of controlling variables should be equal to or not less than the number of variables to be optimized. A long-term stepwise (or harmonic) forecast of disturbances is required to optimize operation of an object for a long time to come.

### **11. Forecasting of Processes Without Models Using Analogues Complexing**

The state of a complex object can be described by a vector of characteristic variables represented in the sample of observation data. To make the forecast, it is sufficient to find one or more close analogues of its state in previous history and to see what happened next. The model of the object is no longer needed since its role is played by the object itself. The forecasting according to analogues from previous history was developed in meteorology for weather forecasting. The modified analogue complexing algorithm involves sorting optimization of discrete values of the following parameters: (a) variants of the set of input variables, (b) the number of moments of time considered in the analogues, (c) the number of analogues to be complexed, and (d) values of weight coefficients with which the analogues are complexed, i.e., summed up [26].

The interconnected parameters (b) and (c) have to be sorted in the plane of two parameters, which increases the number of computations. The sorting of the forecasting algorithm parameters is done once for objects of each new type.

## References

1. Ivakhnenko, A.G. and Stepashko, V.S., *Pomekhoustoichivost' Modelirovaniya* (Noise Immunity of Modeling), Kiev: Naukova Dumka, 1985.
2. Müller, J.A. and Ivakhnenko, A.G., *Selbstorganisation von Vorhersagemodellen*, Berlin: VEB Technik, 1984.
3. Ivakhnenko, A.G. and Yurachkovskij, Yu.P., *Modelirovanie Slozhnykh Sistem po Eksperimental'nym Dannym* (Modeling of Complex Systems from Experimental Data), Moscow: Radio i Svyaz', 1987.
4. Ivakhnenko, A.G., *Nepreryvnost' i Diskretnost'* (Continuity and Discreteness), Kiev: Naukova Dumka, 1990.
5. *Self-Organizing Methods in Modeling*, Statistics: Textbooks and Monographs, Farlow, S.J., Ed., New York: Marcel Dekker Inc., 1984, vol. 54.
6. Madala, H.R. and Ivakhnenko, A.G., *Inductive Learning Algorithms for Complex Systems Modeling*, Boca Raton: CRC Inc., 1994.
7. Sawaragi, Y., Soeda, T., Tamura, H., et al., *Statistical Prediction of Air Pollution Levels Using Non-Physical Models*, Automatica (IFAC), 1979, vol. 15, no. 4, pp. 441-452.
8. Belogurov, V.P., *A Criterion of Model Suitability for Forecasting Quantitative Processes*, Sov. J. Automation Int. Sci., 1990, vol. 23, no. 3, pp. 21-25.
9. Ivakhnenko, A.G. and Müller, J.A., *Parametric and Non-parametric Selection Procedures in Experimental Systems Analysis*, Systems Analysis, Modeling, and Simulation, 1992, vol. 9, pp. 157-175.
10. Ivakhnenko, A.G., *Dolgosrochnoe Prognozirovanie i Upravlenie Slozhnyimi Sistemami* (Long-Term Forecasting and Control of Complex Systems), Kiev: Tekhnika, 1975.
11. Ivakhnenko, A.G. and Osipenko, V.V. *Algorithms of Transformation of Probability Characteristics into Deterministic Forecast*, Sov. J. Automation and Inf. Sci. 1982, vol. 15, no. 2, pp. 7-15.
12. Candel, M., *Vremennyye Ryady* (Time Series), Moscow: Finansy i Statistika, 1981.
13. Ivakhnenko, A.G., Peka, P.Yu., and Vostrov, N.N., *Kombinirovannyj Metod Modelirovaniya Vodnykh i Neftyanykh Polei* (Combined Method of Modeling Water and Oil Fields), Kiev: Naukova Dumka, 1984.
14. Ivakhnenko, A.G. and Müller, J.A., *Problems of Computer Clustering of the Data Sampling of Objects under Study*, Sov. J. Automation Inf. Sci., 1991, vol. 24, no. 1, pp. 58-67.
15. Ivakhnenko, A.G., Ivakhnenko, G.A., and Müller, J.A., *Self-Organization of Optimum Physical Clustering of the Data Sample for Weakened Description and Forecasting of Fuzzy Objects*, Pattern Recognition and Image Analysis, 1993, vol. 3, no. 4, pp. 415-422.
16. Ivakhnenko, A.G., Petukhova, S.A., et al., *Objective Choice of Optimal Clustering of Data Sampling under Nonrobust Random Disturbances Compensation*, Sov. J. Automation Inf. Sci., 1993, vol. 26, no. 4, pp. 58-65.
17. Zholnarskij, A.A., *Agglomerative Cluster Analysis Procedures for Multidimensional Objects: A Test for Convergence*. Pattern Recognition and Image Analysis, 1992, vol. 2, no. 4, pp. 388-390.
18. Vogel, L., Owens, A., and Walsh, M., *Iskusstvennyi Intellekt i Evolyutsionnoe Modelirovanie* (Artificial Intelligence and Evolution Modeling), Moscow: Nauka, 1969.

19. Bukatova, I.L., Mikhasov, Yu. I., and Sharov, A.M., *Evoinformatika: Teoriya i Praktika Evolyutsionnogo Modelirovaniya* (Evoinformatics: Theory and Practice of Evolution Modeling), Moscow: Nauka, 1991.
20. Vapnik, V.N. and Chervonenkis, A.Ya., *Raspoulavanie Obrazov, Statisticheskie Osnovy Teorii* (Pattern Recognition: Statistical Basis of the Theory), Moscow: Nauka, 1974.
21. Ivakhnenko, A.G. and Ivakhnenko, G.A., *Perceptron Synthesis according to Clusterization-Balance Criterion*. Pattern Recognition and Image Analysis, 1995, vol. 5, no. 3, pp. 337-341.
22. Ivakhnenko, A.G., Ivakhnenko, G.A., and Müller, J.A., *Self-Organization of Neural Nets with Active Neurons*, Pattern Recognition and Image Analysis, 1994, vol. 4, no. 2. pp. 185-196.
23. Ivakhnenko, G.A. *Self-Organization of Neuronet with Active Neurons for Effects of Nuclear Test Explosions Forecastings*. System Analysis Modeling Simulation (SAMS), 1995, vol.20, pp.107-116.
24. Triseev, Yu.P., *Approaches to the Solution of Mathematical Programming Problems on the Basis of Heuristic Self-Organization*, Sov. J. Automation Inf. Sci., 1987, vol. 20, no. 2, pp. 30-37.
25. Ivakhnenko, A.G. and Ivakhnenko, G.A., *Simplified Linear Programming Algorithm as Basic Tool for Open-Loop Control*, Systems Analysis, Modeling, and Simulation, (SAMS), 1995, vol.18-19, pp.315-319.
26. Ivakhnenko, A.G., *An Inductive Sorting Method for the Forecasting of Multidimensional Random Processes and Events with the Help of Analogs Forecast Complexing*, Pattern Recognition and Image Analysis, 1991, vol. 1, no.1, pp.99-108.

Aleksey G. Ivakhnenko. Born 1913. Member-Correspondent of National Academy of Sciences of Ukraina. Professor. Doctor of technical sciences. Graduated from the Leningrad Institute of Electric Engineering in 1938. In 1943 he became a candidate in electrical sciences at the Physics Institute of Moscow University, and in 1944 he joined the Ukrainian Academy of Sciences in Kiev. He wrote his doctoral dissertation on invariance theory in 1953. In 1944 he joined the Kiev Polytechnic Institute as Professor in the Department of Electrical Engineering. Since 1964 was Head of the Combined Control Systems Department of the Glushkov Institute of Cybernetics of the Academy of Sciences of Ukraina. Since 1995 he is Directors Advisor of this organization. He was chief editor of the journal *Avtomatika* (Soviet Journal of Automation and Information Sciences). He is author of about 15 monographs and over 300 research papers, specializing in the modeling and pattern recognition of complex systems. About 220 candidates dissertations (from them 27 defend doctor degree) was prepared. Author of the Group Method of Data Handling (GMDH), which is widely used in modeling. Scientific interests: inductive sorting methods of modeling for forecasting random processes in fuzzy systems of ecology, biology, medicine, and economics.

Gregory A. Ivakhnenko. Born 1966. Graduated from the National Technical University (KPI) in 1989. Leading specialist at the National Institute for Strategic Studies of the Security and Defence Council of the Ukraine. Scientific interests: modeling of complex objects, pattern recognition, and diagnostic methods of artificial intelligence. Author of 23 papers.