1.1 OPTIMAL COMPLEXITY OF INDUCTIVE MODELS, REGULARIZATION, SELECTION CRITERIA

Selective Properties of the GMDH Criteria for Inductive Modeling

V.S.Stepashko

International Research and Training Center for Information Technologies and Systems of NAS and MES of Ukraine, Prospect Glushkova 40, 03680 Kyiv, Ukraine

stepashko@irtc.org.ua

Abstract. *The problem of construction (structural identification) of optimum model on the basis of a short data sample in the class of structures linear in parameters is investigated. The choice of a model structure having minimum variance of the forecasting error, or the noise-immunity model, is accepted as a primary objective of the problem solution. The features and regularities of the optimum model construction depending on the noise level and the sample volume are investigated; efficiency of the GMDH external criteria in this problem is studied.*

Keywords

Inductive modelling, GMDH, forecasting error, noise-immunity model, optimal complexity.

1 Introduction

With publishing in 1968 A.G.Ivahnenko's article [1] on GMDH, the new scientific direction named as "inductive self-organization of models from experimental data", or simply "inductive modeling", began to develop.

This was a new approach to the modeling using computer: in place of traditional *deductive* way "from general laws of object functioning – to the concrete mathematical model", an *inductive* method "from the concrete given observations – to the general model" is used. A researcher supplies a data sample, advances a hypothesis about the possible class of models and sets the criterion for choice of the best model in this class. Then a computer program operates and there appears a possibility to minimize the influence of subjective factors and to get the model as an objective result.

GMDH as a realization of the inductive approach is the original method for construction of models from experimental data under conditions of uncertainty. Models of optimum complexity produced by this method represent unknown regularities of performance of the investigated object (process) the information about which is contained implicitly in the given data sample. Principles of automatic generation of variants, nonterminal decisions and successive selection of the best models after external criteria are used in GMDH for the model construction. Such criteria are based on the division of the sample into independent parts where the estimation of parameters and control of models quality are executed on different subsamples. It allows avoiding burdensome a priori assumptions as a sample division enables to take automatically (implicitly) into account different types of a priori uncertainties. Efficiency of the method is verified by the solution of numerous real world problems in ecology, hydrometeorology, economy, technology etc. [2-5].

A qualitative analogy between the problems of modeling from noisy data and passage a signal through a noisy channel was established in [6]. It was found that in the case of stochastic assumptions it is possible to consider the modeling problem from observation data, or structural identification of objects, as the problem of the signal extraction on a noise background. On the ground of this conclusion as well as ideas of the Kotelnikov's theory of noise-immune signal reception, a new approach to the theoretical analysis of modeling problems was offered in [7]. It had allowed beginning to construct the theory of noise-immunity modeling, basic principles of which are reflected in [8, 9]. The main result of this theory is as follows: complexity of the optimum forecasting model depends on the level of noise in data: the higher it is, the simpler the optimum model should be. The GMDH theory began to develop further as a theory of an inductive method of information and knowledge extraction from experimental data [10, 11].

The theoretical part of this paper considers the following items: problem statement; properties of the model error variances (ideal criteria); method of critical variances as a tool for theoretical investigations; technique of comparative analysis of actual criteria for the model choice; brief results of solving some applied tasks of the GMDH theory.

2 Theoretical Part

The class of problems concerning the choice of the best (in some sense) model of an object from the great number of generated models being linear in parameters is considered.

2.1 Problem statement

Let W=(X y) be a data sample with $n\times m$ matrix X and $n\times 1$ vector y containing information on n observations after m independent inputs and one output of a static object. The determined matrix X is considered to be of full rank $rank\, X = m$ and the vector y contains a noise:

$$y = \overset{\circ}{y} + \xi, \quad \overset{\circ}{y} = Ey = X\theta_0, \quad E\xi = 0, \quad E\xi\xi^T = \sigma^2 I_n \;, \tag{1}$$

where $\overset{\circ}{y}$ is an unknown exact (non-noisy) output of the object, θ_0 is an exact (unknown) vector of parameters of the object, ξ is a vector of noise with independent identically distributed components, E is the mathematical expectation for all possible realizations of the noise vector, I_n is $n\times n$ unite matrix, and σ^2 is an unknown finite value of the noise variance.

We consider that the unknown true model (dependence between $\overset{\circ}{y}$ and X) belongs to the class of models linear in parameters. The problem consists in the choice of the model being the best by the given criterion from some great number of models formed by the certain generator in the specified class.

In the simplest case there is a set $\Im_m$ of models being compared in the process of modeling, containing m "nested structures" which corresponds to the successive complexity of a model in one linear member:

$$\Im_m : \overset{\circ}{y}_s = X_s \hat{\theta}_s \,, \; s = \overline{1, m} \,, \tag{2}$$

where s is the number of the estimated parameters, X_s is a submatrix of s arbitrary (for example first) columns of the matrix X, $\hat{\theta}_s$ is the proper LSM-estimation on the sample W. Expression (2) characterizes the arbitrary way of the successive complication of model in the case of the exhaustive search of generated variants; hence the consideration of the restricted set of models $\Im_m$ only does not affect the generality of the problem analysis.

We consider the modeling problem in the framework of the theory of choice of the most *noise-immune model* [7,8], that is the model with minimum error variance determined by the expression:

$$J(s) = \mathrm{E}\left\| \overset{o}{y} - \hat{y}_s \right\|^2 = E\left\| \overset{o}{y} - X_s\hat{\theta}_s \right\|^2 . \tag{3}$$

This "ideal criterion" characterizes the quality of restoration of an exact (true) signal $\overset{\circ}{y}$ from noisy one y by estimation $\hat{y}_s$ and is equal to the general *restoration error variance* for the signal in n points.

Let us divide the W into two subsamples A and B $W = (A^T B^T)^T$, $A = (X_A, y_A)$, $B = (X_B, y_B)$, $n = n_A + n_B$. Similarly to (3), it is natural to characterize the quality of prediction in points X_B of the subsample B for the model of complexity s with coefficients $\hat{\theta}_{As}$ estimated by the LSM on the subsample A, by the value of the expected losses within the prediction interval X_B, or the value of general variance of the prediction error:

$$J_B(s) \overset{\Delta}{=} J_B(s, X_A, X_B) = \mathrm{E}\left\| \overset{\circ}{y}_B - \hat{y}_{Bs} \right\|^2 = \mathrm{E}\left\| \overset{\circ}{y}_B - X_{Bs}\hat{\theta}_{As} \right\|^2 \tag{4}$$

It is obvious that (4) transforms to (3) in the case $X_B = X_A$. We will indicate that criteria (3) and (4) contain the exact (non-measured) vectors of the object output, that is they represent the “ideal goal” of modeling. In practice it is necessary to apply some of their estimations discussed below. Properties of criteria (3), (4) are investigated below as functions of the complexity of models and the level of incompleteness of information.

Let as denominate as *J-optimal* (optimum *restorative*) and *J_B-optimal* (optimum *predicting*) structures of the complexity (number of the estimated parameters), in that

$$s^o = \arg \min_{s=1,m} J(s) \ , \quad s_B^o = \arg \min_{s=1,m} J_B(s) \tag{5}$$

for structures from $\Im_m$ accordingly.

2.2 Properties of the model error variances

Let us consider the properties of $J(s)$ and $J_B(s)$ as functions of the discrete argument s . From (3) it follows that

$$J(s) = J^b(s) + J^v(s) = \left\| \overset{\circ}{y} - X_s \bar{\theta}_s \right\|^2 + \sigma^2 s \tag{6}$$

where $\bar{\theta}_s = E[\hat{\theta}_s]$, by the superscript *"b"* (from “bias”) the value of the average losses $J_b(s)$ from the bias of a model structure is marked, and by the superscript "*v*" (from ”variation”) the value of the functional variation, or the average additional losses J_v(s) because of the noise presence. We will name $J^b(s)$ as the "structural component" and $J^v(s)$ as the "noise component" of the functional $J(s)$.

In the same way, the formula (5) acquires a form [4]:

$$J_B(s) = J_B^b(s) + J_B^v(s) = \left\| \overset{o}{y}_B - X_{Bs}\bar{\theta}_{As} \right\|^2 + \sigma^2 tr\left[\left(X_s^T X_s \right)^{-1} X_{Bs}^T X_{Bs} \right] \tag{7}$$

where $\bar{\theta}_{As} = \mathrm{E}[\hat{\theta}_{As}]$ and by $J_B^b(s)$ and $J_B^v(s)$ the structural and noise components are marked accordingly. But the character of their dependence on s is not obvious therefore it is needed to pass to recurrent in s expressions.

Accepting that $X_{s+1} = (X_s \mid x)$, $\bar{\theta}_{s+1} = (\bar{\theta}_s^T \bar{\vartheta}_{s+1})^T$, one may get the following recurrent expressions [10]:

$$J^b(s+1) = J^b(s) - \bar{\vartheta}_{s+1}^2 \beta_{s+1} = J^b(s) - \alpha_{s+1}^2 / \beta_{s+1} \ ,$$

$$J^v(s+1) = J^v(s) + \sigma^2 \ ,$$

or generally

$$J(s+1) = J(s) - \bar{\vartheta}_{s+1}^2 \beta_{s+1} + \sigma^2 \tag{8}$$

where such denotations are introduced:

$$\bar{\vartheta}_{s+1} = \alpha_{s+1} / \beta_{s+1}, \quad \alpha_{s+1} = x^T D_s \overset{o}{y}, \quad \beta_{s+1} = x^T D_s x \tag{9}$$

thus $D_s = I_n - X_s (X_s^T X_s)^{-1} X_s^T$ is an idempotent matrix: $D_s^2 = D_s$, and $\beta_{s+1} > 0$ for all s = 0, 1, ..., m-1. We see that $J^b(s)$ as an integer function of the argument s decreases strictly monotonically to the zero and $J^v(s)$ increases linearly with the increase of s, therefore $J(s)$ has always a minimum in some point $s^o \in [1, m]$. In the case of increase of the noise level (variance) σ^2 the value of the optimal complexity s^o decreases, that is the optimum structure is simplified [5,6]. The Fig. 1 reflects this regularity of noise-immunity modeling.

The proper formulae for $J_B^b(s)$ and $J_B^v(s)$ was received in [10] taking to account that $X_{A,s+1} = (X_{As} \mid x_A)$, $X_{B,s+1} = (X_{Bs} \mid x_B)$. For $J_B^b(s+1)$ we have the relation:

$$J_B^b(s+1)=J_B^b(s)-2\bar{\theta}_{A,s+1}a_B^T b_B+\bar{\vartheta}^2_{A,s+1}b_B^T b_B \tag{10}$$

where:

$$a_B \overset{\Delta}{=} a_B(s)=\overset{o}{y}_B-X_{Bs}\bar{\theta}_{As}\ ,\ b_B \overset{\Delta}{=} b_B(s)=x_B-X_{Bs}(X_{As}^T X_{As})^{-1}X_{As}^T x_A\,. \tag{11}$$

Recurrent expression for $J_B^v(s)$ is:

$$J_B^v(s+1)=J_B^v(s)+\sigma^2\, b_B^T b_B\big/\beta_{A,s+1} \tag{12}$$

Then generally for variance of the forecasting error we have:

$$J_B(s+1)=J_B(s)-2\bar{\vartheta}_{A,s+1}a_B^T b_B+\bar{\vartheta}^2_{A,s+1}b_B^T b_B+\sigma^2\frac{b_B^T b_B}{\beta_{A,s+1}}\ . \tag{13}$$

Note that from (13) we get (8) in the case $X_B = X$.

So in accordance with (10), (12), components $J_B^b(s)$ and $J_B^v(s)$ as functions of s have such characteristics: the structural component decreases in general case not monotonically from the value $J_B^b(0)=\overset{o}{y}{}_B^T\overset{o}{y}_B$ to the zero and the noise one increases strictly monotonically. This means that, similarly to $J(s)$, the function $J_B(s)=J_B^b(s)+J_B^v(s)$ has a minimum (global) within the interval of change of the discrete argument $s\in[1,m]$ and this minimum corresponds to the progressively less optimum model complexity s_B^o the greater is the variance of noise σ^2. Thus the same Fig.1 characterizes on the whole the behavior of the function $J_B(s)$ too.

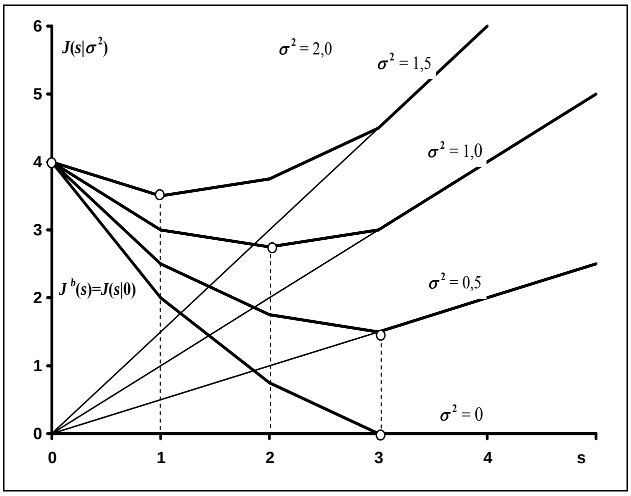

Fig. 1. Model error variance as a function of model complexity ***s*** for various noise variances (circles indicate the minima of the criterion $J(s)$; the true structure corresponds to $\boldsymbol{s}_0 = 3$, total number of inputs is $m = 5$).

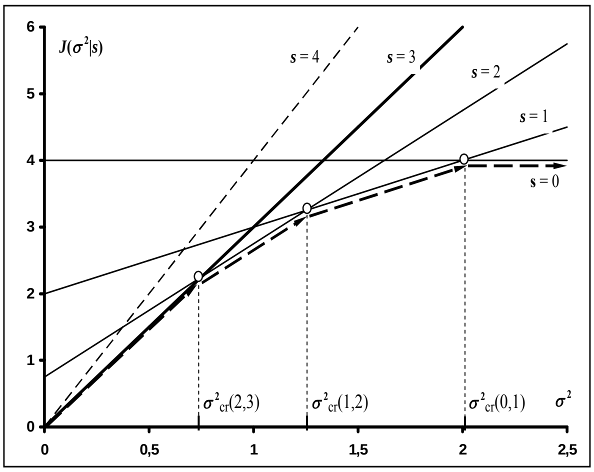

Fig. 2. Model error variance as a function of the noise variance $\boldsymbol{\sigma^2}$ for various complexity of model structures (circles indicate the critical points of changing the model structure; arrows point out the extremal of $J(s)$).

The plots in Fig.1 give qualitative characteristic to principle of noise-immunity modeling but do not allow getting quantitative appraisals of optimum conditions for an arbitrary model structure. That is why it is expedient to pass to more informative representation the dependence of the model error variance on the noise variance σ^2 for different model structures s. As it is obvious from (6), (7), the dependence is linear that simplifies the construction of proper plots.

For example, when constructing dependences $J(\sigma^2\,|\,s)$ it is necessary at first to lay off the structural component ordinates $J^b(s)\equiv J(s\,|\,\sigma^2=0)$ on y-axis of the graph (see Fig.2) and then to draw straight lines corresponding to different s. Slope angles of these lines increase gradually from zero value at $s = 0$ to maximal at $s=m$. Obviously, all lines corresponding to complexities $s \ge s_0$ will pass through the coordinate origin. The line in the Fig.2 corresponding to the correct structure $s_0 = 3$ is represented by the thickened one and the line for s = 4 is dotted. We note that the case s = 0 corresponds to absence of a model or to the model with zero coefficients, that is to the use the given vector of observations y as a “model”, therefore in Fig.1 all graphs begin from one and the same point:

$$\widehat{y}_{s|s=0} \equiv y,\ J(0 \mid \sigma^2) \equiv J(0 \mid 0) = J^b(0) = \left\| \overset{o}{y} \right\|^2 .$$

The graphs in the Fig.2 give some cardinally new information: positions of any line as well as points of their crossing are uniquely determined by properties of the modeled object, more precisely by the data sample, and do not depend on characteristics of noise. In addition, the lower envelope curve of these plots (marked by the dotted lines with arrows) is the extremal of the criterion $J(s)$, that is every point on it equals to the minimum value of the criterion for the proper noise variance. It means that such envelope is the line of “switching” of the optimum structures complexity: in the case of increase of the noise variance (beginning from $\sigma^2 = 0$) firstly the correct structure will be optimal (here $s_0 = 3$), after achievement of the first switching point, structure of one unity less complexity will do (s = 2) etc, to $s = 0$.

2.3 Method of Critical Variances

As points of switching are fixed *critical* points specific for the given sample, it is natural to introduce the so-called ***critical variances*** [3] $\sigma^2_{cr}(s, s+1)$ for denotation of coordinates of these points. As one may see from Fig.2, they are determined as the solution relative to σ^2 of equations

$$\sigma^2_{cr}(s, s+1)\ :\quad J(\sigma^2 \mid s) = J(\sigma^2 \mid s+1)\ ,\ s = 1, 2.,m{-}1\ , \tag{14}$$

which corresponds to the necessary condition of extremum of a function of discrete argument. It is evident that for a given noise variance the optimum structure will conform to the s satisfying the inequality:

$$\sigma^2_{cr}(s-1, s) > \sigma^2 \ge \sigma^2_{cr}(s, s+1) \tag{15}$$

and vice versa: the structure of a complexity s will become optimum for the σ^2 satisfying this inequality.

Variance of the forecasting error $J_B(s)$ also linearly depends on σ^2, therefore Fig.2 is suitable in this case too, so it is possible to define critical variances $\sigma^2_{Bcr}(s, s+1)$ starting from a condition similar to (14). We obtain these critical variances from recurrent relationships (8), (13) (in designations (9), (11)):

$$\sigma^2_{cr}(s, s+1) = \overline{\vartheta}^2_{s+1}\beta_{s+1} \tag{16}$$

$$\sigma^2_{Bcr}(s, s+1) = 2\overline{\vartheta}_{A,s+1}\beta_{A,s+1}\frac{a_B^T b_B}{b_B^T b_B} - \overline{\vartheta}^2_{A,s+1}\beta_{A,s+1}\ . \tag{17}$$

The set of procedures for calculation of the critical variances, the analysis of their properties and interpretation of the obtained results in the problem of optimization of model structures are named as ***the method of critical variances*** served as the basic tool of the theory of noise-immunity modeling including GMDH.

Let us give the general description of the relation between the basic properties of the critical variances values and the behavior of optimum values s^o, s^o_B, that is the extreme properties of the criteria $J(s)$, $J_B(s)$: a) the structural component of a criterion is strictly monotone decreasing if for all $s = \overline{1, m-1}$ the critical variances are strictly *positive*; b) a criterion has *a unique minimum* only if all the values of critical variances *diminish* strictly monotonically:

$$\sigma^2_{cr}(0,1) > \sigma^2_{cr}(1,2) > \ldots > \sigma^2_{cr}(m-1, m)\ \ ; \tag{18}$$

c) in this case the *global* minimum of a criterion is determined by the condition (15); g) if (18) is not hold than (15) is the condition of a *local* minimum and to determine the optimum complexity it is necessary to find the minimum value of the model error variance among all points of local minima.

2.4 Procedure of actual criteria analysis for the model choice

Let us analyze the efficiency of a given criterion $CR(s)$ from point of how the optimum model structure defined as

$$s^*_{CR} = \arg \min_{s=\overline{1,m}} \mathrm{E}[CR(s)] \tag{19}$$

estimates the unknown value s^o_B corresponding to the minimum of the theoretical criterion.

*Definition***.** A criterion $CR(s)$ is named an *optimal* one for the evaluation of models with minimum error variance if $s^*_{CR} = s^o_B$ (the unbiased estimation of s^o_B), an *adequate* one in the case $s^*_{CR} \le s^o_B$ (noise-immune estimation of s^o_B), and an *inadequate* one if $s^*_{CR} > s^o_B$.

Taking this into account, it is reasonable to analyze the efficiency of a given criterion in such a sequence: to find the expected value of it; to check up the compromise character of structural and noise components of the obtained expression; to define the conditions of optimality or adequacy of the criterion. The last task is solved in a formalized way using the method of critical variances.

The critical variance $\sigma^2_{CR}(s,s+1)$ as a theoretical measure of distinguishability of models with complexities s and $s+1$ with respect to the criterion $CR(s)$ equals to the value σ^2 being the solution of the equation:

$$\mathrm{E}[CR(s,\sigma^2)] = \mathrm{E}[CR(s+1,\sigma^2)] \ . \tag{20}$$

It allows analyzing efficiency of the given $CR(s)$ based on such result [11].

Theorem 1. A criterion $CR(s)$ is whether optimum or adequate if accordingly the conditions hold: $\sigma^2_{CR}(s,s+1) = \sigma^2_{Bcr}(s,s+1)$ or $\sigma^2_{CR}(s,s+1) \le \sigma^2_{Bcr}(s,s+1)$; it is inadequate otherwise.

This theorem gives necessary and sufficient conditions for the optimum or adequacy of criteria.

2.5 Analysis of efficiency of the GMDH criteria

Let us consider application of the described method for the analysis of efficiency of one of the basic criteria GMDH, namely the most frequently used regularity criterion $AR(s)$:

$$AR(s) = \left\| y_B - \hat{y}_{Bs} \right\|^2 = \left\| y_B - X_{Bs}\hat{\theta}_{As} \right\|^2 \ . \tag{21}$$

By analogy with (16), one may calculate the expected value of this criterion:

$$\overline{AR}(s) = \mathrm{E}[AR(s)] = J_B(s) + \sigma^2 n_B \ , \tag{22}$$

where $J_B(s)$ equals to the forecasting error variance (7).

From this follows immediately that the criterion $AR(s)$ has the property of noise-immunity (the change of minimum with the increase of the noise variance σ^2) as the necessary property of an effective criterion. Now it is necessary to check up optimum or adequacy of the investigated regularity criterion.

Critical variance for the $AR(s)$ criterion coincides with the expression (17):

$$\sigma^2_{AR}(s,s+1) \equiv \sigma^2_{Bcr}(s,s+1) = 2\overline{\omega}_{A,s+1}\beta_{A,s+1}\frac{a_B^T b_B}{b_B^T b_B} - \overline{\omega}^2_{A,s+1}\beta_{A,s+1} \tag{23}$$

thus the regularity criterion is *optimum* one for construction of the forecasting model.

Like that, the method of critical variances can be applied to analytical verification of fitness of any other GMDH criterion as a computable estimation of the theoretical criterion for structural identification of forecasting models. So, the external criteria GMDH are adequate to construction of models with the minimum variance of the forecasting error.

2.6 Some tasks of the GMDH theory

The method of critical variances can help when solving a range of actual tasks of the GMDH theory: sample division, comparative analysis of criteria, asymptotic analysis etc. Let us consider shortly some of the results in this area.

Sample division task. We call as *optimum* [10] such a division of the sample *W* when the same structure would be built as the result of identification according to the minimum error variance of a model with coefficients calculated on any of subsamples *A* and *B*: $s_A^o = s_B^o$. Here the optimum restoration model s_A^o of the signal corresponds to the minimum error variance on *A*:

$$J_A(s) = \mathrm{E}\left\| \overset{\circ}{y}_A - X_{As}\hat{\theta}_{As} \right\|^2 , \; s_A^o = \arg \min_{s=1,m} J_A(s) , \tag{24}$$

and s_B^o is the same optimum structure as in (5). As it was shown in [10], the strong solving of such problem assumes planning of an experiment for achievement the proportionality of information matrices:

$$\rho_B^2 X_A^T X_A = X_B^T X_B, \quad \rho_B^2 \neq 0, \tag{25}$$

Therefore under the conditions of passive experiment the problem of finding the best sample division of matrix *X* into two submatrices reduces to minimization of some measure of the mismatch of the left and right parts in (25).

Task of comparative analysis of criteria. One may show that subject to the condition (25), equality $\overline{\omega}_{A,s+1} = \overline{\omega}_{B,s+1} = \overline{\omega}_{s+1}$ is held which gives ground for comparison of the GMDH criteria as estimations of the theoretical criterion (3). For example, for the regularity criterion (21) we have

$$\sigma_{AR}^2(s,s+1) = \frac{1}{1+\rho_B^2}\sigma_{cr}^2(s,s+1) \tag{26}$$

that indicates its adequacy for any $\rho_B^2 \neq 0$. The symmetric regularity criterion [8] also appears to be adequate to the problem of the signal restoration on *W* due to

$$\sigma_{AD}^2(s,s+1) = \frac{1}{\rho_B^2 + \frac{1}{\rho_B^2}}\sigma_{cr}^2(s,s+1) \; . \tag{27}$$

With respect to the criteria of unbiasedness *CB*(*s*) and variationality *CV(s)* which belong to the group of coordination criteria [8], under the conditions of the division (25) they are inoperative because in view of $\overline{\omega}_{A,s+1} = \overline{\omega}_{B,s+1} = \overline{\omega}_{s+1}$ the structural components of these criteria are identically zero and accordingly $\sigma_{CB}^2(\cdot) = \sigma_{CV}^2(\cdot) = 0$ so that they choose the simplest models. Therefore for their use one should recommend the choice of the sample division based on the maximal mismatch condition of (25), than criteria of this group become adequate. Such quite different behavior of the regularity and unbiasedness criteria subject to the condition (25) testifies expedience of two-criterion approach [8] to the problem of inductive modeling.

Examples of application of the critical variances method to the analytical comparison of the regularity criterion with other known criteria (Mallows C_p and Akaike *FPE*) can be found in [11].

Asymptotic analysis. If to introduce averaging of criteria values by the number of observations, for example $\bar{J}(s) = \frac{1}{n}J(n,s)$, and the proper optimum complexity $\bar{s}^o$, one may define notion of *asymptotic unbiasedness* of the optimum structure through convergence of optimum values of criteria:

$$\bar{s}^o \to s_o \; : \; \lim_{n\to\infty} \bar{J}(\bar{s}^o) = \lim_{n\to\infty} \bar{J}(s_o) \; . \tag{28}$$

Thus one can prove the following theorem.

Theorem 2. For any given criterion, the divergence of the proper critical variance

$$\sigma_{CR}^2(n,s,s+1) \xrightarrow[n]{} \infty \, . \tag{29}$$

is the necessary and sufficient condition of asymptotic unbiasedness of the optimum structure for the criterion.

Based on the traditional for asymptotic analysis condition of strong regularity of regressors:

$$\lim_{n\to\infty}\frac{1}{n}X_n^T X_n = \overline{H}, \tag{30}$$

where $\overline{H}$ is a nonsingular finite $m \times m$ matrix, critical variances both for an ideal criterion (3) and for the external criteria GMDH are diverged and therefore the criteria are asymptotically optimum.

3 Conclusion

Within the framework of the theory of noise-immunity modeling, since papers [6-8] there are developed theoretical foundations of the GMDH as an inductive method of construction of models with the minimum error variance of restoration and forecasting of the exact (nonnoisy) signal.

The method of critical variances is an analytical tool of the theory which allows investigating in details the regularities of change of optimum structures complexity depending on the level of noise and other indicators of incompleteness of a priori information and also obtaining the theoretical comparative estimations of the criteria efficiency of the GMDH and other methods. In particular, there are specified the conditions of optimality and adequacy for the external criteria of "cross validation" type used in GMDH and based on a sample division.

Due to the division these criteria have the property of implicit (automatic) tradeoff between the optimum model complexity and the uncertainty level in the given observations, which allows operating without burdensome a priori assumptions necessary for application of some other methods of model construction from experimental data.

References

[1] Ivakhnenko A.G.: Group method of data handling as competitor to the method of stochastic approximation. *Soviet Automatic Control*: No. 3, 1968.

[2] Ivakhnenko A.G.: Long-term forecasting and control of complex systems. Kiev: *Tekhnika*, 1975. (In Russian)

[3] Ivakhnenko A.G.: Inductive methods of self-organization of models of complex systems. Kiev: *Naukova dumka*, 1982. (In Russian)

[4] Ivachnenko A.G., Mueller J.A.: Selbstorganization von Vorherzagemodellen. Berlin: *VEB Verlag Technik*, 1984. (In German)

[5] Farlow S.J.: Self-organizing methods in modelling: GMDH type algorithms:/ Edited by .- New-York, Bazel: *Marcel Decker Inc.*, 1984.

[6] Ivakhnenko A.G., Karpinsky A.M.: Self-organization of models by computer in terms of the general communication theory (information theory). *Soviet Automatic Control*: No. 4, 1982.

[7] Stepashko V.S.: The potential noise immunity of modeling on the basis of a combinatorial GMDH algorithm without taking the information about the noise into account. *Soviet Automatic Control*: No. 3, 1983.

[8] Ivakhnenko A.G., Stepashko V.S.: Noise-immunity of modelling. Kiev: *Naukova dumka*, 1985. (In Russian)

[9] Rao H.M., Ivakhnenko A.G.: Inductive learning algorithms for complex systems modelling. London, Tokyo: *CRC Press*, 1994.

[10] Stepashko V.S.: Structural identification of predicting models for a planned experiment conditions. *Journal of Automation and Information Sciences*: No. 1, 1992.

[11] Stepashko V.S.: Efficiency analysis of criteria of structural identification of forecasting models. *Journal of Automation and Information Sciences*: No. 3-4, 1994.