Ministry of Education and Science of Ukraine
National University of Food Technologies

V.V. Samsonov, A.M. Silvestrov, L.Yu. Spinul

Multiple Adaptive System of Identification

scientific publications

Kyiv NUFT 2018

UDC 519.711

Reviewers:
V.N. Azarskov, Doctor of Science, Professor of the Institute of Aerospace Control Systems
A.P. Ladaniuk, Doctor of Science, Professor of National University of Food Technologies
V.Ya. Danilov, Doctor of Science, Profecor of National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

Samsonov V.V., Silvestrov A.M., Spinul L.Yu. Multiple Adaptive System of Identification: scientific publications. - K.: , 2018. – 228 p.

ISBN 978-966-612-208-0

This scientific work aims to represent some elements of the theory of identification that are important for both practical use and further theoretical research in order to build logically complete basic and applied theory of identification as mathematically reasonable theory of knowledge of the cause-and-effect relationship in the objects of the real world.

For those specialists who carry out theoretical and experimental researches (technical, economic, biological, social etc) of the real-world objects with the aim of their optimal adaptive control, diagnostics of state, forecasting the consequences and so on.

It will be useful for students, postgraduates and doctoral research scholars who study the real objects.

Published in the author's edition

ISBN 978-966-612-208-0

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ХАРЧОВИХ ТЕХНОЛОГІЙ

В.В. Самсонов, А.М. Сільвестров, Л.Ю. Спинул

Багаторазово адаптивні системи ідентифікації

Монографія

Київ НУХТ 2018

УДК 519.711

Рецензенти: **В.М. Азарсков,** доктор технічних наук, професор, завідувач кафедри систем управління літальними апаратами Національного авіаційного університету, лауреат Державної премії України в галузі науки і техніки
А.П. Ладанюк, доктор технічних наук, професор , завідувач кафедри автоматизації та інтелектуальних систем керування Національного університету харчових технологій, заслужений діяч науки і техніки України
В.Я. Данилов, доктор технічних наук, професор, професор кафедри математичних методів системного аналізу Навчально-наукового комплексу «Інститут прикладного системного аналізу» КПІ
імені Ігоря Сікорского

Самсонов В.В., Сільвестров А.М., Спінул Л.Ю.
Багаторазово адаптивні системи ідентифікації: монографія. - К.: НУХТ, 2018. – 228 с.

ISBN 978-966-612-208-0

Подано деякі елементи теорії ідентифікації, що є важливими як для практичного застосування, так і для подальших теоретичних досліджень з метою побудови логічно завершеної фундаментальної та прикладної теорії ідентифікації як математично обґрунтованої теорії пізнання причинно-наслідкового зв'язку в об'єктах реального світу.

Для спеціалістів, які здійснюють теоретичні та експериментальні дослідження (технічні, економічні, біологічні, соціальні тощо) об'єктів реального світу з метою їх оптимального адаптивного управління, діагностики стану, прогнозування наслідків тощо.

Може буде корисною для студентів, аспірантів та докторантів, які вивчають реальні об'єкти.

Рекомендовано Вченою радою НУХТ як монографію
(протокол №8 від 1 березня 2018 р.)

Видано в авторській редакції

УДК 519.711

ISBN 978-966-612-208-0

The Introduction

The task of increasing the accuracy of identification of complex dynamic processes, despite the significant development of these issues is still relevant: if the income $\sum^{+}$ (fig. 1) has a finite value even if the model is ideal $\left(\overline{\varepsilon}^{2}=0\right)$, the costs $\left(\sum^{-}\right)$ to obtain this model tends to infinity; so the net profit (Δ) is positive in a limited range of model complexity.

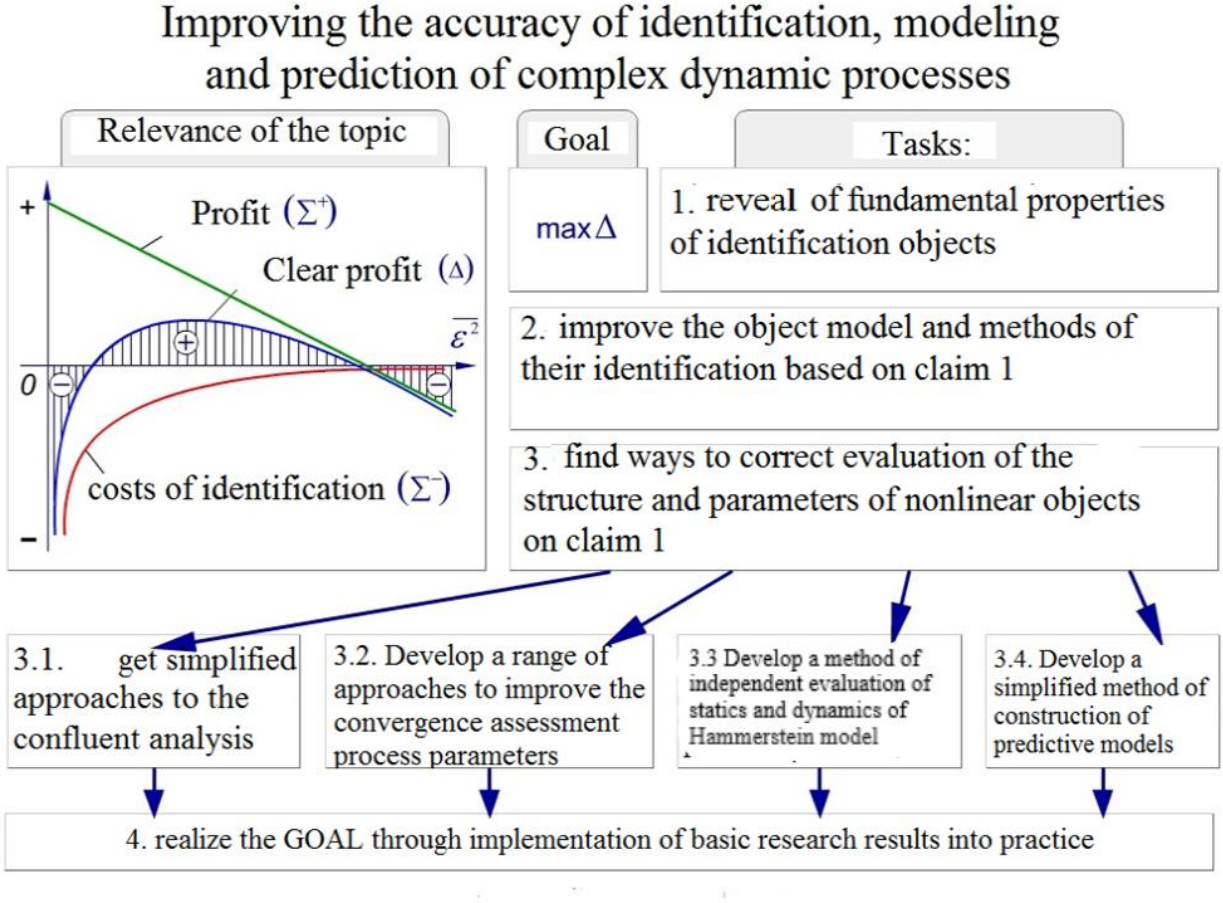


Fig. 1. The models of processes

The task is in maximizing Δ. For this purpose we need to find that fundamental property that is common to all the objects with the help of which you can achieve the goal ($max\Delta$, fig. 1), and on using this property:

1) to improve models and methods of identification;

2) to find methods of correct evaluation of structure and parameters of complex dynamic objects;

3) to get unbiased estimation methods for model parameters in terms of noisiness measurements both output and input;

4) to develop a range of approaches in order to improve the conditioning of the information matrix in terms of active and passive experiment;

5) to develop the method of independent evaluation of nonlinear statics and dynamics of Hammerstein's model in conditions of arbitrary object dynamics;

6) clearly differentiate between task of signal and parametric identification.

The properties of the real world objects are:

– non-autonomy and infinite dimensionality, and as a result there is no state of rest, because all processes are dynamic;

–limited power and natural inertia do not allow immediate change of any coordinate of the object, thus, all processes are **smooth.**

This fundamental property will be the basis of identification:

–there are not two or more identical objects, so averaging on the set gives not precise information about the parameters of a specific object;

–analogously the natural non-stationary of process limits the averaging on time;

–interrelationship and infinite dimensionality of real objects make it impossible to build a model which is isomorphic to object.

Depending on the purpose for which the model is built, infinite dimensional functional space of all state's variables of hypothetical base model (fig. 2) can be divided by the frequency feature in low - $\left(X_{LF}\right)$, middle - $\left(X_{MF}\right)$ and high - $\left(X_{HF}\right)$ frequency. Then there is the only middle frequency component in the partial model (model (1), fig. 2), $\left(X_{LF}\right)$ is considered by quasi-stationary state of the vector $\beta(t)$ of the parameter of this model, $\left(X_{HF}\right)$ is seen as a noise $N(t)$. We get a finite dimensional model (model (3), (4) fig. 2) for scalar *i*-th component f_i from f and limited dimensions $\left(X_{MF}\right)$.

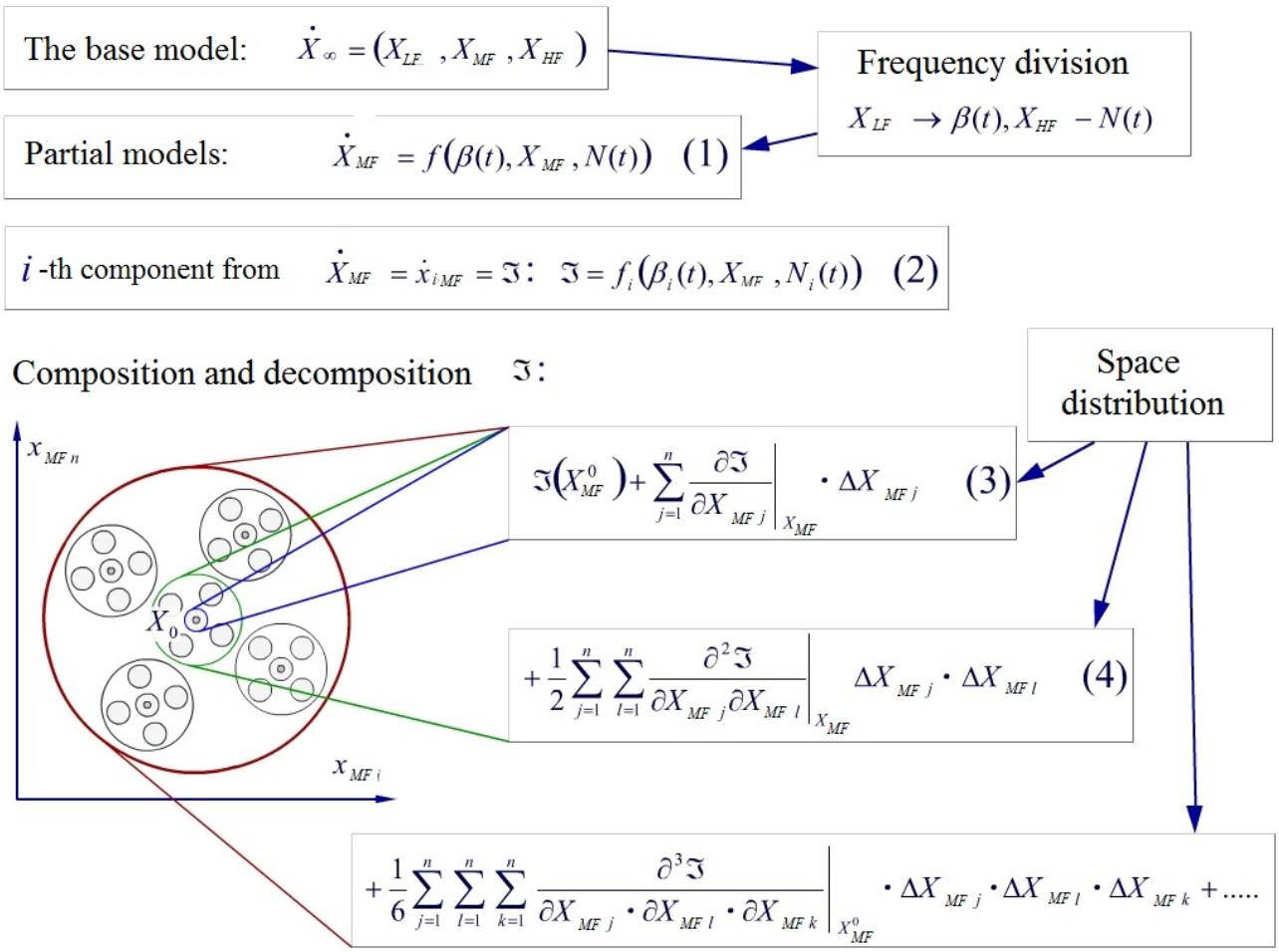


Fig. 2. The hierarchy of models

The system of identification in all is considered as a subsystem of multi-level system, where principles of decomposition, composition and optimization are used. The base model of real processes and its derivative models, modeling and forecasting solutions in the absence or presence of information about the studied process are under examination.

The issue of increasing the accuracy of identification of complex dynamic processes remains relevant. It is therefore important to find a fundamental feature common to all objects with which you can achieve this goal and on its basis to improve models and methods of identification.

The approach to structural-parametric identification of nonlinear multidimensional dependency is considered. It is based on the representation of the series (3), (4), fig. 2 as (14), fig. 3 or (15), where α is (16) for local subregions (fig. 2), it is (17) for combination

of local neighboring regions; and enlarging the region further we have the model (18), next – the model (19) etc.

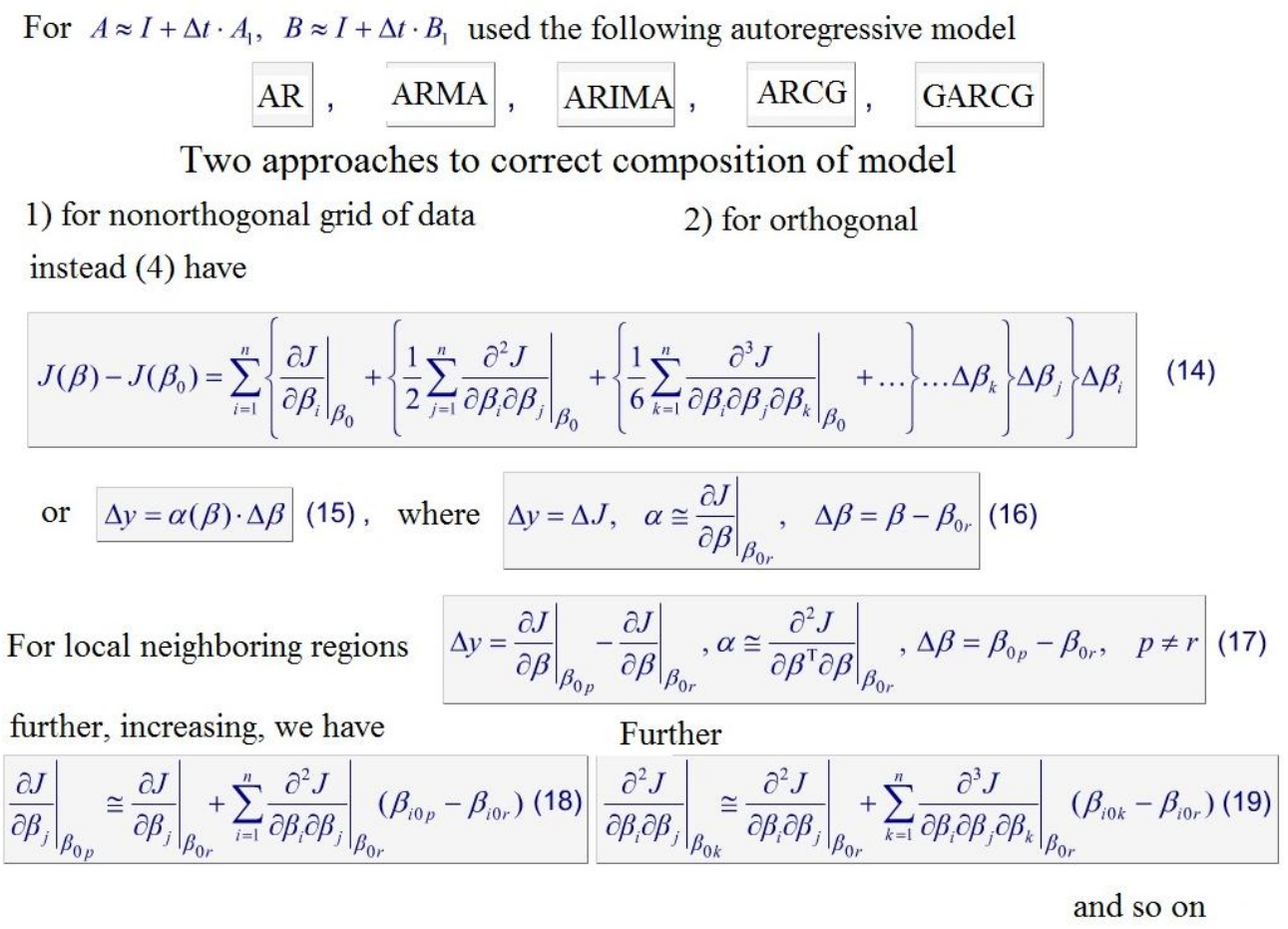


Fig. 3. The composition of models.

The method is used for:

1) determination of the structure and parameters of the test dependence $I(\beta)$ (fig. 4,

p. 1) with minor error $0{,}05\beta_2$ due to proximity of calculation $\frac{\partial I}{\partial \beta}$;

2) construction of analytical dependence of energy of the first half-wave of the discharge current of capacitor *C* in the *RLC*-circle as a function of merit βQ (fig. 4, p. 2), and determination of its optimal value. The error in determining the optimal merit was 0.05% compared to values found on the basis of numerical simulation;

3) definition of multidimensional nonlinear dependencies based on experimental data, presented in the tables (mechanic and energy objects). Here the model is consistently built as a function of one variable, coefficients of which are approximated as functions of the second variable, if there are three variables, the process continues to the third variable.

As a result of such consistent composition, nonessential components in the model are automatically reset, that is the structural identification is correct. At this the search of canonical structure of intermediate one-dimensional models does not create difficulties. The standard regression analysis with brute force of structures is substantially more complex, especially if vector of variable β has large dimension *n*.

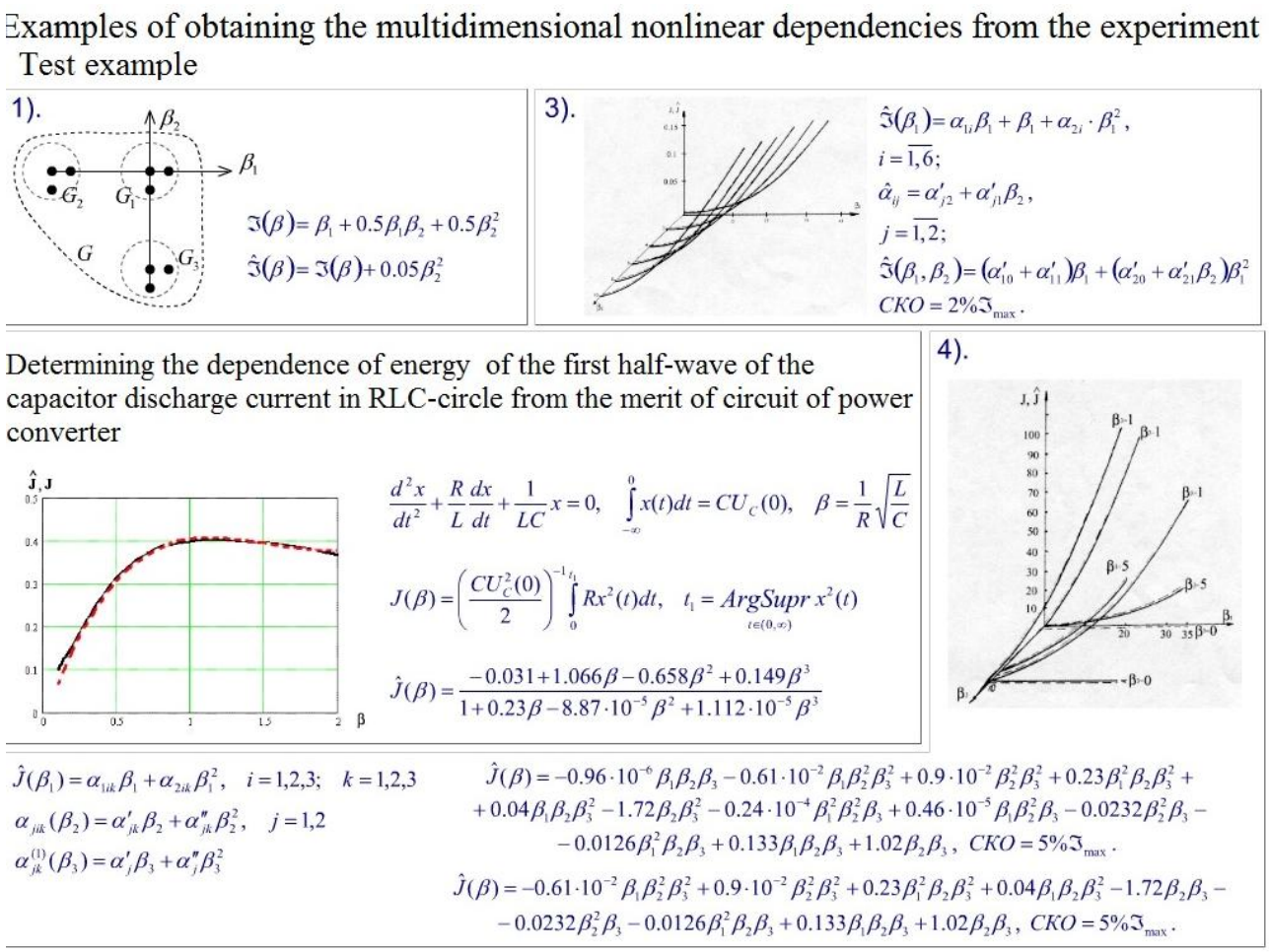
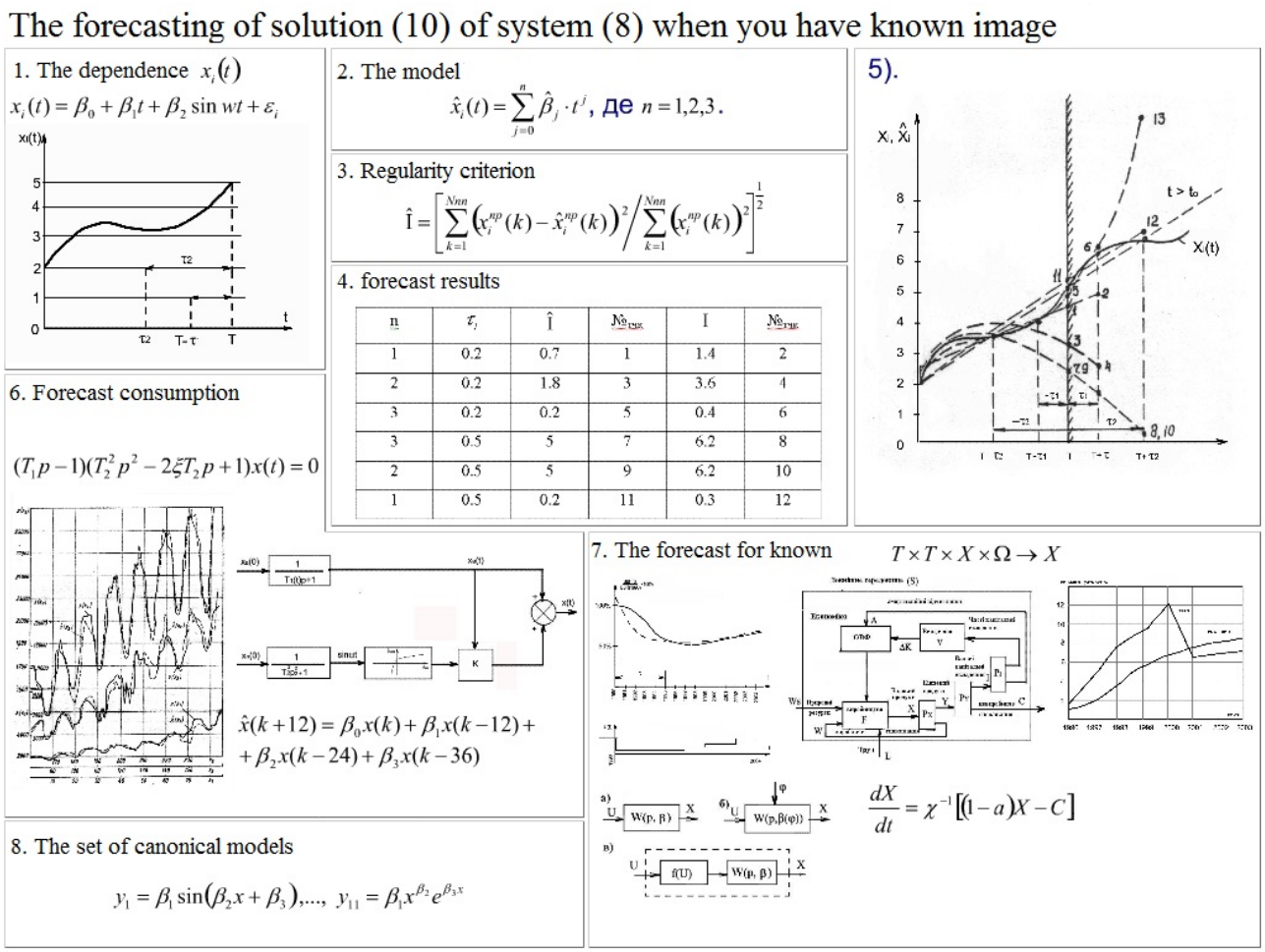


Fig. 4

If the information about the structure of the model is absent, the problem of identification as the problem of approximation has many solutions, but the only one solution will be effective relating to the problem of prediction at a certain time τ.

The test example was considered in fig. 5, p. 1÷4. The unknown dependence is approximated on the interval [0, *T*] by degree polynomials of the Ist, IInd, IIIrd degrees. The higher the degree is, the more accurate approximation is. We can say that the prediction error is proportional to the product of dimension model "*n*" on the interval of the forecast τ: the more "*n*' is', the less τ it needs. The academician A.G. Ivakhnenko proposed to introduce the external criterion to select the structure of forecast model, for example, "regularity" criterion (fig. 5, p. 3).

The perfect *I* and the "external" criterion are compared in the table (fig. 5, p. 4) for the above test. The criterion of "regularity" was not mistaken in choosing the optimal structure for $\tau = 0{,}2$ and $\tau = 0{,}5$. In order to increase the accuracy of the forecast it is proposed to create a base of canonical models ordered on the set of attributes.



n	τ_j	$\hat{I}$	№	I	№
1	0.2	0.7	1	1.4	2
2	0.2	1.8	3	3.6	4
3	0.2	0.2	5	0.4	6
3	0.5	5	7	6.2	8
2	0.5	5	9	6.2	10
1	0.5	0.2	11	0.3	12

Fig. 5

Then the problem of structural identification will be reduced to a problem of image recognition. The nonius approximation is possible with the choice from the base of models at every step of the appropriate specifying supplement (fig. 5, p. 8).

If we have information regarding the structure of the model (fig. 5, p. 6,7), it should be considered in the forecasting models. Thus, the process of energy consumption is determined by the aperiodic trend with seasonal fluctuations that are superimposed on it. This corresponds to a continuous model (fig. 5, p. 6) and its difference analogue in increments of 1 year. The block diagram can be put into compliance with this process (fig. 5, p. 6).

The nonlinear dynamic models, for example Hammerstein's models, take into account the properties of real objects more accurately. In the case of the parameterization of non-linearity $f(u)$ of the model such approach requires the definition of a large number $(n+m+r)$ of interconnected (through an information matrix) unknown parameters.

In order to separate non-parametric estimation of nonlinearity $f(u)$ and dynamic component, the proposed method aims to use the fundamental property - the *smooth* of processes, which is extended to nonlinearity in real objects.

Using the results of arbitrary dynamics of an object we will find the nonparametric model of nonlinear dependence $f(\upsilon)$ or $f(z)$ provided to a minimum of mean square of ($r+1$)-th- derivative from f to υ or to z, or equivalent to its relevant difference of ($r+1$)-th order in accordance to optimized parameters of linear dynamic component. Further, having a model $\hat{f}(\upsilon)$ or $\hat{f}(z)$ we find coefficients of linear component of Hammerstein's models.

The method is used to determine the nonlinear dependences on the dynamic processes of field test of aircraft, electric drive and tare characteristics of thermistor meters TP 100 of gas temperature in main gas pipelines.

The possibility of improving the efficiency of solving the problem of time series' prediction using the main criterion, the extension of the set of identification methods, the use of the set of canonical models are investigated.

The importance of the optimization of natural experiment for the objective determination of the parameters of a mathematical model of dynamic object is presented.

The necessity of a clear division of tasks of the parametric and signal identification is indicated.

The examples of systems of identification and optimization of technological processes of spinning the quartz tube and training on a computer simulator are given.

The following material is only fragments of the theory of identification; it may be effectively used in practice and for the further development of the theory of identification – the mathematically formalized the theory of knowledge of the real world objects.

Chapter 1. The System Approach to Identification

1.1. The Definition of the System

The stage of formalization is the initial step while solving the problem of identification. The task should be set at this stage. The setting of the problem is mapping the input uncertain situation related to the real object into a formalized task which is defined on the set of quantitatively comparable items. These elements are the system, processes that occur in it, the criteria (objectives) and their optimization strategies (achieving objectives).

R. Kalman gave the general mathematical definition of the system [7]. In this book discrete continuous finite linear and smooth dynamical systems definition are fixed in terms of its external behavior. The necessary conditions for the existence of mathematical models (MM) are defined too.

The theorem of state transition function, i.e. the mapping

$$T \times T \times X \times \Omega \to X$$

is the solution of the differential equations, was proved for smooth systems

$$\frac{dx}{dt} = f\left(t,x,u\right), \qquad (1.1)$$

where $x \in X$, $u(t) = \omega\left(t\right) \in \Omega$, $t \in T$, T –is the ordered set of the time moments Ω, T —is the set of the input influences and states.

The system state [7] is that minimum information about the past, which is necessary for complete definition of the future behavior of the system, if the behavior of input variables is known, starting from the current time t_0.

If the set *X* belongs to the finite space, the model (1.1) is a finite system of the nonlinear unsteady first order differential equations.

The linear and (or) stationary equations can be obtained from the system (1.1) with certain assumptions; the discret system is also linear and (or) stationary.

The narrower the range of the variables $\left\{X,\Omega,T\right\}$, the closer the simplified model will be to complete one (1.1) in the space *X*.

If the system (1.1) is smooth there are such $X_1 \in X, \Omega_1 \in \Omega, T_1 \in T$, as the system (1.1) will be an equivalent system of linear differential equations with matrices *A* and *B* with constant coefficients

$$\dot{x}_1(t) = Ax_1(t) + Bu_1(t); \quad y_1(t) = Cx_1(t), \tag{1.2}$$

where $x_1(t)$ - is vector-function $\{x_i(t)\}$, $i = \overline{1,n}$; $u_1(t)$ - is vector-function $\{u_j(t)\}$, $j = \overline{1,m}$; $\dot{x}_1(t)$ - is derivative vector-function $\{\dot{x}_{1i}(t)\}$, $i = \overline{1,n}$; $y_1(t)$ - measurement vector-function $\{y_i(t)\}$, $i = \overline{1,n}$; *; A, B, C* – are matrices $n \times n,$, $n \times m,$ $r \times m$ respectively.

The structure and rank of the matrices *A, B, C* determine the conditions of controllability, observability and identifiability of the system (1.1) [1].

The criterion of optimization or aim which is set by higher level hierarchy system and may be “blurred” in nature.

The formalization of the criterion (aim) is achieved by functional task, which clearly determines the efficiency of system behavior in many cases.

The purpose of the system is providing extreme value of this functional. The functional is the mapping

$$T \times T \times X \times \Omega \to R,$$

where *R* is a set of the real numbers.

The system (1.1) corresponds to a specific real number I from *R* for fixed primary $t_0 \in T$ and ended $t_1 \in T$ time moments, conditions $x_0 \in X$, $x_1 \in X$ system behavior $x(t)$, which caused nonzero initial conditions and control action status $u(t)$.

The control $u(t)$ is optimal if a number I is extremal $I^* \in R$ considering initial conditions, an equation (1.1) and other restrictions for $x(t)$ and $u(t)$. The model (1.1) i.e. the type and parameters of functional mapping f are necessary to find the optimal control strategy $u^*(t)$ by the real object.

The task of identification is the image definition f in (1.1) due to measurement results $x(t)$ and $u(t)$.

Optimum model $\hat{f}$ of the mapping f is evaluated as quality identification criterion J, which is also the functional that reflects the sets of implementations or specific realization $x(t)$ i $x_m(t)$ to a real number J, where $x_m(t)$ is the solution of the model equation:

$$\frac{dx_m}{dt} = \hat{f}(t, x_m, u) \qquad (1.3)$$

for mutual input action with the object $u(t)$.

Since $\hat{f} \neq f$ in general, the value f^* received at the control action on the object $u(t)$ for model (1.3) will be worse than the true optimum value I^*.

The model (1.3) which was built with provision for not only the functional J, but also I, *is* called goal-oriented [35].

If at the set $\{f_i\}$ (i= 1, 2, …n) of the acceptable images, the image $\hat{f}$ (1.3) provides the extreme value for a fixed control $u(t)$, i.e.

$$\hat{f} = \text{argextr} I, \quad \hat{f}_i \in \{\hat{f}_i\},$$

then this model is called optimal goal-oriented.

The nearness of the model to the optimal goal-oriented one is determined by the nearness of measured variables $\hat{x}(t)$, $\hat{u}(t)$ to valid variables $x(t)$ i $u(t)$ of the real object.

On solving the problem of identification in (1.3) and in $J(x, x_m)$ one doesn't put $x(t)$ and $u(t)$, but their estimation $\hat{x}(t)$, $\hat{u}(t)$, the formation of which is the task of object signals identification.

The estimates $\hat{x}(t)$, $\hat{u}(t)$ of the signals $x(t)$ and $u(t)$ in (1.1) are got by using filters f_f:

$$\frac{d\hat{z}}{dt} = f_f(t, \hat{z}, z), \qquad (1.4)$$

where $z(t)$ is the vector-function of measured signals $x(t)$, $u(t)$, $z = \begin{bmatrix} x \\ u \end{bmatrix} + \begin{bmatrix} N_x \\ N_u \end{bmatrix}$, N_x, N_u - are measurement errors x and u respectively.

The filters optimality (1.4) is estimated by filtration quality's criterion J_f which is the functional of signals $z, \hat{z}$ or their spectral characteristics. Since f_f doesn't make perfect conversion z to (x, u), the criterion value J in the problem identification mapping $\hat{f}$ in (1.3) will depend on the criterion J_f in the problem of identification signals $z, \hat{z}$. The value criterion I of the control quality of the real object depends on the criterion value J in the problem identification mapping $\hat{f}$.

Let us introduce the concept of goal-orientation and optimality for filter (1.4).

The filter (1.4) is goal-oriented if the problem of identification $\hat{f}$ was taken into account in its construction in (1.3). If on the set of admissible mappings $\{f_{fi}\}$ the mapping in (1.4) for a fixed $\hat{f}$ provides extreme value J^*, i.e.

$$f_f = \text{argextr} J, \; f_{fi} \in \{f_{fi}\},$$

then the filter will be goal-oriented and optimal.

The functionals' optimization I, J, J_f is linked with a concept of the relaxation process [16]. The set $\{Q_k\}$, $k \geq 0$, each element of which belongs to some (usually convex) region G of the normalized space is called the relaxation process (RP) in relation to optimized functions $F(Q_k)$, if the sequence $F(Q_k)$ is also ordered for the value k. The set of the minimization task $F(Q_k)$ is $F(Q_0) \geq F(Q_1) \geq F(Q_2) \geq \ldots$ RP coincides the functionality if

$$\lim_{k\to\infty}(F(Q_k)-F(Q^*))=0,$$

and simple coincides if

$$\lim_{k\to\infty}Q_k=Q^*,$$

where Q^* - is true value.

Let us expand the definition RP in case if items Q_k are the region *G* of the space, where the concept of standards or distance is not defined. The sign $\geq$ is understood as a symbol of order relation. For example, the estimation of the vector β of the model parameters (Σ,β) RP is not strictly relaxation in the problem of identification while putting in the subset of models structures Σ and methods set *Opt*.

The set $\{Q_k\}, k\geq 0$ will be called RP as to $F(Q_k)$, if for any $k>0$ there is such value $j>0$, as $F_{k+j}\leq F_k$. Such RP is not rigorously relaxation. Its convergence on functional *F* or on Q_k depends on dimension and ordered of the subset non-normalized elements (is (Σ, Opt)).

In common identification system we will call the multiply subset $\{\Sigma,\beta\}$, $\{Opt,\alpha\}$, $\{J\}$, $\{I\}$, $\{T\}$, that is designated as $\{\Sigma,\beta\}$, $\{Opt,\alpha\}$, $\{J\}$, $\{I\}$, $\{T\}$ and have the structure which allows to realize the relaxation process relating to the indicator such as $F_{k+j}\leq F_k$, $k=1, 2\ldots$; $j>0$; $Q_k\in\{\{\Sigma,\beta\},\{Opt,\alpha\}\}$, $F_k\in\{\{J\},\{I\}\}$.

Here $\{\Sigma,\beta\}$ - is the models' subset; $\{Opt,\alpha\}$ is estimation methods subset of the vector models parameters $\{\Sigma,\beta\}$; α - is parameters' vector of the method *Opt*; $\{J\}$ - is the subset of the functionals from $\varepsilon(\beta,t)$ optimized for β by methods $\{Opt,\alpha\}$;

$\varepsilon(\beta,t)$ is the difference of the coordinate measurement of the real object Σ_∞ and the model;

$\{I\}$ -is the subset of the basic functional which determine the quality of solving the basic task for the real system Σ_∞ by using model $\{\Sigma,\beta\}$;

$\{T\}$- is the subset of instants of time t_k, where the real system Σ_∞ is presented by the data set $\{u(t_k),y(t_t)\}$ as a system in terms of its external behavior [7].

Table 1.1

№	type of signal	β	$\{Q_k\}$		Opt	$\{F_k\}$	
			Σ	α		J	I
1	is	$\{\beta_k\}$	Σ	α	Opt	J	-
2	IS	$\{\beta_k\}$	$\{\Sigma\}$	α	Opt	J	-
3	ias	$\{\beta_k\}$	Σ	$\{\alpha_k\}$	Opt	J	-
4	IAS	$\{\beta_k\}$	$\{\Sigma\}$	$\{\alpha_k\}$	Opt	J	-
5	mais	$\{\beta_k\}$	Σ	$\{\alpha_k\}$	$\{\text{Opt}\}$	$\{J_k\}$	I
6	MAIS	$\{\beta_k\}$	$\{\Sigma\}$	$\{\alpha_k\}$	$\{\text{Opt}\}$	$\{J_k\}$	I
7	SOS	$\{\beta_k\}$	$\{\Sigma\}$	$\{\alpha_k\}$	$\{\text{Opt}\}$	$\{J_k\}$	$\{I_q\}$

The more effective identification system is the less prior information is necessary for its work. The required identification quality is achieved by adaptation. The latter is the goal-oriented change of one or all elements of the subset $\{Q_k\}$ for in order to achieve the extremum of the main index *F*. The better adaptation algorithm is the more efficient is the system. If the complexity is ignored, then the wider is the subset $\{Q_k\}$ during the adaptation $\{Q_k\}$, the more efficient is the result.

If the complexity of the system is taken into account in the index *F*, there is the subset potency $\{Q_k\}$ which is optimal for index *F*. Generally the subset $\{Opt,\alpha\}$ consists from one element for method type *Opt* in the adaptive identification system. RP $\{\beta_k\}$ is common for all types of identification systems for index *J*, the subset configuration $\{Q_k\}$ and $\{F_k\}$ is different.

There are identification systems for different $\{Q_k\}$ and $\{F_k\}$ in the table 1.1.

1. In a narrow sense the identification systems (is) implements RP $\{\beta_k\}$ referring to J if Σ, Opt, α are constant and the subset I is empty, namely $Q_k = \{\beta_k\}$, $F=J$.

The identification task is equivalent to the optimization task $J\{\beta_k\}$ which can be solved by theory of linear and nonlinear estimating methods or mathematical programming algorithms, it depends on the structure J.

2. In a wider sense the identification systems (IS) implements RP Σ_k, β_k referring to J if Σ, Opt, α are constant and the subset I is empty, namely $Q_k = \{\Sigma_k, \beta_k\}$, $F=J$.

For example, the dynamic orthogonal vernier or regression models with variable dimension vector β^*, for which the optimal pair $\{\Sigma^*, \beta^*\}$ is determined from the condition

$$\{\Sigma^*, \beta^*\} = \underset{\{\Sigma_k, \beta_k\}}{arg\,min}\, J\{\Sigma_k, \beta_k\},$$ are considered in [37].

3. In a narrow sense the adaptive identification systems (ias) implement RP $\{\Sigma_k, \beta_k\}$ referring to J if Opt is constant and the subset I is empty, i.e. $Q_k = \{\beta_k, \alpha_k\}$, $F = J$.

The systems, using accelerated gradient descent $\{\beta_k\}$ for J, have regularization parameter as α_k, which is optimized with supporting terms of the minimum of the difference of the mean-square values of the errors for two similar models. These models are adjusted identical algorithms Opt, but they are different in parameter α : for the first one it is $\alpha = \alpha_k$, for the second one it is $\alpha = \alpha_k + \Delta$, where $\Delta > 0$, $\alpha_k > 0$.

4. In a wider sense the adaptive identification systems (AIS) implement RP $\{\Sigma_k, \beta_k, \alpha_k\}$ referring to J if Opt is constant and the subset I is empty, i.e. $Q_k = \{\Sigma_k, \beta_k, \alpha_k\}$, $F = J$.

For example, we assort the systems of the “best regression” selection by the algorithm of the ridge regression [5].

5. In a narrow sense the multiply adaptive identification systems (mais) implement RP $\{\beta_k, Opt_k, \alpha_k\}$ referring to the main index I as a composition of the two RPs:

a) RP $\{\beta_{kn}\}, n = 1, 2, ...$, referring to J_k;

b) RP $\{Opt_k, \alpha_k, J_k\}$, $k = 1, 2, \ldots,$ in respect of I; that $Q_k = \{\beta_k, Opt_k, \alpha_k\}$ $F = \{\{J_k\}, \{I\}\}$.

The multiplicity repeats the RP $\{\beta_{kn}\}$ for each element $\{Opt_k, \alpha_k, J_k\}$ of the higher rank RP. The specific method Opt corresponds to each index J_k and it is absent for heuristic algorithms. Then the pair "the method and its parameters" is optimized in reference to I.

As an example we can take the system "MIAS -1". The optimal element $\{\beta^*, Opt^*, \alpha^*\}$ is found from seven methods $\{Opt_k\}$ and their parameters $\{\alpha_k\}$, for one of the five index I (or their weighted sum), which include the regularity and unbiasedness criteria [4], the accuracy of the prediction by the model.

6. In a wider sense the multiply adaptive identification systems (MAIS) implements RP $\{\Sigma_k, \beta_k, Opt_k, \alpha_k\}$ in respect of the main index I as a composition of the two RP:

a) RP $\{\beta_{kn}\}$, $n = 1,2,...,$ referring to J_k;

b) RP $\{\Sigma_k, Opt_k, \alpha_k, J_k\}, k = 1, 2, \ldots,$ in respect of I, that $Q_k = \{\Sigma_k, \beta_k, Opt_k, \alpha_k\}$, $F = \{\{J_k\}, \{I\}\}$, where I – is the higher rank index.

There are systems, which are an example of the MAIS. They were built on the basis of the method of the group accounting arguments [5]. In this methods the structure Σ and parameters β are selected by the condition of the extreme I. The reasonable extension of the set $\{Opt_k\}$ provides obtaining the model which is more effective by the criterion I.

The qualitative difference of the systems of p. 5.6 is their goal orientation on the index I, which regulates and optimizes the selection task $\left\{\Sigma^*,\beta^*,Opt^*,\alpha^*\right\}$. The issue of evaluation of the quality of the model obtained as a result of the identification is not considered for multiply adaptive identification systems. The model quality is estimated for the index I. This is the most objective assessment, because that model is good the use of which provides the best solution to the basic problem, which defined quality of the index I.

7. The self-organizing control system implements RP $\left\{\Sigma_k,\beta_k,Opt_k,\alpha_k,J_k,I_k\right\}$ referring to the some index $\Lambda(I)$ as a composition of three RPs

a) RP, $\left\{\beta_{qkn}\right\}$, $n=1,2,..$ (q and k are constant), referring to J_{qk} with the stationary point

$$\beta_{qk}=\text{argextr}\beta_{qkn}J_{qk}(\beta_{qkn},\Sigma_q,Opt_{qk},\alpha_{qk})=\arg J_{qk}^*;$$

b) RP $\left\{J_{qk}^*(\beta_{qk},\Sigma_{qk},Opt_{qk},\alpha_{qk})\right\}$, $k=1,2,..$ (q is constant), referring to I_q with the stationary point

$$J_q^*=J_q(\beta_q,\Sigma_q,Opt_q,\alpha_q)=\text{argextr}I_q(J_{qk}^*)=\arg I_q^*;$$

c) RP $\left\{I_q^*(\beta_q,\Sigma_q,Opt_q,\alpha_q)\right\}$, $q=1,2,\ldots,$ referring to the higher rank index Λ with the stationary point

$$I^*=\text{argextr}\Lambda(I_q^*)=\arg\Lambda^*,$$

viz $Q_k=\left\{\Sigma_k,\beta_k,Opt_k,\alpha_k\right\}$, $F=\left\{J,I_k\right\}$.

Let us consider the task of improving the automated control process by the complex power aggregate.

Suppose Λ is the enterprise performance index $\left\{I_q\right\}=\left\{I_1,I_2\right\}$; I_1 – is the quality index of stabilization process variables x of the aggregate in the field of the operating

modes x_0, plotted by experts; I_2 – is the index quality of the aggregate's operation; $\{J_{qk}\}$ – is the subset of the quality indicators of identification by the local models $\{\Sigma_{q1},\beta_{q1}\}$ which reflect the control u into x $(k=1)$, and by quality models $\{\Sigma_{q2},\beta_{q2}\}$ which reflect u into I_2 $(k=2)$. The local models $\{\Sigma_{1k},\beta_{1k}\}$ are determined for J_{1k} by the easiest identification methods at the first stage of the automation. The local regulators selected for the models $\{\Sigma_{1k},\beta_{1k}\}$ stabilize x in the range x_0 for the minimum criterion I_1. The optimal element is determined for I_1 in accordance with the RP of p.7,a and 7,b. At the second stage, having accomplished the process of collecting and processing information automated, the complete model $\{\Sigma_{q2},\beta_{q2}\}$ that connects I_2 with variables x, u can be built. The local regulator parameters and the optimal value x_0^* of the operating modes x_0 can be refined for the models $\{\Sigma_{q2},\beta_{q2}\}$, including models $\{\Sigma_{q1},\beta_{q1}\}$. The optimal element $\{\Sigma_2^*,\beta_2^*,Opt_2^*,\alpha_2^*\}$ for I_2 can be determined in accordance with the result RP of p.7,a, 7,b. Thus, the transition from Σ_1 to Σ_2 leads to the replacement (according to RP p.7,c) of the main index I_1^* and I_2^*.

1.2. The Multi-Level Decomposition of Systems

The decomposition is an opportune approach to represent the large dimension system exactly or approximately with complex network of forward and backward linkages as a system with simpler subsystems that are better amenable to formalization.

Imagine the task of designing the system and the optimal adaptive control system of a real object as three subtasks and subsystems corresponding to them:

-optimal control of a real object;

-identification of mapping “input-output” of an object;

- object signals’ identification.

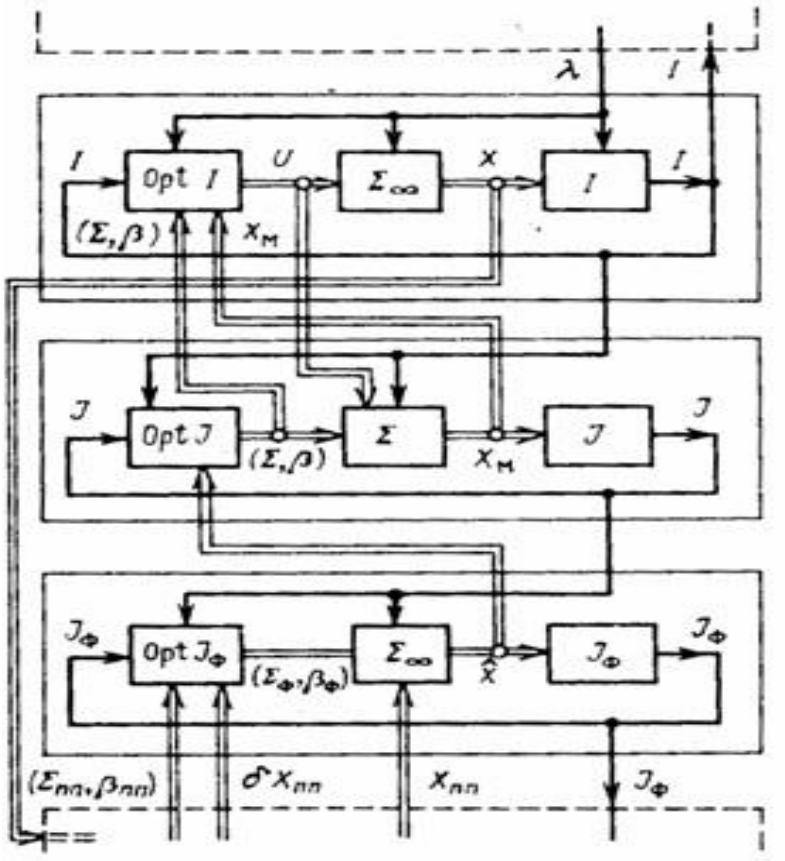


Fig. 1.1 Three levels of the optimization in the multi-level system.

These elements generate the closed optimization system for local quality index inside each subsystem which works as a system with backward linkages:

- issuing the controlling action on the object;
- measurement or calculation the object reaction;
- evaluation of subsystems’ optimality according to the criterion of its level.

The subsystems are combined by direct (bottom-up) and reverse (downward) linkages. The extreme subsystems are united by the same links with the systems of low and high levels which are not considered here. Let us present the designation and physical content of the elements, internal and external direct and feedbacks.

The signal identification system has such elements and linkages:

J_f – is a creator of the identification (filtration) quality index of the vector x of object’s signals;

Σ_f - is the filter which transforms the signal of a primary converter output x_{nn} to its estimated value $\hat{x}$;

$OptJ_f$ - is the optimizer of J_f of the structure Σ_f and the vector β_f of the filter's parameters;

Σ_{nn}, β_{nn} - is the information on the structure and parameters of the primary transducers (sensors) of the physical object's variables, Σ_f is the transfer function of a sensor, static (tare) characteristic etc;

δx_{nn} - is the information on sensor's errors: systematic, random, time; frequency and probability characteristics (distribution law, its parameters, correlation functions etc);

J, J_f -is the information on the optimality criteria of identification systems of an object and its signals: J enters this subsystem through the feedback channel, J_f - enters the subsystem of primary converters.

The subsystem of the object identification consists of:

J – a creator of the quality index of the object's identification;

Σ,β - an object's model with the structure Σ and parameters' vector β;

$Opt\ J$ – an optimizer of J of a structure Σ and parameters' vector β;

$\Sigma_f, \beta_f, \hat{x}$ - the information on the filter and the evaluation $\hat{x}$ of the signal x, taken from signal x_{nn} of the primary converter;

Σ, β, x_M - the information on the model of the controlled object (Σ,β) and the evaluation x_M of the signal x, received in terms of filter signal $\hat{x}$;

I, J – correcting feedbacks from the control system to the identification system and from the object identification system to the identification system.

The subsystem of optimal control of an object includes:

I - a creator of quality control index;

Σ_∞ - a real object (its input and output variables u and x);

Opt I - an optimizer of the index I in terms of control signal u ;

x – an object's output – an input of primary converters' subsystem;

u - an input of an object and a model;

Λ, I - correcting feedbacks from the system of higher level to the control system and from the control system to the identification's system respectively.

Every element (optimizer, model, criterion's creator) has three components in it:

- an element itself resolving its own task;
- the set of elements ordered by their properties;
- a projector (PR) that chooses the optimal element from the elements' set in terms of the higher level quality system index.

Fig.1.2 shows the schematic representation of the tasks of systems design and the systems of the Ist, IInd, IIIrd levels. The indices q, qk, qkn correspond to steps' designation of the composition and relaxation process of the optimization system. For example, q – is a number of iteration of the index's change; qk – are J changes; qkn – an increment number in the optimization β in the system of object's identification.

The whole system consists of 27 elements, among them there are methods of optimization sets $\left\{OptJ_f^q\right\}$, $\left\{OptJ_q\right\},\left\{OptI_q\right\}$, criterion sets $\left\{J_f^q\right\},\left\{J_q\right\},\left\{I_q\right\}$ and models $\left\{\Sigma_f^q\right\},\left\{\Sigma_q\right\},\left\{\Sigma_\infty^q\right\}$ which are composed of the sets of the existing ordered items.

There is a single optimal set of these elements and their parameters for every specific criterion.

The task is to select the "appropriate" set, if the costs for its search are limited. The problem of finding the "appropriate" set may be implemented by sorting options, if the priory information about all elements is unknown.

The complete enumeration guarantees finding the optimal set, if costs for optimum's search are not included in the optimality criterion. If these costs are significant, they affect the optimal solution. In order that this decomposition does not lead to complications, it is necessary to carry out an elaborate analysis of the subsystems and their elements. The use of the projector Pr of each level (fig. 1.2) allows narrowing down the

initial set of the elements to a limited subset of the contenders for the best ones for this particular situation.

The principle of the decomposition is efficient for complex systems and their components. Thus, the infeasibility to formulate all requirements to the designed system using just one functional led to the problem solvable in practice – this informal criterion is decomposed onto the set of criteria that are subjected to strict formalization and decision of multi-criterion optimization problem.

This solution does not give the only result. It only can provide the limited area (Pareto's subset) in the criteria-functionals from the optimized variables. The final version of the solution, which is selected by a designer, belongs to this subset.

The widely accepted is the principle of decomposition for optimization's methods. The easiest method in terms of I of the coordinate optimization (Gauss-Seidel's method) and the method of group relaxation carry out the decomposition of the space's optimizing variables and step-by-step coordinate-wise or group optimization;

for linear stochastic control systems that are optimal in the terms of the quadratic functional, the decomposition of common task into subtasks of the state optimal evaluation and finding the optimal control strategy [13] enable to simplify the difficult task of the dual control [44];

- the division (decomposition) into independent subtasks of signals' evaluation, their covariance matrices and parameters provides a solution, which is closer to optimal, using simple algorithms for the problem of simultaneous estimation of the parameters and states of the stochastic objects (extended Kalman's filter [4], the method of the quasi-linearization and invariant embedding [4]).

The model decomposition's task (1.1), which describes the system of nonlinear unsteady equations

$$\dot{x}_g(t) = f_g(t, x_1, ..., x_n, u_1, ..., u_m), g = \overline{1,n}. \qquad (1.5)$$

and set of linear systems which describe a system's behavior (1.1) with required closeness at the limited areas of determination of the variables t, x, u, is of a particular interest.

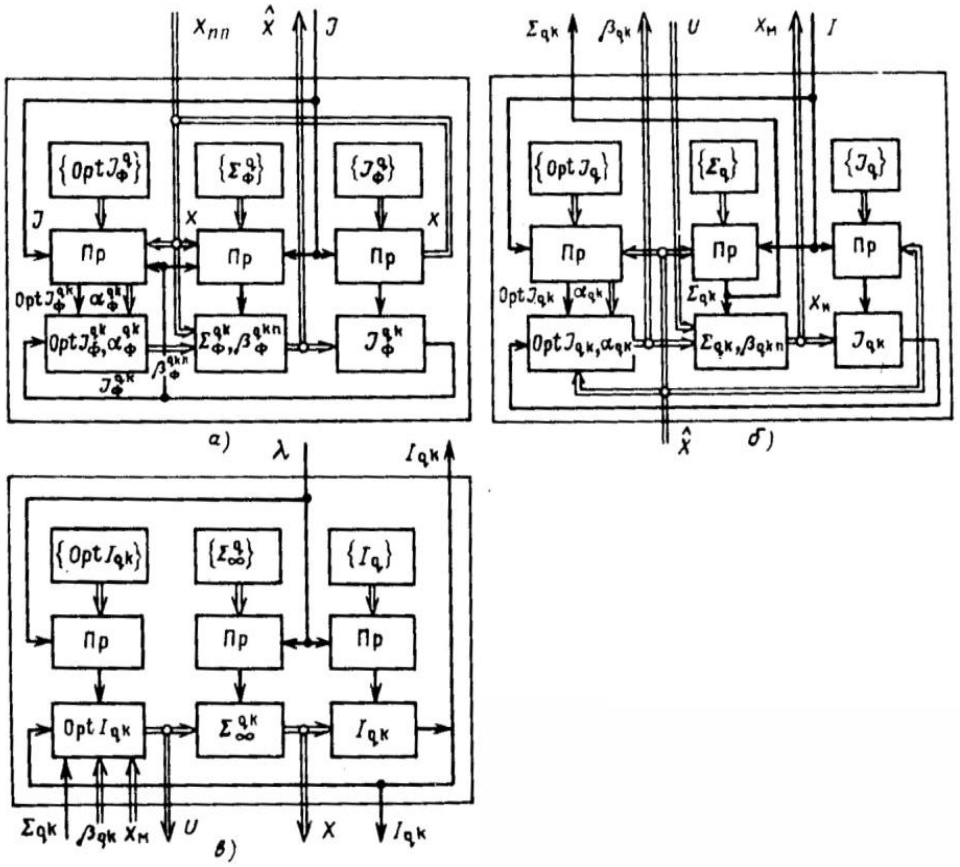


Fig. 1.2. The goal-oriented subsystem of identification of signals (a) and an object (b) and optimal adaptive control of an object.

These local models within the limits linearly proportional systems allow solving the problems of analysis, synthesis and implementation of the optimal control and identification strategies.

The identification problem of the structure of an unknown nonlinear dependence *f* of the model (1.1) can be solved using the local models.

Let us input (1.5) in the denotation of a vector *v* with components ν_i:

$$\nu_i = \begin{cases} x_i, i = \overline{1,n}; \\ u_i, i = \overline{n+1, n+m}; \\ t, i = n+m+1 = s. \end{cases} \qquad (1.6)$$

By reason of the continuity and the i-fold differentiate dependence (1.1), the dependence (1.5) can be represented by Taylor series:

$$y(v(t)) = y(v_0) + \sum_{i=1}^{s} \frac{\partial y}{\partial v_i}\bigg|_{v_0} \Delta v_i(t) = \frac{1}{2}\sum_{i=1}^{s}\sum_{j}^{s} \frac{\partial^2 y}{\partial v_i \partial v_j}\bigg|_{v_0} \Delta v_i \Delta v_j +$$

$$+\frac{1}{6}\sum_{i=1}^{s}\sum_{j=1}^{s}\sum_{k=1}^{s} \frac{\partial^3 y}{\partial v_i \partial v_j \partial v_k}\bigg|_{v_0} \Delta v_i \Delta v_j \Delta v_k + \cdots + R_l(v), \qquad (1.7)$$

where $\Delta v_i(t) = v_i(t) - v_{i0}; v \in G; v_0 \in G;$

$\mathrm{Sup} R_l(v) \le \delta; \delta$ - is the permissible error of the dependence of approximation (1.1) by series (1.7).

Let us divide the region of variables v determination G into subregions so that

a) $\bigcup_{r=0}^{m} G_r = G, v_{r0} \in G_r, G_r \cap G_{r'} = f$, $r \ne r'$;

b) $(\forall G_r)(\exists \beta_r = \underset{\beta_r \in R}{\mathrm{argmin}}(\sup_{v_{r0}} \left\| \frac{dy}{dx} \right|_{v_r 0} \beta_r < h))$

where β_r –is parameters' vector of the linear model

$$\Delta y_1(t) = \beta_r^T \Delta v(t), \qquad (1.8)$$

β_{ri} is a component with number i and evaluation with the accuracy to h of the expansion coefficient $\frac{\partial y}{\partial v}\Big|_{v_0}$ (1.7) around the centre v_{r0}.

Let us find the β_r vector, which is equal to vector $\frac{\partial y}{\partial v}\Big|_{v_0}$ (with the accuracy to h) , i.e. let us obtain estimations of the first derivatives $\frac{\partial y}{\partial v}\Big|_{v_0}$ in (1.7) for the region G_0 that contains the global center v_0 of the expansion (1.7) and uses the model (1.8) and data's $y_1(t), v(t)$.

In a similar way let us calculate the vectors β_r of the partial models (1.8.) for other regions. Now the derivatives of the higher orders can be evaluated using values β_r and ν_{r0}.

Thus, if the subregions neighboring with G_0, merge to extended $G_{01} = U_{r=1}^{m_1} G_r, \quad m_1 \geq s,$ (the regions are numbered with distance from the "central" G_0), such as

$$\sup R_3(\nu) \leq \delta, \quad \nu \in G_{0j},$$

then the decomposition (1.7) will include members of no higher than the second order with the accuracy to δ. Let us differentiate (1.7) with respect to ν_j in the region G_{01}, then with the accuracy to $\partial R_3(\nu) / \partial \nu_j$

$$\frac{\partial y(\nu(t))}{\partial \nu_j} = \left.\frac{\partial y}{\partial y_j}\right|_{\nu_0} + \sum_{i=1}^{s} \left.\frac{\partial^2 y}{\partial \nu_i \partial \nu_j}\right|_{\nu_0} \nu_i, \tag{1.9}$$

where $\left.\frac{\partial y}{\partial y_j}\right|_{\nu_0}$ is found from (1.8) for the region G_0.

If $\Delta \nu_i$ is diminution $\Delta \nu_i = \nu_{roi} - \nu_{0i}$, the derivative is closely related to

$$\left.\frac{\partial y(\nu(t))}{\partial \nu_j}\right|_{r_0} \approx \beta_{rj}.$$

Thus, the expression (1.9) can be represented as

$$\beta_{rj} = \beta_{0j} + \sum_{i=1}^{s} \left.\frac{\partial^2 y}{\partial \nu_i \partial \nu_j}\right|_{\nu 0} (\nu_{r0i} - \nu_{0i}). \tag{1.10}$$

If $r \geq s$ and if s from r vectors $(\nu_{r0} - \nu_0)$ is of linear independency, s of other derivatives is clearly defined $\left.\frac{\partial^2 y}{\partial \nu_i \partial \nu_j}\right|_{\nu_0}$, i= 1, 2,., s.

One must have no less than *s* advanced around G_r^* subregions G_{r1}^*, similar to G_{01} in order to find out all coefficients $\left.\frac{\partial^3 y}{\partial \nu_i \partial \nu_j \partial \nu_k}\right|_{\nu_0}$ of the series (1.7) (the number of such central subregions must be at least s).

The second derivate $\left.\frac{\partial^2 y}{\partial \nu_i \partial \nu_j}\right|_{r_0^*}$ are found for each G_{r1}^* from the equation (1.10), where values obtained for the central subregion G_r^* were taken instead β_0 and ν_{0i}.

If we differentiate twice (1.7), we get

$$\left.\frac{\partial^2 y(t)}{\partial \nu_i \partial \nu_j}\right|_{\nu_{r_0}^*} = \left.\frac{\partial^2 y}{\partial \nu_i \partial \nu_j}\right|_{\nu_0} + \sum_{k=1}^{s} \left.\frac{\partial^3 y}{\partial \nu_i \partial \nu_j \partial \nu_k}\right|_{\nu_0} (\nu_{r_{0k}^*} - \nu_{0k}) \qquad (1.11)$$

The third derivatives are identically found from (1.11) for $r^* \geq s$ and if *s* vectors $(\nu_{r^*} - \nu_0)$ are of linear independency. All derivatives of the series (1.7) are taken in this way and they determine series' structure.

The structure restoration of the those vectors ν_j that changes from region to region is possible in case of the insufficient number of subregions G_r. On conducting active experiments the minimization of the regions' number G_r and optimization of the accuracy of derivatives values in (1.7) are achieved by using composite experimental design that provides orthogonality of vectors $(\nu_{r_0} - \nu_0), (\nu_{r_{r0}^*} - \nu_0)$.

The limited (at every step of expansion) dimension of a vector of unknown parameters, analysis capability at every step of the significance of found coefficients of the series (1.7) and rejection of the insignificant coefficients is the advantage of this approach.

The example 1.1. The unknown dependency (1.6) has the form

$$y(\nu) = \nu_1(t) + 0{,}5\nu_1(t)\nu_2(t) + \nu_2^2(t),$$

Where the variables $y(\nu), \nu_i(t)$ are defined as follows $(\nu_{01} = 0, \nu_{02} = 0) \in G_0$; $(\nu_{11} = -1, \nu_{12} = 0) \in G_1$; $y(0) = 0$.

There are two tests $\nu_i\left(t_{jr}\right)$ in each region, deviations of which from the center are the same for each region $\Delta\nu_1(t_{1r}) = 0{,}1$; $\Delta\nu_2(t_{1r}) = 0$; $\Delta\nu_1(t_{2r}) = 0$; $\Delta\nu_2(t_{2r}) = -0{,}1$. Taking into account these deviations we get:

$\Delta y(t_{10}) = 0{,}1$; $\Delta y(t_{20}) = 0{,}005$; $\Delta y(t_{11}) = 0{,}1$; $\Delta y(t_{20}) = 0{,}055$; $\Delta y(t_{12}) = 0{,}5$; $\Delta y(t_{22}) = 0{,}105$.

The structure and parameters of the unknown dependency $y(\nu)$ are necessary to determine.

The solution of the task:

a) let us make the equation (1.8) for G_0:

$$\Delta y(t_{j0}) = \beta_{01}\Delta\nu_i(t_{j0}) + \beta_{02}\Delta\nu_2(t_{j0});$$

substituting the data, we obtain

$$0{,}1 = \beta_{01} \cdot 0{,}1 + \beta_{02} \cdot 0;\ 0{,}005 = \beta_{01} \cdot 0 + \beta_{02} \cdot (-0{,}1).$$

Hence we have the sought first members of the series (1.7):

$$\beta_{01} = 1;\ \beta_{02} = -0{,}05;$$

b) for G_1:

$$\Delta y(t_{j1}) = \beta_{11}\Delta\nu_1(t_{j1}) + \beta_{12}\Delta\nu_2(t_{j1}),$$

from here $\beta_{11} = 1$; $\beta_{12} = 0{,}55$;

c) for G_2:

$$\Delta y(t_{j2}) = \beta_{21}\Delta\nu_1(t_{21}) + \beta_{22}\Delta\nu_2(t_{j1});$$

from here $\beta_{21} = 0{,}5$; $\beta_{22} = -1{,}05$;

d) having β_{ir}, it is possible to proceed to find the second derivatives in (1.7) according to the equation (1.10):

$$\left.\frac{\partial y}{\partial \nu_1}\right|_{\nu_{r0}} - \left.\frac{\partial y}{\partial \nu 1}\right|_{\nu_0} = \left.\frac{\partial^2 y}{\partial \nu_1^{\,2}}\right|_{\nu_0} (\nu_{r01} - \nu_{01}) + \left.\frac{\partial^2 y}{\partial \nu_1 \partial \nu_2}\right|_{\nu_0} (\nu_{r02} - \nu_{02}),$$

where $(\nu_{r0\,j} - \nu_{0\,j})$, j = 1,2 is the divergence of centers' coordinates for r and 0 region.

The second equation is similar

$$\left.\frac{\partial y}{\partial \nu_2}\right|_{\nu_{r0}} - \left.\frac{\partial y}{\partial \nu_2}\right|_{\nu_0} = \left.\frac{\partial^2 y}{\partial \nu_2^{\,2}}\right|_{\nu_0} (\nu_{r02} - \nu_{02}) + \left.\frac{\partial^2 y}{\partial \nu_1 \partial \nu_2}\right|_{\nu_0} (\nu_{r01} - \nu_{01}).$$

Having substituted the data for the first equation, we obtain the system

$$\left.\begin{aligned} 0 &= \left.\frac{\partial^2 y}{\partial \nu_1^{\,2}}\right|_{\nu_0} \cdot(-1) + \left.\frac{\partial^2 y}{\partial \nu_1 \partial \nu_2}\right|_{\nu_0} \cdot 0; \\ -0{,}5 &= \left.\frac{\partial^2 y}{\partial \nu_1^{\,2}}\right|_{\nu_0} \cdot 0 + \left.\frac{\partial^2 y}{\partial \nu_1 \partial \nu_2}\right|_{\nu_0} \cdot(-1). \end{aligned}\right\}$$

Its solution instantly indicates the lack of the quadratic dependency from x_1:

$$\left.\frac{\partial^2 y}{\partial \nu_2^{\,2}}\right|_{\nu_0} = 0; \quad \left.\frac{\partial^2 y}{\partial \nu_1 \partial \nu_2}\right|_{\nu_0} = 0{,}5.$$

Analogously, from the second equation

$$\left\{\begin{aligned} 0.5 &= \left.\frac{\partial^2 y}{\partial \nu_2^{\,2}}\right|_{\nu_0} \cdot 0 + \left.\frac{\partial^2 y}{\partial \nu_1 \partial \nu_2}\right|_{\nu_0} \cdot(-1); \\ -1 &= \left.\frac{\partial^2 y}{\partial \nu_2^{\,2}}\right|_{\nu_0} \cdot(-1) + \left.\frac{\partial^2 y}{\partial \nu_1 \partial \nu_2}\right|_{\nu_0} \cdot(0); \end{aligned}\right.$$

we obtain the solution

$$\left.\frac{\partial^2 y}{\partial \nu_2^{\,2}}\right|_{\nu_0} = 1; \quad \left.\frac{\partial^2 y}{\partial \nu_1 \partial \nu_2}\right|_{\nu_0} = 0{,}5.$$

We substitute the obtained values of the first and second derivatives in (1.7). Considering that $\nu_0 = 0$, $y_0 = 0$, we find the estimate $\hat{y}(\nu)$ of the unknown dependence $y(\nu)$:

$$\hat{y}(\nu) = \nu_1(t) - 0{,}05\nu_2(t) + 0{,}5\nu_1(t)\nu_2(t) + \nu_2^2(t).$$

The estimate $\hat{y}(\nu)$ differs from $y(\nu)$ at a non-essential value of $-0{,}05\nu_2(t)$ for the region G, which is due to the proximity of calculating first derivatives.

The equally effective principle of decomposition is relevant to the signals' description. We can speak about the temporary and frequency partition of signals. The temporal partitioning can be built on various grounds, such as the selection of slots corresponding to static and dynamic modes of the object, for the independent studies of its statics and dynamics; the selection of the strong and low noise recording areas of the variables for the optimal choice of the intervals and the algorithm of signals' estimation; the selection of the "informative" recording areas (in the sense of object model identification) with the orthogonal or uncorrelated variables.

The frequency partitioning is equivalent to the signal decomposition in Fourier series. The transition to the frequency domain makes it possible to synthesize the optimal Wiener filter for evaluating signals [34,45]. The modern theory of the spectral analysis and synthesis systems [1] is also based on the decomposition of signals and their functional mappings. The decomposition is performed on the system of orthogonal functions. The decomposition of the complex signals oscillatory systems can identify individual tones and build partial simplified aircraft models for them [8]. Thus, the principle of the decomposition is appropriate at all levels and for all elements of the complex systems.

1.3. The Principle of Criteria's Consistency

There is at least one value for the set elements of the hierarchical control system (fig 1.1)

$$\left\{ I, OptI, J, OptJ, \Sigma, \beta, I_f, OptI_f, \Sigma_f, \beta_f, I_{nn}, OptI_{nn}, \beta_{nn} \right\},$$

in which the main index Λ is optimal if losses for searching this value are not taken into account.

The real optimization tasks can't take these losses into account. Therefore, the relaxation process of finding the absolute extremum (for example, minimum) stops at some stage if further increment in losses for search exceeds the reducing of the rest of the functional.

The search of the absolute extremum is performed by brute force if the information on Λ dependence on the set's elements is absent. The general number of steps RP $C_n^m = \frac{n!}{(n-m)!}$ for the real value *m* and *n* is too big.

The task is in many orders easier if each of the functionals $\Lambda, I, ..., J_{nn}$ is optimized on the variable subset of its level: *X, u, .. .,* $(\Sigma_{nn}, \beta_{nn})$. This leads to the composition of conventionally optimal solutions

$$\Lambda^* = ext\Lambda \; x\big|_{I^*=extrI} \; u\big|_{I^*=extrI} \; (\Sigma,\beta)\big|_{J_f^*=extrJ_f} \; (\Sigma,\beta)\big|_{J_{nn}^*=extrJ_{nn}} \; (\Sigma_{mm},\beta_{nn})\big|_{...}, \tag{1.12}$$

where at the left of each vertical bar there is a variable that optimizes the functional of "its" level, and at the right – there is the result of conditional optimization at lower levels.

The optimization problem of each level undergoes quite strict formalization that allows intensifying the process of finding the conditional extremum. The dimension problem is withdrawn, but the problem of efficient solving is still actual. In order that the procedure (1.12) gives the solution close to the absolute extremum of the main functional it is necessary to ensure the consistency of criteria $\Lambda, I, ..., J_{nn}$. The criteria $\Lambda, I, ..., J_{nn}$ are called absolutely consistent if the composition of conventional solutions (1.12) leads to the global extremum Λ. Two functionals neighbouring by level are called

locally consistent, if their variation in a limited region of G space of variables of the lower level subsystems is similar to:

$$\left.\begin{aligned}
&\delta\Lambda(u+\delta u) = k_1\delta I(u+\delta u);\\
&\delta I(\beta+\delta\beta) = k_2\delta J(\beta+\delta\beta);\\
&\delta J(\beta_f+\delta\beta_f) = k_3\delta J_f(\beta_f+\delta\beta_f);\\
&\delta J_f(\beta_{nn}+\delta\beta_{nn}) = k_4\delta J_{nn}(\beta_{nn}+\delta\beta_{nn}).
\end{aligned}\right\} \qquad (1.13)$$

If the intersection of regions G_i variables' existence of the levels i represents a non-empty set that includes a point determining the global extremum for Λ, that their multiple consistent optimization from J_{nn} to Λ and back from Λ to J_{nn} allows to find the global extremum.

However, it is not easy to achieve it. That is because the functionals can't be optimally built from the top down when optimal values of lower levels of upper functional are unknown. Thus, the functional I of the control u on the object Σ_∞ can't be constructed without knowledge of what mathematical model it will be described with, and what level and type of noise is in the estimates obtained according to the filtration subsystem and so on.

It is possible to apply the minimax approach and to build I in order to obtain the best guaranteed solution for the worst model (Σ,β) and estimates X_M of the signal. But there is an identification subsystem in the general system (fig. 1.1) that always gives solutions not worse than the worst as the result of optimization, so on using the minimax approach we will lose on the optimality.

The following algorithm of constructing locally consistent functionals is appropriate:

1. The elementary subsystems of the initial converters of the evaluation signal X, the model (Σ, β) and the optimal control u are accepted on the basis of the prior information on the object Σ_∞, the main functional Λ and existing subsystems' elements of all levels and taking into account the principle of sustainable complications at the first step of the multilevel RP.

2. The selected subsystems of i-level are optimized by brute force of structures and by the method of optimization using a priori adopted criterion of subsystems of the $(i+1)$ level: Σ_{nn} by J_f, $(\Sigma_f, OptJ_f)$ by J, (Σ, Opt) by I, $(OptI)$ by Λ. Since the dimensions of these variables are small, then costs for searching are valid. As a result of this optimization, we obtain the optimal system of the first approximation.

3. The consistent criteria I, J, J_f, J_{nn} are designed by the methods of the sensitivity theory or simulation and experimental design [18].

At first let us consider the approach based on the analytical calculation of the functional's sensitivity. Assume that there is an optimal system of the first approximation, in which the real object Σ_∞ is replaced by its model (Σ, β). Let us find the variation of the basic functional Λ relating to its extreme value $\Lambda^*(u^*)$, which is caused by the variation δu of the optimal control on an object. Assume that

$$u(t) = u^*(t) + \varepsilon\delta u(t), \tag{1.14}$$

where ε -is a small quantity; $u(t) = \begin{bmatrix} u_1(t) \\ \cdot \\ \cdot \\ \cdot \\ u_m(t) \end{bmatrix}$ - is a vector-function of time *t*.

Assume that

$$\Lambda(u(t)) = \int_{\theta_1} \lambda(u^*(t) + \delta u(t)\varepsilon)dt, \tag{1.15}$$

where $\int_{\theta_1}$ - is an integration operator on the time t interval θ_1; $\lambda(\cdot)$ - is a scalar function of a vector argument u.

Let us find the first variation

$$\delta\Lambda = \frac{\partial\lambda}{\partial\varepsilon} = \int_{\theta_1}\frac{\partial\lambda}{\partial\varepsilon}dt = \int_{\theta_1}\left(\frac{\partial\lambda}{\partial u}\right)^T\frac{\partial u}{\partial\varepsilon}dt = \int_{\theta_1}\left(\frac{\partial\lambda}{\partial u}\right)^T\delta u dt,$$

where

$$\frac{\partial\lambda}{\partial u} = \begin{bmatrix}\frac{\partial\lambda}{\partial u_1}\\ \dots \\ \frac{\partial\lambda}{\partial u_m}\end{bmatrix}, \quad \frac{\partial u}{\partial\varepsilon} = \begin{bmatrix}\frac{\partial u_1}{\partial\varepsilon}\\ \dots \\ \frac{\partial u_m}{\partial\varepsilon}\end{bmatrix}.$$

Now let us take the second variation:

$$\delta^2\Lambda = \frac{\partial^2\Lambda}{\partial\varepsilon^2} = \frac{\partial}{\partial\varepsilon}\frac{\partial\Lambda}{\partial\varepsilon} = \frac{\partial}{\partial\varepsilon}\int_{\theta_1}\left(\frac{\partial\lambda}{\partial u}\right)^T\delta u dt =$$

$$= \int_{\theta_1}\frac{\partial}{\partial\varepsilon}\left(\frac{\partial\lambda}{\partial u}\right)^T\cdot\delta u\cdot dt = \int_{\theta_1}\frac{\partial}{\partial u}\left[\left(\frac{\partial\lambda}{\partial u}\right)\delta u\right]^T\frac{\partial u}{\partial\varepsilon}dt, \qquad (1.16)$$

where

$$\frac{\partial^2\lambda}{\partial u^T\partial u} = \begin{bmatrix}\frac{\partial^2\lambda}{\partial u_1\partial u_1} & \dots & \frac{\partial^2\lambda}{\partial u_1\partial u_m}\\ \dots & \dots & \dots \\ \frac{\partial^2\lambda}{\partial u_m\partial u_1} & \dots & \frac{\partial^2\lambda}{\partial u_m\partial u_m}\end{bmatrix}.$$

If Λ is determined on the $x(t)$, but not on the $u(t)$, so acting by the same procedure, we get

$$\delta^2\Lambda = \int_{\theta_1}\delta x^T\frac{\partial^2\lambda}{\partial x^T\partial x}\delta x dt,$$

or

$$\delta^2\Lambda = \int_{\theta_1} \partial u^T \left(\frac{\partial x}{\partial u}\right)^T \frac{\partial^2 \lambda}{\partial x^T \partial x} \frac{\partial x}{\partial u} \delta u dt, \qquad (1.17)$$

where

$$\delta x = \frac{\partial x}{\partial u} \delta u,$$

$$\delta x = \begin{bmatrix} \delta x_1(t) \\ \cdots \\ \delta x_n(t) \end{bmatrix}, \quad \delta u = \begin{bmatrix} \delta u_1(t) \\ \cdots \\ \delta u_m(t) \end{bmatrix}, \quad \frac{\partial x}{\partial u_1} = \begin{bmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} \cdot\cdot & \cdot \frac{\partial x_1}{\partial u_m} \\ \cdots & \cdots & \cdots \\ \frac{\partial x_n}{\partial u_1} & \frac{\partial x_n}{\partial u_2} \cdot\cdot & \cdot \frac{\partial x_n}{\partial u_m} \end{bmatrix}.$$

Comparing (1.16) and (1.17), we see that

$$\frac{\partial^2 \lambda}{\partial u^T \partial u} = \left(\frac{\partial x}{\partial u}\right)^T \frac{\partial^2 \lambda}{\partial x^T \partial x} \frac{\partial x}{\partial u}. \qquad (1.18)$$

The first variation is equal to zero on the extremal $x^*(t), u^*(t)$, the second variation (1.16) or (1.17) determines the etalon surface (for the functional $I(u)$) in the functional variations space $\delta u(t)$. Assume $\delta u(t) = 1(t)$ - is an ordinary vector-function from t, then (1.17) is an etalon matrix A_1 of the real numbers for I

$$\delta^2\Lambda = A_1 = \int_{\theta_1} \left(\frac{\partial x}{\partial u}\right)^T \frac{\partial^2 \lambda}{\partial x^T \partial x} \frac{\partial x}{\partial u} dt. \qquad (1.19)$$

When constructing the functional I we will seek the coincidence of the arguments of the extremums Λ and I, and the second variations should be similar.

Example 1.2. Assume

$$I = \int_{\theta_2} (x^T G x + u^T Q u) dt, \qquad (1.20)$$

where G, Q – are weight matrices; θ_2 - is a control interval.

Then

$$\frac{\partial I}{\partial \varepsilon} = \int_{\theta_2} \left[\delta u^T \left(\frac{\partial x}{\partial u} \right)^T G x + x^T G \frac{\partial x}{\partial u} \delta u + \delta u^T Q u + u^T Q \delta u \right] dt$$

$$\frac{1}{2} \frac{\partial^2 I}{\partial \varepsilon^2} = \frac{\partial}{\partial \varepsilon} \frac{\partial I}{\partial \varepsilon} = \int_{\theta_2} \left[\delta u^T \left(\frac{\partial x}{\partial u} \right)^T G \frac{\partial x}{\partial u} \delta u + \delta u^T Q \delta u \right] du = \qquad (1.21)$$

$$\int_{\theta_2} \delta u^T \left[\left(\frac{\partial x}{\partial u} \right)^T G \frac{\partial x}{\partial u} + Q \right] \delta u dt.$$

The functional (1.20) will be consistent with the main functional for those ordinary variations δu that are in (1.19), if the variations (1.19) and (1.21) are similar:

$$\int_{\theta_2} \left[\left(\frac{\partial x}{\partial u} \right)^T G \frac{\partial x}{\partial u} + Q \right] dt = k_1 A_1, \qquad (1.22)$$

or in a scalar mapping

$$\sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} \int_{\theta_2} \frac{\partial x_i}{\partial u_j} \frac{\partial x_j}{\partial u_k} dt + \int_{\theta_2} q_{lk} dt = k_1 \alpha_{lk}^1, \quad l,k = \overline{1,m}; \qquad (1.23)$$

where k_1 - is a similarity coefficient; α_{lk}^1 - is a matrix A element in (1.19); g_{ij}, q_{lk} - are matrix elements G, Q respectively.

Since A_1 is symmetric, then the total amount of equations (1.23) is $0{,}5m(m+1)$, and the total amount of the sought coefficients g_{ij}, q_{lk}, is $0{,}5n(n+1)$ and $0{,}5m(m+1)$ respectively. This allows them to satisfy the $m(m+1)$ requirements of consistency I to Λ and a series of other requirements which are determined by the optimization system I. The tasks of matrices G, Q and the task [4] become much easier while using this approach to construct the functional I.

The next step is to design of the criterion J of the identification subsystem. Assume that the functional I (1.20) which is consistent with Λ, was obtained at the previous step with weight G^* and Q^* that satisfy the condition (1.23). Let us find the second variation

I at parameters β space of the model (Σ,β) and define it as the standard for the functional J: $\frac{1}{2}\frac{\partial^2 I}{\partial\varepsilon^2}=\int\limits_{\theta_2}\delta u^T\left[\left(\frac{\partial x}{\partial u}\right)^T G^*\frac{\partial x}{\partial u}\delta u+Q^*\right]\delta u dt,$

where

$$\delta u=\frac{\partial u}{\partial\beta}\delta\beta,\quad \frac{\partial u}{\partial\beta}=\begin{bmatrix}\frac{\partial u_1}{\partial\beta_1} & \frac{\partial u_{r1}}{\partial\beta_2} & \cdots & \frac{\partial u_1}{\partial\beta_r}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial u_n}{\partial\beta_1} & \frac{\partial u_n}{\partial\beta_2} & \cdots & \frac{\partial u_n}{\partial\beta_r}\end{bmatrix},\ \beta=\begin{bmatrix}\beta_1\\ \cdots\\ \beta_r\end{bmatrix}.$$

Then

$$\delta^2 I(\beta)=\int\limits_{\theta_2}\delta\beta^T\left(\frac{\partial u}{\partial\beta}\right)^T\left[\left(\frac{\partial x}{\partial u}\right)^T G^*\frac{\partial x}{\partial u}+Q^*\right]\frac{\partial u}{\partial\beta}\delta\beta dt=$$

$$=\delta\beta^T\left[\int\limits_{\theta_2}\left(\frac{\partial u}{\partial\beta}\right)^T\left[\left(\frac{\partial x}{\partial u}\right)^T\times G^*\frac{\partial x}{\partial u}+Q^*\right]\frac{\partial u}{\partial\beta}dt\right]\delta\beta=\delta\beta^T A_2\delta\beta.$$

(1.24)

At the functional

$$J=\int\limits_{\theta_3} i(\hat{x},x_M,\alpha)dt \qquad (1.25)$$

at the point $\hat{x}=x_M$, corresponding to the estimate $\hat{\beta}$ that is equal to β, the first variation J should be zero, and the second one

$$\delta^2 J(\beta)=\delta\beta^T\left[\int\limits_{\theta_3}\left(\frac{\partial x_M}{\partial\beta}\right)^T\frac{\partial^2 i(\alpha)}{\partial x_M^T\partial x_M}\frac{\partial x_M}{\partial\beta}dt\right]\delta\beta$$

is similar to (1.24), viz

$$A_3=\int\limits_{\theta_3}\left(\frac{\partial x_M}{\partial\beta}\right)\frac{\partial^2 i(\alpha)}{\partial x_M^T\partial x_M}\frac{\partial x_M}{\partial\beta}dt=k_2 A_2, \qquad (1.26)$$

where A_2 has $0{,}5r(r+1)$ of various elements α_{ij}^2.

The functional (1.25) should contain at least $0{,}5r(r+1)$ of controlled parameters α in order to satisfy (1.26) the functional (1.25) absolutely.

Let the model (Σ, β) be presented in the form of

$$x_M(t) = \sum_{i=1}^{n} \beta_i W_i(p) u(t), \qquad (1.27)$$

where $W_i(p)$ - are assigned operators from $p = \frac{d}{dt}$.

Assign that is in (1.24)

$$A_2 = \mathrm{diag}\left[\alpha_{ij}^2\right].$$

From the condition $k_2\alpha_{ij}^2 = \alpha_{ij}^3$, let us find $u(t)$ and present it by the set of orthogonal functions $\varphi_k(t)$

$$u(t) = \sum_{k=1}^{m} \alpha_k \varphi_k(t), \qquad (1.28)$$

where m – is the amount of components $\varphi_k(t)$, that provide the adequate representation $u(t)$ on the identification interval $\theta_3 = [0 \quad T]$; $\varphi_k(t), k = \overline{1,n}$ - is the system of orthogonal functions, for example cosine series

$$\varphi_k(t) = \cos\left(\frac{2k-1}{2}\right)\frac{\pi}{T}t. \qquad (1.29)$$

If $J = \int_0^T \varepsilon^2(t)dt$, where $\varepsilon(t) = \hat{x}(t) - x_M(t)$, then in (1.26) there are elements

$$\alpha_{ij}^2 = \frac{\alpha_{ij}^3}{k_2} = \int_0^T \left[W_i(p)u(t)\right]\left[W_j(p)u(t)\right]dt.$$

Let us denote

$$W_i(p)u(t) = y_i(t), \quad i = \overline{1,n},$$

taking (1.28), (1.29) into account

$$y_i(t) = \sum_{k=1}^{m} \alpha_k \left[W_i(p) \varphi_k(t) \right] = \sum_{k=1}^{m} \alpha_k y_{ij}(t);$$

$$y_{ij}(t) = W_i(p) \varphi_k(t).$$

Then

$$\frac{\alpha_{ij}^3}{k_2} = \int_0^T \left(\sum_{k=1}^{m} \alpha_k y_{ik}(t) \right) \left(\sum_{k=1}^{m} \alpha_k y_{jk}(t) \right) dt =$$

$$= \sum_{k=1}^{m} \sum_{q=1}^{m} \alpha_k \alpha_q \int_0^T y_{ik}(t) y_{jq}(t) dt = \sum_{k=1}^{m} \sum_{q=1}^{m} \alpha_k \alpha_q y_{ik},$$

where $y_{ik} = \int_0^T y_{ik}(t) y_{jq}(t) dt$ are calculated previously.

The optimal value α will be found from the condition

$$\alpha^* = \arg\min_{\alpha \in R} \left\| \frac{\alpha_{ij}^3}{k_2} - \sum_{k=1}^{m} \sum \alpha_k \alpha_q y_{ij}^{kq} \right\|, \quad i, j = \overline{1,n}, \qquad (1.30)$$

where all variables, except α, are calculated previously. If the dimension is equal or more than $0{,}5r(r+1)$, then the least value of the norm in (1.30) is zero and the functionals' variations J and I are similar.

In a similar way to the previous ones, having obtained J and having defined its variation from the vector β_f of filter's parameters, let us select the quality index J_f of the filter (Σ_f, β_f) from the similarity condition (1.24), and the quality index J_{nn} of primary converters is similar to J_f, obtained at the previous step.

Taking into account the complexity of the analytical solution of the functional' concordance, the approach of planning the experiment will be constructive.

The creation of computer aided design (CAD) of hierarchical systems' criteria with appropriate mathematical software facilitates the issue. The computer consistently issued subsystems of the two neighbouring levels, starting from the top; the experiment aimed to

determine only the second functionals' variations relating to their extreme value is performed; the additional functional [type (1.30)] is formed, which determines the proximity (similarity) of the second variations of given (upper level) and constructed (lower) functionals; the additional functional is optimized by adjusting options of the lower level subsystem. In order to reduce the options it is appropriate to apply the optimal plans of the experiment [14,18].

It is necessary and sufficient to perform a "star experiment" in order to determine the second variation only if extreme value is known and the experiment is performed to study the region of the functional's extremum. The "star experiment" is a successive deviation on the "star shoulder" magnitude $\pm\alpha$ each of the variables for which the second variation of the functional is found. If the coordinates of functional's extremum are unknown, the optimal experiment performed to design the type of quadratic model

$$y(x) = \alpha_0 + \sum_{i=1}^{n} \alpha_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \alpha_{ij} x_i x_j + \sum_{i=1}^{n} \alpha_{ij} x_i^2 . \qquad (1.31)$$

(1.31) is Box's central composite plan [18]. It is possible to perform this plan as orthogonal. The normalized variable models (1.31) take five values $0; +1; -1; +\alpha; -\alpha$.

The plan of the experiment is as follows:

1) the complete factor experiment 2^n or fractional replicas if $n >> 5$ [18];

2)"star plan"

3) point in the plan's center (N experiments only).

In order to design the orthogonal plan the model (1.31) is transformed to the type

$$y(x) = b_0 + \sum_{i=1}^{n} \alpha_i x_i + \sum_{i=1}^{n-1} \alpha_{ij} x_i x_j + \sum_{i=1}^{n} \alpha_{ij} \tilde{x}_i^2 , \qquad (1.32)$$

where

$$b_0 = \alpha_0 + \sum_{i=1}^{n} \alpha_{ij} x_i^2 , \quad \tilde{x}_i^2 = x_i^2 - c, c = \frac{1}{N} \sum_{k=1}^{N} x_i^2(k),$$

k –is the number of the experiment.

The new variables $\tilde{x}_i^2$ will be displaced on the value –c. This allows to select α so that all the columns of the experiment matrix become of pairwise orthogonality.

For example, for $n = 3$ and the model (1.31) the matrix of the Box's plan is depicted in the table 1.2, where $c = (1/15)2\sum_{k=1}^{15} x_i^2(k) = (8 + 2\alpha^2)/15$ for any i.

The "star shoulder" α is selected according to the table 1.2, from the orthogonal condition $\tilde{x}_i^2$:

$$\sum_{k=1}^{N} \tilde{x}_i^2(k)\tilde{x}_j^2(k) = 8(1-c)^2 - 4c(\alpha^2 - c) + 3c^2 = 0.$$

Substituting the expression for c, we get

$$\alpha = \sqrt{30} - 4 \approx 1{,}5,\ c \approx 0{,}83.$$

The model's (1.36) coefficients $b_0, \alpha_i, \alpha_{ij}, \alpha_{jj}$ are determined by the method of least squares independently because of mutual orthogonality.

From these coefficients it is easy to access the models' coefficients (1.31) and from there – to the models with nonnormalized variables. For this α_j is divided by the modulus of variation's step of the variable i, and α_{ij} is divided by the product of these moduli and variables i and j, $i = \overline{1,n}, j = \overline{1,n}, i = j$.

If there are functionals' models (1.31) of the upper and lower levels, the lower level is designed on upper by variations of structure and parameters of the lower subsystem from the condition of the minimum of the additional functional of the deference between coefficients $\alpha_i, \alpha_{ij}, \alpha_{ii}$ of models (1.31) of two levels.

The number of different options can be cut on using the experiment planning in order to construct the dependence model of this additional functional from structures and parameters of lower level's subsystems. Having made subsystems' criteria of all levels consistent and optimized, we obtain the optimal system of the second approximation. Having repeated this procedure of criteria concordance and having optimized them, we get the third approximation. The concordance process stops when the results of further iterations differ little.

Table 1.2

Plan's type	Normalized variables $x(k)$										
	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$\tilde{x}_1^2$	$\tilde{x}_2^2$	$\tilde{x}_3^2$	$y(k)$
1	1	-1	-1	-1	1	1	1	1-c	1-c	1-c	$y(1)$
	1	1	-1	-1	-1	-1	1	1-c	1-c	1-c	$y(2)$
	1	-1	1	-1	-1	1	-1	1-c	1-c	1-c	$y(3)$
	1	1	1	-1	1	-1	-1	1-c	1-c	1-c	$y(4)$
	1	-1	-1	1	1	-1	-1	1-c	1-c	1-c	$y(5)$
	1	1	-1	1	-1	1	-1	1-c	1-c	1-c	$y(6)$
	1	-1	1	1	-1	-1	1	1-c	1-c	1-c	$y(7)$
	1	1	1	1	1	1	1	1-c	1-c	1-c	$y(8)$
2	1	$-\alpha$	0	0	0	0	0	$\alpha^2 - c$	$-c$	$-c$	$y(9)$
	1	α	0	0	0	0	0	$\alpha^2 - c$	$-c$	$-c$	$y(10)$
	1	0	$-\alpha$	0	0	0	0	$-c$	$\alpha^2 - c$	$-c$	$y(11)$
	1	0	α	0	0	0	0	$-c$	$\alpha^2 - c$	$-c$	$y(12)$
	1	0	0	$-\alpha$	0	0	0	$-c$	$-c$	$\alpha^2 - c$	$y(13)$
	1	0	0	α	0	0	0	$-c$	$-c$	$\alpha^2 - c$	$y(14)$
	1	0	0	0	0	0	0	$-c$	$-c$	$-c$	$y(15)$

Ensuring criteria' consistency in theory or by simulative computer modeling brings the conditional optimization task into proximity with the task of finding the best solution (the determination of the global extremum of the main criterion Λ).

1.4. The Composition of the Subsystems of Different Levels into the Single System

The third and final stage of designing the structure and algorithms of functioning of multilevel optimization system close to the optimal at Λ is the stage of composition of individual subsystems into the one complex.

The formulated definitions of a system, relaxation processes, and the principle of the rational complication [43] allow determining the main approaches to the composition:

- sorting the elements of the models sets and optimization methods;
- a choice of the first approximation optimal system;
- an organization of the relaxation processes of complication and improvement of certain elements of systems, and perhaps the structure of the system itself, beginning with the optimal system of the first approximation and ending up with the optimal to Λ.

The sorting of the subsystems' elements is implemented by projectors PJ, narrowing at the first step the initial set of elements to acceptable subsets, and then to the candidates' subset for the optimal elements.

For example, the nonlinear identification methods that don't require the complete observability, are acceptable at the incomplete observability of the object Σ_∞. At the first level of projectors PJ1 select those elements (methods and models) that can be applied in this situation, which is determined by the signals $\hat{u}, \hat{x}$ under observation and by given criteria J_f, J, I, Λ.

The projectors PJ are multidimensional discriminators that allow outputting those elements which properties are fully consistent with the factors, determining the actual situation. At the second level projectors PJ2 select two or three candidates for optimality. For this in the projector PJ2 there are regressive dependencies of the relevant criteria $\hat{J}_f, ..., \hat{\Lambda}$ (if criteria can't be realized a priori) for each element, criteria taken from the quantitatively calculated vector of factors θ, characterizing the specific situation.

The factors for the elements-methods of the linear evaluation are the correlation of the signal - noise on the amplitude and width of the spectrums, signals correlation, the model's approximation (the value of the final approximation's error) and so on. Having determining several factors that are the most distinctive for the practice of regression analysis, we construct the regression dependency of the appropriate criterion $J_f, ..., \Lambda$ or its estimate $\hat{J}_f, ..., \hat{\Lambda}$ from θ.

Then projector PJ1 determines the region, and the PJ2 determines the criterion value for concrete θ for each element of this region. The task of projectors constructing refers to the phase of the design and it requires as much costs as accurate and wider are the regions and dimensions of the factors θ.

The choice of the optimal system of the first approximation was described above. The basic process at this stage is the organization of the system's structure.

Let us consider the version of the optimization system structure J – the identification's system. In the table 1.1 identification's subsystems are structurally systemized from the simplest to self-organizing structural ones: we get the structure of the MAIS as acceptable in order to create the identification subsystem in the multilevel optimization system Λ (if Σ^* and Opt^* are prior independent).

For this identification system, contrary to the simple ones, the closedness at the basic criterion I and the nonidentity of the element Σ are typical (as for mais). Let us make up the functional scheme, determining its interconnection with the subsystem of upper and lower levels (fig. 1.3). The peculiar parts of the system (subsystem) are encircled by a dotted line and numbered: 1 – is the optimization Λ, 2 – is the optimization I, 3 – is the optimization J that adds the simplest identification's system (part 4) to the multiadaptive one (parts 3 and 4), 5 – is the optimization J_f.

The part 4 of the system contains the functional J_{qk}, which is determined above, the model Σ_{qk} with the vector β_{qkn} of adjusted parameters, the optimizer (Opt_{qk}, α_{qk}). Here the RP $\{\beta_{qkn}\}$ is realized in terms of J_{qk}, using the optimizer (Opt_{qk}, α_{qk}). The

information on the structures Σ_{qk} optimal on J_{qk} and vectors β_{qk} of the model's parameters enters the subsystems of the upper levels.

The part 3 of the system contains three sets $\left\{\Sigma_{qk_1}\right\}$, $\left\{Opt_{qk_1}\right\}$, $\left\{I_{qk_1}\right\}$; the first level projectors PJ1 that restrict these sets to subsets $\left\{\Sigma_{qk}\right\}$, $\left\{Opt_{qk}\right\}$, $\left\{I_{qk}\right\}$ of feasible elements; the second level projectors PJ2 that choose among acceptable candidates on the optimal elements.

For a specific type I_q and required value I_{qk} that is determined above, the projectors PJ1, PJ2 define the elements J_{qk}, Σ_{qk}, Opt_{qk}, α_{qk} of the part 4, where RP $\left\{\beta_{qkn}\right\}$ get the model (Σ_{qk},β_{qk}) optimal on J_{qk} as a result. In the subsystem of the part 2 the criterion value I_{qk} is estimated on the real object Σ_∞ or its model (Σ_{qk},β_{qk}). If it does not satisfy the required one, the candidate in one of the projectors PJ2 is substituted. The process $\left\{\beta_{qkn}\right\}$ repeats again for as long as I_{qk} is no worse than required, or until the process of sorting all candidates is completed.

The structure and parameters of the lower level subsystem can be changed while changing J_{qk} in the parts 4 and 5. In the parts 3 and 4 there is the composition of two RPs $\left\{\beta_{qkn}\right\}$ on n and $\left\{J_{qk}\right\}$ on k that corresponds to the definition MAIS.

In the part 2 there is a real object Σ_∞, its optimizer $\underset{u}{Opt}\, I_q$, the set $\left\{I_q\right\}$ and the projector PJ that determines (based on values I_{qk} and Λ) type I_q of the criterion I_{qk}. The signal x from the object's Σ_∞ output enters the lower level subsystem. The estimate $\hat{x}$ gets to the system of identification and control from the lower level subsystem's output.

The optimizer $\underset{u}{Opt}\, I_q$ based on the information $\Sigma_{qk},\beta_{qk},I_{qk},\hat{x}$, produces the control action u, entering on the object or its model (at the stage of searching the optimal

system's elements). If the optimal value I_q on the k does not satisfy the part 1 of the system after the completion of RP $\{\beta_{qk}\}$, then the projector PJ can receive the command from the upper level system to change on the q type of the functional I_q. The step on q leads to the repetition of the steps on k, next k on n in the MAIS, viz. there is the composition of three RPs in the parts 1-4. The composition of three RPs is a self-organizing system (SOS).

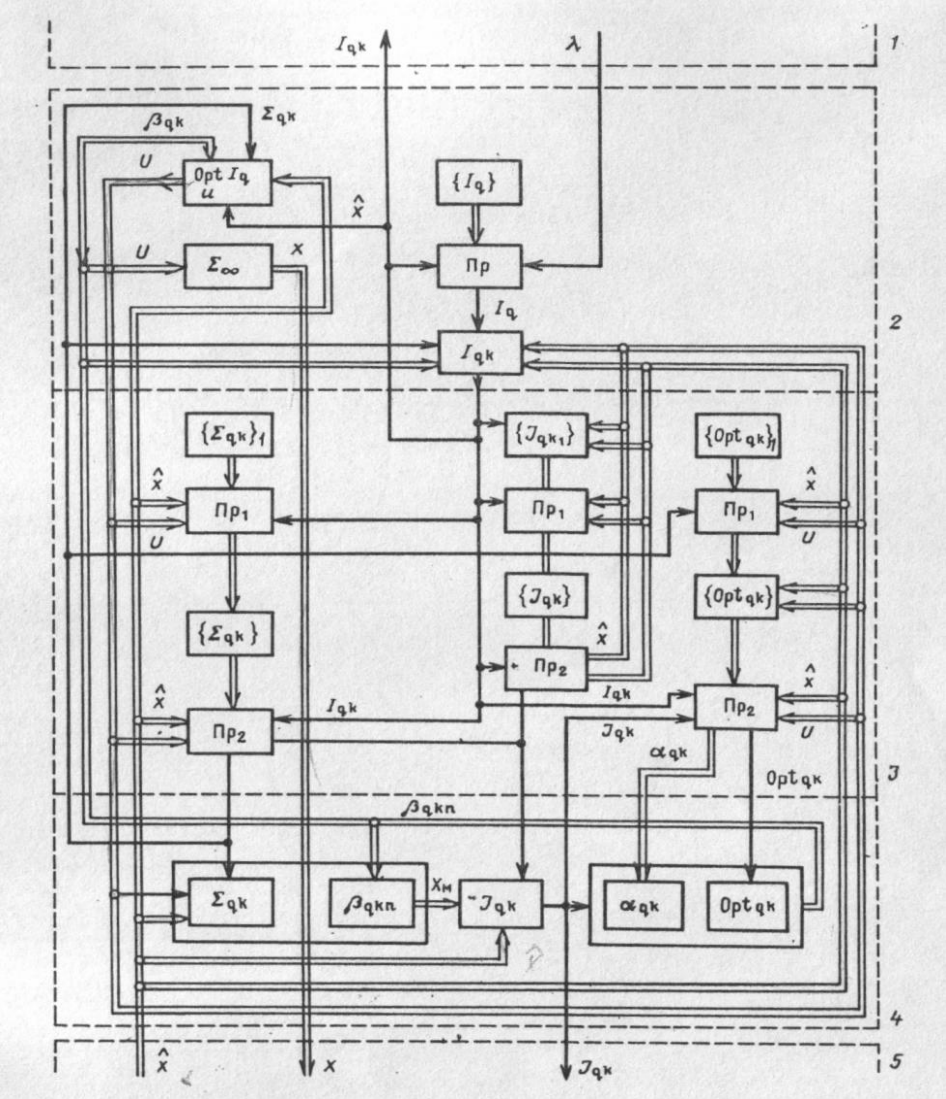


Fig. 1.3 The functional scheme of the multiple adaptive subsystem of identification

The strong feedback with the criteria from Λ to I, from I to J, from J to J_f, which optimizes and orients the structure and parameters of the all level to the goal, accomplishes the regularization in a broad sense.

The multiple adaptive systems of identification have more opportunities for regularization in comparison with conventional ones.

$$\inf I_q(\Sigma, Opt, \alpha, u)\Big|\beta = \arg\min J\langle \inf I_q(u)\Big|\begin{matrix}\Sigma = const\\ (Opt,\alpha) = const\\ \beta = \arg\min J,\end{matrix} \qquad (1.33)$$

The MIAS efficiency is as obvious as far as obvious the inequality

$$\inf I_q(\Sigma,Opt,\alpha,u)\Big|\beta = \arg\min J < inf\ I_q(u)\Big|\begin{matrix}\Sigma = const\\ (Opt,\alpha) = const\\ \beta = \arg\min J,\end{matrix} \qquad (1.33)$$

where the right side corresponds to minimization's systems with a conventional identifier (only the part 4 of the scheme in fig. 1.3), the left side corresponds to the systems with MAIS- identifier (the parts 2, 3, 4).

If I_q doesn't take losses for searching $\inf I_q$ into account (as in the example of the prognosis), then the inequality is strict. If I_q does take, that the inequality can't be perform for the elements Σ, Opt, α of large dimensionality and bad sorting by projectors PJ1, PJ2. The principle of rational complication is discolated in such systems. The complications (the extension of elements' dimensionality) of multilevel system with MAIS- identifier must be that that the inequality (1.33) is not weakened but intensified. Only this complication is appropriate.

Chapter 2. The Basic Model of the Real Processes and its Derivative Models

2.1. The Basic Mathematical Models of Real Processes

The processes that occur in time in real material objects are subjected to the fundamental laws of nature. As we said before in the Chapter 1:

– all processes and their elements as a particle of matter can't be completely autonomous, all is interconnected with everything;

– there is no absolute state of rest (statics) in the real objects, all objects are dynamic due to the infinity of the material world and the direct or indirect interrelation with its components;

– the presence of the response rate (for example, the mass in the mechanical objects) and real delimitation of the power of control actions on the object do not allow an instant change in time of any coordinate of the real world's object, viz. all variables' of the object's state are smooth functions of time;

– two or more identical objects do not exist in the nature, so an ensamble averaging of roughly identical objects does not allow to define accurately their characteristics or coefficients, because each object has its own ones;

– in the same way the nonstationarity of processes in the real objects makes impossible to use the time averaging (Chebyshev's theorem on the large numbers is accomplished only approximately in practice: suppose you toss a coin a hundred times, you will change coin's geometry by fingers' friction and, as a result, the instantaneous average value of the correlation "heads-tails");

– the total interralation of the objects and their natural infinite-dimensionality does not allow to construct an accurate finite-dimensional model of any process; that is all models are approximate, the task of the researcher is to select the best one for specific application (prognosis, control, internal parameters' control, etc).

In accordance with fundamental Kalman's theorem [7], taking the smoothness of real processes into account, it can be said:

"For smooth systems, the transition function of which is a mapping $\mathrm{T}\times\mathrm{T}\times\mathrm{X}\times\Omega\to\mathrm{X},$ it is also a differential equations solution at the same time

$$\dot{x}=f(x,u,t), \qquad (2.1)$$

where $\dot{x}$ - is a derivative from x at the time t, T – is an ordered set of time t moments, Ω - is the set of input infuences, X – is the set of states, $x\in X,\ u\in\Omega,\ t\in T$ ".

Thus, any process that occurs in time t_0, can be represented by the model (2.1) exactly (if the dimension x,u is directed to the infinity) and approximately if their dimension is finite in the real world.

Let us add, that the process can be presented also by the system's (2.1) solution $x(t)$ - this is the direct modeling task, or we have the inverse task – the task of the identification's mapping f with samplings $x(t)$ $u(t)$, if there are the experimentally obtained dependences $x(t),u(t)$ and the function f should be identified (parametrically or non-parametrically). We construct the dependence $x(t)$ as the function t or the values $x(t)$ that are previous time (as a rule, these are polynomial and autoregressive models), not searching for the interconnection of the component $x_i(t),\ x_j(t),\ i,j=\overline{1,n},\quad x_i(t)$ with $u_k(t),\ k=\overline{1,m}$.

It is clear that for the finite deviations' values ΔX, ΔU, Δt and the acceptable approximation's error ε, the nonlinear non stationary model (2.1) can be represented as a linear stationary system:

$$\dot{x}(t)=A_1x(t)+B_1u(t), \qquad (2.2)$$

where $x,\ \dot{x},\ u$ – are deviations from the basic regime, $x(t)=[x_1(t),\ldots,x_n(t)]^{\mathrm{T}}$, $u(t)=[u_1(t),\ldots,u_m(t)]^{\mathrm{T}}$, $A_1=[a_{ij}]_{j=1,n}^{i=1,n}$, $B_1=[b_{ik}]_{k=1,m}^{i=1,n}$.

For the inverse task the measurements x and u are perturbated by the noise $N(t)$, which is considered as a random process if the dimensions X, U and the time's t interval T_1 are limited. The presence $N(t)$ causes the difficulty of the inverse task: the simpler

model (smaller n and m) is, the greater the error ε of its approximation of the real process will be; but if the model (2.2) is more complex (and potentially more accurate), the dimensionality $(n+m)$ of the coefficients a_{ij}, b_{ij} in the inverse task is larger, the conditionality of the information matrix [1] is worse, and, as a result, the solution of the inverse task is less correct: the small perturbations $N(t)$ in the measurements $x(t)$ and $u(t)$ lead to large errors $\delta A,\ \delta B$ in coefficients of matrices A and B of the model (2.2).

Then we use the property of *smoothness* for the more accurate solution of the direct (simulation) and inverse (identification) tasks. In order to avoid the mentioned incorrectness of the inverse problem's solution, we use the principles of *decomposition* and correct *composition* in constructing mathematical models of the real processes. For this purpose we will divide the infinite large number of variables $x,\ u$ interconnected by the hypothetical model (2.1) into:

1. $x_1,\ u_1$ - are essential for solving the main task for which the model is constructed (their dimensionality is small);

2. the other variables $x_2,\ u_2$, which are divided into variables $x_2^{\mathrm{AF}},\ u_2^{\mathrm{AF}}$, that have the spectrum band close to the band $x_1,\ u_1$ (AF –is an average frequency), and the amplitude insignificant towards to $x_1,\ u_1$ (power, variance) and the variables, the frequency spectrum of which lies on the axis of frequencies above ($x_1^{\mathrm{HF}},\ u_1^{\mathrm{HF}}$) or below ($x_1^{\mathrm{LF}},\ u_1^{\mathrm{LF}}$) of the average frequency band of the significant variables $x_1,\ u_1$.

Then for the simplified model (2.2) the unmeasured variables $x_2^{\mathrm{AF}},\ u_2^{\mathrm{AF}}$ create a systematic error $\varepsilon_c(t)$ while determining coefficients of the matrices A_1 and B_1, so far as the variables $x_2^{\mathrm{AF}},\ u_2^{\mathrm{AF}}$ can be represented in the basis $x_2,\ u_2$ and the orthogonal additive $x_2^0(t),\ u_2^0(t)$: $x_2^{\mathrm{AF}} = k_1 x_1(t) + k_2 x_1^0(t),\ u_2^{\mathrm{AF}} = k_3 u_2(t) + k_4 u_2^0(t).$

This yields that instead of sought matrices $A_1,\ B_1$ at the identification we get their estimates: $\hat{A}_1 = A_1 + k_1,\ \hat{B}_1 = B_1 + k_3$ shifted on $k_1,\ k_3$. But it does not interfere with

prognosis and modeling tasks. Instead, the estimates A_1, B_1 take into account the impact of variables non-included in the model that correlated with included ones.

The unaccounted low-frequency components x_2^{LF}, u_2^{LF}, which are in the nonlinear model (2.1) slowly transfer the mapping point $\{x,u\}$ in the space $\{X \times U\}$, the point of decomposition (2.1) into the series (2.2), from one value to another, slowly changing the value of matrices A_1, B_1 at time as derivatives from the basic nonlinearity f of the model (2.1): $[a_{1ij}] = \left[\frac{\partial f_i}{\partial x_j}\right]_{i,j=\overline{1,n}}$, $[b_{1ie}] = \left[\frac{\partial f_i}{\partial u_e}\right]_{i,e=\overline{1,m}}$.

They create the error $\varepsilon_i(t)$ from the nonstationarity of the coefficients of matrices A_1, B_1 which in a limited period T_k of the time t are taken as stationary: $\hat{A}_1 \neq A_1(t)$; $\hat{B}_1 \neq B_1(t)$.

However, we can correctly receive the nonstationary model (2.2) based on the composition of partial models with constant matrices $\hat{A}_{1k}$, $\hat{B}_{1k}$ while defining $\hat{A}_{1k}$, $\hat{B}_{1k}$ for certain intervals T_k limited in time and further approximation $\hat{A}_{1k}$, $\hat{B}_{1k}$ to the whole interval T that consists of subintervals T_k.

Finally, the impact of high frequency components x_2^{HF}, u_2^{HF} can be significantly reduced both in the direct and in the inverse tasks (modeling tasks and tasks of identification respectively) by the averaging (filtering) allowable for spectral band of the basic variables x_1, u_1.

Let us combine the variables $x^{\mathrm{LF}}(t)$, $u^{\mathrm{LF}}(t)$, $x^{\mathrm{AF}}(t)$, $u^{\mathrm{AF}}(t)$, $x^{\mathrm{HF}}(t)$, $u^{\mathrm{HF}}(t)$ that are unrecorded in (2.2) and included in the hypothetical exact equation (2.1) with unknown final value's coefficients into one variable-remainder $\varepsilon(t)$. Let us also denote the set (X, U) through z, the matrices (A, B) through C, $\dot{x}$ through y. Then, instead of the approximate equation (2.2), we have the hypothetical exact

$$y = z \cdot C + \varepsilon. \qquad (2.3)$$

Hence LSM-estimate $\hat{C}$ [35] of the matrix C

$$\hat{C} = (z^{\mathrm{T}} z)^{-1} z^{\mathrm{T}} (y - \varepsilon), \tag{2.4}$$

will consist of the sought exact value

$$C = (z^{\mathrm{T}} z)^{-1} z^{\mathrm{T}} y \tag{2.5}$$

and of the error

$$\Delta C = -(z^{\mathrm{T}} z)^{-1} z^{\mathrm{T}} \varepsilon, \tag{2.6}$$

where the scalar product $z^{\mathrm{T}}\varepsilon$ is non-zero for the nonorthogonal to z components (these are x^{AF} and u^{AF}). Lower-frequency components will be orthogonal to z, if at first we place z in the centre that is to take

$$z^0 = z - \overline{z}, \tag{2.7}$$

where $\overline{z}$ -is value z average for the interval T. Then $(z^0) \cdot (x^{\mathrm{НЧ}}, u^{\mathrm{НЧ}}) > \equiv 0$ as x^{LF} and u^{LF} can be considered as constant on the interval T. The high-frequency components are orthogonal to z and will affect only on the variance of the LSM estimate [2]:

$$\mathrm{cov}\left\{\hat{C}\right\} = (z^{\mathrm{T}} z)^{-1} z^{\mathrm{T}} Q z (z^{\mathrm{T}} z)^{-1}. \tag{2.8}$$

If we assume x^{HF}, u^{HF} as “white noise”, that covariance matrix is converted into the diagonal:

$$\mathrm{cov}\left\{\hat{C}\right\} = \sigma^2 (z^{\mathrm{T}} z)^{-1}, \tag{2.9}$$

where σ^2 – is a “white noise” variance.

Considering the total association of variables, it can be argued that with further increase of dimensionalities n and m and $(n+m)$ the functional space of variables $z(t)$ space for z, the condition of the linear independence of components $z(t)$ will become worse and, consequently, the norm of inverse matrix $(z^{\mathrm{T}} z)^{-1}$ will be increased. This leads to the increase of the covariance’s (2.9) estimates (2.5). However, the larger is the dimensionality $(n+m)$ of the model (2.3), the more variables are accounted, the less is the remainder ε and, as a consequence, σ^2 in (2.9). Thus, it can be said that in the series

of the models which are ordered by complexity (dimensionalities $(n+m)$) there is an optimal one on condition that the norm of the covariance matrix (2.8) or (2.9) is minimum [37].

2.2. The Discrete Time Mathematical Models that Correspond to the Basic Continuous Model Exactly or Approximately (2.2)

As so far natural is the continuity in time for real processes, so natural is time discreteness for real measurements of these processes. Thus, there is the issue: how to find out the continuity's dependence (2.2) using the discrete time measurements $x(k), u(k),\ k = 0,1,\ldots,M$ and knowing the step constant Δt or variable $\Delta_k t$ in time. A separate issue is to do it in a way that the perturbations $N(k)$ influence its solution as less as possible.

The one-valued transition from the continuous model (2.2) with the initial conditions $x(0) = x_0$ to discrete one

$$x(t_{k+1}) = Ax(t_k) + Bu(t_k),\ \ x(t_0) = x_0, \tag{2.10}$$

can be obtained through the system's (2.2) solution for discretes t_k of the time *t*:

$$x(t_k) = \Phi(t_k - t_0)x(t_0) + \int_0^{t_k} \Phi(t-\tau)B_1 u(\tau)d\tau, \tag{2.11}$$

where $\Phi(\cdot)$ - is the transition (2.2) matrix of the system, which is equal to[16]

$$\Phi(t) = e^{tA_1} = \sum_{k=0}^{\infty} A_1^k \frac{t^k}{k!}, \tag{2.12}$$

or through the Laplace transform $X(s) = (s \cdot I - A_1)^{-1}x(t_0) + (s \cdot I - A_1)^{-1}Bu(s)$, from which

$$\Phi(t - t_0) = Z^{-1}\left[(sI - A_1)^{-1}\right], \tag{2.13}$$

where *s* – is a Laplace variable, Z^{-1} - is an inverse Laplace transform.

If the inverse task is solved, the signal $u(\tau)$ in (2.4) will be represented as a sample of measurements $u(\tau_k)$, which (for the calculation $x(t_k)$ in (2.4)) should approximate by the analytical dependency. Then the solution (2.4) can be obtained analytically. As it will be shown below, in many cases of the inverse problem's solution, the step of time shift Δt of variables of the difference equation (2.10) is larger than the step $\Delta_1 t$ of the discreteness's measurements. Let us represent within the step Δt the experimental dependence $u(\tau)$ as a power-polinomial, for example of the third order:

$$u(\tau) = \alpha\tau^3 + \beta\tau^2 + \gamma\tau + \delta, \qquad (2.14)$$

where $\tau \in [0, \Delta t]$. If the dependence is simpler, then the part of coefficients will be zero.

According to the scalar recording of the least squares' method

$$R_k = \sum_{j=0}^{n} u(\tau_j) \cdot \tau_j^k, \quad k = 0,1,2,3;$$

$$\alpha = D^{-1} \sum_{k=0}^{3} A_{k3} \cdot R_k, \qquad \beta = D^{-1} \sum_{k=0}^{3} A_{k2} \cdot R_k,$$

$$\gamma = D^{-1} \sum_{k=0}^{3} A_{k1} \cdot R_k, \qquad \delta = D^{-1} \sum_{k=0}^{3} A_{k0} \cdot R_k,$$

$$D = det \begin{bmatrix} n+1 & \sum_{j=0}^{n}\tau_j & \sum_{j=0}^{n}\tau_j^2 & \sum_{j=0}^{n}\tau_j^3 \\ \sum_{j=0}^{n}\tau_j & \sum_{j=0}^{n}\tau_j^2 & \sum_{j=0}^{n}\tau_j^3 & \sum_{j=0}^{n}\tau_j^4 \\ \sum_{j=0}^{n}\tau_j^2 & \sum_{j=0}^{n}\tau_j^3 & \sum_{j=0}^{n}\tau_j^4 & \sum_{j=0}^{n}\tau_j^5 \\ \sum_{j=0}^{n}\tau_j^3 & \sum_{j=0}^{n}\tau_j^4 & \sum_{j=0}^{n}\tau_j^5 & \sum_{j=0}^{n}\tau_j^6 \end{bmatrix},$$

here $A_{ks}, \quad k,s = 0,1,2,3$ - are algebraic additions of the matrix $[D]$.

The matrix recording (2.11) of the vector process $x(t_k)$ is the sum of the scalar processes, which are (depending on the dimensionality of the matrix A_1) the solutions of

the typical differential equations of the corresponding order. Let us consider the series of these equations which are ordered by complexity and their analytical solutions. Taking up the operator $\frac{d}{dt}$ as p, we have

$$px_1(\tau) = b_0\hat{u}(\tau), \tag{2.15}$$

$$(p + a_0)x_2(\tau) = b_0\hat{u}(\tau), \tag{2.16}$$

$$(p + a_0)x_3(\tau) = (b_1 p + b_0)\hat{u}(\tau), \tag{2.17}$$

$$(p^2 + a_1 p + a_0)x_4(\tau) = b_0\hat{u}(\tau), \tag{2.18}$$

$$(p^2 + a_1 p + a_0)x_5(\tau) = (b_1 p + b_0)\hat{u}(\tau), \tag{2.19}$$

$$(p^2 + a_1 p + a_0)x_6(\tau) = (b_2 p^2 + b_1 p + b_0)\hat{u}(\tau), \tag{2.20}$$

where $\hat{u}(\tau)$ corresponds to (2.14). The analytic solutions of the equations (2.15)÷(2.20) are mentioned in the appendix A.

If the step Δt is constant, the matrix exponent e^{tA_1} in the equation (2.5) can be approximated by the linear component of the series $\sum_{k=0}^{\infty} A_1^k \frac{t^k}{k!}$ or the derivate $\dot{x}(t_k)$ can be substituted by the relative difference $\left[x(t_{k+1}) - x(t_k)\right] \cdot \Delta t^{-1}$, and the differential equation (2.2) becomes directly difference (2.10), where $A \cong I + \Delta t \cdot A_1, \quad B \cong I + \Delta t \cdot B_1$.

The discrete models which are the most frequently encountered in the contemporary tasks of modeling and prognosis will be discussed in the following sections:

- autoregressive (AR);
- autoregression with moving average (ARMA);
- autoregression with integrated moving average (ARIMA);
- autoregressive conditionally heteroscedastic model (ARCH);
- generalized (GARCH).

In addition to the models listed above, we obtained and used these combined models for prognosis the time series: the aperiodic trend approximated by power-polinomial (Taylor series) and the oscillatory components approximated by autoregression.

2.3. The Determining the Structure of the Nonlinear Operator f of the Model (2.1) by Composing the Local Models (2.2) into the Global one

Let us denote the variables (x,u,t) which are the argument of the vector function f in (2.1), through β, $\dot{x}_i$ through J. Then, for i-line of the system (2.1) $\dot{x}_i$ from x,u,t will be represented as $J(\beta)$. Next, as we know from Weierstrass theorem, every smooth dependency can be represented by Taylor series and, if the region β is limited, and series are endless, that this representation is exact:

$$J(\beta)=J(\beta_0)+\sum_{i=1}^{n}\left.\frac{\partial J}{\partial\beta_i}\right|_{\beta_0}\Delta\beta_i+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left.\frac{\partial^2 J}{\partial\beta_i\partial\beta_j}\right|_{\beta_0}\Delta\beta_i\Delta\beta_j+$$

$$+\frac{1}{6}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left.\frac{\partial^3 J}{\partial\beta_i\partial\beta_j\partial\beta_k}\right|_{\beta_0}\times\Delta\beta_i\Delta\beta_j\Delta\beta_k+\cdots \qquad (2.21)$$

where β_0 - is a global centre of decomposition, $\Delta\beta=\beta-\beta_0$.

Apart from the global center β_0 of the region G let us introduce (according to the capabilities of the experiment) a number of local centers β_{0r} $(r=\overline{1,m})$ of the limited subregions G_r. The dimension G is selected under condition that the dependency $J(\beta)$ with the permissible error δ is approximated by the linear part of decomposition (2.21), and the distance between centers of neighboring subregions is such that $J(\beta)$ is approximated by the linear-quadratic part series (2.21) with an accurate δ.

Analogously, the distance between the centers of the groups of neighboring subregions should satisfy the conditions of approximation by the first three terms of series

(2.21) and so on. Thus, there is a gradual increase of the dimension of the region and of the polynomial order (2.21) respectively. In this case the components of series for which the coefficients $\left.\frac{\partial^p J}{\partial\beta_i \ldots \partial\beta_n}\right|_{\beta_{0p}}$ are insignificant, are automatically discarded at the step p ($p = 1, 2, \ldots$).

This approach enables to determine the final set of coefficients α at each step using the linear model only

$$\Delta y = \alpha \cdot \Delta\beta, \tag{2.22}$$

where Δy determines for local subregions r

$$\Delta y = \Delta J, \quad \alpha \cong \left.\frac{\partial J}{\partial\beta}\right|_{\beta_{0r}}, \quad \Delta\beta = \beta - \beta_{0r}; \tag{2.23}$$

Δy is to increase the region by combining neighbouring subregions

$$\Delta y = \left.\frac{\partial J}{\partial\beta}\right|_{\beta_{0p}} - \left.\frac{\partial J}{\partial\beta}\right|_{\beta_{0r}}, \quad \alpha \cong \left.\frac{\partial^2 J}{\partial\beta^{\mathrm{T}}\partial\beta}\right|_{\beta_{0r}}, \quad \Delta\beta = \beta_{0p} - \beta_{0r}, \quad p \neq r; \tag{2.24}$$

and so on.

Genuinely, for local neighboring subregions r the first derivative from J on β into (2.21) is approximately equal to

$$\left.\frac{\partial J}{\partial\beta_j}\right|_{\beta_{0p}} \cong \left.\frac{\partial J}{\partial\beta_j}\right|_{\beta_{0r}} + \sum_{i=1}^{n} \left.\frac{\partial^2 J}{\partial\beta_i\partial\beta_j}\right|_{\beta_{0r}} (\beta_{i0p} - \beta_{i0r}), \tag{2.25}$$

this yields the dependence (2.24);

- for increasing neighboring regions

$$\left.\frac{\partial^2 J}{\partial\beta_i\partial\beta_j}\right|_{\beta_{0k}} \cong \left.\frac{\partial^2 J}{\partial\beta_i\partial\beta_j}\right|_{\beta_{0r}} + \sum_{k=1}^{n} \left.\frac{\partial^3 J}{\partial\beta_i\partial\beta_j\partial\beta_k}\right|_{\beta_{0r}} (\beta_{i0k} - \beta_{i0r}), \tag{2.26}$$

and so on.

If in certain direction l any derivative p is close to zero, the corresponding term of the series (2.21) is discarded while determining the coefficients α of the linear model

(2.22). That is, the sought structure of the polynomial (2.21) is automatically found without sorting all the options.

The coefficients α of the polynomial that is obtained in this way can be specified while minimizing of the appropriate norm of the proximity $J(\beta)$ and the polynomial $J(\alpha,\beta)$. Euclidean's mean square or Chebyshev's norm of the uniform approximation are commonly used. In order to increase the accuracy in the corresponding subregions the weighted mean square norm can be used.

A slightly different approach to the approximation $J(\beta)$ by analytical dependence arises from the following presentation of Taylor series, which is isomorphic to the expression (2.21):

$$J(\beta)-J(\beta_0)=$$

$$=\sum_{i=1}^{n}\left\{\left.\frac{\partial J}{\partial\beta_i}\right|_{\beta_0}+\left\{\frac{1}{2}\sum_{j=1}^{n}\left.\frac{\partial^2 J}{\partial\beta_i\partial\beta_j}\right|_{\beta_0}+\left\{\frac{1}{6}\sum_{k=1}^{n}\left.\frac{\partial^3 J}{\partial\beta_i\partial\beta_j\partial\beta_k}\right|_{\beta_0}+\ldots\right\}\ldots\Delta\beta_k\right\}\Delta\beta_j\right\}\Delta\beta_i\,, \tag{2.27}$$

or
$$\Delta y=\alpha(\beta)\cdot\Delta\beta\,. \tag{2.28}$$

This points out the possibility of defining the unknown structure $J(\beta)$ by constructing a partial model, if the latest $\beta_i \quad (i\neq 1)$ are constant. Next the coefficients α_1 as functions β_2, are approximated by relevant dependencies $\alpha_1(\beta_2,\alpha_2)$, next α_2 as function β_3 and eventually we get the sought dependence $\alpha(\beta_1,\ldots,\beta_n)$ for the model (2.28).

In this approach the dependencies' unidimensionality that are approximated simplifies the task of selecting the structure of partial models. We can create a database of canonical models (the models with minimum dimensionality of the vector α of the unknown coefficients) and arrange it in order by using the table of properties. Then the candidate for the best model will be that one, the set of properties of which coincide the most with a similar set of properties of the experimental curve.

The following examples illustrate these methods.

The test example. The unknown dependence $J(\beta) = \beta_1 + 0.5\beta_1\beta_2 + 0.5\beta_2^2$ defined by the table 2.1,

Table 2.1

Region	G_1			G_2			G_3		
№ point	1	2	3	4	5	6	7	8	9
J	0	0.1	0.005	-1	-0.9	-0.945	1	0.61	0.605
β_1	0	0.1	0	-1	-0.9	-1	0	0.1	0
β_2	0	0	-0.1	0	0	-0.1	-1	-1	-1.1

where the total region G consists of three subregions G_1, G_2, G_3, each has a centerpoint 1, 4, 7 respectively, and two points with a deviation of ± 0.1 from the centerpoint. Using the equation (2.22) let us determine the first derivatives $\left.\frac{\partial J}{\partial \beta_i}\right|_{\beta_{0r}}$, $i = 1,2; r = 1,2,3;$ next for the whole region G – the second derivatives (2.24); having substituted them in the equation (2.21), we obtain the model's equation: $\hat{J}(\beta) = \lfloor \beta_1 + 0.5\beta_1\beta_2 + 0.5\beta_2^2 \rfloor + 0.05\beta_2$, that coincides with the sought one, after discarding the insignificant component which was arisen as a result of substituting the derivative by difference .

The example 2. The determination of the analytical dependence of the energy of the first half-wave of the current discharge to the capacitor C in the circle with the inductance L and the resistance R depending on the quality factor β of the contour line of the energy converter's power.

The current x with the R, L, C parameters of circle is connected with the equation

$$\frac{d^2x}{dt^2} + \frac{R}{L}\frac{dx}{dt} + \frac{1}{LC}x = 0, \quad \int_{-\infty}^{0} x(t)dt = CU_C(0), \quad \beta = \frac{1}{R}\sqrt{\frac{L}{C}}, \qquad (2.29)$$

where $U_C(0)$ - are initial conditions for the voltage on the capacitor, $x(0) = 0$.

The index

$$J(\beta)=\left(\frac{CU_C^2(0)}{2}\right)^{-1}\int_0^{t_1} Rx^2(t)dt,\quad t_1=\underset{t\in(0,\infty)}{\text{ArgSupr}}\ x^2(t). \tag{2.30}$$

The task (2.29), (2.30) was solved by numerical method. The dependence $J(\beta)$ was obtained where $\beta-$ is a scalar approximated by fractional rational function from β (fig.2.1)

Fig.2.1. Dependence $J(\beta)$ (a dotted line) and its model (2.2.31) $\hat{J}(\beta)$ (a continuous line)

$$\hat{J}(\beta)=\frac{-0.031+1.066\beta-0.658\beta^2+0.149\beta^3}{1+0.23\beta-8.87\cdot10^{-5}\beta^2+1.112\cdot10^{-5}\beta^3}. \tag{2.31}$$

Hence the estimate $\hat{\beta}^*$ of the optimum value β^* is found under condition $\frac{\partial\hat{J}(\beta)}{\partial\beta}=0,\quad \frac{\partial^2\hat{J}(\beta)}{\partial\beta^2}<0$. The error $\frac{\beta^*-\hat{\beta}^*}{\beta^*}\cdot100\%$ is 0.05%.

The example 3. The obtaining of the multidimensional nonlinear dependences at the power facilities according to the natural experiments' data.

The charts of the experimental $J(\beta)$ and analytical $\hat{J}(\beta_1,\beta_2)$ dependencies are represented in the fig. 2.2. The dependence $\hat{J}(\beta_1,\beta_2)$ is obtained by the approximation of the coefficients α_i of the local i- models $\hat{J}(\beta_1)=\alpha_{1i}\beta_1+\alpha_{2i}\beta_1^2,\quad i=\overline{1,6}$ in terms of the

linear functions β_2 $\alpha_{ij}(\beta_2) = \alpha'_{j2} + \alpha'_{j1}\beta_2$, $j = 1,2$. The mean square error is 2% from J_{max}.

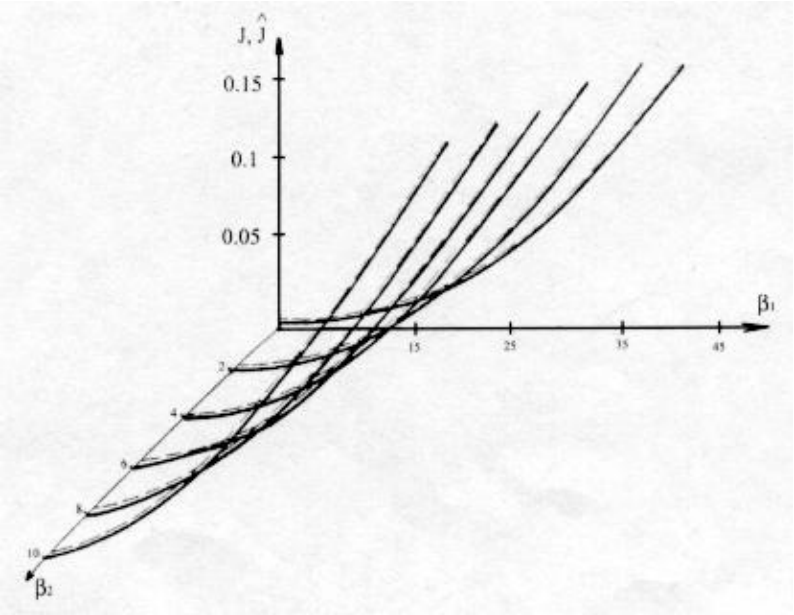


Fig. 2.2 The dependence $J(\beta_1,\beta_2)$ (a dotted line) and its model $\hat{J}(\beta_1,\beta_2)$ (a continuous line).

In the fig. 2.2 we see the charts of the experimental $J(\beta_1,\beta_2,\beta_3)$ and analytical $\hat{J}(\beta_1,\beta_2,\beta_3)$ dependencies, which are obtained in the same way:

$$\hat{J}(\beta_1) = \alpha_{1ik}\beta_1 + \alpha_{2ik}\beta_1^2, \quad i = 1,2,3; \quad k = 1,2,3; \tag{2.32}$$

$$\alpha_{jik}(\beta_2) = \alpha'_{jk}\beta_2 + \alpha''_{jk}\beta_2^2, \quad j = 1,2; \tag{2.33}$$

$$\alpha^{(1)}_{jk}(\beta_3) = \alpha'_j\beta_3 + \alpha''_j\beta_3^2. \tag{2.34}$$

Substituting the coefficients (2.34) into (2.33), (2.33) into (2.32), we obtain

$$\hat{J}(\beta) = -0.96\cdot10^{-6}\beta_1\beta_2\beta_3 - 0.61\cdot10^{-2}\beta_1\beta_2^2\beta_3^2 + 0.9\cdot10^{-2}\beta_2^2\beta_3^2 + 0.23\beta_1^2\beta_2\beta_3^2 +$$

$$+0.04\beta_1\beta_2\beta_3^2 - 1.72\beta_2\beta_3^2 - 0.24\cdot10^{-4}\beta_1^2\beta_2^2\beta_3 + 0.46\cdot10^{-5}\beta_1\beta_2^2\beta_3 - 0.0232\beta_2^2\beta_3 +$$

$$-0.0126\beta_1^2\beta_2\beta_3 + 0.133\beta_1\beta_2\beta_3 + 1.02\beta_2\beta_3.$$

The approximation error does not exceed 0.5% from the maximum value *J*. If the permissible error is 5%, the expression simplified

$$\hat{J}(\beta) = -0.61\cdot10^{-2}\beta_1\beta_2^2\beta_3^2 + 0.9\cdot10^{-2}\beta_2^2\beta_3^2 + 0.23\beta_1^2\beta_2\beta_3^2 + 0.04\beta_1\beta_2\beta_3^2 - 1.72\beta_2\beta_3 -$$

$$-0.0232\beta_2^2\beta_3 - 0.0126\beta_1^2\beta_2\beta_3 + 0.133\beta_1\beta_2\beta_3 + 1.02\beta_2\beta_3$$

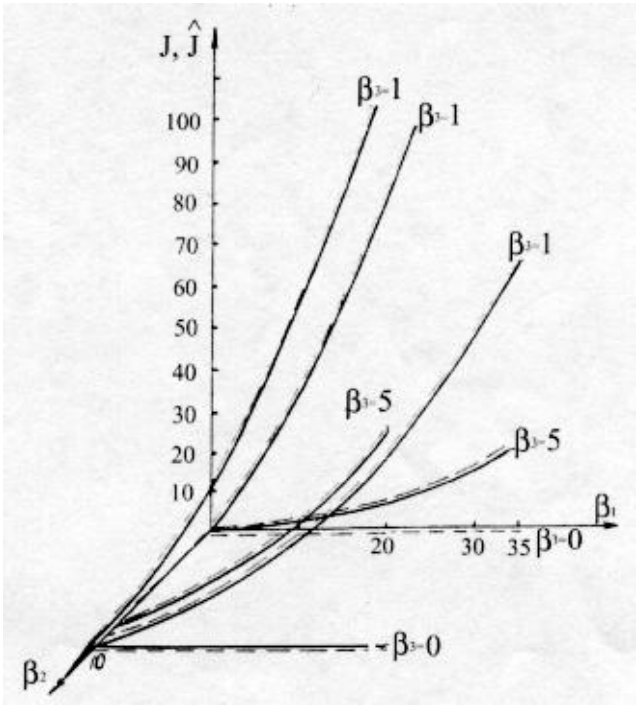


Fig. 2.3 The dependence $J(\beta_1,\beta_2,\beta_3)$ (a dotted line) and its model $\hat{J}(\beta_1,\beta_2,\beta_3)$ (a continuous line).

Thus, the rational use of the mathematical or numerical modeling or experimental research and the methods of the identification theory makes it possible to obtain the analytical dependence which connects the appropriate parameters of the complex system that is tested or designed. Next the obtained analytical dependence $\hat{J}(\beta)$ is used for solving tasks of systems' analysis and synthesis, for interpolation and extrapolation for regimes, for which the experimental data are unknown.

2.4. The Modeling and Forecasting of the Solution (2.11) of the System (2.1) if the Information on the Mapping $\mathbf{T}\times\mathbf{T}\times\mathbf{X}\times\mathbf{\Omega}\to\mathbf{X}$ is Missed

Under condition that the information on the mapping of the set Ω of the controlling action $u(t)$ and the previous values of the state's X variables are missed the modeling task comes to the approximation task of the data extraction $x(k),\ \ k=0,1,2,\ldots,M$, which are perturbed by noise $N(k)$. The task of identification in order to approximate the

aperiodic processes is simple with provision for the fundamental process's property, its smoothness, that is known a priori, and taking into account two Weierstrass's theorems on approximation of the aperiodic processes by Taylor series, and the periodic processes – by Fourier series.

The other situation is at the approximation for forecasting. Among the model's set which describe the process $X(k)$ with a given accuracy you should select one that corresponds to the real hidden regularity, because only this can provide the sufficient accuracy of the forecasting process $X(k)$ in the future. This is especially actual for short data $x(k)$ extractions $[k=\overline{1,m}]$.

The example. The unknown solution (2.1) of the unknown system (2.2) for one from the coordinates x_i of the vector-function $x(k)$ has the unknown type

$$x_i(t)=\beta_0+\beta_1 t+\beta_2 \sin wt+\varepsilon_i, \qquad (2.35)$$

where the coefficients $\beta_0=2;\beta_1=3/t;\beta_2=0.7;\quad w=\dfrac{2\pi}{T}$, T – is the interval of the solution's observation $x_i(t)$, which is like an ordinary degree function at the $[0,T]$. Therefore, as a rule, in these conditions the researcher and automated system will select the model $\hat{x}(t)$ as a degree polynomial for the approximation $x_i(t)$.

$$\hat{x}_i(t)=\sum_{j=0}^{n}\hat{\beta}_j\cdot t^j \text{, where } n=1,2,3. \qquad (2.36)$$

For the models' selection (degrees n) it is necessary to use the external criterion, for example, the "regularity" criterion

$$\hat{I}=\left[\sum_{k=1}^{Npr}\left(x_i^{pr}(k)-\hat{x}_i^{pr}(k)\right)^2\Big/\sum_{k=1}^{Npr}\left(x_i^{pr}(k)\right)^2\right]^{\frac{1}{2}}, \qquad (2.37)$$

where $x_i^{pr}(k),\quad k=\overline{1,N_{pr}}$ – are values of the time series $x_i(k), k=\overline{1,M}$ selected in the checking sequence; $\hat{x}_i^{pr}(k),\quad k=\overline{1,N_{pr}}$ – is an estimate's forecasting $x_i(k)$ at the

checking sequence, obtained on the model constructed at the operating sequence $M - N_{pr}$ of the points k $(k = 1,2,\ldots, M - N_{pr})$.

Let us take $[0, T - \tau_j]$, $j = 1,2$, $\tau_1 = 0.2T$, $\tau_2 = 0.5T$ as an operating range, where we obtain LSM- estimates $\hat{\beta}_j$ of the model (2.36) for various degrees n. As the criterion of "regularity" (2.37) we take the module of the difference of real and forecasting values $x_i(t)$ in the point T:

$$\hat{I} = \left| x_i(T) - \hat{x}_i^{pr}(T) \right|. \qquad (2.38)$$

So far as this is a test task, let us calculate the forecasting accuracy criterion that is not physically implemented,

$$\hat{I} = \left| x_i(T + \tau_j) - \hat{x}_i^{pr}(T + \tau_j) \right|, \qquad (2.39)$$

in order to compare it with the physically implemented criterion $\hat{I}$.

In the table 2.2 there are criterions' I i $\hat{I}$ values for $\tau = 0.2T$ and $\tau = 0.5T$, numbers of the relevant points of the graph (fig. 2.5) for the models (2.36) of various degrees n: $1, 2, 3$.

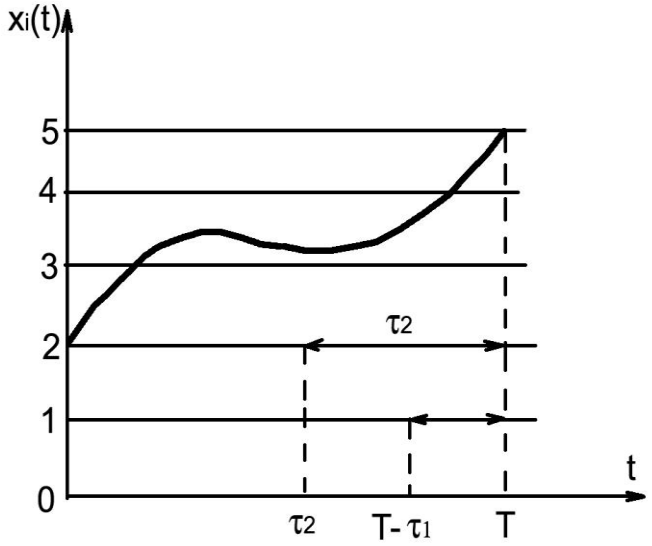


Fig. 2.4. The dependence $x_i(t)$

Table 2.2

The forecasting results

n	τ_i	$\hat{I}$	№pnt	I	№pnt
1	0.2	0.7	1	1.4	2
2	0.2	1.8	3	3.6	4
3	0.2	0.2	5	0.4	6
3	0.5	5	7	6.2	8
2	0.5	5	9	6.2	10
1	0.5	0.2	11	0.3	12

As it follows from the table and figure 2.5, the model (2.36) is optimal for $\tau = 0.2T$ in terms of I and $\hat{I}$ at $n = 3$; the model (2.36) is optimal for $\tau = 0.5T$ at $n = 1$. The proximity of values I and $\hat{I}$ confirms the validity of the criterion of regularity in the forecasting task. Though different models will be optimal for different intervals τ. If we use the optimal model for $\tau = 0.2T$ while forecasting at $\tau = 0.5T$ (point 13), the error of the forecast increases significantly. Thus, it is desirable that checking and forecasting intervals will be the same. The effect of regularization by external criterion (2.38) is that for the same data, the larger τ is, the simpler is the model selected by the criterion (2.38): the criterion $\hat{I}$ discarded the models of the second and third orders for $\tau = 0.5T$.

Let us consider the case in which the presence of a periodic component time series is obvious, for example, the monthly volume of electric power consumption. Although this index varies from year to year, but within each year, that is strictly periodically in time, it increases in winter and decreases in summer. In the fig. 2.6 there is the sample at 312 points of average values of electricity consumption in Ukraine in a period of 26 years (1960 ÷ 1986).

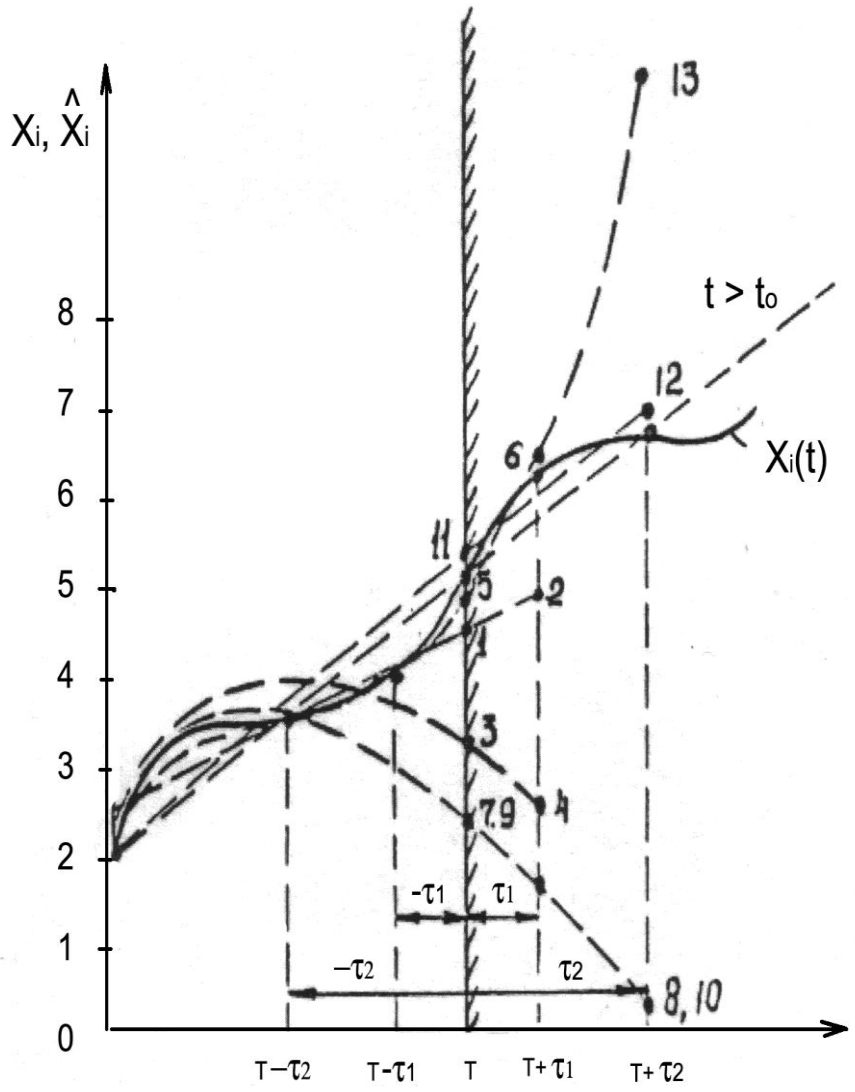


Fig. 2.5. The graph $x_i(t)$ and its forecasting models.

The lack of input variable $u(t)$ (mainly of the growth of industrial capacity) makes it impossible to construct the model "input-output" or "cause-effect". Analyzing the consequence, we have the opportunity to represent it as a homogeneous differential equation only

$$(T_1 p - 1)(T_2^2 p^2 - 2\xi T_2 p + 1)x(t) = 0, \qquad (2.40)$$

$x(0) \cong 2000;\ \ px(0) \cong 70 / \text{month}$, $p = \dfrac{d}{dt}$ – is a differentiation's operator.

For that time this model corresponds to the growth tendency of the average annual consumption ($p_1 = \dfrac{1}{T_1}$ is the root) and gradual increase of vibrations.

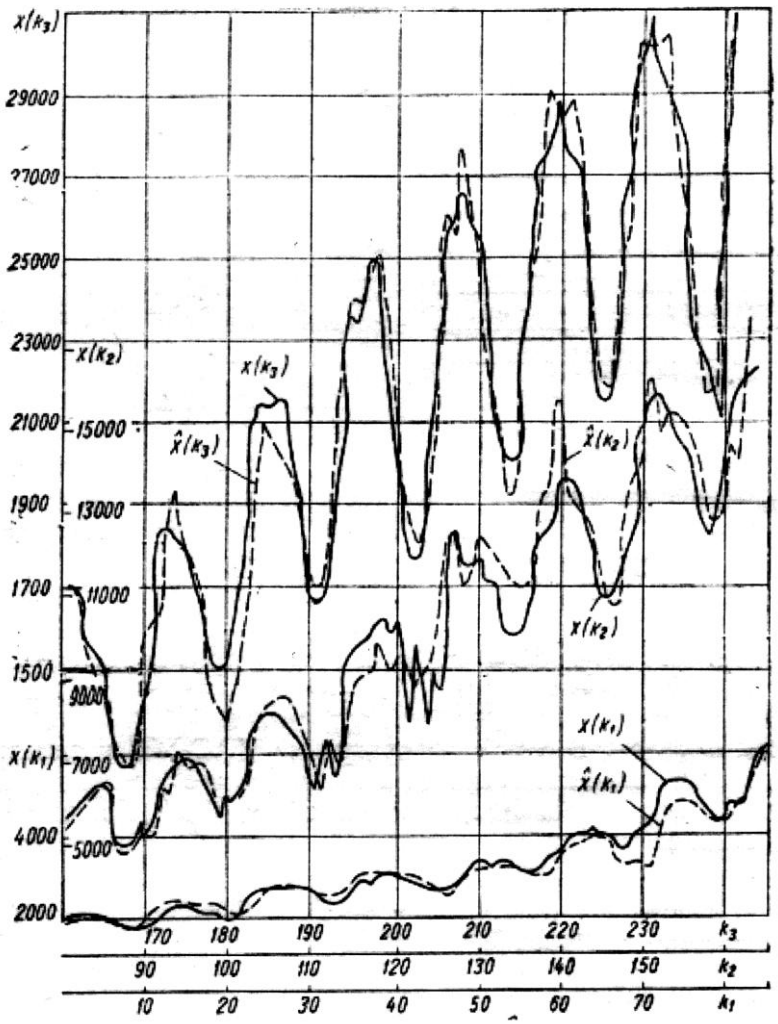


Fig. 2.6. The predicted $\hat{x}(k_i)$ and measured $x(k_i)$ values of the series, $i = 1, 2, 3$.

$$(p_{2,3} = \frac{\xi}{T} \pm j\sqrt{\frac{\xi^2 - 1}{T^2}}).$$

However, there is the exponential growth tendency for $k = \overline{0,220}$. Next the process goes into the stationary mode. Thus, the model with time variant coefficients $T_1(t)$ and $\xi(t)$ will be more accurate, while T_2 is unchanged and equal to 12 months. Within that set up the predicted value $\hat{x}^{pr}(T+\tau)$ is calculated by the model (2.40) as its solution (2.11) for predicted values $T_1(T+\tau)$ and $\xi(T+\tau)$.

In the fig.2.6 there are measured $x(k_i)$ and predicted $\hat{x}(k_i)$ values of electricity consumption for one year to come. The forecast is carried out by the nonstationary

difference model in increments of a year (12 months), which was calculated for the previous three years

$$\hat{x}(k+12) = \beta_0 x(k) + \beta_1 x(k-12) + \beta_2 x(k-24) + \beta_3 x(k-36). \qquad (2.41)$$

This model is a difference equivalent of the continuous model (2.40) with time variant coefficients. The estimates β_i, $i = \overline{0,3}$ of the model (2.41) are determined by the least squares' method.

As it is highlighted in the work [5], the problem of constructing the models with the use of the criterion (2.37) is the abnormal criticality of the models to the method of selecting points for checking and operating sequences. For greater efficiency of models' selection, the criterion (2.37) is completed with criteria of unbiasedness or balance of the variables. This makes it possible to construct models of rather high quality for medium term and long term forecasting of time series.

In general the time series' models which are uniform differential or autoregressive difference equations can be considered as nonuniform ones, where the input variable $u(t)$ is represented by the additional system of differential equations of the order m: $\dot{u} = cu$, $u(0) = u_0$. Then we get the uniform system $\dot{Z}(t) = DZ(t)$ of extended dimensionality $n+m$, instead of the system (2.2) of the order n where $Z = (X, U)$, $D = D(A,B,C)$.

It is observed in the description of the time series. Thus, the homogeneous equation (2.40) which is artificially extended to the third order due to the lack of controlling action $u(t)$, such as gross domestic product, the total production volume, which creates aperiodic (annual average) series' component; heating costs that make up the vibrational component, and their connection with power consumption is noninertial, or it is of the response rate of the first order maximum and a nonidentity coefficient of the transfer (due to losses).

It is obvious that the more accurate are the models that take into account the fundamental cause-effect connections in the examined object. If the information on these connections is absent, it is necessary to select the models' structures at the external criterion $\hat{I}$, beginning with the simplest ones (with minimum dimensionality of the

evaluated parameters), but those ones that take into account the geometric properties of the approximated time series. Let us give some of these models and their properties (characteristics) as an example:

$y_1 = \beta_1 \sin(\beta_2 x + \beta_3)$ - is a sine wave extracted at β_1 y times, compressed at β_2 x times, shifted along the axis at a segment β_3/β_2, with a period $T = 2\pi/\beta_2$ and has zero at the points $x = \dfrac{n\pi - \beta_3}{\beta_2}$;

$y_2 = \beta_1 \mathrm{tg}(\beta_2 x + \beta_3)$ has a period π/β_2, is discontinuous at the y_2 in the points $x = \dfrac{(n+1)}{2}\dfrac{\pi}{\beta_2}$;

$y_3 = \beta_1 \arcsin(\beta_2 x)$ - increases monotonically from $-\pi/2$ to $+\pi/2$, if x changes from $-1/\beta_2$ to $+1/\beta_2$;

$y_4 = \beta_1 arctg(\beta_2 x)$ - increases monotonically from $-\pi/2$ to $+\pi/2$, if x changes from $-\infty$ to $+\infty$;

$y_5 = \beta_1 x^{\beta_2}$ - increases monotonically with acceleration if $\beta_2 > 1$ and with deceleration, if $\beta_2 < 1$;

$y_6 = \beta_1 e^{\beta_2 x} = \beta_1 \cdot \beta_3^x$ - increases monotonically from 0 to ∞, if $\beta_2 > 0$ or $\beta_3 > 1$, decreases monotonically from 0, if $\beta_2 < 0$ or $\beta_3 < 1$;

$y_7 = \beta_1 \log_{\beta_2} x \ \ (\beta_2 > 0)$ - is a mirror image of y_5 in relation to a line, if $\beta_1 = 1$;

$y_8 = \beta_1 e^{-(\beta_2 x)^2}$ - increases monotonically from 0 to β_1 and decreases from β_1 to 0 symmetrically in relation to $x = 0$;

$y_9 = \beta_1 e^{\beta_2 x} + \beta_3 e^{\beta_4 x}$ - changes the configuration monotonically depending on the signs and quantities $\beta_1 \div \beta_4$;

$y_{10} = \beta_1 e^{\beta_2 x + \beta_3 x^2} = y_6 y_8$ - is increasing nonsymmetrically, if $\beta_3 > 0$ or decreasing to β_1, if $|\beta_3| > |\beta_2|, \quad \beta_3 < 0$;

$y_{11} = \beta_1 x^{\beta_2} e^{\beta_3 x}$, $x > 0$ - is increasing, if $\beta_1 > 0$, $\beta_2 > 0$, $\beta_3 > 0$, it is extreme, convex, nonsymmetric, bell-shaped with the shift of the extremum on the axis OX , the shift as large, as larger is β_3.

Analyzing the examined series with the use of characteristics provided for functions $y_1(x) \div y_{11}(x)$, where x for time series is t, we can select the model that is optimal by criterion of minimum complexity. This analysis is easily automated using the standard contemporary algorithms of images' recognition which are based on the theory of fuzzy sets and neuron shaped structures or any other statistical methods.

2.5. The Modeling, Forecasting and Diagnostics of the Solution (2.11) of the System (2.1) if the Information on the Mapping $T \times T \times X \times \Omega \to X$ is Available

In the section 2.2 we considered the direct modeling and forecasting, under condition that the mapping "input – output" is known and using the cubic approximation of input signals and analytical calculation of the solution of an appropriate differential equation.

If the mapping "input-output" is unknown, but input $u(t)$ and output $x(t)$ variables are available, there is an inverse task, viz the task of the identification of mapping $f(u,x,t)$ in (2.1). This task is the most complicated, but it is the most important one. There are several options:

1) to determine the linear operator $W(p,\beta)$ at a given point (x_0, u_0), $p = \dfrac{d}{dt}$ of the dynamic connection of small deviations Δu, Δx, using an input $u(t)$, an output $x(t)$, which are usually perturbated by the noise $N(t)$. It is necessary for the analyzis of the object's stability and the synthesis of the process $x(t)$ controlling system.

2) to find out the fundamental (more stable) static, usually nonlinear connection of sustained values $\bar{u}$ and $\bar{x}$, viz $\bar{x}(\bar{u})$ (balancing dependances) from the process dynamics for the input $u(t)$ and the output $x(t)$ perturbated by noise $N(t)$.

3) to determine the shape of the input signal or parametric variable using the output $x(t)$ and the apriori known operator $W(p,\beta)$; to establish its physical meaning: to diagnose the cause of the transition process in the $x(t)$ using the methods of expert evaluation. This is the task of causes' analysis.

Let us illustrate the first two tasks by the previous example of the electricity consumption. Let us represent the process $x(t)$ as the sum of periodic $x_p(t)$ and aperiodic $x_a(t)$ components of the object's output which has this structure (fig. 2.7).

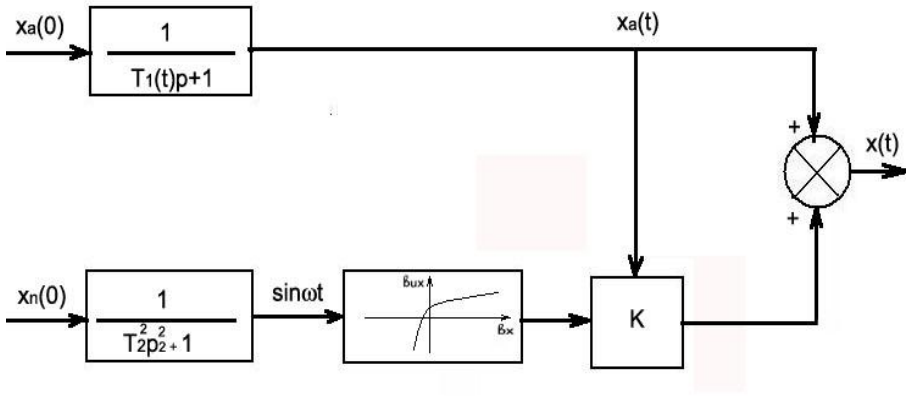


Fig. 2.7. The block diagram of the process $x(t)$ model.

Here $T_1(t)$, changing in time from negative to positive respectively, provides the coincidence of aperiodic component $x_a(t)$ with the current average value $\bar{x}(t)$. The operator $\left(T_2{}^2 p + 1\right)^{-1}$ is the generator of the sine signal $\sin\omega t$, where $\omega = \dfrac{2\pi}{T}$, T = are 12 months. The signal $\sin\omega t$ passes through the nonlinear transformer, its output is amplified in proportion to the estimate of average $x_a(t)$, forming a periodic component $x_p(t)$ of the signal $x(t)$.

The first task is to define T_1 using the window $x_a(0)$ current values, for example, at the beginning of each year and reactions $x_a(t)$ at it within 12 months.

The second task is to find out the nonlinear type, which transforms the sine signal into $x_p(t)$. The fundamental (stable) property is found in nonlinearity's asymmetry: in winter the half-wave of oscillations has a larger area than in summer, and depends on the oscillation's amplitude from $x_a(t)$.

From the model's analysis follows that managing impact onto the system of electric power consumption of Ukraine, which will reduce the peak load in winter, may be a change of tariffs for industrial power consumption proportionally to $x_p(t)$ (in the peak winter months the power is expensive, in summer it is cheaper). This may force the company to redistribute capacity at time in order to minimize the cost of the consumed energy. As a result the peak winter load decreases.

As an example of partially the first task and partially the third task, let us consider the schedule curve of the change of the gross national product (GNP) (fig. 2.8).

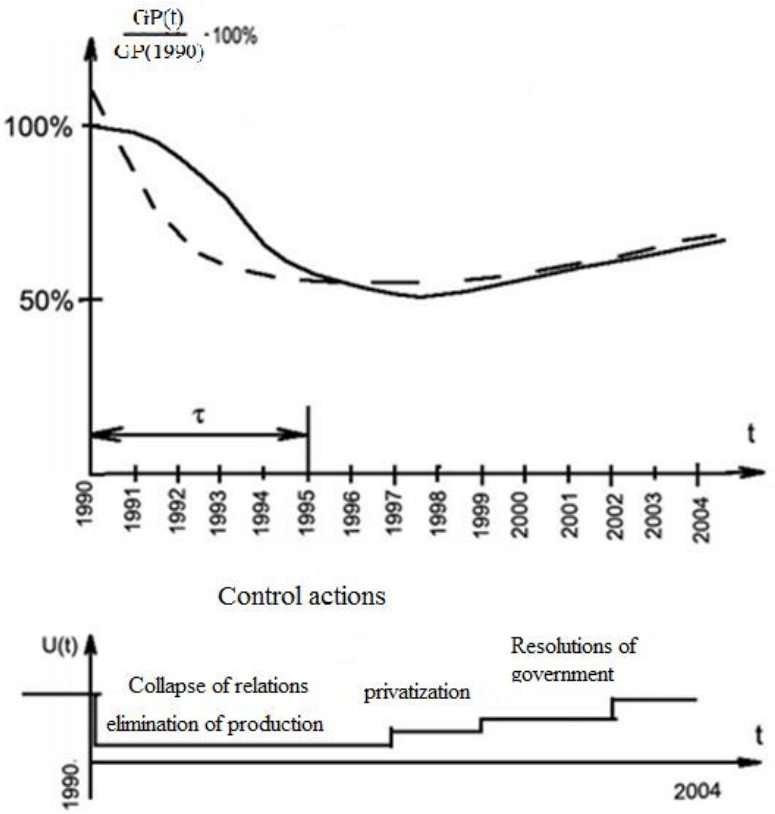


Fig. 2.8. The curve of the gross national product in period of "rebuilding" ("perestroika") and probable controlling actions $u(t)$.

The sustained value of GNP begins to decrease rapidly in 1989–1990. In 1995 the rate drops and in 1997-2000 we have the consistently low level of GNP. Since 2001 to nowadays we observe the slow rate of GDP's growth. In order to determine the reasons for such dynamic of GDP, we use classical macroeconomic model of the economy in the industrial and technological interpretation (fig. 2.9).

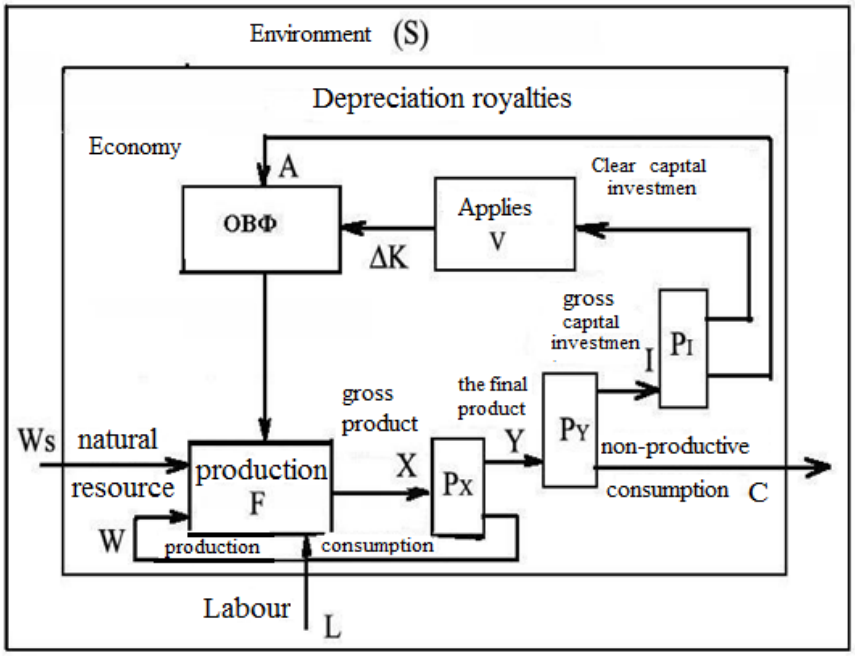


Fig. 2.9. The dynamic macroeconomic model.

The production process variables are:

L –is a labor; K – are capital goods (the main production assets); $\tilde{W}$ – are objects of labor which consist of natural resources W and labor object $\tilde{W}$, returned into production as a part of the gross national product.

The output variables of the production process are: X –is a gross product, which is distributed in the subsystem P_x on the production consumption W and the final product Y. Then, in the subsystem of distribution P_y the product Y is divided into the gross capital investment I and productive consumption C. The subsystem P_I divides the investments I into amortization deductions A and net capital investments.

It follows from the scheme (fig.2.9) that $X = W + Y$, $Y = I + C$. Assuming that the investment I is used for the increment of growth ΔK of basic productive assets K over the same year and also for A, we obtain the discrete model

$$I_t = q\Delta K_t + A, \quad A = \mu K_t, \tag{2.41}$$

or

$$\Delta K_t = q^{-1}(I - \mu K), \tag{2.42}$$

where q – is the model's parameter; μ – the amortization factor ; K_t – are basic production assets in t - year.

If instead of the year we take the increment growth per dt, we obtain the continuous model of the first order

$$\frac{dK}{dt} = q^{-1}(I - \mu K). \tag{2.43}$$

If we assume that production costs W are proportional to the gross domestic product X $(W = aX)$, $X_t = W_t + qDK_t + mK_t + C_t$, then, going to the continuous model, we obtain the expression

$$\frac{dK}{dt} = \frac{1}{q}\left[(1-a)X - \mu K - C\right]. \tag{2.44}$$

According to the absolute Lavrentiev model, assume that the increment growth the gross product $DX_t = X_{t+1} - X_t$ (where t – are years) is proportional to the capital investments I_t, $I_t = \chi \Delta X_t$.

Then we get the gross product model from the previous equations:

$$\frac{dX}{dt} = \chi^{-1}\left[(1-a)X - C\right]. \tag{2.45}$$

From the structural scheme (fig. 2.9) and the equation (2.49) it follows that the reducing of the investments I_t (t = 1990, ..., 1997) has led to an exponential decrease with

a time constant $\tau = \frac{q}{\mu}$ of the basic production assets K and as a result, of the gross domestic product X.

For the normal course of economic development $X(t)$ must grow, viz the root $P_1 = \chi^{-1}(1-a)$ of the characteristic polynomial $P_1 - \chi^{-1}(1-a) = 0$ of the equation (2.45) must be positive. But in the period of "rebuilding" the coefficient a exaggerated the unit and in the equation (2.45) we have the process *x(t)* damped in the interval [1990 - 1997]. Approximately approximating it by the exponent (a dotted line in fig. 2.8), define the time constant $\tau = P_1^{-1} = \frac{\chi}{a-1} \approx 5$ years.

Thus, the conclusion on both parametric and signal perturbations, which arose in 1990, follows from the graph (fig. 2.8), the model (fig. 2.9) and the formula (2.45). The signal perturbation has the reduction of W_t, K, W, L; the parametric one has the change of relations of a (between W and X) and χ (between I_t and ΔX_t): a increased, χ decreased.

This example does not provide the detailed analysis of all the problems of social production, distribution, exchange and consumption, but it confirms the continuity of processes in the economy and, consequently, the possibility of presenting their dynamics by the continuous or difference equations.

The presence of inertia, continuity and, consequently, the smoothness of dependences of the economic index on the time should be taken into account when planning and forecasting its development in time. Otherwise the purpose and the reality will vary significantly, for example, as it took place in the program of conserving the primary energy resources PER of Ukraine and its implementation (2.10).

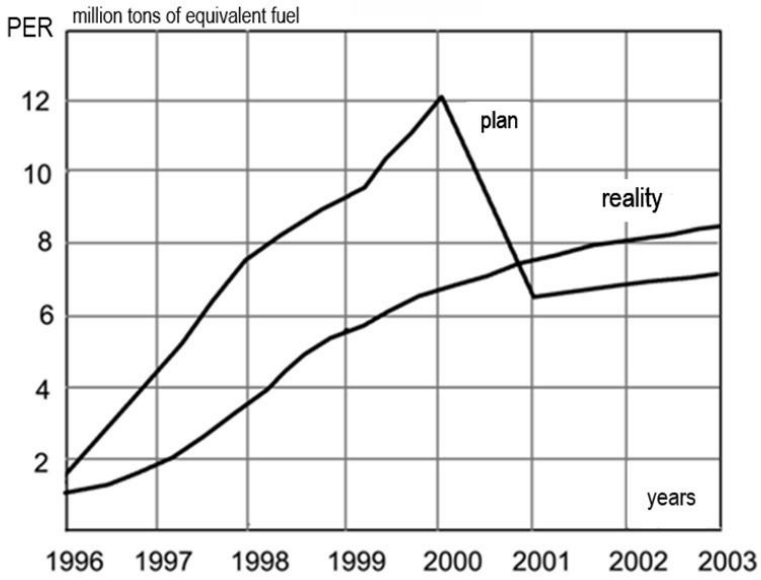


Fig. 2.10. The costs of PER in Ukraine.

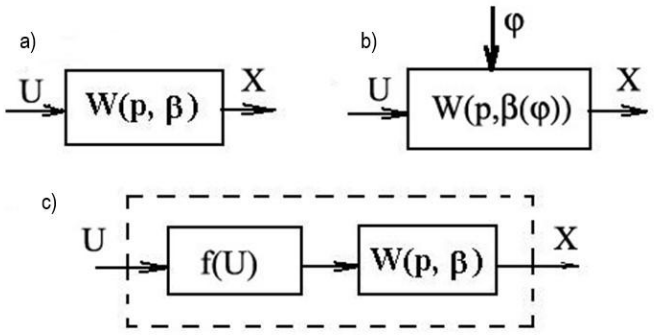


Fig. 2.11. The models of the mappings “input-output”.

These objects have the measured input variables U, output variables X, that are related to the input ones:

- by the linear stationary operator $W\left(p,\beta\right)$ (fig. 2.11,a);

- by the linear nonstationary operator $W\left(p,\beta\left(\varphi\right)\right)$ (fig. 2.11,b);

- by the nonlinear stationary Hammerstein’s operator $W\left(p,\beta\right)\ \ f\left(U\right)$ (fig. 2.11,c), where β is the parameters vector, φ is the variable of the time t or of the input U or of the output X, $f\left(U\right)$ is the static nonlinear dependence.

The variables *U*, *X* and the operators *W*, *f* can be both vectors and scalars.

Chapter 3. The Analysis of the Simplified Methods of the Confluent Analysis

3.1. The Least Square Method's Analysis if the Input and Output Signals are Noisy

The setting up of the practical problem contains a substantial proportion of the uncertainty of statistical properties of the **noises** at the measurements of both input *X* and output *Y* data about the examined object, the model of which can be represented as

$$Y^* = X^*\beta^* + \varepsilon^*, \tag{3.1}$$

where де Y^*, X^*, ε^* – are accurate output and input variables and deficiency, under condition that the estimate β^* for these values is obtained by the least square method

$$\beta^* = (X^{*T}X^*)^{-1}X^{*T}Y^* = C^*Y^*,\ (X^{*T}X^*)^{-1}X^{*T} = C^*. \tag{3.2}$$

In case of:

$$\beta^* = \operatorname{argmin}\varepsilon^{*T}\cdot\varepsilon^*. \tag{3.3}$$

In practice the LSM-estimate is obtained using the perturbated noises N_x and the data N_y:

$$X = X^* + N_x, \quad Y = Y^* + N_y, \tag{3.4}$$

where $X = \begin{bmatrix} x_1(1) & x_2(1)\ldots & x_i(1)\ldots & x_n(1) \\ x_1(2) & x_2(2)\ldots & x_i(2)\ldots & x_n(2) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_1(\text{“}) & x_2(\text{“})\ldots\ldots & x_i(\text{“})\ldots\ldots & x_n(\text{“}) \end{bmatrix}$,

$Y^{\mathrm{T}} = [y(1)\ y(2)\ldots y(M)]$, $x_i(j) = x_i^*(j) + N_{x_i}(j)$, $y_i(j) = y_i^*(j) + N_{y_i}(j)$, $j = \overline{1,M}$.

Then LSM-estimate $\hat{\beta}$ of the vector β^* is found from the condition (3.3), but for the real data (3.4) in the usual

$$\hat{\beta} = (X^T X)X^T Y = C \cdot Y, \; C = (X^{\mathrm{T}} X)X^{\mathrm{T}} \tag{3.5}$$

or in the recursive type with the initial conditions $P(0)$, $\hat{\beta}(0)$

$$\hat{\beta}(j+1) = \hat{\beta}(j) + P(j+1)X(j)\left[Y(j) - X^T(j) \cdot \hat{\beta}(j) \right],$$

$$P(j+1) = P(j) - P(j)X(j)\,[X^{\mathrm{T}}(j)P(j)X(j) + 1]^{-1} X^{\mathrm{T}}(j)P(j). \tag{3.6}$$

If the initial conditions are unknown, we take $\hat{\beta}(0)=0$, $P(0)=\sigma_{\beta(0)}{}^2 \cdot I$, $\sigma_{\beta(0)}{}^2 \to \infty$.

In order to simplify the analysis, let us take that noises N_x and N_y are white Gaussian noises, and consequently they are auto- and mutually uncorrelated. Let us determine the shift $\Delta\hat{\beta}$ of the estimate (3.5) with respect to the estimate (3.2):

$$\begin{aligned}
&\Delta\hat{\beta} = M\left\{\hat{\beta}\right\} - \beta^* = M\left\{\left[C^* + \delta C(N)\right] \cdot \left[Y^* + \varepsilon^* + N_y\right]\right\} - \beta^* = \\
&= M\left\{\left[C^* + \delta C(N)\right] \cdot Y^*\right\} - \beta^* = M\left\{C^* + \delta C(N)\right\} \cdot Y^* - \beta^* = \\
&\cong \left[X^{*T} X^* + M\left\{N_x^T N_x\right\}\right]^{-1} \cdot X^{*T} Y^* - \beta^* = \\
&= \left[X^{*T} X^* + \mathrm{diag}\sigma_i^2 \cdot M \cdot I \right]^{-1} \cdot X^{*T} \cdot Y^* - \beta^*.
\end{aligned} \tag{3.7}$$

Let us denote $X^{*T} X^* = A^*$, $M\left\{N_x^T N_x\right\} = \delta A$, $X^{*T} Y^* = B^*$. Then (3.2) and (3.7) are $A^*\beta^* = B^*$ $\left[A^* + \delta A\right]\hat{\beta} = B^*$ respectively. Hence $\delta A\hat{\beta} = -A^*\left(\hat{\beta} - \beta^*\right) = -A^*\Delta\hat{\beta}$, viz $\delta A\left(\beta^* + \Delta\hat{\beta}\right) = -A^*\Delta\hat{\beta}$ or

$$\Delta\hat{\beta} = -\left(A^* + \delta A\right)^{-1} \cdot \delta A \cdot \beta^*. \tag{3.8}$$

As we can see, the estimate $\hat{\beta}$ will be reduced with respect to the true β^* by the value (3.8) if "white noises" N_x and N_y are uncorrelated. $\Delta\hat{\beta}$ tends to zero if the norm $||\delta A|| \to 0$ and $\hat{\beta}$ tends to β^*; if the norm $||\delta A|| \to \infty$ that $\Delta\hat{\beta}$ tends to $-\beta^*$ and $\hat{\beta}$ - to zero.

Under conditions mentioned above and taking that the norm $||N_X{}^T \cdot \varepsilon||$ is much less than $||X^{*T} \cdot \varepsilon||$ or $||N_X{}^T \cdot Y^*||$, the estimate (3.5) covariance is approximately equal to

$$\begin{aligned} \operatorname{cov}\left[\hat{\beta}\right] &\cong M\left\{\left[C_1 \cdot \varepsilon + C_2 N_x\right] \cdot \left[C_1 \cdot \varepsilon + C_2 N_x\right]^T\right\} = \\ &= C_1 \cdot M\left\{\varepsilon \cdot \varepsilon^T\right\} \cdot C_1^T + C_2 \cdot M\left\{N_x N_x^T\right\} \cdot C_2^T, \end{aligned} \tag{3.9}$$

where $\varepsilon = \varepsilon^* + N_y$,

$$C_1 = \left[X^{*T} X^* + M\left\{N_x N_x^T\right\}\right]^{-1} \cdot X^{*T},$$

$$C_2 = \left[X^{*T} X^* + M\left\{N_x N_x^T\right\}\right]^{-1} \cdot Y^{*T}.$$

The first component of the expression (3.9) with increasing of the level N_X decreases, at the second one – C_2 decreases and $M\{N_X N_X{}^T\}$ increases, but C_2 enters the expression (3.9) quadratically, while $M\{N_X N_X{}^T\}$ does it linearly.

Then for the "white noise" when $M\left\{N_x N_x^T\right\} = \sigma_{N_x}^2 \cdot M \cdot I$, the estimate $\hat{\beta}$ covariance with the increasing σ_{N_x} will be reduced.

Thus, LSM has tendency to regularization of the normal system equations, similarly to Tikhonov's regularization [14]. The last one consists in minimization of the ordinary quadratic functional $I = \varepsilon^T \varepsilon$ with regularizing additive $\alpha\hat{\beta}^T\hat{\beta}$, where α - the regularization's parameter.

$$\boldsymbol{I} = \varepsilon^T \varepsilon = (Y - X^* \hat{\beta})^T (Y - X^* \hat{\beta}) + \alpha \hat{\beta}^T \beta,$$
$$\frac{\partial I}{\partial \hat{\beta}} = 0 = 2\left(X^{*T} X^* \hat{\beta} - X^{*T} Y + \alpha \hat{\beta} \right), \qquad (3.10)$$
$$\hat{\beta} = \left(X^{*T} X^* + \alpha \cdot I \right)^{-1} \cdot X^{*T} \cdot Y.$$

Comparing (3.7) and (3.10) we see that the Tikhonov's parameter in LSM is $\mathrm{diag}\sigma_i^2 \cdot M$.

The characteristic curve of the normed values $\frac{\left\|\hat{\beta}\right\|}{\left\|\beta^*\right\|}$ and $\frac{\left\|\mathrm{cov}\hat{\beta}\right\|}{\left\|\mathrm{cov}\beta^*\right\|}$ (curve a) and $\frac{\left\|\Delta\hat{\beta}\right\|}{\left\|\beta^*\right\|}$ (curve b) is represented in the fig. 3.1.

Fig. 3.1. The dependence of the shift and the covariance of the LSM-estimates from relation "noise-signal" in X.

3.2. The Generalized LSM and its Practical Implementation

In this method LSM-estimates are found using weighted (filtered) data:

$$\hat{X} = U^{-1} X, \qquad \hat{Y} = U^{-1} Y.$$

This is equivalent to minimizing the functional

$$\boldsymbol{I} = 0{,}5 \sum_{j=1}^{M} \left\| y(j) - x(j)\hat{\beta} \right\|^2 \cdot Q^{-1}, \qquad (3.11)$$

where Q –is the weight matrix of each j measurement: $Q = (\mathrm{cov}\hat{\beta}) \cdot U^T$.

Then, the estimate weighted by weight Q^{-1} of the generalized LSM (GLSM) and obtained with minimum (3.11) is equal to

$$\hat{\beta} = (X^T Q^{-1} X)^{-1} \cdot X^T Q^{-1} Y. \qquad (3.12)$$

The estimate (3.12) covariance is

$$\mathrm{cov}\hat{\beta} = (X^T Q^{-1} X)^{-1} X^T Q^{-1} M\{N_y N_y^T\} \cdot Q^{-1} X (X^T Q^{-1} X)^{-1}. \qquad (3.13)$$

The estimates (2.3.12) have the minimal dispersion if

$$Q = M\{N_y N_y^T\}.$$

Then

$$\mathrm{cov}\hat{\beta} = \left[X^T M\{N_y N_y^T\} X \right]^{-1} \cdot. \qquad (3.14)$$

For the uncorrelated "white noise" in the measurements Y $Q = \sigma^2 y(j) I$. Then the recurrent formula of the GLSM coincides with the weighted LSM

$$\begin{aligned} \hat{\beta}(j+1) &= \beta(j) + P(j+1) X(j) \frac{1}{\sigma_y^2(j)} \left[y(j) - X^T(j) \beta(j) \right], \\ P(j+1) &= P(j) - X(j) \left[X^T(j) P(j) X(j) + \sigma_y^2(j) \right]^{-1} X^T(j) P(j). \end{aligned} \qquad (3.15)$$

The inaccuracy of the definition or prior task of the matrix Q^{-1} leads to a significant optimality loss of the estimates (3.12) or (3.15).

This algorithm is not robust [21]: the statistically insignificant inadequacy of the covariance matrix Q of noises N_y is dictated, for example, by some failures in the data, which form so-called "heavy tails" [21] at the noise N_y distribution law, while estimating the vector $\hat{\beta}$ by the algorithm (3.12) can lead to a significant error. Thus, the quasi-optimal GLSM will be more convenient and reliable on implementing:

- at the first stage there is the quasi-optimal estimation of the signals X, Y by smoothing their noisy samples $X(j)$, $Y(j)$, $j = \overline{1,M}$ with the linear filters;

- at the second stage there is a LSM-estimation of the vector $\hat{\beta}$ using estimates $\hat{X}$ and $\hat{Y}$ of the accuracy value X^*, Y^* signals obtained at the first step.

For this approach the noises N_x, N_y may be mutually correlated. In addition, since the model (3.1) is linear, and under condition that all the variables $X_i(t)$, $Y_i(t)$ are filtering by one filter, the non-coincidence of the smoothed values $\hat{X}^*$, $\hat{Y}^*$ to the true X^*

, Y^* does not lead to the bias of the estimates $\hat{\beta}$ with respect to β^*. Indeed the sign of equality in (3.1) is not violated if we act upon its left and right parts by the linear filter operator W_f:

$$W_f = \{Y^*\} = W_f\{X^*\beta^* + \varepsilon^*\} = W_f\{X^*\} \cdot \beta^* + W_f\{\varepsilon^*\}. \qquad (3.16)$$

The mathematical expectation of the estimation $\hat{\beta}$ is

$$M\left\{\hat{\beta}\right\} = \left[\hat{X}^{*T} \cdot \hat{X}^* + M\left\{N_x N_x^T\right\}\right]^{-1} \cdot \hat{X}^* \hat{Y}^*. \qquad (3.17)$$

If for simplicity we take that $M\left\{N_x N_x^T\right\} = \mathrm{diag}\left\{\sigma^2_{\hat{N}_S}\right\} \cdot I$, and put that matrix eigenvalues $\left(\hat{X}^{*T}\hat{X}^*\right)^{-1} \cdot \mathrm{diag}\left\{\sigma^2_{\hat{N}_i}\right\}$ are less than one, then the expression $\left[I + \left(\hat{X}^{*T}\hat{X}^*\right)^{-1} \cdot \mathrm{diag}\left\{\sigma^2_{\hat{N}_i}\right\}\right]^{-1}$ can be represented by a series. Then

$$\left(\hat{X}^{*T}\hat{X}^*\right) \cdot \hat{\beta} = \left[I + \left(\hat{X}^{*T}\hat{X}^*\right)^{-1} \cdot \mathrm{diag}\left\{\sigma^2_{\hat{N}_i}\right\}\right]^{-1} \cdot \hat{X}^{*T}\hat{Y}^* \cong$$

$$\cong \left[I - \left(\hat{X}^{*T}\hat{X}^*\right)^{-1} \cdot \mathrm{diag}\left\{\sigma^2_{\hat{N}_i}\right\}\right] \cdot \hat{X}^{*T}\hat{Y}^*.$$

Therefore

$$\hat{\beta} \cong \left(\hat{X}^{*T}\hat{X}^*\right)^{-1} \hat{X}^{*T}\hat{Y}^* - \left(\hat{X}^{*T}\hat{X}^*\right)^{-2} \cdot \mathrm{diag}\left\{\sigma^2_{\hat{N}_i}\right\} \hat{X}^{*T}\hat{Y}^*.$$

Taking into account that

$$\hat{\beta} \cong \left(\hat{X}^{*T}\hat{X}^*\right)^{-1} \hat{X}^{*T}\hat{Y}^* = \left(\hat{X}^{*T}\hat{X}^*\right)^{-2} \hat{X}^{*T}\hat{Y}^*,$$

we obtain the expression for the shift:

$$\Delta\hat{\beta} = \hat{\beta} - \beta = -\mathrm{diag}\left\{\sigma^2_{\hat{N}_i}\right\} \cdot \left(\hat{X}^{*T}\hat{X}^*\right)^{-1} \cdot \hat{\beta}^*. \qquad (3.18)$$

So far as the norm $\Delta\beta$ is not larger than the product of norms of the components of the right part of (3.18), we have the inequality

$$\frac{\left\|\Delta\hat{\beta}\right\|}{\left\|\hat{\beta}^{*}\right\|} \le \left\|\mathrm{diag}\left\{\sigma^{2}_{\hat{N}_i}\right\}\right\| \cdot \left\|\hat{X}^{*T}\hat{X}^{*}\right\|^{-1}. \qquad (3.19)$$

The stronger is filtering X, Y by filter operator W_f, the less is the norm of the perturbations' matrix $\sigma^{2}_{\hat{N}_i}$. Thus, the filter suppresses noises without violating the equation (2.16). Smoothing the components $X_i(t)$ of the vector-function $X(t)$ by filter, however, reduces their frequency spectra and, consequently, reduces their linear independence. The condition number of the matrix $\hat{X}^{*T}\hat{X}^{*}$ reduces and, consequently, the inverse matrix $\left(\hat{X}^{*T}\hat{X}^{*}\right)^{-1}$ norm increases.

The high-grade picture of the dependence of the normalized values $\frac{\left\|\Delta\hat{\beta}\right\|}{\left\|\hat{\beta}\right\|}$ (curve a), $\frac{\left\|\mathrm{diag}\sigma^{2}_{\hat{N}_i}\right\|}{\left\|\mathrm{diag}\sigma^{2}_{N_i}\right\|}$ (curve b), $\frac{\left\|\left(\hat{X}^{*T}\hat{X}^{*}\right)^{-1}\right\|}{\left\|\left(X^{*T}X^{*}\right)^{-1}\right\|}$ (curve c) as the function τ of the filter' W_f inertia for a fixed relation of "noise-signal" is represented in fig. 3.2.

As it follows from the graph, for each case there is an optimal value of the smoothing effect of the filter W_f, which norm of displacement (3.19) will be minimal.

The covariance of the estimate $\hat{\beta}$ by GLSM method assuming the mutual uncorrelated noises, is similar to LSM, but the matrices $M\left\{\varepsilon\cdot\varepsilon^{T}\right\}$, $M\left\{N_x N_x^{T}\right\}$ are not diagonal anymore. The further separated are the spectra of signals and noises, the better are the estimates of the quasi-optimal GLSM.

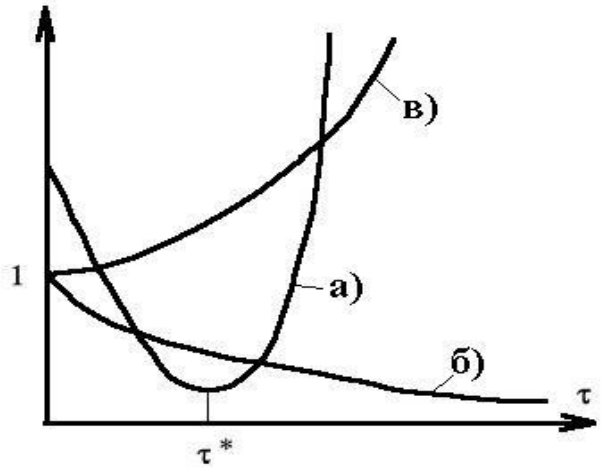


Fig. 3.2. The dependence of the normalized shift (a), noises' covariance (b) and the information matrix (c) on the parameter τ of the inertial filter.

3.3. The Analysis of the Integral-Correlation Criterion and Method of its Minimization (the Integrated LSM)

3.3.1 The Method's Main Point

The LSM and GLSM-estimates of the previous methods are found as a minimum point coordinate of the corresponding functionals $\varepsilon^T\varepsilon$ and $\hat{\varepsilon}^T\hat{\varepsilon}$. We take the minimum point as a coordinate's value of zero point of the functional derivative of β. So far as the functional is averaged on the finite interval T by the value of square ε or $\hat{\varepsilon}$ which is a mixture of useful signal $Y^* - X^*\beta$ and random perturbations $N_y - N_x\beta$, then, it is not accurate as a function of β.

It is known [22] that the operation of differentiation of the noisy function $\varepsilon^T\varepsilon$ is ill-posed. It is responsible for the low-precision of the LSM-estimates on the short heavily noised data X, Y samples. GLSM slightly improves the accuracy due to the smoothing noises at the X, Y. However, the uncertainty of boundary conditions influences on the precision of filtration on the short samples. Moreover the smoothing reduces the matrix

$\hat{X}^T\hat{X}$ conditionality and that is equivalent to reducing of the functional steepness. Then the uncertainty's region G of the estimate $\hat{\beta}$ increases again.

As it follows from the fig. 3.3, it is desirable to reduce variations the functional's values without reducing its curvature in the extremum's zone. This can be done for unsmoothed X, Y by additional averaging over the set of quasi statistically independent functionals that are close to the average quadratic one.

The average products can be shifted in time t on the interval Θ by those functionals, $\frac{1}{T}\int_0^T \varepsilon(t)\varepsilon(t+\Theta)dt$. Averaging them on the interval $[-\tau_1, \tau_1]$, we obtain the sought integral-correlation criterion:

$$I = \frac{1}{2}\int_{-\tau_1}^{\tau_1} \eta(\Theta)\int_0^T \varepsilon(t)\varepsilon(t+\Theta)dt\ d\Theta, \qquad (3.20)$$

where $\eta(\Theta)$ – is a weight function, unit in the simplest form.

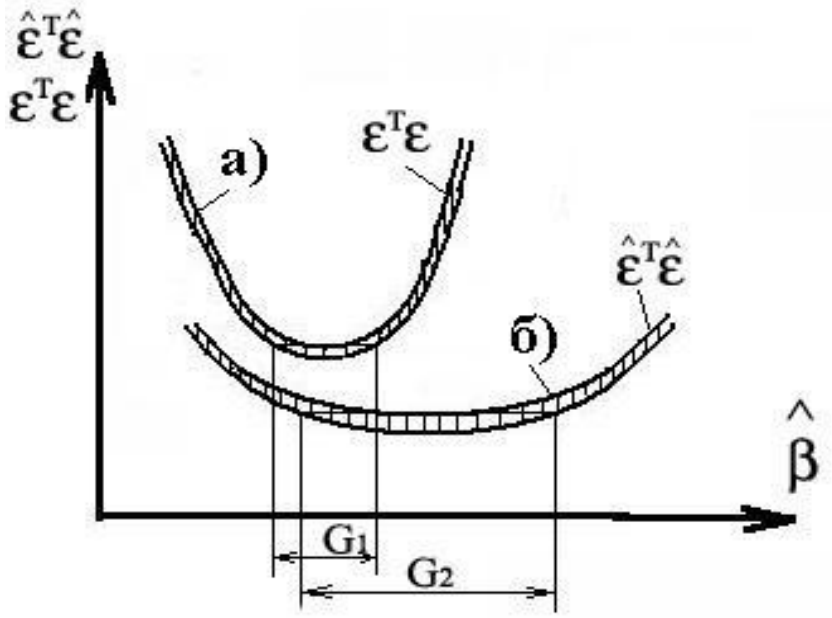


Fig. 3.3. The uncertainty's region G of the optimal value of LSM (a) and GLSM (b) of the estimate $\hat{\beta}$.

The necessary condition for a minimum I in $\hat{\beta}_k$, $k=\overline{1,n}$,

$$\frac{\partial I}{\partial \beta_k}=\frac{1}{2}\int_{-t_1}^{t_1}\eta(\Theta)\int_0^T\left[\frac{\partial\varepsilon(t)}{\partial\beta_k}\varepsilon(t+\Theta)+\varepsilon(t)\frac{\partial\varepsilon(t+\Theta)}{\partial\beta_k}\right]dtd\Theta=$$

$$=\int_{-t_1}^{t_1}\eta(\Theta)\int_0^T(-x_k(t))\left[y(t+\Theta)-\sum_{i=1}^{n}\beta_i x_i(t+\Theta)\right]+ \qquad (3.21)$$

$$+(-x_k(t+\Theta))\left[y(t)-\sum_{i=1}^{n}\beta_1 x_1(t)\right]=0$$

From the expression (3.21) follows the standard equations system:

$$A\cdot\hat{\beta}=B, \qquad (3.22)$$

where A – is a matrix $n\times n$ with elements a_{ik}; B – is a matrix-column $n\times 1$ with elements b_k;

$$a_{ik}=\int_{-t_1}^{t_1}\eta(\Theta)\int_0^T\left[x_i(t+\Theta)\cdot x_k(t)+x_i(t)\cdot x_k(t+\Theta)\right]dtd\Theta, \qquad (3.23)$$

$$b_k=\int_{-t_1}^{t_1}\eta(\Theta)\int_0^T\left[y(t+\Theta)\cdot x_k(t)+y(t)\cdot x_k(t+\Theta)\right]dtd\Theta. \qquad (3.24)$$

For discretely time t samples let us substitute the integrals by the relevant sums. Then

$$a_{ik}=\sum_{l=-p}^{p}\eta(l)\cdot\sum_{j=1}^{M}\left[x_i(j+l)\cdot x_k(j)+x_i(j)\cdot x_k(j+l)\right], \qquad (3.25)$$

$$b_k=\sum_{l=-p}^{p}\eta(l)\cdot\sum_{j=1}^{M}\left[y(j+l)\cdot x_k(j)+y(j)\cdot x_k(j+l)\right]. \qquad (3.26)$$

The system's solution (3.22) gives the sought estimate $\hat{\beta}$:

$$\hat{\beta}=A^{-1}\cdot B \qquad (3.27)$$

3.3.2. The Analysis of the Method Components

Let us analyze how this estimate differs from the true β^* (3.2). For this, firstly we consider one from the components of the sum in I, for example, if $I=-m$. There is a shift $\tau=-m\Delta t$, where Δt - is a step of discrete measurements X, Y at time t. Then, without considering the weight $\eta(m)$, where m – is the component of the discrete model

$$I=\frac{1}{2}\sum_{l=-1}^{p}\eta(l)\sum_{j=1}^{M}\varepsilon(j)\varepsilon(j+l) \tag{3.28}$$

of the integral – correlation criterion (3.20) will be equal to

$$\begin{aligned} I_m &= \frac{1}{2}\varepsilon^T\varepsilon_{-\tau}=\frac{1}{2}\left(Y_0-X_0\hat{\beta}\right)^T\left(Y_{-\tau}-X_{-\tau}\hat{\beta}\right)= \\ &=\frac{1}{2}Y_0^TY_{-\tau}-\frac{1}{2}\hat{\beta}^T\left(X_0^TY_{-\tau}+X_{-\tau}^TY_0\right)+\frac{1}{2}\hat{\beta}^TX_0^TX_{-\tau}\hat{\beta}, \end{aligned} \tag{3.29}$$

where the index τ is the shift $x_i(t)$, $y_i(t)$ at the $\tau=m\Delta t$:

$$X_{-\tau}=\begin{bmatrix} x_1(1) & \dots & x_n(1) \\ x_1(2) & \dots & x_n(2) \\ & \dots & \\ x_1(M-m) & \dots & x_n(M-m) \end{bmatrix},\quad Y_{-\tau}=\begin{bmatrix} y(1) \\ y(2) \\ \dots \\ y(M-m) \end{bmatrix},$$

$$X_0=\begin{bmatrix} x_1(m+1) & \dots & x_n(m+1) \\ x_1(m+2) & \dots & x_n(m+2) \\ & \dots & \\ x_1(M) & \dots & x_n(M) \end{bmatrix},\qquad Y_0=\begin{bmatrix} y(m+1) \\ y(m+2) \\ \dots \\ y(M) \end{bmatrix}.$$

If the shift τ in (3.29) is larger than correlation time of the noises N and $\tilde{\varepsilon}$, that the mathematical expectation I_m will be invariant to those noises.

$$M\{I_m\} = \frac{1}{2}\varepsilon_0^{*T}\varepsilon_{-\tau}^{*}.$$

The necessary of the minimum I_m:

$$\frac{\partial I_m}{\partial \hat{\beta}^T} = -\frac{1}{2}(X_0^T Y_{-\tau} Y_0) + \frac{1}{2}(X_0^T X_{-\tau} + X_{-\tau}^T X_0)\hat{\beta}_2 = 0. \qquad (3.30)$$

The sufficient condition:

$$\frac{1}{2}\det\left(X_0^T X_{-\tau} + X_{-\tau}^T X_0\right) > 0. \qquad (3.31)$$

Contrary to LSM the linear independence of the functions $x_i(k)$ is insufficient for the positive definiteness of the matrix $\left(X_0^T X_{-\tau} + X_{-\tau}^T X_0\right)$. Let us determine the upper border τ_{bd} for the shift τ, for which the inequality (3.31) becomes the equality, viz the system becomes degenerate.

So far as the noise N only improves the conditionality of the matrix $X^T X$, we will take $N=0$, $X = X^*$ while developing τ_{bd}. Let us represent $X^*_{-\tau}$ by Taylor series

$$X_{-\tau}^{*} = X_0^{*} - \tau\dot{X}_0^{*} + R_1(\ddot{X}_0^{*}), \qquad (3.32)$$

where $\dot{X}_0^{*} = \frac{dX_0^{*}}{dt}$, $R_1(\ddot{X}_0^{*})$ - the matrix of remainders in the decomposition (3.32). With an accuracy to $R_1(\ddot{X}_0^{*})$ we have:

$$\begin{aligned}&\frac{1}{2}(X_0^{*T}X_{-\tau}^{*}) + X_{-\tau}^{*T}X_0^{*} \doteq \frac{1}{2}\left[X_0^{*T}(X_0^{*} - \tau\dot{X}_0^{*}) + (X_0^{*} - \tau\dot{X}_0^{*})^T X_0^{*}\right] = \\ &= X_0^{*T}X_0^{*} - \frac{\tau}{2}\left(X_0^{*T}\dot{X}_0^{*} + \dot{X}_0^{*T}X_0^{*}\right) = \Gamma - \frac{\tau}{2}\Gamma_1,\end{aligned} \qquad (3.33)$$

where Γ, Γ_1 - are symmetric matrices. Γ is the Gram matrix, that is why it is positively identified.

It is known [11] that two symmetric matrices, one of which is positive and defined by transformation T, can be reduced to the diagonal form so that there are only unity elements on the diagonal matrix Γ, there are eigenvalues λ_i of the matrix $\Gamma^{-1}\Gamma_1$ (for 2.33 -$\frac{\tau}{2}\Gamma^{-1}\Gamma_1$) on the diagonal Γ_1:

$$T^{-1}\left(\Gamma-\frac{\tau}{2}\Gamma_1\right)T=\begin{bmatrix}1\cdots 0\\ \ddots\\ 0\cdots 1\end{bmatrix}-\frac{\tau}{2}\begin{bmatrix}\lambda_{11}\cdots 0\\ \ddots\\ 0\cdots\lambda_{1n}\end{bmatrix}. \qquad (3.34)$$

The border value τ_{bd} of the shift is defined under condition

$$1-\frac{\tau_p}{2}\max_i\lambda_{1i}=0, \quad \text{or} \quad \tau_{bd}=\frac{2}{\max_i\lambda_{1i}}, \qquad (3.35)$$

where λ_{1i} - are eigenvalues of the matrix $\frac{\tau}{2}\Gamma^{-1}\Gamma_1$. It is clear that the larger τ is, the worse conditioned is the matrix Γ (the larger $\left\|\Gamma^{-1}\right\|$), the larger $\max_i\lambda_{1i}$ and, as a consequence, the less possible is the shift τ_{bd}.

3.3.3. The Definition of the Maximum Shift

Let us consider how symmetry of the displacement $\pm\tau$ will affect on the border value τ_{bd}. Here

$$\begin{aligned}\boldsymbol{I}&=\varepsilon_0^T\left(\varepsilon_\tau+\varepsilon_{-\tau}\right)=\left(Y_0-X_0\hat{\beta}\right)^T\left(Y_\tau+Y_{-\tau}-\left(X_\tau+X_{-\tau}\right)\hat{\beta}\right)=\\ &=Y_0^T\left(Y_\tau+Y_{-\tau}\right)-\hat{\beta}^T\left[X_0^T\left(Y_\tau+Y_{-\tau}\right)+\left(X_\tau+X_{-\tau}\right)^T Y_0\right]+\\ &+\hat{\beta}^T X_0^T\left(X_\tau+X_{-\tau}\right)\beta^T,\end{aligned} \qquad (3.36)$$

$$X_0 = \begin{bmatrix} x_1(m+1) & \dots & x_n(m+1) \\ x_1(m+2) & \dots & x_n(m+2) \\ \hdashline x_1(M-m) & \dots & x_n(M-m) \end{bmatrix}, \qquad Y_0 = \begin{bmatrix} y(m+1) \\ y(m+2) \\ \vdots \\ y(M-m) \end{bmatrix},$$

$$X_\tau = \begin{bmatrix} x_1(2m+1) & \dots & x_n(2m+1) \\ x_1(2m+2) & \dots & x_n(2m+2) \\ \hdashline x_1(M) & \dots & x_n(M) \end{bmatrix}, \qquad Y_\tau = \begin{bmatrix} y(2m+1) \\ y(2m+2) \\ \vdots \\ y(M) \end{bmatrix},$$

$$X_{-\tau} = \begin{bmatrix} x_1(1) & \dots & x_n(1) \\ x_1(2) & \dots & x_n(2) \\ \hdashline x_1(M-2m) & \dots & x_n(M-2m) \end{bmatrix}, Y_{-\tau} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(M-2m) \end{bmatrix}.$$

The necessary condition of the minimum I:

$$\frac{\partial \boldsymbol{I}}{\partial \hat{\beta}^{\boldsymbol{T}}} = \varepsilon_{\mathbf{0}}^{\boldsymbol{T}} \left(\varepsilon_\tau + \varepsilon_{-\tau}\right) = \left[X_0^T \left(Y_\tau + Y_{-\tau}\right) + \left(X_\tau + X_{-\tau}\right)^T Y_0 \right] + \\ + \left[X_0^T \left(X_\tau + X_{-\tau}\right) + \left(X_\tau + X_{-\tau}\right)^T X_0 \right] \hat{\beta} = 0, \tag{3.37}$$

The sufficient condition:

$$\det \frac{\partial^2 \boldsymbol{I}}{\partial \hat{\beta}^{\boldsymbol{T}} \partial \hat{\beta}} = \left(X_0^T \left(X_\tau + X_{-\tau}\right) + \left(X_\tau + X_{-\tau}\right)^T X_0 \right) > 0. \tag{3.38}$$

Analogously to the derivation (3.35) we find the border value τ_{bd}. from the condition that (3.18) is zero. Let us consider three terms of series (3.32):

$$X_\tau^* = X_0^* + \tau \dot{X}_0^* + \frac{\tau^2}{2} \dddot{X}_0^* + R_3\left(\dddot{X}^*, \tau\right), \\ X_{-\tau}^* = X_0^* - \tau \dot{X}_0^* + \frac{\tau^2}{2} \ddot{X}_0^* + R_3\left(\dddot{X}^*, -\tau\right). \tag{3.39}$$

With an accuracy to $R_3\left(\dddot{X}^*, \pm\tau\right)$, we obtain

$$\frac{1}{4}X_0^{*T}\left[\left(X_\tau^*+X_{-\tau}^*\right)+\left(X_\tau^*+X_{-\tau}^*\right)^T X_0^*\right]$$

$$\approx X_0^{*T}X_0^*+\frac{\tau^2}{4}\left[X_0^{*T}\ddot{X}_0^*+\ddot{X}_0^{*T}X_0^*\right]= \qquad (3.40)$$

$$=\Gamma+\frac{\tau^2}{4}\Gamma_{2,}$$

where Γ and Γ_2 - are symmetric matrices, Γ is positively identified Gram matrix. Then, as (3.34), the expression (3.40) is given by the transformation T to the form

$$\mathrm{T}^{-1}(\Gamma+\frac{\tau^2}{4}\Gamma_2)\mathrm{T}=\begin{bmatrix}1 & 0\\ 0 & \ddots 1\end{bmatrix}+\frac{\tau^2}{4}\begin{bmatrix}\lambda_{21} & 0\\ 0 & \ddots \lambda_{2n}\end{bmatrix}, \qquad (3.41)$$

where λ_{2i} - are eigenvalues of the matrix $\frac{\tau^2}{4}\Gamma^{-1}\Gamma_2 \quad (i=\overline{1,n}).$

The border value τ_{bd} of the displacement τ will be defined under condition

$$1+\frac{\tau_{bd}^2}{4}\max_i\left|\lambda_{2i}\right|=0$$

or

$$\tau_{bd}=\frac{2}{\sqrt{\max_i\left|\lambda_{2i}\right|}}, \qquad (3.42)$$

where, contrary to (3.35) for the asymmetric displacement, $\max_i\left|\lambda_{2i}\right|$ is taken over the set $\left\{\lambda_{2i}\right\}$ of the negative eigenvalues of the matrix $\frac{\tau^2}{4}\Gamma^{-1}\Gamma_2$. Continuing the decomposition (3.39) and substituting it into (3.38), we see that all odd terms of the decomposition (3.38) are cancelled out.

Thus, for the precise X_0^*, the matrix $X_0^{*T}\left(X_\tau^*+X_{-\tau}^*\right)+\left(X_\tau^*+X_{-\tau}^*\right)^T X_0^*$ differs from the matrix $X_0^{*T}X_0^*$ of LSM by the pairing terms of the decomposition (3.39). Considering that $x_0(t)$ is smooth, this difference is significantly less than in (3.1) for the

precise X^*, so far as all components of the decomposition are present there (3.39). Thus, the border value (3.42) is larger than (3.35), and at the same τ the index (3.36) at inaccurate $X = X^* + N$ is closer to the LSM index (at inaccurate X^*) than the index (3.29). To confirm this, let us consider the displacement and covariance of the estimate $\hat{\beta}$:

$$\hat{\beta} = \left[X_0^T \left(X_\tau + X_{-\tau} \right) + \left(X_\tau + X_{-\tau} \right)^T X_0 \right]^{-1} \cdot \\ \cdot \left[X_0^T \left(Y_\tau + Y_{-\tau} \right) + \left(X_\tau + X_{-\tau} \right)^T Y_0 \right]. \tag{3.43}$$

3.3.4. The Displacement Estimation of the Vector's $\hat{\beta}$ Estimate

The displacement $\Delta\hat{\beta}$**:**

$$\Delta\hat{\beta} = M\left\{ \hat{\beta} - \hat{\beta}^* \right\} = M\left\{ \left[X_0^T (X_\tau + X_{-\tau}) + (X_\tau + X_{-\tau})^{-1} X_0 \right] \cdot \right. \\ \left. \cdot \left[X_0^T (Y_\tau + Y_{-\tau}) + (X_\tau + X_{-\tau})^T Y_0 \right] - (X_0^{*T} X_0^*)^{-1} X_0^{*T} Y_0^* \right\} \tag{3.44}$$

or

$$\Delta\hat{\beta} = \mathrm{M}\{ \left[X_0^T (X_\tau + X_{-\tau}) + (X_\tau + X_{-\tau})^T X_0 \right]^{-1} \times \\ \times \mathrm{M}\left\{ X_0^T (Y_\tau + Y_{-\tau}) + (X_\tau + X_{-\tau})^T Y_0 \right\} - (X_0^{*T} X_0^*)^{-1} X_0^{*T} Y_0^* = \\ = [X_0^{*T} (X_\tau^* + X_{-\tau}^*) + (X_\tau^* + X_{-\tau}^*)^T + \\ + \mathrm{M}\left\{ N_0^T (N_\tau + N_{-\tau}) + (N_\tau + N_{-\tau})^T N_0 \right\}]^{-1} \cdot \\ \cdot [X_0^T (Y_\tau^* + Y_{-\tau}^*) + (X_\tau^* + X_{-\tau}^*)^T Y_0 + \\ + \mathrm{M}\left\{ X_0^{*T} (\varepsilon_\tau + \varepsilon_{-\tau}) + (X_\tau^* + X_{-\tau}^*)^T \varepsilon_0 \right\}] - (X_0^{*T} X_0^*)^{-1} X_0^* Y_0^*. \tag{3.45}$$

Using the decomposition(3.39) and the formula (3.40), we obtain

$$\Delta\hat{\beta} = [(4\Gamma + \tau^2 \Gamma_2) + 4(\mathrm{M} - 2m) R_{NN}(\tau)]^{-1} \cdot \\ \cdot [4 X_0^{*T} Y_0^* + \tau^2 (X_0^{*T} \ddot{Y}_0^* + \ddot{X}_0^{*T} Y_0^*) + 2(\mathrm{M} - 2m) \cdot \\ \cdot (R_{X^*\varepsilon^*}(\tau))] - \Gamma^{-1} X_0^{*T} Y_0^*, \tag{3.46}$$

where $R_{NN}(\tau)$ - is the matrix ($n \times m$) of correlation functions $R_{N_i N_j}(\tau)$ $i, j = \overline{1,n}$;

$R_{X^*\varepsilon^*}(\pm\tau)$ - is the column vector ($n \times 1$) of correlation functions $R_{X_i^* \varepsilon_j^*}(\pm\tau)$, $i, j = \overline{1,n}$.

It can be shown that the shift $\Delta\hat{\beta}$ is caused by members $(\mathrm{M} - 2m)R_{NN}(\tau)$ and $(\mathrm{M} - 2m) \cdot (R_{X^*\varepsilon^*}(\tau) + R_{X^*\varepsilon^*}(-\tau))$. The first one is equal to zero if there is τ longer time of noises correlation. The second one can be represented as:

$$(\mathrm{M} - 2m)[R_{X^*\varepsilon^*}(\tau)] \approx \frac{\tau^2}{2}(X_0^{*T}\ddot{\varepsilon}_0^* + \ddot{X}_0^{*T}\varepsilon_0^*), \qquad (3.47)$$

Then $\Delta\hat{\beta} = (2\Gamma + \frac{\tau^2}{2}\Gamma_2)^{-1}(\mathrm{M} - 2m)(R_{X^*\varepsilon^*}(\tau) + (R_{X^*\varepsilon^*}(-\tau)). \qquad (3.48)$

From here we can see, that the displacement $\Delta\hat{\beta}$ is reduced if

$$\left(\frac{\mathrm{M} - 2m}{\mathrm{M} - m}\right)\left\|\left(2\Gamma + \frac{\tau^2}{2}\Gamma_2\right)^{-1}\right\| < \left\|(2\Gamma - \tau\Gamma_1)^{-1}\right\|. \qquad (3.49)$$

3.3.5. The Variance Estimation of the Vector's $\hat{\beta}$ Estimate

The covariance of the estimates $\hat{\beta}$

$$\mathrm{cov}\left[\hat{\beta}\right] = \mathrm{M}\left\{(\hat{\beta} - \mathrm{M}\{\hat{\beta}\})(\hat{\beta} - \mathrm{M}\{\hat{\beta}\})^T\right\}.$$

Let us represent $\hat{\beta}$ as

$$\begin{aligned}\hat{\beta} = {} & [X_0^{*T}(X_\tau^* + X_{-\tau}^*) + (X_\tau^* + X_{-\tau}^*)^T X_0^* + 4(\mathrm{M} - 2m)R_{NN}(\tau) + \\ & + \xi(X^*, N, \tau)]^{-1}[(X_0^{*T} + N_0^T)(Y_\tau^* + C + Y_{-\tau}^* + \varepsilon_{-\tau}) + \\ & + (X_\tau^* + N_\tau + X_{-\tau}^* + N_{-\tau})^T(Y_0^* + \varepsilon_0)].\end{aligned} \qquad (3.50)$$

For large M, ignoring the random part and random components $N_0^T(\varepsilon_\tau + \varepsilon_{-\tau})$, $(N_\tau + N_{-\tau})\varepsilon_0$ of the second order of smallness, denoting

$$\frac{1}{4}C_1' = [X_0^{*T}(X_\tau^* + X_{-\tau}^*) + (X_\tau^* + X_{-\tau}^*)^T X_0^* + 4(\mathrm{M} - 2m)R_{NN}(\tau)]^{-1} X_0^{*T},$$

$$\frac{1}{4}C_1'' = [X_0^{*T}(X_\tau^* + X_{-\tau}^*) + (X_\tau^* + X_{-\tau}^*)^T X_0^* + 4(\mathrm{M} - 2m)R_{NN}(\tau)]^{-1} X_{-\tau}^{*T},$$

$$\frac{1}{4}C_1''' = [X_0^{*T}(X_\tau^* + X_{-\tau}^*) + (X_\tau^* + X_{-\tau}^*)^T X_0^* + 4(\mathrm{M} - 2m)R_{NN}(\tau)]^{-1} X_\tau^{*T},$$

$$\frac{1}{4}C_2' = [X_0^{*T}(X_\tau^* + X_{-\tau}^*) + (X_\tau^* + X_{-\tau}^*)^T X_0^* + 4(\mathrm{M} - 2m)R_{NN}(\tau)]^{-1} Y_0^*,$$

$$\frac{1}{4}C_2'' = [X_0^{*T}(X_\tau^* + X_{-\tau}^*) + (X_\tau^* + X_{-\tau}^*)^T X_0^* + 4(\mathrm{M} - 2m)R_{NN}(\tau)]^{-1} Y_{-\tau}^*,$$

$$\frac{1}{4}C_2''' = [X_0^{*T}(X_\tau^* + X_{-\tau}^*) + (X_\tau^* + X_{-\tau}^*)^T X_0^* + 4(\mathrm{M} - 2m)R_{NN}(\tau)]^{-1} Y_\tau^*.$$

we obtain:

$$\begin{aligned}
&\mathrm{cov}\left[\hat{\beta}\right] \equiv \mathrm{M}\{[\frac{1}{4}C_1'(\varepsilon_\tau + \varepsilon_{-\tau}) + \frac{1}{4}\left(C_1'' + C_1'''\right)\varepsilon_0 + \frac{1}{4}C_2'(N_\tau + N_{-\tau}) + \\
&+\frac{1}{4}\left(C_2'' + C_2'''\right)N_0][\frac{1}{4}C_1'(\varepsilon_\tau + \varepsilon_{-\tau}) + \frac{1}{4}\left(C_1'' + C_1'''\right)\varepsilon_0 + \frac{1}{4}C_2'(N_\tau + N_{-\tau}) + \\
&+\frac{1}{4}\left(C_2'' + C_2'''\right)N_0]^T\} = \frac{1}{16}C_1'\mathrm{M}\{(\varepsilon_\tau + \varepsilon_{-\tau})(\varepsilon_\tau + \varepsilon_{-\tau})^T\}C_1'^T + \\
&+\frac{1}{16}C_1'\mathrm{M}\{(\varepsilon_\tau + \varepsilon_{-\tau})\varepsilon_0^T\}\left(C_1'' + C_1'''\right)^T + \frac{1}{16}\left(C_1'' + C_1'''\right)\mathrm{M}\{\varepsilon_0(\varepsilon_\tau + \varepsilon_{-\tau})^T\}C_1'^T + \\
&+\frac{1}{16}\left(C_1'' + C_1'''\right)\mathrm{M}\{\varepsilon_0\varepsilon_{\tau 0}^T\}\left(C_1'' + C_1'''\right)^T + \frac{1}{16}C_1'\mathrm{M}\{(N_\tau + N_{-\tau})(N_\tau + N_{-\tau})^T\}C_1'^T \\
&+\frac{1}{16}C_1'\mathrm{M}\{(N_\tau + N_{-\tau})N_0^T\}\left(C_1'' + C_1'''\right)^T + \frac{1}{16}\left(C_1'' + C_1'''\right)\mathrm{M}\{N_0(N_\tau + N_{-\tau})^T\}C_1'^T + \\
&+\frac{1}{16}\left(C_1'' + C_1'''\right)\mathrm{M}\{N_0N_{\tau 0}^T\}\left(C_1'' + C_1'''\right)^T. \qquad (3.51)
\end{aligned}$$

or

$$cov\left[\hat{\beta}\right] = C_5'[\frac{1}{16}\mathrm{M}\{\left(\varepsilon_\tau + \varepsilon_{-\tau}\right)\left(\varepsilon_\tau + \varepsilon_{-\tau}\right)^T\} +$$
$$+\frac{1}{8}\mathrm{M}\{\left(\varepsilon_\tau + \varepsilon_{-\tau}\right)\varepsilon_0^T\} + \frac{1}{8}\mathrm{M}\{\varepsilon_0\left(\varepsilon_\tau + \varepsilon_{-\tau}\right)^T\} +$$
$$+\frac{1}{4}\mathrm{M}\{\varepsilon_0\varepsilon_{\tau 0}^T\}]C_5^T + C_6[\frac{1}{16}\mathrm{M}\{\left(N_\tau + N_{-\tau}\right)\left(N_\tau + N_{-\tau}\right)^T\} + \qquad (3.52)$$
$$+\frac{1}{8}\mathrm{M}\{\left(N_\tau + N_{-\tau}\right)N_0^T\} + \frac{1}{8}\mathrm{M}\{N_0\left(N_\tau + N_{-\tau}\right)^T\} +$$
$$+\frac{1}{4}\mathrm{M}\{N_0 N_{\tau 0}^T\}]C_6^T = C_5 Q_3 C_5^T + C_6 F_3 C_6^T,$$

where $C_1' \approx C_1'' \approx C_1''' = C_5$, $\quad C_2' \approx C_2'' \approx C_2''' = C_6$.

Considering the correlation (3.49), it can be expected that at the $|\tau| > 0$ $\|C_3\| > \|C_5\| > \|C_1\|$, $\|C_4\| > \|C_6\| > \|C_2\|$ while $\|Q_3\| = \|Q_2\|$, $\|F_3\| = \|F_2\|$

due to the fact that in (3.52) there is the correlation of diagonal and nilpotent matrices. The displacement and the covariance of the estimates $\hat{\beta}$ at $\pm\tau$ will be less than for -τ if τ are alike. That is to say, the symmetrical displacement moves the criterion $\varepsilon_0^T(\varepsilon_\tau + \varepsilon_{-\tau})$ closer to the ideal $\varepsilon_0{}^{*T}\varepsilon_0{}^*$.

Thus, each m component of the integral-correlation criterion (3.28) within the displacement $\pm m\Delta t = \pm\tau$, which is smaller in magnitude than τ_{bd} (3.42), in its minimization gives the estimate $\hat{\beta}$ (3.43), that is close to the true β^*, if the displacement is greater than the noises' correlation time T_{cor} and less than the border τ_{bd}.

Then it is simpler to define the weight function $\eta(\Theta)$ from the condition

$$\eta(\Theta) = \begin{cases} 0, T_{cor} \geq |\Theta| \geq \tau_{bd} \\ 1, T_{cor} < |\Theta| < \tau_{bd} \end{cases}. \qquad (3.53)$$

If noises N_x are the "white noise", at the displacement in one step the correlation is absent, the errors of the partial estimates $\hat{\beta}$ are statistically independent for *I* component of

the criterion (3.28). Then, according to the law of statistics [19], the average value $\hat{\beta}$ in p partial estimates (the estimate at zero displacement is rejected), will have at $\sqrt{p}$ less variation and zero displacement.

.3.6. The Recursive Form of Calculations

The recursive LSM at the normal distribution of prior estimates β_a and noises ε leads to Bayesian estimation [4]. But the presence of noises N_x in X brings a shift into the estimates $\hat{\beta}$. Therefore, we consider the recursive form of the calculation for M measurements by the step Δt, i.e. $t_k = k\Delta t, \quad k = \overline{1,\mathrm{M}}$:

$$\boldsymbol{I} = \sum_{k=1}^{\mathrm{M}} \left[y(k) - \sum_{i=1}^{n} \hat{\beta}_{i2} x_i(k) \right] \left[y(k+m) - \sum_{i=1}^{n} \hat{\beta}_{i2} x_i(k+m) \right]. \tag{3.54}$$

From the condition (3.30) we obtain the system:

$$\begin{aligned} &\sum_{k=1}^{\mathrm{M}} \left[y(k) x_j(k+m) + y(k+m) x_j(k) \right] = \\ &= \sum_{i=1}^{n} \hat{\beta}_i \sum_{k=1}^{\mathrm{M}} \left[x_i(k) x_j(k+m) + x_j(k+m) x_j(k) \right], \ j = \overline{1,n}, \end{aligned} \tag{3.55}$$

or in a vector form

$$\left[X_0^T X_m + X_m^T X_0 \right] \hat{\beta} = X^T Y_m + X_m^T Y_0. \tag{3.56}$$

let us denote

$$P_2 = \left[X_0^T X_m + X_m^T X_0 \right]^{-1}, \tag{3.57}$$

then

$$\hat{\beta}_2 = P_2 \left(X^T Y_m + X_m^T Y_0 \right). \tag{3.58}$$

Let us set (3.57) at the bloc type for the (k+1) measurement

$$P_{k+1} = \left[X_{0k}^T X_{mk} + X_{mk}^T X_{0k} \right]^{-1} =$$

$$= \left[\begin{bmatrix} X_{0(k-1)} \\ x_{0k}^T \end{bmatrix}^T \cdot \begin{bmatrix} X_{m(k-1)} \\ x_{mk}^T \end{bmatrix} + \begin{bmatrix} X_{m(k-1)} \\ x_{mk}^T \end{bmatrix}^T \begin{bmatrix} X_{0(k-1)} \\ x_{0k}^T \end{bmatrix} \right]^{-1} = \tag{3.59}$$

$$= \left[\left[X_{0(k-1)}^T X_{m(k-1)} + X_{m(k-1)}^T X_{0(k-1)} \right] + \left[x_{0k} x_{mk}^T + x_{mk} x_{0k}^T \right] \right]^{-1} =$$

$$= \left[P_k^{-1} + \left[x_{0k} x_{mk}^T + x_{mk} x_{0k}^T \right] \right]^{-1},$$

where $x_{0k}^T = \left[x_1(k), x_2(k), \cdots, x_n(k) \right],$

$x_{mk}^T = \left[x_1(k+m), x_2(k+m), \cdots, x_n(k+m) \right].$

From (3.59) we find

$$P_{k+1}^{-1} - P_k^{-1} = x_{0k} x_{mk}^T + x_{mk} x_{0k}^T. \tag{3.60}$$

Then, we obtain $\hat{\beta}$ for the k measurement

$$\hat{\beta}(k+1) = P_{k+1}(X_{0k}^T Y_{mk} + X_{mk}^T Y_{0k}) =$$

$$= P_{k+1} \left[\left[X_{0(k-1)}^T, x_{0k} \right] \begin{bmatrix} Y_{m(k-1)} \\ y(k+m) \end{bmatrix} + \left[X_{m(k-1)}^T, x_{mk} \right] \begin{bmatrix} Y_{0(k-1)} \\ y(k) \end{bmatrix} \right] = \tag{3.61}$$

$$= P_{k+1} \left[P_k^{-1} \hat{\beta}(k) + x_{0k} y(k+m) + x_{mk} y(k) \right] =$$

$$= \hat{\beta}(k) + P_{k+1} \left[x_{0k} y(k+m) + x_{mk} y(k) + (P_k^{-1} - P_{k+1}^{-1}) \hat{\beta}(k) \right].$$

Considering (3.60), we obtain from (3.61):

$$\hat{\beta}(k+1) = \hat{\beta}(k) + P_{k+1}[x_{0k} y(k+m) + x_{mk} y(k) -$$

$$-(x_{0k} x_{mk}^T + x_{mk} x_{0k}^T) \hat{\beta}_2(k)] = \tag{3.62}$$

$$= \hat{\beta}(k) + P_{k+1} \left[x_{0k} (y(k+m) - x_{mk}^T \hat{\beta}(k)) + x_{mk} (y(k) - x_{0k}^T \hat{\beta}(k) \right] =$$

$$= \hat{\beta}(k) + P_{k+1} \left[x_{0k} \varepsilon(k+m) + x_{mk} \varepsilon(k) \right],$$

where P_{k+1} – is defined with respect to (3.59).

For the criterion (3.36) from the minimum condition we obtain the system

$$\sum_{k=1}^{M}\Big[(y(k+m)+y(k-m)x_j(k)+y(k)(x_j(k+m)+x_j(k-m))\Big]= \\ =\sum_{i=1}^{n}\hat{\beta}_i\Big[(x_i(k+m)+x_i(k-m))x_j(k)+x_i(k)(x_j(k+m)+x_j(k-m))\Big] \tag{3.63}$$

or in a vector form:

$$\Big[X_0^T Z+Z^T X_0\Big]\hat{\beta}_3 = X_0^T(Y_m+Y_{-m})+Z^T Y_0, \tag{3.64}$$

where $Z = X_m + X_{-m}$.

Let us denote

$$\Big[X_0^T Z+Z^T X_0\Big]^{-1} = P_3. \tag{3.65}$$

Then

$$\hat{\beta} = P_3\Big[X_0^T(Y_m+Y_{-m})+Z^T Y_0\Big]. \tag{3.66}$$

Analogously to (3.59) ÷ (3.61) we obtain

$$P_{k+1} =\Big[X_{0k}^T Z_k + Z_k{}^T X_{0k}\Big]^{-1} = \\ =\left[\begin{bmatrix}X_{0(k-1)}\\ x_{0k}^T\end{bmatrix}^T\begin{bmatrix}Z_{k-1}\\ x_{mk}^T+x_{-mk}^T\end{bmatrix}+\begin{bmatrix}Z_{k-1}\\ x_{mk}^T+x_{-mk}^T\end{bmatrix}^T\begin{bmatrix}X_{0(k-1)}\\ x_{0k}^T\end{bmatrix}\right]^{-1} = \\ =[P_k^{-1}+x_{0k}(x_{mk}+x_{-mk})^T+(x_{mk}+x_{-mk})x_{0k}^T]^{-1}; \tag{3.67}$$

$$P_{k+1}^{-1}-P_k^{-1} = x_{0k}(x_{mk}+x_{-mk})^T+(x_{mk}+x_{-mk})x_{0k}^T;$$

$$\hat{\beta}(k+1) = P_{k+1}\Big[X_{0k}^T(Y_{mk}+Y_{-mk})+Z_k{}^T Y_{0k}\Big]= \\ = P_{k+1}\left[\Big[X_{0(k-1)}^T, x_{0k}\Big]\begin{bmatrix}Y_{m(k-1)}+Y_{-m(k-1)}\\ y(k+m)+y(k-m)\end{bmatrix}+\Big[Z_{k-1}^T, x_{mk}+x_{-mk}\Big]\begin{bmatrix}Y_{0(k-1)}\\ y(k)\end{bmatrix}\right]= \\ = P_{k+1}[P_k^{-1}\hat{\beta}(k)+x_{0k}(y(k+m)+y(k-m))+(x_{mk}+x_{-mk})y(k)]= \\ =\hat{\beta}(k)P_{k+1}[x_{0k}(y(k+m)+y(k-m))+(x_{mk}+x_{-mk})y(k)+ \\ +(P_k^{-1}+P_{k+1}^{-1})\hat{\beta}(k)]; \tag{3.68}$$

Let us denote

$$u(k) = y(k+m) + y(k-m),\ z_k = x_{mk} + x_{-mk}.$$

Then

$$\begin{aligned}
&\hat{\beta}(k+1) = \hat{\beta}(k) + P_{k+1}[x_{0k}u(k) + z_k y(k) - (x_{0k}z_k^T + z_k x_{0k}^T)\hat{\beta}(k)] = \\
&= \hat{\beta}(k) + P_{k+1}[x_{0k}(u(k) - z_k^T\hat{\beta}(k)) + z_k(y(k) - x_{0k}^T\hat{\beta}(k))] = \\
&= \hat{\beta}(k) + P_{k+1}[\,x_{0k}(\varepsilon(k+m) + \varepsilon(k-m)) + z_k\varepsilon(k)].
\end{aligned} \tag{3.69}$$

The estimation of the recursive integrated LSM is equal to the sum of estimates, averaged at $m = \overline{1,P}$. It is possible to use robust algorithms for calculating the mean. For example, it is Tukey's algorithm [12], where we take the distribution median of the values $\hat{\beta}_i$ at m, $i = \overline{1,n}$ instead of the mean. In order to do this the estimates $\hat{\beta}_i(m)$ are ranked by value, the lowest and highest values are dropped, the remainders are averaged out. The parameters $\eta(\Theta)$, P of the integrated LSM may be optimized by the external criterion [Nakhnenko] (stability, forecast accuracy and others).

3.4. The Method of the Auxiliary Variable (MAV)

3.4.1. The Method's Main Point

In the cases when the model accurately reflects the behavior of the examined object (process), that ε^* is insignificant, the noises N_x in the measurements X are significant and $x_i(t)$, ($i = \overline{1,n}$) is alternative, the method of the auxiliary variable $U_{i,}$ defined as $\operatorname{sign} x_i(t)$, can be relatively simple and accurate. In order to reduce partially the noise's N_X influence, the function $U_i(t)$ are defined as

$$U_i(t) = \begin{cases} \operatorname{sign} x_i(t),\ |x_i(t)| \geq \Delta i, \\ 0,\ |x_i(t) < \Delta i|. \end{cases} \tag{3.70}$$

Then, according to the method of the auxiliary variable there is the estimate $\hat{\beta}$ from the equation

$$U^T Y = U^T X\hat{\beta}\ , \tag{3.71}$$

which is equal to

$$\hat{\beta} = (U^T X)^{-1} U^T Y, \tag{3.72}$$

where

$$U^T X = \begin{bmatrix} \sum_{k_1=1}^{M_1} x_1(k_1)U_1(k_1), & \sum_{k_1=1}^{M_1} x_2(k_1)U_1(k_1), & \cdots & \sum_{k_1=1}^{M_1} x_n(k_1)U_1(k_1) \\ \sum_{k_2=1}^{M_2} x_1(k_2)U_2(k_2), & \sum_{k_2=1}^{M_2} x_2(k_2)U_2(k_2), & \cdots & \sum_{k_2=1}^{M_2} x_n(k_2)U_2(k_2) \\ \hdotsfor{4} \\ \sum_{k_n=1}^{M_2} x_1(k_n)U_n(k_n), & \sum_{k_n=1}^{M_2} x_2(k_n)U_n(k_n), & \cdots & \sum_{k_n=1}^{M_2} x_n(k_n)U_n(k_n) \end{bmatrix},$$

$$U^T Y = \begin{bmatrix} \sum_{k_1=1}^{M_1} y(k_1)U_1(k_1) \\ \sum_{k_2=1}^{M_2} y(k_2)U_2(k_2) \\ \hdots \\ \sum_{k_n=1}^{M_2} y(k_n)U_n(k_n) \end{bmatrix},$$

$\{k_i\}$ - is the discrete set, where the condition (3.70) is implemented.

3.4.2. The Shift of the Estimates (3.72)

If the noises N_x are mutually uncorrelated with X^*, Y^* and small, then approximately

$$U_i(t) \approx \operatorname{sign} x_i^*(t),$$

and the estimate $\hat{\beta}$ at $\varepsilon^* = 0$ remains constant, so far as

$$M\{U^T X\} = U^{*T} X, \qquad M\{U^T Y\} = U^{*T} Y.$$

The shift can take place if $\varepsilon \neq 0$

$$\Delta\hat{\beta} = (U^{*T} X^*)^{-1} U^{*T} \varepsilon^*, \tag{3.73}$$

So far as from the condition $X^{*T}\varepsilon^* = 0$ in a general case does not leak out that

$$U^{*T}\varepsilon^* = 0.$$

In total, the shift (3.73) is less than the shift of LSM estimates due to the noises N_x, if the influence $\varepsilon \neq 0$ on (3.73) is less than the influence N_x on the shift of LSM estimates.

3.4.3. The Covariance of the Estimates (3.72)

$$\operatorname{cov}\left[\hat{\beta}\right] = \mathrm{M}\{[\hat{\beta} - \mathrm{M}\{\hat{\beta}\}][\hat{\beta} - \mathrm{M}\{\hat{\beta}\}]^T\}.$$

Putting the uncorrelatedness of the random errors at $\left(U^{*T} X\right)^{-1}$ and $U^T Y$ and decomposing in a series

$$\hat{\beta} \approx \mathrm{M}\{\hat{\beta}\}[\delta(U^T X)^{-1}][U^{*T} Y^*] + (U^{*T} X^*)^{-1}[\delta U^T Y^* + U^{*T} N_y], \tag{2.74}$$

we obtain:

$$\begin{aligned}
&\operatorname{cov}[\beta] \approx \mathrm{M}\{[[\delta(U^T X)^{-1}][U^{*T} Y^*] + (U^{*T} X^*)^{-1}[\delta U^T Y^* + U^{*T} N_y]] \times \\
&\times [[\delta(U^T X)^{-1}][U^{*T} Y^*] + (U^{*T} X^*)^{-1}[\delta U^T Y^* + U*^T\ N_y]]^T\} = \\
&= \mathrm{M}\{[\delta(U^T X)^{-1}][U^{*T} Y^*] \cdot [[U^{*T} Y^*]^T [\delta(U^T X)^{-1}]^T\} + \\
&+ [U^{*T} X^*]^{-1} Y^{*T} \mathrm{M}\{\delta U \cdot \delta U^T\}]^T [\delta(U^T X)^{-1}]^T\} + \\
&+ [U^{*T} X^*]^{-1} Y^{*T} \mathrm{M}\{\delta U \cdot \delta U^T\}[(U^{*T} X^*)^T]^{-1} Y^* + \\
&+ [U^{*T} X^*]^{-1} U^{*T} \mathrm{M}\{N_y N_y^T\}[(U^{*T} X^*)^T]^{-1} U^*,
\end{aligned} \tag{3.75}$$

where the first and the second components depend on the thresholds Δ_i (3.70), noises N_x , the third one depends on the noises N_y. There is an optimal vector Δ^* of the thresholds Δ_i, $i=\overline{1,n}$, where the covariance (3.75) norm is minimum. If the noises N_x, N_y are uncorrelated and $\tilde{\varepsilon}^* \to 0$, then the estimate (3.72) will be unshifted. The estimation (3.72) effectiveness can be optimized by the external criterion by the vector's selection Δ.

The estimates (3.72) in MAV can be calculated by a recursive algorithm:

$$\begin{aligned} &\hat{\beta}(k+1) = \hat{\beta}(k) + P_{k+1}[U(k)\varepsilon(k)], \\ &P_{k+1} = [P_k^{-1} + U(k)x^T(k)]^{-1}, \\ &\varepsilon(k) = y(k) - x^T(k)\hat{\beta}(k), \end{aligned} \qquad (3.76)$$

where $U(k)$ can be not only relay one (3.70), but it can be any system of linearly independent, or even orthogonal functions. In order that the estimate (3.76) approximates to the optimal (LSM estimate for precise X^*, Y^*), it is desirable that the auxiliary variable is close to X^*. This will provide a positive definition of the matrix $U^T X = \hat{X}^T X$, but smoothing X makes the condition number of the matrix $\hat{X}^T X$ slightly worse. Thus, the algorithms of the data X quasidiagonalization are important; they will be considered next.

Chapter 4. The Increasing of Data Informativeness and, as a Consequence, of the Accuracy of the Estimates of Parameters of the Examined Objects

4.1. The Increasing of Informativeness of Data Samples in Terms of the Passive Experiment

If the precise values of X^*, Y^* are powerful not at all sites of the sample and have sites with linearly independent components $x_i(t)$, $i=\overline{1,n}$, then, with increasing length

M of such sample the reducing of covariance $\mathrm{cov}\hat{\beta}$, dictated by the noise N_y, may be irrelevant, and the shift, dictated by the noise N_x may even increase. Therefore, for sufficiently large numbers *M* it is appropriate to conduct the weighing of data in for their full or partial orthogonalization.

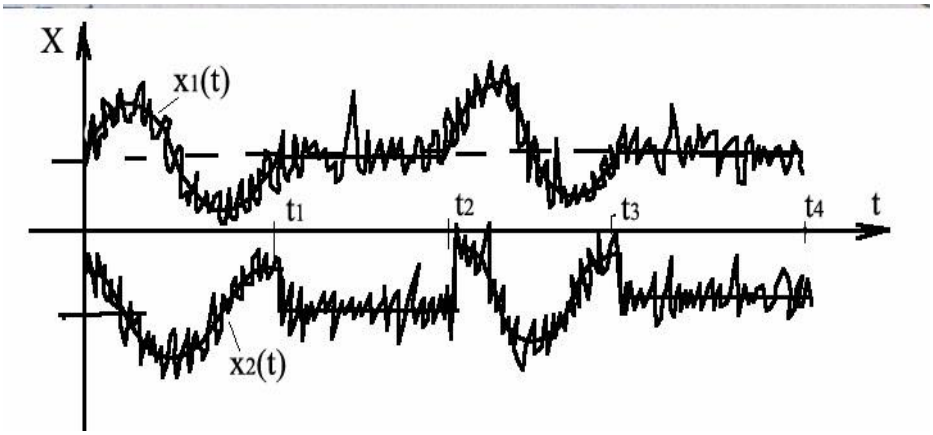


Fig. 4.1. The informative $([0,t_1],\ [t_2,t_3])$ and uninformative $([t_1,t_2],\ [t_3,t_4])$ sites of the sample $X^T = [x_1, x_2]$.

In the fig. 4.1 it is shown the simplest illustration of the necessity of that weighing of samples. If we introduce the weight function

$$\eta(t) = \begin{cases} 1, & t \in [0,t_1],\ [t_2,t_3] \\ 0, & t \in [t_1,t_2],\ [t_3,t_4] \end{cases},$$

the informative matrix $X^T X$ will be diagonal as a result of the orthogonality $x_1(t)$ and $x_2(t)$ at the sites where $\eta(t) = 1$. If you do quite the reverse, the $X^T X$ will be degenerate.

Let us construct some algorithms of the quasidiagonalization of the matrix $X^T X$.

The algorithm 1. At the recursive LSM we introduce the weight function that is adaptive to the current value *X*

$$\eta_j(k) = \frac{x_j^2(k)}{\sum_{i=1}^{n} \left| x_i(k) x_j(k) \right|}. \qquad (4.1)$$

Then, the measurements weight will increase for those *k*, where the correlation of the diagonal elements of the matrix $X^T X$ to the nondiagonal ones is greater. The weighted matrix will not be symmetrical, but it will be better determined:

$$\hat{\beta}(k+1) = \hat{\beta}(k) + P_{k+1} \cdot X_B(k)\varepsilon(k),\ P_{k+1} = [P_k^{-1} + X_B(k) \cdot X_B^T(k)]^{-1}, \qquad (4.2)$$

where $X_B^T(k) = [\eta_1(k) \cdot x_1(k), \cdots, \eta_n(k) \cdot x_n(k)]$.

The algorithm 2. You can obtain the more qualitative result, if you introduce the weight function $\eta_i(t)$ from the condition of the maximum accuracy of estimating the *i* parameter $\hat{\beta}_i$, $i = \overline{1,n}$. only. Thus, for LSM, GLSM, MAV the estimate $\hat{\beta}_i$ can be presented as a solution of the equation

$$B = A \cdot \hat{\beta}, \qquad (4.3)$$

which is the correlation of determinant Δ_i of the matrix A_i (the matrix *A*, where the first column is substitutes by the vector *B*) to determinant Δ of the matrix *A*:

$$\hat{\beta}_i = \frac{\Delta_i}{\Delta}. \qquad (4.4)$$

The differential of the expression (4.4) with respect to the accurate values (*) is equal to $d\hat{\beta}_i = \frac{d\Delta_i}{\Delta^*} - \frac{d\Delta \cdot \Delta_i^*}{\left(\Delta^*\right)^2}$, or

$$\frac{d\beta_i}{\hat{\beta}_i^*} = \frac{d\Delta_i}{\Delta_i^*} - \frac{d\Delta}{\Delta^*}. \qquad (4.5)$$

From here we obtain the expression of the relative variance

$$\left(\frac{\sigma_{\delta\beta i}}{\hat{\beta}_i^*}\right)^2 = \frac{\sigma_{\Delta_i}^2}{(\Delta_i^*)^2} + \frac{\sigma_{\Delta}^2}{(\Delta^*)^2} - \frac{2\sigma_{\Delta_i\Delta}^2}{\Delta_i^*\Delta^*}. \qquad (4.6)$$

In order to reduce the variance of the relative error $\frac{\delta\beta_i}{\hat{\beta}_i^*}$ in the estimate $\hat{\beta}_i$, it is necessary to maximize not only Δ, but Δ_i. From here we obtain the optimal weigh function $\eta_i^*(k)$:

$$\eta_i^*(k) = \arg \max_{\eta \in \{-1,1\}} |\Delta \cdot \Delta_i|, \; k = \overline{1,M}. \qquad (4.7)$$

In other words we add the variables ($\eta_i(k) = 1$) or subtract ($\eta_i(k) = -1$), trying to get execution (4.7). The multistep process of defining $\eta_i^*(k)$ may be constructed using the scheme:

the elements a_{ij}, b_i of the matrices A and B are calculated in (4.3), next Δ, Δ_i and the value $|\Delta \cdot \Delta_i|$ for $\eta_i^*(k) = 1$, $k = \overline{1,M}$;

for all rows $b_i = \sum_{j=1}^{n} a_{ij}\hat{\beta}_j$

of the equation (4.3) consistently on the every *k*-point all possible combinations are sorted from *n* to 2 (so far as $\eta_i \in \{-1;\ +1\}$) of the products $x_i(k) \cdot x_j(k)$ with a plus sign;

or with a minus sign $(\eta = -1)$ and something is left on the every *k*-point for which (4.7) is maximum, and we do the same for all $k = \overline{1,M}$;

- the process repeats *I* times until the values $|\Delta \cdot \Delta_i|$ difference on the *I* and (*I*-1) steps becomes insignificant with respect to $|\Delta \cdot \Delta_i|$.

Then, the estimate $\hat{\beta}_i$ is calculated. The process repeats for (i+1) estimation. The performance of this algorithm is not high; for it is necessary to count the value $|\Delta \cdot \Delta_i|$ $M/2^n$ times. If X^*, Y^* are smooth functions, the enumeration $\eta_i(k)$ can be carried out

at the intervals $\left[\left(k-m\right),\ \left(k+m\right)\right]$ from 2*m* points. The number of calculations $\left|\Delta\cdot\Delta_i\right|$ reduces to $\frac{\mathrm{M}}{2m}\cdot l\cdot 2^n$.

The algorithm 3. The previous algorithms of diagonalization of the matrix *A* of the system of standard equations do not provide the symmetry *A*. This algorithm involves the invariance of the symmetry of the matrix *A* at its diagonalization. Assume the system is formed as a result of usage LSM, GLSM or MAV

$$B = A\cdot\hat{\beta}, \tag{4.8}$$

$$A=[a_{ij}]_{i,j=1}^{n},\ a_{ij}=\sum_{k=1}^{\mathrm{M}}a_{ij}(k),$$

where $a_{ij}(k)$ - is determined depending on the method, for example, for LSM:

$$a_{ij}(k)=x_i(k)x_j(k);\ B=[b_i]_{i=1}^{n},\ b_i=\sum_{k=1}^{\mathrm{M}}b_i(k);\quad b_i(k)=y(k)\cdot x_i(k).$$

Let us introduce the weight function into the system (4.8)

$$\left[\sum_{k=1}^{\mathrm{M}}\eta_i(k)a_{ij}(k)\right]_{i,j=1}^{n}\cdot\hat{\beta}=\left[\sum_{k=1}^{\mathrm{M}}\eta_i(k)b_i(k)\right]_{i=1}^{n}. \tag{4.9}$$

The condition number (cond*A*) of the symmetric matrix *A* is equal to the ratio of the maximum value λ_A of the matrix *A* to a minimum [14]

$$\mathrm{cond}A=\frac{\max\lambda_A}{\min\lambda_A}. \tag{4.10}$$

For the diagonal matrix *A* cond*A* = 1, for degenerate one – the infinity. The inequality is known [25]:

$$\max\left|\lambda_i-\lambda_j\right|\le\sqrt{2\left|\mathrm{Sp}AA^T\right|-\frac{1}{n}\left(\mathrm{Sp}A\right)^2}, \tag{4.11}$$

where $\mathrm{Sp}A$ -is the trace of matrix *A*, $\left|\mathrm{Sp}AA^T\right|=\left\|A\right\|$ - is the Euclidean norm of the matrix *A*, $i,j=\overline{1,n};\ i\ne j$.

We see from the expression (4.11), that on minimizing the right part at the limit on the trace *A*: $\mathrm{Sp}A = C_n$ or on the every element a_{ij} of the trace:

$$\sum_{k=1}^{\mathrm{M}} \eta_i(k) a_{ii}(k) = C,\ i = \overline{1,n}, \tag{4.12}$$

if a_{ij} and a_{ji} are symmetrical"

$$\sum_{k=1}^{\mathrm{M}} \eta_i(k) a_{ij}(k) = \sum_{k=1}^{\mathrm{M}} \eta_j(k) a_{ji}(k),\ i=\overline{1,n},\ j = \overline{1,n},\ i \neq j, \tag{4.13}$$

it is possible to provide the minimum $\left|\lambda_i - \lambda_j\right|$ at the minimum value λ_i, λ_j, which is fixed due to (4.12), viz. it is possible to provide the minimum cond*A* (4.10). If you put *C*=1, then at the full diagonalization of the matrix *A*:

$$\hat{\beta}_i = \sum_{k=1}^{\mathrm{M}} \eta_i(k) b_i(k),\ i = \overline{1,n}, \tag{4.14}$$

where

$$\sum_{k=1}^{\mathrm{M}} \eta_i(k) a_{ij}(k) = 0,\ i = \overline{1,n}, j = \overline{1,n},\ i \neq j. \tag{4.15}$$

Thus, it is necessary to execute *n* conditions (4.12), $\frac{1}{2}(n^2 - n)$ – (4.13) or $\frac{1}{2}(n^2 - n)$ – (4.15), $\frac{1}{2}(n^2 + n)$ conditions in all, changing $M \cdot n$ of the values $\eta_i(k)$ ($i = \overline{1,n};\ k = \overline{1,\mathrm{M}}$).

So far as $M >> n$, the task can be solved. The solution can be simplified if we parameterize the functions $\eta_i(k)$

$$\eta_i(k) = \sum_{l=1}^{\mathrm{m}} \gamma_{il} \varphi_l(k),\ i = \overline{1,n},\ k = \overline{1,\mathrm{M}}, \tag{4.16}$$

where $\varphi_l(k)$ – are basic functions.

Then, $m \cdot n$ of coefficients γ_{il} are defined by $\frac{1}{2}(n^2+n)$ conditions (4.12), (4,13) or (4.1). Thus, putting C=1, from (4.12), (4.15) and (4.16) we obtain the system of the equations that are linear with respect to γ_{il}

$$\sum_{k=1}^{M}\left(\sum_{l=1}^{m}\gamma_{il}\varphi_l(k)\right)a_{ii}(k)=1,\ i=\overline{1,n};$$

$$\sum_{k=1}^{M}\left(\sum_{l=1}^{m}\gamma_{il}\varphi_l(k)\right)a_{ij}(k)=0,\ i=\overline{1,n-1};\ j=\overline{i+1,n};$$

or

$$\sum_{l=1}^{m}\gamma_{il}\left(\sum_{k=1}^{M}\varphi_l(k)\right)a_{ii}(k)=1,\ i=\overline{1,n};$$

$$\sum_{l=1}^{m}\gamma_{il}\left(\sum_{k=1}^{M}\varphi_l(k)\right)a_{ij}(k)=0,\ i=\overline{1,n-1};\ j=\overline{i+1,n}.$$

In the matrix form:

$$F\cdot\gamma=G \qquad (4.17)$$

where F– is the netting matrix $(m\cdot n)\times(m\cdot n)$; γ – is the vector γ_{il} of dimensionality $(m\cdot n)\times 1$; G – is the vector with single and zero elements.

The system (4.1) is solved by LSM. LSM-is the solution of the system (4.17):

$$\hat{\gamma}=\left(F^TF\right)^{-1}F^TG \qquad (4.18)$$

will satisfy the conditions accurately (4.12), (4.15), and the estimates $\hat{\beta}_i$ will be defined from (4.14). If m is an even number, you can assume the different numbers m_i for different i so that $\sum_{i=1}^{n}m_i=0{,}5(n^2+n).$ Then (4.18) will definitely satisfy (4.12), (4.15). The matrix F^TF is non-degenerate, if $x_i(k)$ and $\varphi_l(k)$ form the systems $\{x_i(k)\}$, $\{\varphi_l(k)\}$ of the linearly independent functions. Namely, the matrix A in the (4.8) should be non-degenerate. Any system of linearly independent functions, better orthogonal ones,

can be defined as $\{\varphi_l(k)\}$; Haar functions [11] and other which take values ±1, will be convenient for calculating.

4.2. The Improving of the Convergence

This property is characteristic for identification methods: the more uncertain is the problem, the easier are the solutions. This is LSM in the algorithms of the chapter 3, where $\hat{\beta}$ enters linearly; this is the gradient procedure [42] in the algorithms of nonlinear estimation at the uncertainty of the statistical characteristics of the surface $I(\beta)$:

$$\hat{\beta}(q+1) = \hat{\beta}(q) - \lambda(q)\frac{\partial \boldsymbol{I}(\hat{\beta}(q))}{\partial \hat{\beta}}. \qquad (4.19)$$

For the **convergence of the** algorithm (3.16) it is necessary that $I(\hat{\beta})$ is continuously **differentiable** by Frechet on $\hat{\beta}$:

$$\lim_{\|\Delta\beta\|\to 0}\left\|\frac{\partial \boldsymbol{I}(\hat{\beta}+\Delta\hat{\beta})}{\partial(\hat{\beta}+\Delta\hat{\beta})} - \frac{\partial \boldsymbol{I}(\hat{\beta})}{\partial \hat{\beta}}\right\| = 0, \qquad (4.20)$$

and Lipschitz condition is satisfied for the gradient from *I*:

$$\lim_{\|\Delta\beta\|\to 0}\left\|\frac{\partial \boldsymbol{I}(\hat{\beta}_1)}{\partial \hat{\beta}_1} - \frac{\partial \boldsymbol{I}(\hat{\beta}_2)}{\partial \hat{\beta}_2}\right\| \le L\left\|\hat{\beta}_1 - \hat{\beta}_2\right\|, \qquad (4.21)$$

where $L > 0$, $\|\cdot\|$ –is a norm in a Banach space [26].

Considering (4.29) we can write:

$$\boldsymbol{I}(\hat{\beta}+\Delta\hat{\beta}) - \boldsymbol{I}(\hat{\beta}) = \frac{\partial \boldsymbol{I}(\hat{\beta})}{\partial \hat{\beta}^T}\Delta\beta + \varepsilon\left\|\Delta\hat{\beta}\right\|, \qquad (4.22)$$

where

$$\lim_{\Delta\beta\to\mathbf{0}}\frac{\varepsilon\left\|\Delta\hat{\beta}\right\|}{\left|\Delta\hat{\beta}\right|} = 0,\ \Delta\hat{\beta} = \hat{\beta}(q) - \hat{\beta}^*.$$

On performing (4.21), (3.19) there are [11] such $\lambda(q)$ in which the algorithm (4.19) converges to a stationary point, that is equal to $\hat{\beta}^*$ for the strongly convex functional. Thus, for the strongly convex functional

$$\boldsymbol{I}(\hat{\beta}(q)) = \boldsymbol{I}_0 + \Delta\hat{\beta}^T \frac{\partial^2 \boldsymbol{I}}{\partial\hat{\beta}\partial\hat{\beta}^T}\Delta\hat{\beta},$$

$$\Delta\hat{\beta} = \hat{\beta}(q) - \hat{\beta}^*,$$

taking the derivative on $\hat{\beta}$:

$$\frac{\partial \boldsymbol{I}(\hat{\beta})}{\partial\hat{\beta}} = \frac{\partial^2 \boldsymbol{I}}{\partial\hat{\beta}\partial\hat{\beta}^T}\Delta\hat{\beta},$$

and substituting it in (4.19) we obtain

$$\Delta\hat{\beta}(q+1) = -\lambda(q)\frac{\partial^2 \boldsymbol{I}}{\partial\hat{\beta}\partial\hat{\beta}^T}\Delta\hat{\beta}(q), \qquad (4.23)$$

where $\frac{\partial^2 \boldsymbol{I}}{\partial\hat{\beta}\partial\hat{\beta}^T}$ – is the strongly positive [27] operator.

Then, there are $\lambda(q) = \mathrm{diag}\{\lambda_i(q)\}$ which are not identically zero, that the system (4.23) will be asymptotic stable [42]. In practice the gradient is inaccurately calculated due to noises.

The low rate of convergence in “ravine” situations is a serious weakness of the gradient algorithm. The modification of the “ravine method” is proposed for accelerating the convergence; it involves the application of *n* identical models with the initial values $\hat{\beta}^{(l)}(0)$, $(l = \overline{1,n})$ of *i*-th model parameters that are given in the *l*-th top of *n*-dimensional cube with the value $\hat{\beta}^*$ inside of it. Then, the index $\boldsymbol{I}_1^{(l)}$ with the additive accelerating the convergence can be constructed for each *l*-th model.

$$\boldsymbol{I}_2^{(l)} = \alpha_1(\Delta\hat{\beta}_H^{(l)})^T\Delta\hat{\beta}_H^{(l)}, \qquad (4.24)$$

where α_1 – is the algorithm parameter,

$\left(\Delta\hat{\beta}_H^{(l)}\right)^T = \left(\frac{\hat{\beta}_1^{(l)} - \overline{\beta}_1}{\overline{\beta}_1}, \cdots, \frac{\hat{\beta}_n^{(l)} - \overline{\beta}_n}{\overline{\beta}_n}\right)$, $\overline{\beta}_i = \frac{1}{n}\sum_{l=1}^{n}\hat{\beta}_i^{(l)}$ – i-th component for the l-th model:

$$\frac{\partial \boldsymbol{I}^{(l)}}{\partial \hat{\beta}_i^{(l)}} \approx \frac{\partial \boldsymbol{I}_1^{(l)}}{\partial \hat{\beta}_{ii}^{(l)}} + \frac{\hat{\beta}_i^{(l)} - \overline{\beta}_i}{\overline{\beta}_i}. \qquad (4.25)$$

Substituting (4.25) into (4.19), we obtain the gradient algorithms with accelerated convergence in "ravine" situations. There is the gradient setting of n models at the $\alpha_1 = 0$. Then, taking into account the spread of stationary points between the models we can judge about the spread of estimates $\hat{\beta}$:

$$\mathrm{cov}\left[\hat{\beta}\right] \approx \left[\hat{\sigma}^2_{\hat{\beta}_i\hat{\beta}_j}\right]_{i,j=1}^{n}, \qquad (4.26)$$

where

$$\hat{\sigma}^2_{\hat{\beta}_i\hat{\beta}_j} = \frac{1}{n-1}\sum_{l=1}^{n}\left(\hat{\beta}_i^{(l)} - \overline{\beta}_i\right)\left(\hat{\beta}_j^{(l)} - \overline{\beta}_j\right).$$

If $I_1 \to \infty$, all n models immediately reconstruct their parameters to the average value relative to the initial value, viz to $\hat{\beta}_0$, of the corresponding center of the initial values' hypercube. It is advisable to set α_1 increasing from zero. Then, at the first steps there is the independent motion of the parameters of each model to the "bottom of the ravine" and next the contraction of the estimates $\hat{\beta}^{(l)}$ on the ravine to the average $\overline{\beta}$, which also changes, approximating to the stationary point $\hat{\beta}^*$. The number of models can be both larger and less than n. Thus, applying two models only we can construct the algorithm with the desired law of change of the index $\boldsymbol{I}$ at the time t:

$$\boldsymbol{I} = \boldsymbol{I}_1^{(1)} + \boldsymbol{I}_1^{(2)} + \alpha_1\left(\hat{\beta}^{(1)} - \hat{\beta}^{(2)}\right)^T\left(\hat{\beta}^{(1)} - \hat{\beta}^{(2)}\right). \qquad (4.27)$$

Let us define the desired law

$$f(t) = \frac{d\boldsymbol{I}}{dt} < 0, \qquad (4.28)$$

where $f(t)$ – is the desired function of time, for example, the exponent, the constant etc.

Let us substitute from (4.27) the value $\frac{d\boldsymbol{I}}{dt}$ into (4.28). Then, we obtain:

$$f(t)=\left(\frac{\partial \boldsymbol{I}_1^{(1)}}{\partial\hat{\beta}^{(1)}}\right)^T\cdot\frac{d\hat{\beta}^{(1)}}{dt}+\left(\frac{\partial \boldsymbol{I}_1^{(2)}}{\partial\hat{\beta}^{(2)}}\right)^T\cdot\frac{d\hat{\beta}^{(2)}}{dt}+$$

$$+2\alpha_1\left(\hat{\beta}^{(1)}-\hat{\beta}^{(2)}\right)^T\left(\frac{d\hat{\beta}^{(1)}}{dt}-\frac{d\hat{\beta}^{(2)}}{dt}\right),$$

or

$$\left[\left(\frac{\partial \boldsymbol{I}_1^{(1)}}{\partial\hat{\beta}^{(1)}}\right)+2\alpha_1\left(\hat{\beta}^{(1)}-\hat{\beta}^{(2)}\right)\right]^T\frac{d\hat{\beta}^{(1)}}{dt}=$$

$$=f(t)-\left[\left(\frac{\partial \boldsymbol{I}_1^{(2)}}{\partial\hat{\beta}^{(2)}}\right)-2\alpha_1\left(\hat{\beta}^{(1)}-\hat{\beta}^{(2)}\right)\right]^T\frac{d\hat{\beta}^{(2)}}{dt}.$$

Let us denote

$$A_1=\left[\left(\frac{\partial \boldsymbol{I}_1^{(1)}}{\partial\hat{\beta}^{(1)}}\right)+2\alpha_1\left(\hat{\beta}^{(1)}-\hat{\beta}^{(2)}\right)\right],$$

$$A_2=\left[\left(\frac{\partial \boldsymbol{I}_1^{(2)}}{\partial\hat{\beta}^{(2)}}\right)-2\alpha_1\left(\hat{\beta}^{(1)}-\hat{\beta}^{(2)}\right)\right],$$

then,

$$\left.\begin{aligned}\frac{d\hat{\beta}^{(1)}}{dt}&=(A_1\cdot A_1^T)^{-1}\cdot A_1\left(f(t)-A_2^T\frac{d\hat{\beta}^{(2)}}{dt}\right)\\ \frac{d\hat{\beta}^{(2)}}{dt}&=(A_2\cdot A_2^T)^{-1}\cdot A_2\left(f(t)-A_1^T\frac{d\hat{\beta}^{(1)}}{dt}\right)\end{aligned}\right\}\qquad(4.29)$$

If the functional (4.27) is convex, the algorithms (4.29), constructed under condition (4.28), guarantee the convergence not only by the functional (4.27), but also by the parameters $\hat{\beta}^{(1)}$, $\hat{\beta}^{(2)}$ to $\hat{\beta}^{*}$ at the given law of the reduction $\boldsymbol{I}_1$.

If the signals, noises and parameters are non-stationary, the additional filtering is required, for example, by the inertial filter of the first order $W_f(p) = (\alpha p + 1)^{-1}$, where $p = \frac{d}{dt}$, α - is a filter constant.

Then, the gradient algorithm (4.19) is transformed into the method of "heavy ball" [4]:

$$\alpha \frac{d^2\hat{\beta}}{dt^2} + \frac{d\hat{\beta}}{dt} = -\lambda \frac{\partial \boldsymbol{I}(\hat{\beta}, t)}{\partial \hat{\beta}}, \qquad (4.30)$$

in which the parameter α "ball mass" may be adaptive: the "mass" should be that as $\hat{\beta}$ is smoothly "rolling" down the non-stationary stochastic surface to the minimum. For the stationary objects (according to the method of stochastic approximation [24]) we can require that $\alpha(t)$ indefinitely increases. The factor λ significantly affects on the convergence of the algorithm (4.30). Let us rewrite (4.30) with regard to (4.23) in order to get the appropriate choice of its structure:

$$\alpha \frac{d^2\Delta\hat{\beta}}{dt^2} + \frac{d\Delta\hat{\beta}}{dt} + \lambda \frac{\partial^2 \boldsymbol{I}(t)}{\partial \hat{\beta} \partial \hat{\beta}^T} \Delta\hat{\beta} = 0, \qquad (4.31)$$

where $\frac{\partial^2 \boldsymbol{I}(t)}{\partial \hat{\beta} \partial \hat{\beta}^T}$ – is the stochastic nonstationary symmetric matrix of instantaneous values.

The matrix $\frac{\partial^2 \boldsymbol{I}(t)}{\partial \hat{\beta} \partial \hat{\beta}^T}$ should be diagonal. It is the sufficient condition for the system's strength (4.31). In order to do this it is not necessarily to count the matrix $\left(\frac{\partial^2 \boldsymbol{I}(t)}{\partial \hat{\beta} \partial \hat{\beta}^T} \right)^{-1}$, especially as it does not always exist for the non-averaged values. Assume

$\frac{\partial^2 \boldsymbol{I}(t)}{\partial\hat{\beta}\partial\hat{\beta}^T} = A(t) = \left[a_{ij}(t) \right]_{i,j=1}^{n}$, then $\left(\frac{\partial^2 \boldsymbol{I}(t)}{\partial\hat{\beta}\partial\hat{\beta}^T} \right)^{-1} = \left[\frac{A_{ji}(t)}{\det A(t)} \right]$, where $A_{ji}(t)$ – is an algebraic addition of the *ji*-th element of the matrix $A(t)$.

It is enough to set $\lambda = \alpha' \left[A_{ji}(t) \right] = \alpha' \det A \cdot A^{-1}$ for the orthogonalization (4.31), then

$$\alpha \frac{d^2 \Delta\hat{\beta}}{dt^2} + \frac{d\Delta\hat{\beta}}{dt} + \alpha' \det A \cdot \Delta\hat{\beta} = 0,$$

where α' – is an algorithm parameter.

This algorithm is adaptive to the "informativeness": if the Fisher's matrix A is not informative (degenerate), that $\det A(t)$ is close to zero and $\Delta\hat{\beta} = 0$ (and $\alpha = const$); when the useful information occurs, there is an independent adjustment of each element $\hat{\beta}_i$ of the vector $\hat{\beta}$:

$$\frac{d\Delta\hat{\beta}}{dt} = -\alpha \left[\hat{A}_{ij}(t) \right] \cdot \frac{\partial \hat{I}(\hat{\beta}, t)}{\partial\hat{\beta}}, \qquad (4.32)$$

The algorithm is realized in practice

$$\frac{d\Delta\hat{\beta}}{dt} = -\alpha \left[\hat{A}_{ij}(t) \right] \cdot \frac{\partial \hat{I}(\hat{\beta}, t)}{\partial\hat{\beta}}, \qquad (4.32)$$

where $\hat{A}_{ij}(t)$, $\frac{\partial \hat{I}(\hat{\beta}, t)}{\partial\hat{\beta}}$ – are the values $A_{ij}(t)$, $\frac{\partial I(\hat{\beta}, t)}{\partial\hat{\beta}}$ smoothed by the filter W_f.

The functional diagram of the algorithm (4.32) of the orthogonal setting with adaptation to the "informativeness" of the object's signals is represented in the fig. 4.2.

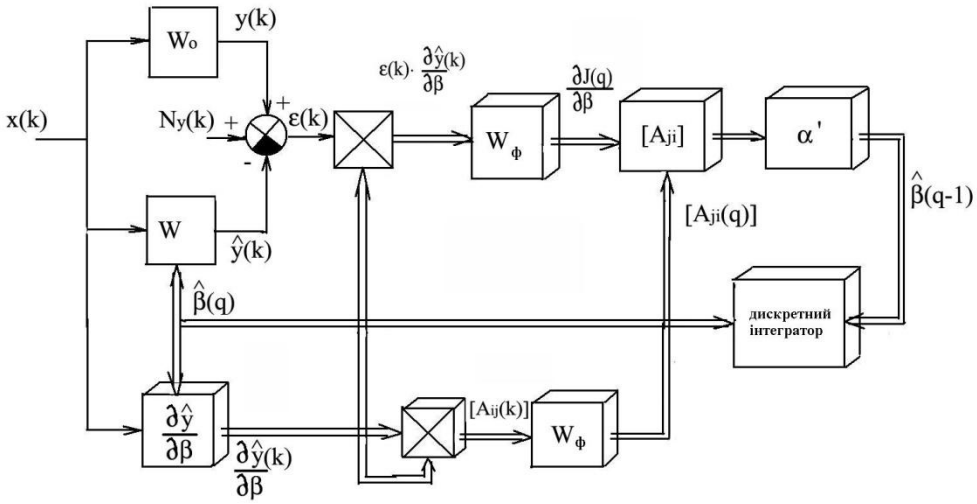


Fig. 4.2. The block diagram of the identification system with the gradient algorithm (4.32).

4.3. The Increasing of the Informativeness of Data Samples in Terms of the Active Experiment

4.3.1. The Theoretical Argumentation of the Expediency of the Coordinatewise Quasi-diagonalization

If the functional $I(\beta)$ is of the "ravine" character, viz it is not strictly convex and in some ways has very small changes, then, the system,

$$\boldsymbol{I}''(\beta^*)\cdot\Delta\hat{\beta} = -\boldsymbol{I}'(\beta^*), \qquad (4.33)$$

where $\boldsymbol{I}''$ – is the second derivative, $\boldsymbol{I}'$ – is the first derivative from I on $\hat{\beta}$ in the point β^*, will be ill-conditioned. Then, the solution:

$$\hat{\beta} = \left[\boldsymbol{I}''(\beta^*)\right]^{-1}\cdot\boldsymbol{I}'(\beta^*), \qquad (4.34)$$

will be excessively inaccurate as a consequence of the impropriety [46] of calculations under the expression (4.34).

The significant separation of the matrix spectrum into two groups that correspond to the large and small eigenvalues λ

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{k-1} \geq \sigma\lambda_k \geq \ldots \geq \sigma\lambda_n;\ \sigma >> 1 \qquad (4.35)$$

is the main condition for the "ravinity" $I(\beta)$.

The linear membrane of the eigenvectors that correspond to small eigenvalues can be taken as a bottom of the "ravine" *Q* for the system (3.29).

$$\begin{aligned} Q &= \left\{ x \in R^n \left(< x - x^*,\ u_i \geq 0 \quad (i = \overline{1,n-r}) \right\}, \right. \\ Q_\delta &= R^n;\ \sigma \approx \lambda_1 \Big/ \left|\lambda_{n-r+1}\right|. \end{aligned} \qquad (4.36)$$

The condition of the "ravinity" (4.35) is realized here.

The condition number of the matrix $\boldsymbol{I''}(\beta^*)$ for the model (4.35) (4.36) of the "ravine" functional is defined by the correlation:

$$\mathrm{cond}\left(\boldsymbol{I''}(\beta^*)\right) = \left\|\boldsymbol{I''}(\beta^*)\right\| \cdot \left\|\left[\boldsymbol{I''}(\beta^*)\right]^{-1}\right\| = \lambda_{max} \Big/ \lambda_{min} \geq \sigma >> 1. \qquad (4.37)$$

Moreover, the greatest error of the solution (4.34) is concentrated in the subspace spanned on the eigenvectors of the matrix $\boldsymbol{I''}(\beta^*)$, that correspond to small eigenvalues. Let us show this.

The lemma. Let the condition (4.35) is implemented for $\boldsymbol{I''}(\beta^*)$. Then, the solution's (4.34) errors satisfy the relation

$$\Delta d_i < \sigma^{-1} \Delta d_j;\ \sigma >> 1;\ i = \overline{1,k-1};\ j = \overline{k,n}. \qquad (4.38)$$

It is known [33], that the solutions (4.34) of the system (4.33) can be represented as $\Delta\beta = \sum d_i \cdot u_i$, where u_i - are the matrix $\boldsymbol{I''}(\beta^*)$ eigenvectors.

The coefficients d_i are equal to

$d_i = C_i \big/ \lambda_i$, $i = \overline{1,n}$, where λ_i – are the matrix $\boldsymbol{I}''(\beta^*)$ eigenvalues that correspond to the vectors u_i; C_i – are coefficients of the vector's $I(\beta)$ decomposition at the basis $\{u_i\}_{i=1}^{n}$.

Expanding the error $\delta\beta$ at the same basis

$$\Delta\beta + \delta\beta = \sum_{i=1}^{n} a_i \cdot u_i,$$

where $a_i = d_i + \Delta d_i = \dfrac{C_i + \Delta C_i}{\lambda_i + \Delta\lambda_i} = \dfrac{C_i}{\lambda_i} + \dfrac{\Delta C_i \cdot \lambda_i - C_i \cdot \Delta\lambda_i}{(\lambda_i + \Delta\lambda_i)\cdot\lambda_i}$, we obtain that $\Delta d_i = \dfrac{\Delta C_i \cdot \lambda_i - C_i \cdot \Delta\lambda_i}{(\lambda_i + \Delta\lambda_i)\cdot\lambda_i}$, $i = \overline{1,n}$. In accordance with (4.36), the eigenvectors u_k …, u_n set up the ravine's bottom where the gradient's $|\boldsymbol{I}'(\beta)|$ vector norm is defined by the small eigenvalues and it is significantly lower than in any other part of the parameters β space, then,

$$C_i < C_j;\ i = \overline{1,k-1};\ j = \overline{k,n}; \qquad (4.38)$$

at the same time

$$\lambda_i \geq \sigma\lambda_j C_j;\ i = \overline{1,k-1};\ j = \overline{k,n}. \qquad (4.39)$$

That means we have $\Delta d_i < \sigma^{-1} \cdot \Delta d_j;\ \sigma >> 1$. The lemma is proved.

It follows from the lemma that the axis of the most rational coordinate system while minimizing the quadratic functional ravine structure coincide with the eigenvectors of the matrix $\boldsymbol{I}''(\beta^*)$. The dependence of the error $\delta\beta$ on the relative orientation of the coordinate and orthonormalized bases is shown in the fig. 4.3

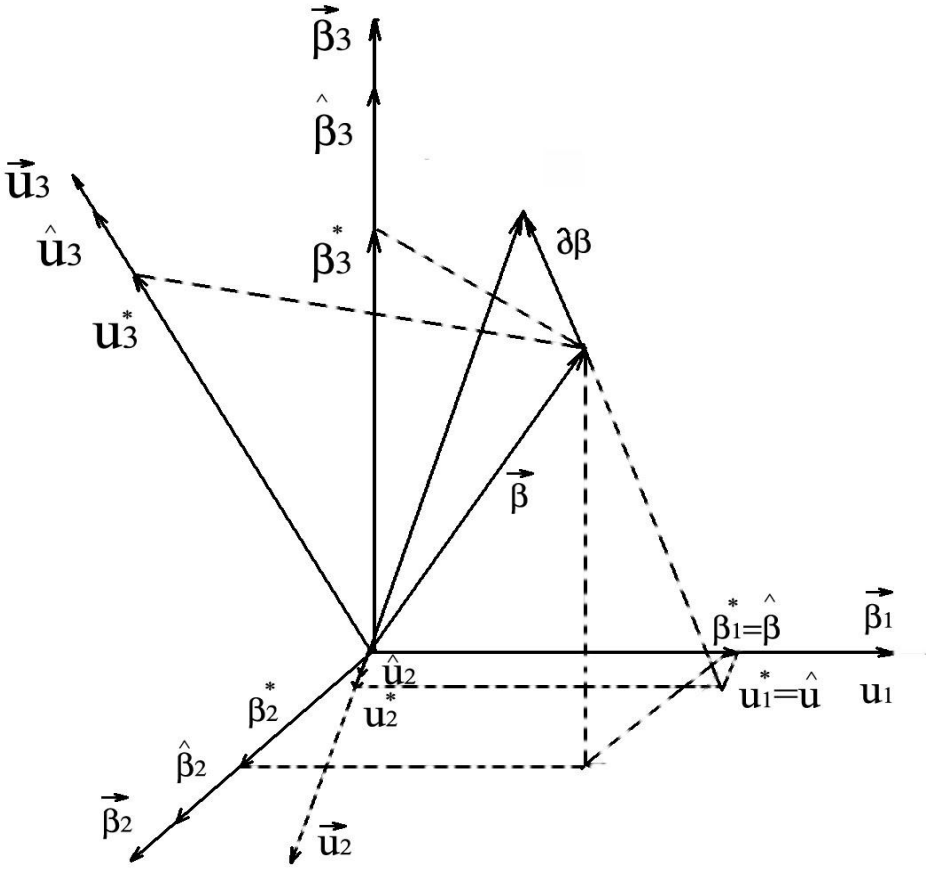


Fig. 4.3. The dependence of the error δβ on the choice of the basis.

If the unit vector coincides with the eigenvector u_i, the error $\delta\beta_1$ of the estimate $\hat{\beta}_1$ is minimum. In the contrary case, even for large eigenvalues (β_2 and u_2) the error $\delta\beta_2$ of the estimate $\hat{\beta}_2$ is significant. The error $\delta\beta_3$ is maximum for small eigenvalues (β_3 i u_3). If the coordinate basis coincides with the orthonormalized basis, the quadratic functional:

$$\boldsymbol{I}(\beta) \cong \boldsymbol{I}(\beta^*) + \boldsymbol{I}'(\beta^*) \cdot \Delta\beta + \frac{1}{2}\Delta\beta^T \boldsymbol{I}''(\beta^*) \cdot \Delta\beta, \qquad (4.40)$$

takes the separable form

$$\boldsymbol{I}(\beta) = \sum_{i=1}^{n} \boldsymbol{I}(\beta_i), \qquad (4.41)$$

the estimate $\hat{\beta}_i$ are found independentl and the smallness of the error $\delta\beta_j$ is guaranteed for those β_j that have large eigenvalues.

The task is significantly simpler, if it is required to ensure that the coordinate unit vectors β_i coincides with its eigenvectoru_i. Moreover, the errors $\delta\beta_j$ at other directions $u_j, (i \neq j)$ have little impact on $\delta\beta_j$.

The required properties $\boldsymbol{I}(\beta)$ can be ensured by the appropriate selection of the controlling actions, i.e. the active experiment on the examined object. We will prove this with the help of the assertion that follows.

The assertion. Let the matrix $\boldsymbol{I}''(\beta) = A$ has structure:

$a_{ij} \neq 0,\ a_{kj} = a_{jk} = 0;\ \ i, j \neq k$, where *k* – is the fixed number, 1≤ *k* ≤ *n*, then, $a_{kk} = \lambda_k(A)$ and the eigenvector $u_k(A)$ coincides with the coordinate unit vector $\vec{\beta}_k$.

In accordance with the theorem on the spectral decomposition [42], let us represent the matrix *A* as follows

$$\Lambda = U^T A U = U^T A_1 U + U^T A_2 U\,, \qquad (4.42)$$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_n);\ A_1 = A - A_2,\ A_2$ - is the matrix $[n \times n]$ in which all elements are zero except a_{kk}, *U* – is the orthogonal matrix, the columns of which are the eigenvalues of the matrix *A*.

So far as the basis is orthogonal

$$U^T A_2 U = a_{kk}\vec{U}_k \cdot \vec{U}_k = \mathrm{diag}(0,\ldots,a_{kk},\ldots,0)\,, \qquad (4.43)$$

$$U^T A_2 U = \mathrm{diag}(\lambda_1,\ldots,\lambda_{k-1},0,\lambda_{k+1},\ldots,\lambda_n). \qquad (4.44)$$

then, it follows from (4.42)

$$\Lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_{k-1},0,\lambda_{k+1},\ldots,\lambda_n). \qquad (4.45)$$

from here

$$\lambda_k = a_{kk}\,. \qquad (4.46)$$

According to a spectral representation of the operator A [21] for any vector $\vec{\beta}$, $\vec{\beta} = C_1\vec{U}_1 + \ldots + C_n\vec{U}_n$ there is a relation:

$$A\beta = \sum_{i=1}^{n} \lambda_i C_i \vec{U}_i = \lambda_1 C_1 \vec{U}_1 + \ldots + \lambda_n C_n \vec{U}_n. \qquad (4.47)$$

Considering (4.46) for the k unit coordinate vector of the matrix A we obtain

$$a_{kk}\vec{\beta}_k = \lambda_k C_k \vec{U}_k. \qquad (4.48)$$

From here it follows, that the vector $\vec{\beta}_k$ coincides with the vector $\vec{U}_k$. The assertion is proved.

Let us apply the obtained result. Let us set the criterion of the signal synthesis as any matrix norm $\|\cdot\|$ of the difference of the optimized $\boldsymbol{I}''(\beta)$ and the wanted A matrices:

$$\Im(\boldsymbol{I}''(\beta)) = \|\boldsymbol{I}''(\beta) - A\|. \qquad (4.49)$$

From the inequality [19] $\|A - B\| \geq \sum_{i=1}^{n} [\lambda_1(A) - \lambda_i(B)]^2$ it follows that the functional (4.49) with Euclidean norm $\|\cdot\|$ minimizes the mean square deviation of the eigenvalues . It is convex, monotonous and uniform on the set of the Hesse's matrices. The next conditions are met for this:

$$\Im[aA + (1-a)B] \leq a\Im(A) + (1-a)\Im(B), \quad 0 < a < 1,$$

$$\Im(A) \leq \Im, \quad A \leq B,$$

$$\Im(\alpha A) = \alpha\Im(A)$$

These conditions are met in accordance with the axiomatics of the matrix norm.

$$\|A + B\| \leq \|A\| + \|B\|, \ \|\alpha A\| \leq \|\alpha\| \cdot \|A\|.$$

If we connect the functional with one k order of the matrix $\boldsymbol{I}''(\beta)$, then, instead of (4.49) we obtain:

$$\Im(\boldsymbol{I}''(\beta)) = \| j''_{ki}(\beta) - a_{ki} \|, \qquad (4.50)$$

where j''_{ki} - is an element i of the k row of the matrix $\boldsymbol{I}''(\beta)$, $a_{ki} = a_{kk} \neq 0$, $i \neq k$.

After this optimization of the active identification modes, the minimization task of the ravine functional $I(\beta)$ becomes the task of coordinate-wise optimization, which is represented with respect to β in the separable form:

$$\min_{\beta} \boldsymbol{I}(\beta) = \min_{\beta} \sum_{i=1}^{n} \boldsymbol{I}_i(\beta), \qquad (4.51)$$

where $\boldsymbol{I}_i(\beta) = \boldsymbol{I}_i(\beta_i / C_i\beta = d)$.

The matrix C_i of the linear restraints has the form $C_i'^{T} = [I_{n-1}, 0]$.

4.3.2. The Synthesis of the Testing Input Actions on the Examined Object

Some assumptions should be taken into account during the synthesis of the optimal signals that test the object:

1. The signals should be physically executed at the object.

2. They should not introduce the object into the area of emergency operations.

3. Their amplitude should be such as to meet the requirements for the ratio “signal – noise”, but not so large that the phenomena, not provided by the model, appeared at the object (nonlinearity, etc).

4. The signals should provide the optimality of the relevant criterion (4.49) or (4.50). The pulse input actions’ sequence can satisfy these restrictions.

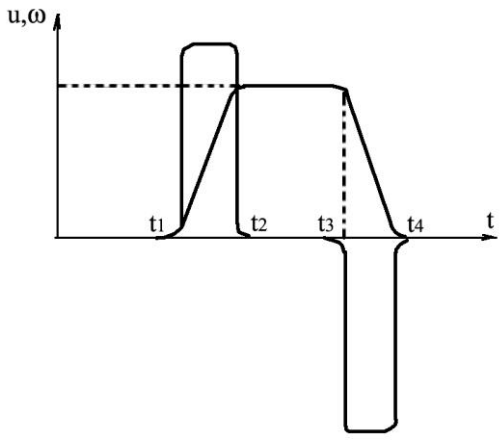


Fig. 4.4. The testing pulse $u(t)$ and its derivative $\omega(t)$.

$$u(t)=u(t_1,t_2,t_3,t_4)=\begin{cases}0; & t<t_1\\ \mu(t-t_1); & t_1\le t<t_2\\ U_{max}; & t_2\le t<t_3\\ U_{max}-\mu(t-t_1) & t_3\le t<t_4\\ 0; & t_4\le t<t_5\end{cases} \qquad (4.52)$$

or

$$u(t)=\sum_{k=1}^{2}(-1)^{k+1}\sum_{j=1}^{2}(-1)^{j+1}S(t-t_{j,k})\ . \qquad (4.53)$$

The sequence of pulses of trapezoid form (fig. 4.3) is quite simply executed at the real objects of technology and economy:

$$S(t-t_{j,k})=\varphi\left[\omega(t-t_{j,k})\right], \qquad (4.54)$$

$$\varphi(x)=\begin{cases}0, & x<0\\ x, & x\ge 0\end{cases}, \qquad (4.55)$$

where $\omega=\dfrac{du}{dt}$ – is the slope of signal fronts.

The sequence of controlling signals of *i* control is as follows:

$$u^i(t)=\sum_{l=1}^{L}(-1)^{l+1}u^i(t,t_{l,1},t_{l,2},t_{l,3}). \qquad (4.56)$$

The derivative of the single pulse

$$\frac{du}{dt}=\omega\sum_{k=1}^{2}(-1)^{k+1}\sum_{j=1}^{2}(-1)^{j+1}1(t-t_{jk}), \qquad (4.57)$$

where $1(t)$ – is a single function $1(t)=\begin{cases}0, & t<0\\ 1, & t\ge 0\end{cases}$.

The limitations to the fronts slope, the length and the amplitude signals are

$$\max\left(\frac{du}{dt}\right)=\omega, \tag{4.58}$$

$$t_{min}\le t_{j2}-t_{j1}\le t_{max}, \tag{4.59}$$

$$U_{max}\ge\omega(t_{j2}-t_{j1})\ . \tag{4.60}$$

So far as the signals are completely determined by the moments $t_{i,k}$ and the fronts slope ω, it is not hard, if it is necessary, to impose the additional restrictions on the relative position of signals at their synthesis. The task of synthesis of optimal testing signals is formulated as the task of optimal control is that: knowing the required equations of the object,

$$\dot{x}=Ax+Bu, \tag{4.61}$$

the class of signals (4.53)–(4.57) and the restrictions (4.58)–(4.60), to find the program control at which the functional (4.49) or (4.50) takes the minimum value. The optimization of control at the classic functional:

$$I=V_3\left[x(t_k)\right]+\int_{t_9}^{t_k}L\left[x(t),u(t),t\right]dt, \tag{4.62}$$

where L, V_3 - are set scalar functions of the vector arguments x and u, is complex. This is due to high power of sets at the numerical solution of the dual point boundary problem of the functional.

The application of the semidefinite

$$I=V_3\left[x(t_k)\right]+\frac{1}{2}\int_{t_0}^{t_k}Q(x,t)dt+\frac{1}{2}\int_{t_0}^{t_k}(U^TKU+U_{onm}^TK^{-1}U_{onm})dt, \tag{4.63}$$

allows to solve the problem easier and properly. This problem is solved by the method of forecasting models for the synthesis of continuous control in the work [13].

The solution in this case. We assume the switching points in (4.53) (4.56) as components of the generalized object, obtained by association (4.60) i (4.64):

$$\dot{T}_p=U;\ T_p=\left\{t_{ij}\right\};\ j=\overline{1,l};\ i=\overline{1,m}. \tag{4.64}$$

The equation (4.64) describes the restructuring of the *j*-th switching moment of *i* component of the control vector T_p.

$$U_{opt} = -K\frac{\partial V^T}{\partial T}, \tag{4.65}$$

where $\frac{\partial V}{\partial t}$ – is a partial derivative of scalar function $V(x,T,t)$, satisfies the equation:

$$\frac{\partial V}{\partial t} - \frac{\partial V}{\partial x}(Ax + Bu) = -Q(x,t), \tag{4.66}$$

with the boundary condition $V(t_k) = V_3$.

The application of

$$\begin{cases} \dot{x}_M = Ax_M + Bu_M \\ T_p = 0 \end{cases}, \tag{4.67}$$

allows to find the function $V(x,T,t)$ at the point corresponding to the current state τ_k of the trajectory

$$V(x,t,\tau) = V_3(x_M,\tau_k) + \int_{\tau_0}^{\tau_K} Q(x_M,\tau)d\tau, \tag{4.68}$$

where τ – is a current time for the models (4.67).

Substituting (4.68) in (4.65), taking into account the rules of differentiation, we get:

$$\begin{aligned} U_{opt} &= -k\left[\frac{dV_3(x_M,\tau_k)}{dT_p} + \int_{\tau_0}^{\tau_k}\frac{dQ(x_M,\tau)}{dT_p}\right] = \\ &= -k\left\{\begin{array}{l} z^T(\tau_k)\dfrac{\partial V_3(\tau_k)}{\partial x_M(\tau_k)} + \dfrac{\partial V_3(x\tau_k)}{dT_p} \\ \displaystyle\int_{\tau_0}^{\tau_k}\left[z^T(\tau)\frac{\partial Q^T(\tau)}{\partial x_M(\tau)} + \frac{dQ^T(\tau)}{dT_p}\right]d\tau \end{array}\right\}, \end{aligned} \tag{4.69}$$

where $z(\tau) = \frac{\partial x}{\partial T_p}$ - is the matrix function of sensitivity of the extended object (4.63), (4.64), which is obtained by differentiation at the vector T_p:

$$\frac{\partial}{\partial T_p}\left(\frac{dx}{dt}\right)=\frac{d}{dt}\left(\frac{\partial x}{\partial T_p}\right)=\dot{z}=A\frac{\partial x}{\partial T_p}+B\frac{\partial u}{\partial T_p}=Az+B\frac{\partial u}{\partial T_p}$$

or

$$\dot{z}=Az+B\frac{\partial u}{\partial T_p}. \qquad (4.70)$$

The matrix of the partial derivatives $\frac{\partial u}{\partial T_p}$ with regard for (4.53)

$$\frac{\partial u}{\partial T_p}=\left\|\frac{\partial u^i(t)}{\partial t^i_{jk}}\right\|, \qquad (4.71)$$

where

$$\frac{\partial u^i(t)}{\partial t^i_{jk}}=\omega(-1)^{k+j}1(t-t^i_{jk}), \qquad (4.72)$$

$1(t)$ – is the Heaviside function.

Solving (4.70) relatively to z and simultaneously integrating (4.72), we obtain the vector of gradient which allows to determine where to shift the switching points along the time axis in order to achieve the functional's I extremum at the end of one cycle of modeling.

The dimensionality of the vector T_p can be reduced. Thus, if the amplitude of signals is fixed, then, the dimensionality T_p is reduced by half. The signal is completely defined by the moments t^i_{j1}, t^i_{j3}, and the sequence of signals of i control is as follows:

$$u^i(t)=\sum_{j=1}^{L_i}(-1)^{j+1}S_i(t-t^i_j), \qquad (4.73)$$

where L_i - is the number of inclusions of the *i-th* control organ, and $S_i(t-t^i_j)=\varphi\left[\omega(t-t^i_j)\right]$,

$$\varphi(x)=\begin{cases}0; & x<0\\ x; & 0\le x<1.\\ 1; & x\ge 1\end{cases} \qquad (4.74)$$

The matrix of partial derivatives:

$$\frac{\partial u^i(t)}{\partial t_j^i}=\omega(-1)^{j+1}\left[1(t-t_j^i)-1\left(t-t_j^i-\frac{1}{\omega}\right)\right]. \qquad (4.75)$$

The furthest reducing of the dimensionality of the vector T_p is possible, if the test signals have fixed shape. Such signals are completely determined by the moment of their injection. It is necessary only to define their relative position at the time axis. The sequence of the heteropolar pulses can be described by the equation:

$$u_i^i(t)=\sum_{j=1}^{L_i}(-1)^{j+1}S_i(t-t_j^i), \qquad (4.76)$$

where $S(t-t_j^i)=\varphi\left[\omega(t-t_j^i)\right]$, $\varphi(x)=\begin{cases}0; & t<t_j\\ x; & t_j\le t<t_j+c_1\\ u_M=c_1; & t_j+c_1\le t<t_j+c_2\\ u_M-x; & t_j+c_2\le t<t_j+c_2+c_1\\ 0; & t\ge t_j+c_2+c_1\end{cases}$.

The restrictions on signals are connected with their relative position and duration of the experiment:

$$t_j^i-t_j^k\ge c_{3;}\ \forall i,k;\ t_k-t_1^1\le T_{exp}. \qquad (4.77)$$

The criterion (4.63) has the terminal component:

$$V_3[t_k]=\Im[J''(\beta)] \qquad (4.78)$$

and, with the restrictions like (4.58)-(4.60), it has the integral component:

$$Q(x,t)=\begin{cases}c; & u\in G_u\\ \alpha^T t_c+\beta; & u\notin G_u\end{cases}. \qquad (4.79)$$

Then,

$$\frac{\partial Q(x,t)}{\partial T_p} = \begin{cases} 0; & u \in G_u \\ \alpha; & u \notin G_u \end{cases}, \qquad (4.80)$$

where α – is the vector of coefficients, the value of which depends on the restriction.

The designed method of synthesis of test signals allows to construct the gradient algorithm of the correction of the switchings vector T_p:

$$T_p^0 = T_p(t_k^0),\ k = 0, \qquad (4.81)$$

that allows to identify the local extremums of the functional (2.64) in the space of controlled variables. The condition for stopping may be the demand for the relative or absolute accuracy:

$$\left|I_{end} - I_k\right| \le \alpha_1 \left|I_k\right| + \alpha_2, \qquad (4.82)$$

where α_1, α_2 – are set values of the relative and absolute accuracy of the optimum's localization on the functional.

It is necessary to carry out the check-ups on the argument at the slow approximating to the extremum

$$\left\|T_p^{k+1} - T_p^k\right\| \le \varepsilon_1 \left\|T_p^k\right\| + \varepsilon_2. \qquad (4.83)$$

Due to the multiple experimental properties of the functional (4.78) in the space of controlled variables T_p, it is possible to have a case where the localized extremum of the functional does not provide the significant improvement of the identification conditions. In case that computations stop, it is required to verify the condition $\Phi_k \le \Phi_0$ and repeat the searching for the optimal T_p at other starting conditions T_p^0.

The algorithm of optimization:

1. The initial conditions are formed for (4.67), (4.70) $u(t) = u^0(t, T_p)$, $T_p = T_p^0$, $z(t_0) = 0$, where T_p, T_p^0 - are the initial conditions for the switching points and the sensitivity function.

2. The object's movement is modeled using (4.67), and simultaneously the matrix equation of sensitivity is integrated using (4.70).

3. Using the results of modeling with the help of (4.69) we define the gradient of the minimized functional by the controlled variable T_p. We obtain the algorithm of correction of the vector T_p component on the iteration *j*: $T_p{}^j = T_p{}^{j-1} + kv_j$, experimentally selecting the value of *k* in (4.65).

4. If the rule of algorithm stopping is not executed, that we should move to p.2 for the next iteration of search T_p. The rule of stopping:

$$A_1 = \left|\Im_j - \Im_{j-1}\right| \le \varepsilon_1 \left|\Im_{j-1}\right| + \varepsilon_2,\ A_2 = \left\|T_j - T_{j-1}\right\| \le \varepsilon_1 \left\|T_{j-1}\right\| + \varepsilon_2,$$

where ε_1, ε_2 -are the predefined small numbers.

5. The condition $\Im_j \le \Im_{def}$ is verified, if it is implemented, that should move to p.1 with the formation of new initial conditions.

4.4. The Method of the Separate Estimation of Static Nonlinear and Dynamic Linear Components of the Hammerstein's Model

Hammerstein's model describes the real system at the input (x_{in}) - output (x_{out}) by the combination of linear dynamic operator

$$\beta'_n \frac{d^n x_{out}}{dt^n} + \beta'_1 \frac{dx_{out}}{dt} + x_{out}(t) = \gamma_m \frac{d^m x_{in}}{dt^m} + \cdots + \gamma_1 \frac{dx_{in}}{dt} + \gamma_0 x_{in}(t) \qquad (4.84)$$

with the static nonlinear $f(u)$ that is decomposed by the system of linearly independent (preferably orthogonal) functions $\varphi_k(u)$:

$$f(u) = \sum_{k=1}^{r} l_k \cdot \varphi_k(u). \qquad (4.85)$$

According to the non-linearity's position (fig. 4.5), denoting the operator $\frac{d^i}{dt^i} = p^i$,

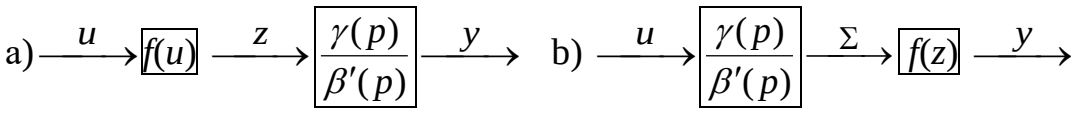


Fig. 4.5. Hammerstein's models: *a*) – the nonlinearity at the input,
b) – the nonlinearity at the output.

and substituting the decomposition (4.85) and dynamics (4.84) into the corresponding structures (4.5), we obtain the model for the variants of the nonlinearity at the input:

$$\beta'(p)\cdot y(t) = \sum_{i=1}^{m}\gamma_i p^i\left(\sum_{k=0}^{r}\alpha_k\varphi_k(u)\right) \qquad (4.86)$$

we obtain the model for the variants of the nonlinearity *b)* at the output:

$$\sum_{i=0}^{n}\beta'_i p^i\left(\sum_{k=0}^{r}\alpha_k\varphi_k(y(t))\right) = \gamma(p)\cdot u(t), \qquad (4.87)$$

where $\beta'(p) = \beta'_n p^n + \cdots + \beta'_i p + 1, \quad \gamma(p) = \gamma_m p^m + \cdots + \gamma_i p + \gamma_0.$

Then, the (n+m+r+2) parameters α_i, β_i, γ_i are determined in the identification task. The parameters γ_i, α_k are the part of the model (4.86) as a set their products, and the parameters β_i, α_k are the part of the model (4.87) as a set of their products too. This complicates the solution of the estimation's problem. Moreover, if $u(t)$ that provides the orthogonality $\varphi_k(u(t))$, that is difficult to implement, the orthogonality is discolated as a result of differentiation of functions $\varphi_k(u(t))$ in time.

For example, if φ_k is sine-cosine series, then, the first derivative from φ_k will be linear dependent function from φ_{k+1}. Using the exponential polynomial:

$$f(u(t)) = \sum_{k=0}^{r}\alpha_k u^k(t), \qquad (4.88)$$

does not creates the orthogonality. All this worsens the practical application of the models (4.86), (4.87), especially in the noisiness of signals, which have to be differentiated.

Therefore, putting the limitations of the band of the signals spectrum $x_{in}(t)$, $x_{out}(t)$ and assuming the finite error Δ of the approximation of the differential equation (4.84) by the equation

$$\beta_n \frac{d^n x_{out}}{dt^n} + \cdots + \beta_1 \frac{dx_{out}}{dt} + x_{out}(t) = x_{in}(t) + \Delta(t), \qquad (4.89)$$

we obtain the simplified Hammerstein's models: instead of (4.86), we obtain the model:

$$\beta(p) \cdot y(t) = \sum_{k=0}^{r} \alpha_k \varphi^k(u), \qquad (4.90)$$

instead of (4.87) we obtain the model

$$\sum_{i=0}^{n} \beta_i p^i \left(\sum_{k=0}^{r} \alpha_k \varphi_k(y(t)) \right) = u(t). \qquad (4.91)$$

The last model may be represented as (4.90), if we consider the inverse dependence $u(y)$ instead of direct one $y(k)$.

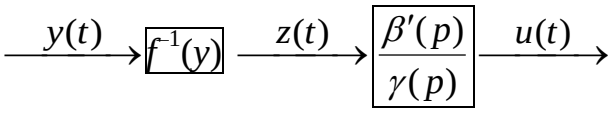


Fig. 4.6. The model that is inverse to the model (4.87).

Then, the task of identification of both models is reduced to the identification task of the model (4.90) with the $n+r+1$ unknown variable. Further simplification of the task is accomplished by taking into account the fundamental laws of smooth mappings, namely $f(u)$.

The criteria of approximation, that take into account the smooth of sought dependence

The root-mean-square criterion of proximity, provided by Legendre-Gauss in 1806-1809, makes it possible to get the best approximation to the mathematical expectation of the sought dependence, if noises are normally distributed, and the data sample is large

enough. If the data samples are not sufficient and the normality of the law of data distribution is broken, this method loses its actuality [22].

The situation may be improved, if we take into account the additional information:

– we discard the anomalous data

– we control the nature smoothness of the sought dependencies

The anomalous data are discarded or corrected at the stage of robust filtering of signals, and the property of smoothness was first taken into account when determining the models of technological processes in the work [23]. Instead of a minimum of the mean square error, it is offered to minimize the mean square value of r+1 finite difference $\Delta^{r+1}\varepsilon$ according to the standard procedure:

$$\boldsymbol{I} = \sum(\Delta^{r+1}\varepsilon)^2 \to min, \qquad (4.92)$$

where, if r=1, then, the approximation of smoothness takes place, if r=2, – the approximation on curvature and so on .

The generalization criterion is proposed in the [23], where instead of the square, any degree is taken. Taking into account the demand of robustness [22], it is not advisable to increase the degree. The order of the difference r+1 is set a priori, if the model's order is known(4.85), or it is gradually sought, starting with r=1.

The method of determining the smooth static nonlinearity from the dynamic of control

Let the dynamic of the object be described by the linear differential equation

$$\sum_{k=0}^{n} a_{n-k} \frac{d^{n-k} y(t)}{dt^{n-k}} = f\left[x(t)\right], \qquad (4.93)$$

where $f\left[x(t)\right]$ - is the static nonlinearity, on the input of which the testing signal $x(t)$ arrives; the output of this nonlinearity influences on the dynamics of the object, the model of which is the transfer function $W(a, p)$.

The task is to recover (estimate) the static nonlinearity and therefore to offset its impact on the dynamics of the object, taking into account the measured noisy values of the output $\hat{y}(t_k)$.

The classical methods of solving problems of this involve applying of method of least squares (LSM) for estimating parameters of the model of the object's dynamic (parameters of the transfer function).

The unknown nonlinearity as a function of the input exposure $f\left[x(t)\right]$ is approximated by the polynomial:

$$f\left[x(t)\right] = b_0 + b_1 x(t) + b_2 x^2(t) + ... + b_m x^m(t). \qquad (4.94)$$

the discrepancy is formed:

$$\varepsilon(t) = \sum_{k=0}^{n} a_k \frac{d^k y(t)}{dt^k} - f\left[x(t)\right] \qquad (4.95)$$

and the functional is minimized

$$J = \frac{1}{T}\int_0^T \varepsilon^2(t)dt\,. \qquad (4.96)$$

The parameters $b_k, k = \overline{1,m}; a_k, k = \overline{1,n}$ are estimated under condition of the minimum of this functional:

$$\frac{\partial J}{\partial b} = 0;\ \frac{\partial J}{\partial a} = 0. \qquad (4.97)$$

The practice of application of the least squares method to solve such problems shows that it is characterized by fairly significant weaknesses, connected with the high requirements for testing (input) signal, and the low accuracy of the estimates of parameters of the nonlinear dynamic model as a result of the inadequacy of the model structure of the control object, that restrict the application of the classical least squares method for solving those tasks.

The attempts to improve the adequacy of the model by increasing its order lead to the sharp increasing of estimation's time, which is generally unacceptable in a limited time of the control (especially while solving problems in a real time) and the noisiness of the measured output data. This leads to the systematic errors that significantly limit the application of the classical LSM.

Having regard to the above, the problem of finding methods of estimation of the static nonlinearities which affect on the dynamic of the control object and estimation of parameters of model of the transfer function of the control object is important and relevant.

In order to determine the nonparametric model $\hat{f}\left[x(t)\right]$ of the static nonlinearity $f\left[x(t)\right]$, let us define the compensated output of the control object define as:

$$y_{cor}\left(t\right)=\hat{y}\left(t\right)-\beta_1\frac{d\hat{y}\left(t\right)}{dt}-\beta_2\frac{d^2\hat{y}\left(t\right)}{dt^2}, \tag{4.98}$$

where parameters β_1,β_2 are determined under condition of the mean square *r* of the derivative from $\hat{y}\left(t\right)$ on *x*,

$$\left(\beta_1,\beta_2\right)=\operatorname{argmin}\sum_{k=1}^{N}\frac{d^r y_{cor}\left(t_k\right)}{dx^r}. \tag{4.99}$$

Taking into account the discreteness of measurements of the output value *y* instead of the y r-th derivative we can use the appropriate discrete sequence $x(t_k)$ which is measured with a constant time step Δt at the sequence with the constant increment Δx, but with the variable time-step. This task can be solved in this way.

Firstly we should smooth of noisy input and output sequences of the measurements. In order to solve this task the smoothing of the measured values is performed by using by smoothed splines.

Then, we sort the value of the input variable $x(t_k)$ in ascending order. Next in order to determine the values t_j that correspond to the input variable, that are changed with the constant step, let us perform the interpolation of the obtained sequence using the interpolation splines.

$$S_2\left(t\right)=x_k+m_k\left(t-t_k\right)+c_k\left(t-t_k\right)^2,\ t\in\left[t_k,t_{k+1}\right]. \tag{4.100}$$

We find the values t_j that correspond to $j\Delta x$ by solving the equation spline

$$t_{j_k}=t_k+\frac{1}{2c_k}\left(-m_k+\sqrt{m_k^2+4jc_k\Delta x}\right), \tag{4.101}$$

where $j=\overline{1,l_k}$, $l_k=\left[x\left(t_{k+1}\right)-x\left(t_k\right)\right]/\Delta x$.

Having determined all t_{jk} we count the values $\hat{y}(t_{jk}), d\hat{y}(t_{jk})/dt$.

So far as the finite differences are used instead of the derivatives, we have

$$\Delta^r y_{cor}=\Delta^r\left[\hat{y}(t)\text{-}\beta_1\frac{d\hat{y}(t)}{dt}-\beta_2\frac{d^2\hat{y}(t)}{dt^2}\right]. \qquad (4.102)$$

In particular, for $r=2$ we have

$$\Delta^2 y_{cor}=\left[y_{cor}\left(t_{k+2}\right)-2y_{cor}\left(t_{k+1}\right)+y_{cor}\left(t_k\right)\right]/\left(\Delta x\right)^2. \qquad (4.103)$$

For $r=3$:

$$\Delta^3 y_{cor}=\left[y_{cor}\left(t_{k+3}\right)-3y_{cor}\left(t_{k+2}\right)+3y_{cor}\left(t_{k+1}\right)-y_{cor}\left(t_k\right)\right]/\left(\Delta x\right)^3. \qquad (4.104)$$

Thus, the minimization of the functional is in solving this system of linear equations with respect to β_1,β_2

$$\begin{gathered}\sum_{k=1}^{n}\left[\Delta^r\hat{y}(t_k)\text{-}\Delta^r\frac{d\hat{y}(t_k)}{dt}\beta_1-\Delta^r\frac{d^2\hat{y}(t_k)}{dt^2}\beta_2\right]\Delta^r\frac{d\hat{y}(t_k)}{dt}=0,\\ \sum_{k=1}^{N}\left[\sum_{k=1}^{N}\left[\Delta^r\hat{y}(t_k)\text{-}\Delta^r\frac{d\hat{y}(t_k)}{dt}\beta_1-\Delta^r\frac{d^2\hat{y}(t_k)}{dt^2}\beta_2\right]\right]\Delta^r\frac{d^2\hat{y}(t_k)}{dt^2}=0.\end{gathered} \qquad (4.105)$$

Having identified the parameters β_1,β_2 , we define the value of the smooth static nonlinearity by the formula:

$$f\left[x(t)\right]=\hat{y}(t)\text{-}\beta_1\frac{d\hat{y}(t)}{dt}-\beta_2\frac{d^2\hat{y}(t)}{dt^2}. \qquad (4.106)$$

The algorithm of smoothing the experimental data.

Let the value of the function $f(x)$ be obtained as a table for values of the argument $x=\left[x_1,x_2,...,x_n\right]^T$ as a result of observations. The process of smoothing is used in order to reduce the random errors and obtain the smooth function. The process of smoothing is in replacing the values that are obtained as a result of observations, by the values that are obtained as a result of the smoothing treatments.

The problem of smoothing is formulated that way.

It is necessary to find the function $f^*(x) \in L_2^n[a,b]$ in order to achieve the minimum of the functional:

$$J[t] = \sum_{i=0}^{n} p_i \left[y_i - f(x_i) \right]^2 + p \int_a^b \left[f^{(n)}(t) \right] dt, \qquad (4.107)$$

where $p > 0$ - is an auxiliary parameter, p_i - are given numbers (weight coefficients).

If $p = 0$ the task is reduced to the task of approximation at LSM. Having all $p = 0$ the task turns into the task of interpolation.

The smoothing spline is sought in the form

$$S_3(x) = \alpha + \beta x + \sum_{i=0}^{3} c_k^{[i]} x^i, \quad x_k < x < x_{k+1} \qquad (4.108)$$

In order to construct a smoothing spline and to specify the system of equations, it is necessary to apply the additional conditions at the sites of “stitching” of splines. These additional conditions are commonly called boundary conditions. They depend on the nature of the data that are measured, and on the conditions that should be satisfied at points of “stitching”. While smoothing we distinguish: the frequency terms and the conditions for the first or the second derivatives at these points [11].

There is a system for the definition of the coefficient of smoothing spline in the nonperiodic case

$$\begin{aligned}
&a_0 M_0 + b_0 M_1 + c_0 M_2 = g_0, \\
&b_0 M_0 + a_1 M_1 + b_1 M_2 + c_1 M_3 = g_1, \\
&c_{k-2} M_{k-2} + b_{k-1} M_{k-1} + a_k M_k + b_k M_{k+1} + c_k M_{k+2} = g_k; k = \overline{2, N-2}; \qquad (4.109) \\
&c_{N-2} M_{N-3} + b_{N-2} M_{N-2} + a_{N-1} M_{N-1} + b_{N-1} M_N = g_{N-1}, \\
&c_{N-1} M_{N-2} + b_{N-1} M_{N-1} + a_N M_N = g_N.
\end{aligned}$$

The coefficients of this system are determined by formulas

$$a_k = \frac{1}{2}\left(b_{k-1} + h_k\right) + \frac{1}{h_{k-1}^2} p_{k-1} + \left(\frac{1}{h_{k-1}} + \frac{1}{h_k} \right)^2 + \frac{1}{h_k^2} p_{k-1}; k = \overline{1, N-2};$$

$$b_k = \frac{h_k}{6} - \frac{1}{h_k}\left[\left(\frac{1}{h_{k-1}} + \frac{1}{h_k}\right)^2 + \frac{1}{h_{k-1}^2} p_{k-1} + \left(\frac{1}{h_{k+1}} + \frac{1}{h_k}\right) p_{k+1}\right]; k = \overline{1, N-2};$$

$$c_k = \frac{1}{h_k h_{k+1}} p_{k-1}; k = \overline{1, N-3};$$

$$g_k = \frac{z_{k+1}^0 - z_k^0}{h_k} - \frac{z_k^0 - z_{k-1}^0}{h_{k-1}} p_{k-1}; k = \overline{1, N-1}. \qquad (4.110)$$

If the spline that smooths, satisfies the boundary conditions $S''(a) = S''(b) = 0$, then

$$a_0 = a_N = 1; \quad b_0 = c_0 = b_{N-1} = c_{N-1} = g_N = g_0 = 0.$$

In the periodic case the system of the equations has the form

$$c_{k-2} M_{k-2} + b_{k-1} M_{k-1} + a_k M_k + b_k M_{k+1} + c_k M_{k+2} = g_k,\ k = \overline{1, N}. \quad 4.111)$$

where the matrix coefficients are determined by formulas (4.110) for all *k*.

The systems of equations listed above are solved by the method of non-monotonic sweep. After determining the parameters M_k , the value z_k are determined by correlations

$$z_k - z_k^0 = -p_k D_k, k = \overline{0, N}. \qquad (4.112)$$

Moreover

$$D_0 = \frac{1}{h_0}\left(M_1 - M_0\right),$$

$$D_k = \frac{1}{h_k}\left(M_{k-1} - M_k\right) - \frac{1}{h_k}\left(M_k - M_{k-1}\right), k = \overline{1, N-1},$$

$$D_N = \frac{1}{h_{N-1}}\left(M_N - M_{N-1}\right). \qquad (4.113)$$

In the periodic case $h_0 = h_N; M_0 = M_N; M_1 = M_{N-1}$, all value D_k are determined by formulas (4.113) for $k = \overline{1, N}$.

The determination of the weight parameters p_k is the most important aspect while constructing a smoothing spline. The errors of the determination of the magnitude z_k^0 are

usually known in practical applications, viz $\left|z_k - \overline{z}_k\right| \le \delta_k$ are known, where $\overline{z}_k$ - are the exact values of the measured value. In this situation it is natural to demand that the smoothing spline satisfies the conditions

$$\left|z_k - z_k^0\right| \le \delta_k, k = \overline{0,N}, \tag{4.114}$$

or the conditions

$$p_k \left|D_k\right| \le \delta_k, k = \overline{1,N}.$$

These restrictions are used to calculate the weighting factors p_k. Let us construct the iterative process, the implementation of which will provide us with the unknown factors M_k and multipliers p_k.

$$\left(A + 6HR^{(m)}H^T\right)M^{(m)} = 6Hz^0. \tag{4.115}$$

$$p_k^{(m)} = \begin{cases} \dfrac{\delta}{\left|D_m^{(m)}\right|}, \text{if } D^{(m+1)} \ne 0; \\ 0, \text{if } D^{(m+1)} = 0. \end{cases} \tag{4.116}$$

where *m* – is the number of iteration.

The matrices *A*, *H* are

$$A = \begin{bmatrix} 1 & 0 & . & . & \dots & . & . \\ 0 & r_1 & h_1 & . & \dots & . & . \\ . & h_1 & r_2 & . & \dots & . & . \\ .. & .. & .. & .. & .. & .. & .. \\ . & . & . & h_{N-3} & r_{N-2} & h_{N-2} & . \\ . & . & . & . & h_{N-2} & r_{N-1} & 0 \\ . & . & . & . & . & 0 & 1 \end{bmatrix}, \tag{4.117}$$

where $r_k = 1\left(h_{k-1} + h_k\right)$

$$H=\begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ h_0^{-1} & -\left(h_0^{-1}+h_1^{-1}\right) & h_1^{-1} & \dots & 0 \\ .. & .. & .. & ... & .. \\ . & h_{N-1}^{-1} & -\left(h_{N-2}^{-1}+h_{N-1}^{-1}\right) & ... & h_{N-1}^{-1} \\ . & 0 & 0 & ... & 0 \end{bmatrix},$$

$$R=\begin{bmatrix} p_0 & & \\ & \ddots & \\ & & p_N \end{bmatrix}. \qquad (4.118)$$

The connection M_k with z_k is defined by the matrix equation

$$AM = 6Hz. \qquad (4.119)$$

As an initial approximation we take $p_0^0 = 0$ that corresponds to the interpolation spline with the values $D_k = D^{(0)}$. The iterative process continues until the value of spline z_k does not appear in the "corridor".

The Algorithm of Interpolation by Parabolic Splines

Let $f(x) \in C_{[a,b]}, \{a,b] \in R_n; a < b,$ and two sets of knots Δ_n, Δ'_n are presented

$$\Delta'_n: \quad \bar{x}_0 = a < \bar{x}_1 < ... < \bar{x}_{n+1} = b. \qquad (4.120)$$

Assume that $x_{i-1} < \bar{x}_i < x_i, i = \overline{1,n}$. The function $S_2(x; f)$ is the interpolation parabolic spline for the function $f(x)$ if

$$S_2(x) \in P_2; \quad x \in \left(\bar{x}_i, \bar{x}_{i+1}\right); i = \overline{1,n};$$

$$S_2(x) \in C^{(1)}_{[a,b]};$$

$$S_2(x) = f(x_i). \qquad (4.121)$$

The numbers $\bar{x}_i$ are knots of spline, and x_i are knots of interpolation. The knots of spline are the points of a possible gap of the senior derivative (in this case - the second

one). The spline $S_2(x; f)$ depends on $n+3$ parameters, thus, it contains two free parameters. Therefore, two additional restrictions are imposed on the parabolic interpolation spline.

If the function $f(x)$ is $(b-a)$ periodic, we require the spline to be $(b-a)$ also periodic and to have the continuous first derivative, and the point $x_0 = a$ don't be the knot spline.

$$S_2^{(i)}(a) = S_2^{(i)}(b), \quad i = 1,2. \tag{4.122}$$

Generally, the following conditions are the most widespread:

$$S_2'(a) = a_n; \quad S_2'(b) = b_n; \tag{4.123}$$

$$S_2'(a) = A_n; \quad S_2'(b) = B_n, \tag{4.124}$$

where a_n, b_n, A_n, B_n - are given real numbers.

The particular choice of these numbers depends on the task under solution. For example, if the function has corresponding derivatives, we can put $a_n = f'(a), b_n = f'(b), A_n = f''(a), B_n = f''(b)$ or replace them by the appropriate approximated values of the derivatives (the finite differences). If the choice of boundary conditions is difficult we can require the spline to have the continuous second derivative, viz

$$S_2'(z-0) = S_2'(z+0); \quad z = \bar{x}_i, (i = 1, i = n) \tag{4.125}$$

at the points $\bar{x}_1, \bar{x}_n$.

Let $m_i = S_2'(x_1); i = \overline{0,n}; M_i = S_2''(x_i)$. So far as $S_2''(x)$ - a piecewise constant function, that

$$S_2''(x_i) = M_i; \ \bar{x}_i < x < \bar{x}_{i+1}, \ i = \overline{0,n}. \tag{4.126}$$

Let us denote

$h_i = x_{i+1} - x_i, \bar{h}_i = x_{i+1} - \bar{x}_{i+1}, i = \overline{0,n-1};$

$$\sigma_i = f\left(x_{i+1}\right) - f\left(x_i\right)h_i = y_{i+1} - h_i y_i;$$

$$u_i = h_i h_{i+1} + h_i; \quad \gamma_i = \bar{h}_i h_i - \bar{h}_i;$$

$$\nu_i = h_{i-1}h_{i-1} + h_i; \quad \bar{u}_i = u_i\gamma_i; \bar{\nu}_i = \nu_i\gamma_i; i = \overline{0,n-1}. \qquad (4.127)$$

Then, for $x \in \left[x_k, x_{k+1}\right]$ we have

$$S_2\left(x\right) = y_i + m_i\left(x - x_i\right) + c_i\left(x - x_i\right)^2 + d_i\left(x - x_{i+1}\right)^2. \qquad (4.128)$$

We will require that $S_2(x_{k+1}) = f(x_{k+1})$ and $S_{2.}'(x_{k+1}) = m_{k+1}, k = \overline{0,n-2}$. Then, for the coefficients c_k, d_k we obtain

$$c_i = \frac{m_i - 1}{2}\left(\frac{1}{h_i} - \frac{\bar{\nu}_i}{\bar{h}_i}\right) - \frac{m_i}{2}\left(\frac{1}{h_i} + \frac{\bar{\nu}_i}{\bar{h}}\right) + \frac{\sigma_i}{\bar{h}_i}\bar{\nu}_i, \qquad (4.129)$$

$$d_i = \frac{\gamma_i h_i}{\bar{h}_i^2}\left(m_i + \frac{m_{i+1}}{2} - \left(y_{i+1} - y_i\right)\right). \qquad (4.130)$$

So far as the equality $S_2''(x_{k+1} - 0) = S_2''(x_{r+1} + 0)$ should be implemented in the point x_{i+1} we obtain the equality $c_k + d_k = c_{k+1}$ or

$$\frac{1}{\gamma_k \bar{h}_k}m_k + \left(\frac{h_k + \bar{h}_k}{h_k\bar{h}_k} + \frac{1}{h_{k+1}} + \frac{\gamma_k}{\bar{h}_{k+1}}\right)m_{k+1} + \frac{1}{\gamma_{k+1}h_{k+1}}m_{k+2} =$$

$$= 2\frac{\sigma_k}{\bar{h}_k} + 2\frac{\sigma_{k+1}}{h_{k+1} + \bar{h}_{k+1}}. \qquad (4.131)$$

In the periodic case $k = \overline{0,n-2}$, at this $m_0 = m_n$; $m_1 = m_{n+1}$; $h_0 = h_n, h_{n-1} = h_1$, that we have the system of equations from *n* unknown $m_1, m_2, ..., m_n$.

m_0, m_n are known in the case of boundary conditions (4.111), thau, the system (4.123) contains $(n-1)$ the equation with $(n-1)$ unknowns. For the boundary conditions (4.131) we have

$$2c_0 = A_n, 2c_{n-1} + 2d_{n-1} = B_n.$$

For the boundary conditions (4.125):

$$d_0 = d_{n-1} = 0;$$

$$d_k = e_k = 2u_k \frac{h_{k-1}}{\bar{h}_{k-1}} \sigma_{k-1} + 2\nu_k \gamma_k \frac{h_k}{\bar{h}_k} \sigma_k; \quad k = \overline{1, n-1}. \tag{4.132}$$

In the periodic case the system of equations (4.131) relatively to m_k takes the form

$$\bar{u}_k m_{k-1} + \left(\bar{u}_k + \nu_k + 1\right) m_k + \sigma_k m_{k+1} = g_k; k = \overline{1, n}. \tag{4.133}$$

$$m_0 = m_n; \; m_{n+1} = m_1; \; h_n = h_0, \bar{h}_n = \bar{h}_0.$$

For the boundary conditions (4.123) the system of equations takes the form

$$Am = g, \tag{4.134}$$

$$m = \left[m_1, m_2, ..., m_{n-1}\right];$$

$$g = \left[g_1 - \bar{u}_1 a_n, g_2, g_3, ..., g_{n-2}, g_{n-1} - \bar{\nu}_{n-1} b_n\right]. \tag{4.135}$$

For the boundary conditions (4.124) the system of equations takes the form

$$\left(2 - \frac{\bar{h}_0}{h_0}\right) m_0 + \frac{\bar{h}_0}{h_0} m_1 = 2\sigma_0 - A_n \left(h_0 - \bar{h}_0\right);$$

$$\bar{u}_k m_{k-1} + \left(\bar{u}_k + \bar{\nu}_k + 2\right) m_k + \bar{\nu}_k m_{k+1} = g_k; k = \overline{1, n}; \tag{4.136}$$

$$\frac{1}{\gamma_{n-1}} m_{n-1} + \frac{h_{n-1} + \bar{h}_{n-1}}{h_{n-1}} = 2\sigma_{n-1} + B_n \bar{h}_{n-1}.$$

The Least Squares Method

The minimization of the criterion (4.104) is the determination of the estimates of the parameters b_k, $k = \overline{1, m}$ and a_i, $i = \overline{1, n}$, and it is implemented by solving the system of linear algebraic equations

$$\begin{cases} \left\{\left[\sum_{i=1}^{m} b_i x_k^i - \sum_{i=1}^{n} a_i \frac{d^i y_k}{dt^i}\right] x_k^l\right\} = \sum_{k=1}^{N} y_k x_k^l, \quad l = \overline{0, m}; \\ \left\{\left[\sum_{i=1}^{m} b_i x_k^i - \sum_{i=1}^{n} a_i \frac{d^i y_k}{dt^i}\right] \frac{d^l y_k}{dt^l}\right\} = \sum_{k=1}^{N} y_k \frac{d^l y_k}{dt^l}, \quad l = \overline{0, n}; \end{cases} \tag{4.137}$$

Let us denote

$$c=\left[b,a\right]^T,\ e_l=\sum_{k=1}^{N}y_k x_k^{l-1},\quad l=\overline{0,m};\ e_l=\sum_{k=1}^{N}y_k\frac{d^l y_k}{dt^l},\quad l=\overline{0,n}; \tag{4.138}$$

$$d_{1,1}=N,\ d_{j,l}=d_{l,j}=\sum_{k=1}^{N}x_k^j x_k^l,\ l=\overline{0,m};$$

$$d_{l,m+j}=\sum_{k=1}^{N}\frac{d^j y_k}{dt^j}x_k^{l-1};\ l=\overline{0,m},\ j=\overline{0,n}; \tag{4.139}$$

$$d_{l+m,j}=\sum_{k=1}^{N}\frac{d^l y_k}{dt^l}x_k^{j-1};\ l=\overline{0,m};\ j=\overline{0,m};$$

$$d_{l+m,j+n}=\sum_{k=1}^{N}\frac{d^j y_k}{dt^j}\frac{d^l y_k}{dt^l};\ l=\overline{0,n};\ j=\overline{0,n}.$$

Then, the system of linear algebraic equations for determining the unknown parameters b_k, $k=\overline{1,m}$ and a_i, $i=\overline{1,n}$, has the "classic" form

$$Dc=e. \tag{4.140}$$

The example. *The definition of the nonlinearity automated electric drive*

In order to control the stability and accuracy of the automated electric drive it is necessary to determine the size and asymmetry of the dead zone, the slope and the levels of saturation of the nonlinear dependence of the rotational speed of the output shaft of the system on the voltage U_a at the shaft of the armature of the DC motor in the reverse mode, viz the speed of transition from $-\Omega_{max}$ to $+\Omega_{max}$ [38]. The differential equation corresponds to dynamics of this process:

$$a_2\frac{d^2\Omega(t)}{dt^2}+a_1\frac{d\Omega(t)}{dt}+\Omega(t)=f\left[U_a(t)\right], \tag{4.141}$$

where U_a changes at the interval from $-U_{max}$ to $+U_{max}$, it is stepwise:

$$U_a(t)=U_{max}\left[-1+\frac{1}{q/2}\sum_{k=1}^{q}1(-k\Delta t)\right],$$

where $q=16;\ 1(t)=\begin{cases}1, t>k\Delta t;\\ 0, t<k\Delta t.\end{cases}$

The values of parameters a_1, a_2 are unknown. The parameters of testing impact are equal to $U_{max}=120\text{B}$, $\Delta t=1c.$ The initial condition: $\Omega(0)=-\Omega_{max}=-300$ rad/s, $d\Omega(0)/dt=0.$

In order to model the process, let us take the nonlinear dependence as

$$\Omega\left[U_a(t)\right]=3\left[U_a(t)\right]-60\sin\left(0{,}065U_a(t)\right),$$

that corresponds to the condition of smoothness of dependence with dead zones and saturation zones. The measurements of the output signal of the drive are executed with the step $\Delta t=1c$ at $N=200$, $q=16.$

So far as the measurements of the output variable are performed under the influence of interference measurements, for the process modeling we impose a 10% error of measurements in the form of “white noise”, that is the random normal process with zero mean and unit variance.

The process modeling includes the following steps:

1. Smoothing of the input $U_a(t)$ and output $\hat{y}_k$, $k=\overline{1,N}$ values that are measured.

2. Sorting the values $U_a(t)$ in the order of their increasingness.

3. The interpolation of the values using the interpolation splines.

4. The determination of the values t_{kj} that correspond to the uniform change $U_a(t_k)$ with the constant step ΔU.

5. The calculation of the values $\hat{y}(t_{kj})$ and derivatives of the first and the second order from these values (the finite differences).

6. The minimization of the functional (4.96) by composing and solving the systems of the equations (4.140).

The simulation of the process of determination of the static nonlinearity is performed for several values m in order to define the optimal order of the polynomial dependence of the static nonlinearity on the input influence. In the fig. 4.7 we see that the optimal value is $m=5$. The mean square error is $\varepsilon_5 = 26{,}72$.

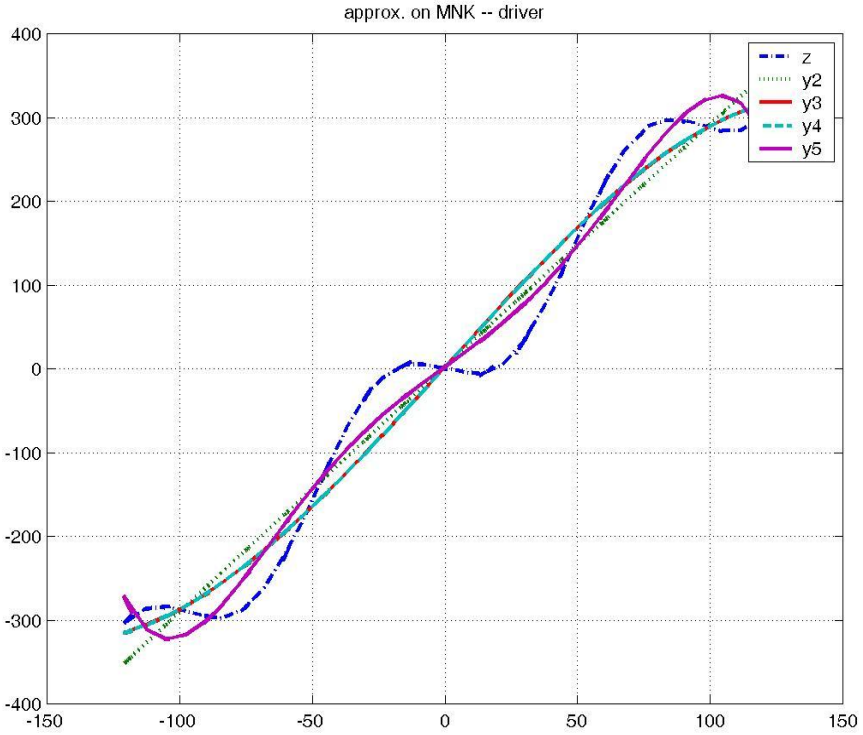


Fig. 4.7. The estimation of the nonlinearity for $m = 2,3,4,5$ by LSM.

Let us solve the same task applying the method of compensation of dynamics of the control object.

The first three steps are the same. Having performed the steps 4 and 5, viz the the definition of time intervals that correspond to the constant increment of the values of the testing signal and determination of the values of the output signal that correspond to the value of time t_{kj}, instead of the step 5 we do

7. The minimization at β_1, β_2 of the functional (4.99) in this form

$$\min J_{cor} = \min \sum_{k=1}^{N} \left(\frac{\partial^r \hat{y}_{cor}(t_k)}{\partial U^r(t)} \right),$$

provides the estimates $\hat{\beta}_1 = 0{,}0196,\ \hat{\beta}_2 = 0{,}000134.$ Then, the sought nonlinearity is defined as follows:

$$f\left[U_a(t) = y(t)\text{-}0{,}0196\frac{dy(t)}{dt} - 0{,}000134\frac{d^2y(t)}{dt^2}\right].$$

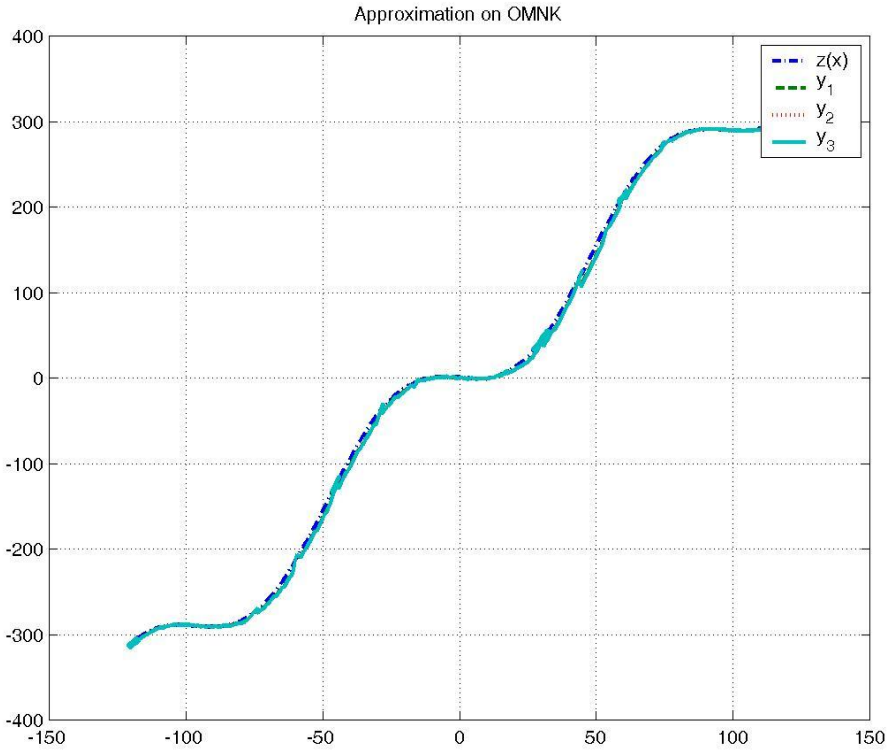


Fig. 4.8. The estimation of the static nonlinearity by the method of compensation of dynamics for $r = 1{,}2$ and $r = 3.$

The results of modeling for $r = 1{,}2{,}3$ are represented in the fig. 4.8.

The final error is the least for $r = 2$ and is equal to $\varepsilon_2 = 3{,}067$, that is less than 0,5%.

Chapter 5. Constructing of the Models of the Deterministic Processes, the Measurements of which are Perturbated by Random Interferences

5.1. Modeling, Identification and Forecasting of Economy Performance by Methods of the Simplified Confluent Analysis

The indicators of the economy (power industry, etc) as a function of time may have the most diverse structure, longitude of a range, accuracy, the type of hidden conformity of development (change) in time, the step in time, the interval of the forecast, etc [26].

The elements that make up the model of time series can be analytic functions of time t (degree: t^j, $-\infty < j < \infty$; periodic trigonometric: $\sin(\omega t + \varphi)$; combined: $\sin^i(\omega t + \varphi)$ and other exactly given functions: $f(t)$) or suspended in the time t values of the series $y(t)$ ($y(t-T), 0 < T < t_k$, where t_k - the finite value of time). As a rule, the time t, represented by the discretes t_k, k=0,1,2,…, is not always the uniform step.

If the structure of the model of series is unknown, we can choose the best structure from the different variants of structures constructed on these elements (according to the main criterion *I* (the criterion of the accuracy of forecast)). The models with the exponential elements are generally better used for the short series and the autoregressive models are used – for the long ones (where the uncertainty does not influence much on the initial conditions).

Let us represent the index *I* of the forecasting accuracy (which is physically realized) as the weighted sum of the partial indexes I_i (i = 1,2,3) that are responsible for the quality of individual properties of the model of the series. The index I_1

$$I_1 = \frac{1}{n}\mathrm{tr}\left[\mathrm{diag}\frac{\left|\hat{\beta}_i^e - \hat{\beta}_i^o\right|}{\left|\hat{\beta}_i\right|}\right], \quad i = \overline{1,n}; \tag{5.1}$$

where $\hat{\beta}_i^e$, $\hat{\beta}_i^o$ and $\hat{\beta}_i$ - are the estimates of *i* parameters of the model, that are received by sorting even, odd, and all discrete *k* of the time t_k; it is so called the parametric index of regularity.

The index I_2 :

$$I_2 = \left(\varepsilon^T \varepsilon\right) \cdot \left(x^T x\right)^{-1}, \tag{5.2}$$

$$\varepsilon^T = \left[\varepsilon(1),...,\varepsilon(M)\right],\ \hat{x}^T = \left[\hat{x}(1),...,\hat{x}(M)\right],$$

where $\varepsilon(k)$ - is an error of approximation of signal $x(k)$ by the corresponding model in the k point of series; it is the signal indicator of the unbiasedness or the accuracy of the series simulated by the model.

The index I_3:

$$I_3 = |1-K|, \qquad K = \frac{\sum_{i=1}^{L} \eta_i \left| x(M-i) \cdot \hat{x}(M-i) \right|}{\sum_{i=1}^{L} \eta_i \left| x(M-i) \right| \cdot \sum_{i=1}^{L} \eta_i \left| \hat{x}(M-i) \right|}. \qquad (5.3)$$

Here η_i - is the coefficient of distribution of the desired accuracy of forecast by the L last points of sampling $x(k), \quad k = \overline{1,M}, \quad \sum_{i=1}^{L} \eta_i = 1; \ \hat{x}(M-i)$ - is a predictive value $(M-i)$, received from the model that is constructed on a shortened L of the last points of the sample $k = \overline{1,M-L}.$

Since it is considered that the estimated number $x(k)$ consists of the hidden determinate component which is smoothed in time and the random component that is close to the Gaussian "white noise", then, in the variation series of models that are ordered by complexity (the dimensionality of the vector β of unknown parameters), the indexes I_1 and I_3 restrict the dimensionality n of the vector β, while the index I_2 decreases with the increasing n. Depending on the purpose of identification the coefficients of the weight g_i of the weighted sum of these three indicators vary:

$$I = \sum_{i=1}^{3} g_i I_i, \quad \sum_{i=1}^{3} g_i = 1, \quad g_i \geq 0. \qquad (5.4)$$

For the task of the control of parameters β_i of the model of the known structure, the maximum weight is g_1; for the task of the accurate approximation of series $x(k)$ by the model $\hat{x}(k)$, the maximum weight is g_2; for the forecasting task, the maximum weight is

g_3. The set of indexes I_1, I_2, I_3 provides the compromise between the stability of the model's estimates, the accuracy $\varepsilon^T\varepsilon$ of approximation and the accuracy of forecast.

As an example of the real time series that have 43 discretes $x(k)$ with a uniform step $\Delta t = 4$ months (one of the indexes in power industry of Ukraine), the fig. 5.1. and the table. 5.1., let us consider the solution of the forecasting task $x(k)$, $k = \overline{1,37}$ for the last 6 points, taking them as unknown. This formulation of the problem makes it possible to implement the objective index *I* of relative accuracy of the forecast for these 6 points that was physically unimplemented in the forecast for the future:

$$I = \frac{[\varepsilon(38),...,\varepsilon(43)]\cdot[\varepsilon(38),...,\varepsilon(43)]^T}{[x(38),...,x(43)]\cdot[x(38),...,x(43)]^T}, \qquad (5.5)$$

viz the relative mean square deviation $\varepsilon(k) = \hat{x}(k) - x(k)$, $k = \overline{38,43}$, of the forecasting values $\hat{x}(k)$ from the known $x(k)$. In the criterion we have $\eta_i = \frac{1}{L} = \frac{1}{6}$.

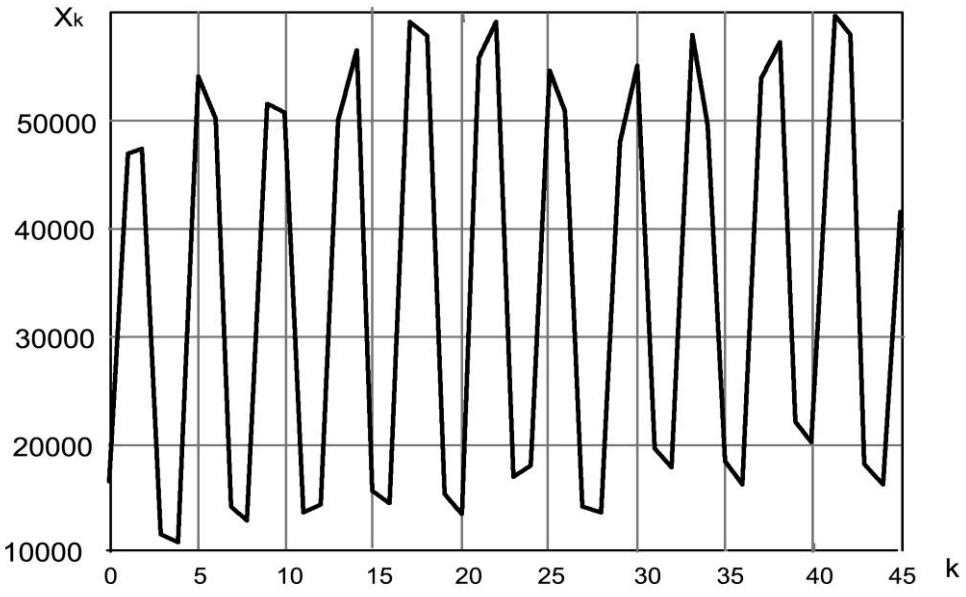


Fig. 5.1. The time series of one economic indicator in the energy sector of Ukraine.

The weight coefficients in the generalized criterion (5.4) are $g_1 = 0,9$; $g_2 = 0,4$; $g_3 = 0,4$.

Table 5.1

The dependence $x(k)$ on k

k	X	k	x	k	x	k	x
1	10550	12	13600	23	59250	34	57800
2	47070	13	14550	24	16850	35	49630
3	47350	14	49900	25	17830	36	18780
4	11500	15	56750	26	54800	37	16000
5	10900	16	16050	27	50700	38	53950
6	54700	17	14680	28	14150	39	57500
7	50000	18	59300	29	13550	40	22000
8	14320	19	57700	30	47430	41	19650
9	12900	20	15500	31	56350	42	59900
10	51650	21	13350	32	19450	43	57550
11	50740	22	55600	33	17600		

The mean value of series is 35100, the mean square deviation of series from the mean is 19550, the coefficient of variation is 0,55. The various mathematical models were associated with the series that are represented in the table 5.1.

1) the models as a different order of the degree polynomial from *t*, viz from the discretes *k* of time:

$$\hat{x}(k) = \beta_0 + \beta_1 k, \tag{5.6}$$

$$\hat{x}(k) = \beta_0 + \beta_1 k + \beta_2 k^2, \tag{5.7}$$

$$\hat{x}(k) = \beta_0 + \beta_1 k + \beta_2 k^2 + \beta_3 k^3, \tag{5.8}$$

$$\hat{x}(k) = \beta_0 + \beta_1 k^{\frac{1}{4}} + \beta_2 k^{\frac{1}{3}} + \beta_3 k^{\frac{1}{2}} + \beta_4 k^{\frac{3}{2}}, \quad (5.9)$$

$$\hat{x}(k) = \beta_0 + \beta_1 k + \beta_2 k^{-1} + \beta_3 k^{-3}, \quad (5.10)$$

2) the autoregressive model from *k* with the constant and variable step:

$$\hat{x}(k) = \beta_0 + \beta_1 x(k-1), \quad (5.11)$$

$$\hat{x}(k) = \beta_0 + \beta_1 x(k-1) + \beta_2 x(k-2), \quad (5.12)$$

$$\hat{x}(k) = \beta_0 + \beta_1 x(k-1) + \beta_2 x(k-2) + \beta_3 x(k-3), \quad (5.13)$$

$$x(k) = \beta_0 + \beta_1 x(k-4), \quad (5.14)$$

$$\hat{x}(k) = \beta_0 + \beta_1 x(k-1) + \beta_2 x(k-2) + \beta_3 x(k-3) + \beta_4 x(k-4), \quad (5.15)$$

$$\hat{x}(k) = \beta_0 + \beta_1 x(k-1) + \beta_2 x(k-4), \quad (5.16)$$

$$\hat{x}(k) = \beta_0 + \beta_1 x(k-1) + \beta_2 x(k-4) + \beta_3 x(k-8), \quad (5.17)$$

3) the combined polynomial autoregressive models:

$$\hat{x}(k) = \beta_0 + \beta_1 k + \beta_2 x(k-1), \quad (5.18)$$

$$\hat{x}(k) = \beta_0 + \beta_1 k + \beta_2 x(k-4), \quad (5.19)$$

$$\hat{x}(k) = \beta_0 + \beta_1 k + \beta_2 x(k-1) + \beta_3 x(k-4). \quad (5.20)$$

The efficiency of application of the physically realizable criterion (5.4) in terms of its proximity to the physically unrealizable criterion (5.5) was tested at a set of structures (5.6) ÷ (5.20) of the models of time series (table 5.1.) and a set of methods (LSM, GLSM, ILSM,) that were suggested in the previous chapter for the forecasting tasks. Here, the efficiency should be read as a right choice of the best method by the criterion (5.5) that is selected by the criterion (5.4).

The results of numerical simulation are represented in the table .2., where 15 lines of the table represent:

in columns

1 –the types of models (degree (5.6)-(5.10), autoregressive (5.11)-(5.17), combined (5.18)-(5.20));

2 –the relative mean-square error of simulation of series by corresponding model in its $k = \overline{1,37}$ identification by LSM;

3 –the physically unrealizable ideal criterion (5.5) by LSM;

4 –the physically realizable criterion (5.4) in the identification of the model by LSM;

5 –the best (by criterion (5.4)) method of identification for the model that corresponds to the row;

6 –the value of the ideal criterion (5.5) for chosen by real criterion (5.4) for the appropriate method to model line;

7 – the value of the criterion (5.4) for the best method of identification, chosen with its help for appropriate to row model;

8 –the best method of identification the appropriate to row model, chosen by the ideal criterion (5.5);

9 –the value of the ideal criterion (5.5) for the best method of identification the model of appropriate row, chosen by this criterion;

10 –the value of the real criterion (5.4) for the best method of identification of model the appropriate row, chosen by this criterion (5.5).

Table 5.2.

The results of modeling

№	1	2	3	4	5	6	7	8	9	10	11
1	5.6	0,49	0,47	0,25	GLSM	0,41	0,242	MAV	0,363	0,26	1,3
2	5.7	0,484	0,593	0,26	ILSM	0,415	0,223	ILSM	0,415	0,223	1,43
3	5.8	0,476	0,883	0,41	ILSM	0,38	0,2	ILSM	0,38	0,2	2,32
4	5.9	0,485	0,593	0,27	GLSM	0,43	0,226	ILSM	0,365	0,235	1,62
5	5.10	0,488	0,49	0,25	ILSM	0,45	0,23	GLSM	0,425	0,237	1,15
6	5.11	0,49	0,435	0,24	GLSM	0,42	0,235	GLSM	0,415	0,235	1,05
7	5.12	0,62	0,58	0,28	GLSM	0,56	0,262	GLSM	0,558	0,262	1,04
8	5.13	0,123	0,143	0,04	LSM	0,143	0,04	MAV	0,096	0,048	1,49
9	5.14	0,133	0,1	0,03	LSM	0,1	0,03	MAV	0,088	0,126	1,13
10	5.15	0,113	0,122	0,037	MAV	0,092	0,03	MAV	0,092	0,03	1,33
11	5.16	0,131	0,103	0,034	MAV	0,091	0,031	MAV	0,091	0,031	1,13
12	5.17	0,087	0,092	0,015	MAV	0,063	0,011	MAV	0,063	0,011	1,46
13	5.18	0,488	0,47	0,245	MAV	0,489	0,225	GLSM	0,411	0,237	1,14
14	5.19	0,132	0,108	0,035	LSM	0,108	0,035	MAV	0,081	0,038	1,33
15	5.20	0,131	0,111	0,036	LSM	0,111	0,037	LSM	0,111	0,037	1

Let us analyze the results of calculations:

The autoregressive model (5.17) with the variable delay at *k*-1, *k*-4 and *k*-8 steps is the best one by the ideal criterion (5.5) at a set of 15 structures and 4 methods of identification for specific row (table 5.1); the best method is MAV. The same result is obtained by the real criterion (5.4).

Generally, the optimal method of identification by the real criterion (5.4) was selected correctly (rows 2, 3, 6, 7, 10, 11, 12, 15, the table 5.2) in 8 of 15 considered cases, viz it coincided with the method selected by the ideal criterion. In other 7 cases (rows 1, 4, 5, 8, 9, 13, 14) the ideal index (5.5) for the method, selected by the real index (5.4), is slightly worse than this index for the optimal method by the ideal index (columns 6,9, fig. 5.2).

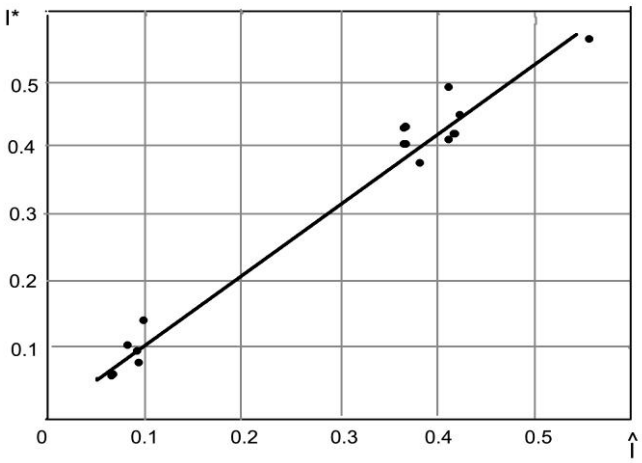


Fig. 5.2. The dependence of the ideal index I^* (5.5), obtained for the optimal method of identification on the same index for the optimal method of identification by the real index $\hat{I}$.

The conclusion that there is a strong correlation of indexes (5.4) and (5.5) and, as a consequence, the possibility of the effective application of the physically realizable criterion (5.4) clearly follows from the fig. 5.2.

With the complication of the models (5.6) (5.7) (5.8), that are degree series, the index (5.2) (the column 2 of the table 5.2) of the mean square error of approximation of the

series (the table 5.1) by models (5.6)÷(5.8) decreases. It follows from the first Weierstrass theorem [11] on the approximation by degree polynomials (Taylor series).

At the same time the ideal criterion (4.5) of the forecasting accuracy at the complication of models worsens (rows 1,2,3 of the third column in the table 5.2). This confirms the biased nature of the internal approximative criterion (5.2) and the incorrectness of its application for the task of forecast.

Another situation occurs for autoregressive and mixed autoregressive polynomial models (5.11)÷(5.20). Here due to the regularizing properties of the LSM, when variables are noisy, the internal criterion (5.2) of the mean-square measure of the proximity in the area of approximation and the external one, both ideal (5.5) and actual (5.4), criteria are rather strongly correlated (the fig. 5.3).

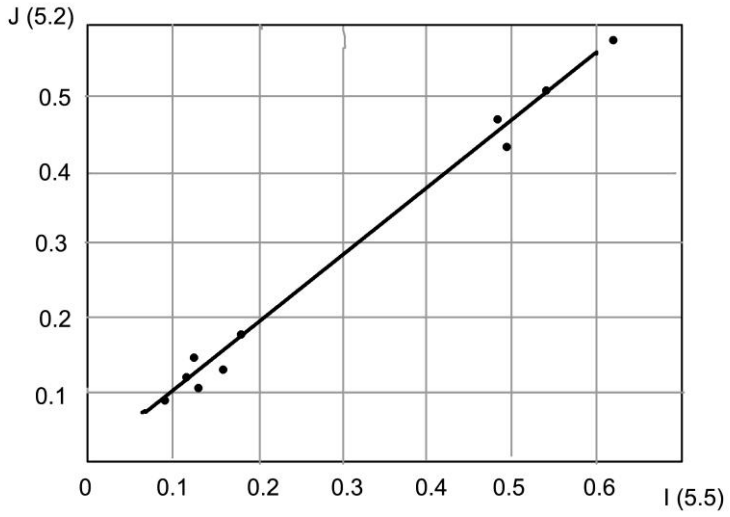


Fig. 5.3. The regressive dependence J (5.2) on I (5.5).

In other words, the application of the approximative criterion (5.2) for this class of models in the forecasting task to the points (38÷43) at the noisy data in the points (1÷37) is less critical. We have the self regularization here. The more complex is the autoregression, the worse is the conditionality of the information matrix LSM for the accurate data. But for the data, noisy by the uncorrelated obstacle, the diagonal elements of this matrix increase and, as a result, the LSM-estimates of the coefficients of the model reduced (by the module), thus simplifying the model (regularizing by Tikhonov [14]).

Let us compare the value of ideal criterion (5.5) for the models, obtained by the LSM (the column 3) and one of the proposed methods (the column 6), with the optimization at the real criterion (5.4). The index (5.5) is slightly lower only for the model (5.18) from these 15 models. Thus, only here the MAV was mistakenly selected by the criterion (5.4) instead of the LSM. In other 14 cases the method, that was found at a condition of minimum of the physically realizable criterion (5.4) of the forecast accuracy, gives better results than the LSM or it gives the same results if in (5.4) the LSM was selected as the best one (the columns 6 and 3 of the table 5.2).

Within the framework of one method of identification, for example, the MAV (the column 6, the lines 10 ÷ 13), the dispersion of the ideal criterion (5.5), depending on the structure of the model, is from 0.063 to 0.489, that confirms the relevance of the choice of the model's structure.

Within the framework of one model (5.17), for example, optimal by the criterion (5.5), the optimization of the solution at a set of four methods (LSM, MAV, GLSM, ILSM) gives the gain of 1.5 times (0.092 - 0.063 for the LSM and 0,063 for the MAV, as the best method).

In the large, the optimization at a set of methods and models gives significant gains in the accuracy of forecast. Let us define this gain as the ratio criterion (5.5) for the model with the coefficients that are determined by the LSM (the column 3, the table 5.2), to the value of the same criterion (5.5) for the same model, with the coefficients that are defined by the optimal (5.5) method (the column 9, the table 5.2). The column 11 of the table 5.2 represents the ratio that lies between 1 to 2.32. Upon the average it is 1.33.

5.2. The Numerical Modeling of the Task of Active Identification of the Parameters of Discrete Dynamical Systems

The theoretical principles on the advantages of the method of active experiment with the coordinate-wise optimization, discussed in the section 5.1, were also confirmed by the numerical simulation that is considered below. As it is known, the continuous system

$X(t) = A_1 X(t) + B_1 U(t)$ that describes the dynamics of the real process (the electromechanical object) at a set of the discretes of time t_k, provided with the uniform step Δt, can be displayed by the discrete system:

$$X(k+1) = AX(k) + BU(k),\ Y(k) = X(k) + N(k), \qquad (5.21)$$

where $k = \overline{1,M}$; $X(k)$ - are the variables of the state, $U(k)$ -are the control influences; $N(k)$ - are the errors of measurements of $Y(k)$ of variable $X(k)$. In the example below $X(k)$, $Y(k)$, $N(k)$- are three-dimensional vector functions from k, $k = \overline{1,300}$; $U(k)$ - are two-dimensional functions k; $N(k)$ - the sequence of "white noises", the ratio of "noise – signal" is10%. The scalar form of the equation (5.21):

$$x_i(k+1) = \sum_{j=1}^{3} a_{ij} x_j(k) + \sum_{l=1}^{2} b_{il} u_l(k), \quad i = 1,2,3 \qquad (5.22)$$

The numerical value of the coefficients of the mathematical model (5.21) of the electromechanical object is shown in the table 5.3.

Table 5.3.

The numerical value of the coefficients

	a_{ij}			b_{il}	
	$j=1$	$j=2$	$j=3$	$l=1$	$l=2$
$i=1$	-1,27	0,01	0,05	-1	0,8
$i=2$	-0,34	-0,13	1	-0,01	-12,5
$i=3$	-16	-3	0,18	-1	0

We should find the estimates of parameters a_{ij}, b_{il}, ($i = 1,2,3$, $l = 1,2$) using the integrated method of the least square (ILSM), the methods of synthesis of the optimal testing signals $U1(k)$, $U2(k)$ and the technology of the coordinate-wise optimization of the plan of experiment that were discussed in the section 4.3.

The identification is performed row by row. But for each i-th row of the system (5.22) the matrices of the second derivatives of the minimized functional are the same and form the information matrix ILSM:

$$M = \sum_{-\tau}^{+\tau} \left[F^T \left(F_{-\tau} + F_{+\tau} \right) + \left(F_{-\tau} + F_{+\tau} \right)^T F \right], \qquad (5.23)$$

where F - [300×5] is the matrix of the plan of experiment, $F_{\pm\tau}$ -is the same matrix, shifted in time to the discrete k, $k = 1,2,3$.

Each of the testing signals $U1(k)$ and $U2(k)$ is defined as a sequence of two heteropolar pulses of the same form. At this the first pulse on the time axis was fixed, but the position of the last three pulses varied discretely. The dependence of the functionals F_j, $j = 1,5$ of the coordinate-wise optimization (4.46) on the number of the variant of the plan of experiment (8 options are taken) is presented as an example in the fig. 5.4 for the coefficients of the first row of the table 5.

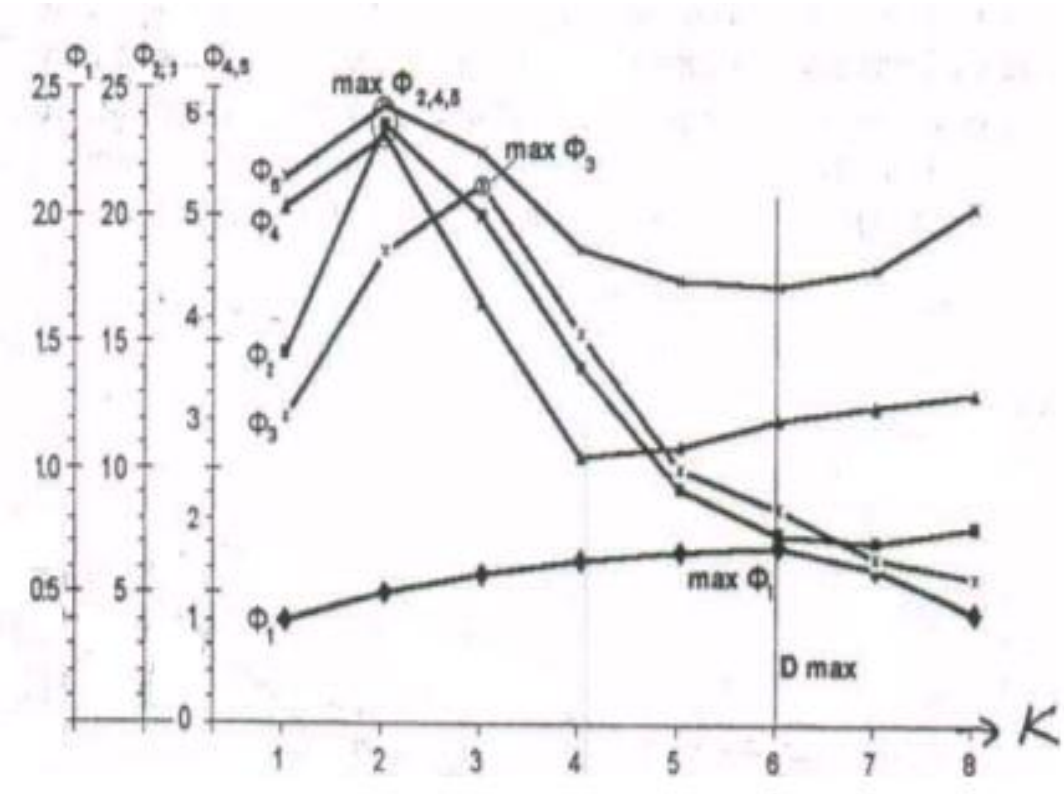


Fig. 5.4. The graphs of the change of functionals F_j of the coordinate-wise optimization on the number k of the plan of experiment.

As it is shown in the figure, the maximums of the functionals of optimization F_{ij} and *D*-criterion [44], viz the normalized determinant of the matrix (5.23), are significantly different for all i , except i = 1. This confirms the effectiveness of the algorithm of the coordinate-wise optimization.

The estimates of the parameters $\hat{a}_{ij}$, $\hat{b}_{il}$ of the model (5.22) are shown in the table 5.4. They are obtained by the ILSM using the results of the experiment which is optimal by *D*-criterion. Let us compare these estimates with the true values of the coefficients a_{ij} , b_{il} (the table 5.3). As it follows from the comparison, the optimality by D-criterion does not guarantee the equivalence on accuracy for each of the ratios.

Table 5.4

The numerical values of the estimates of the coefficients for the optimal plan by *D*-criterion

	$\hat{a}_{ij}$			$\hat{b}_{il}$	
	$j=1$	$j=2$	$j=3$	$l=1$	$l=2$
$i=1$	-1,38	0,007	0,063	-0,99	0,83
$i=2$	-0,3	-0,17	0,93	-0,011	-12,52
$i=3$	-16,24	-3,63	0,179	-1,04	0,0014

The estimates of the coefficients $a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}$, a_{32}, b_{32} are significantly different.

The estimates of the parameters of the model (5.22), obtained by the ILSM as the results of the active experiment on using the coordinate-wise optimization procedure are represented in the table 5.5. These estimates $\hat{a}_{ij}$, $\hat{b}_{il}$ are the best approximated to the precise (table 5.3) values of the coefficients.

Table 5.5

The numerical value of the estimates of the coefficients for the plan of coordinate-wise optimization

	$\hat{a}_{ij}$			$\hat{b}_{il}$	
	$j=1$	$j=2$	$j=3$	$l=1$	$l=2$
$i=1$	-1,28	0,008	0,052	-0,99	0,83
$i=2$	-0,32	-0,14	0,99	-0,011	-12,52
$i=3$	-15,8	-3,28	0,1804	-1,03	0,0013

Several algorithms have been tested when optimizing the information matrix *M* (5.23) by the methods of nonlinear programming. The working of the gradient algorithm with simultaneously descent in all directions and the constant step is presented in the tables 5.6, 5.7. The same algorithm is presented in the table 5.8, but the length of step changes when the sign gradient changes. The table 5.9 presents the working of the algorithm with the constant small step in all directions.

The same algorithm is presented in the table 5.10, but with the shifting of the initial conditions for one of the coordinates. The convergence to the local extremum is obvious. If we substantially change the initial conditions in the same coordinate, then there will be the convergence to another local extremum (the table 5.12). The algorithm of the gradient descent is presented in the table 5.13. The working of the gradient algorithm is illustrated in the table 5.14, where the length of step is taken as a proportion to the gradient with the integrated limit $Q(x)=x^T x$. The working of the gradient algorithm of the minimization of criterion coordinate-wise in the form $\Phi_i = \sqrt{\sum_{j=1}^{n} a_{ij}^2}$, for maximum (the table 5.15) and minimum (the table 5.16) diagonal elements of the matrix M (5.23) is represented in the tables 5.15 and 5.16.

The results of the simulation of the gradient algorithms of optimization of the plan of experiment showed the efficiency of the gradient procedures. For further practical application it is reasonable to use the algorithm with the simultaneous descent in all

directions with the small constant step of descent. However, it is necessary to take into account the possible multiexperimentality of the functional. In order to find the global extremum we should carry out the gradient process several times with the different initial conditions.

Table 5.6

The optimization of the determinant of the information matrix for different initial conditions

№	Gradients			Control			*Det (M)*
	V_1	V_2	V_3	k_2	k_3	k_4	
1	2	3	4	5	6	7	8
1	0,0567	0,0341	0,0522	100	50	200	0,96
2	0,1702	0,1268	0,1304	88	38	188	3,67
3	0,2610	0,0648	0,1628	76	26	176	10,17
4	0,3110	0,2027	0,2587	64	14	164	16,05
5	- 0,2111	0,4469	0,1051	52	2	152	20,21

Remark: 1. The algorithm is stopped due to the approximation of the moment of the pulse delivery before starting the counting time; 2. The step of the algorithm is $\Delta=12$.

Table 5.7

1	2	3	4	5	6	7	8
1	0,0556	0,0424	0,0530	100	150	200	0,74
2	0,1189	0,1118	0,0976	88	138	188	2,16
3	0,1721	0,0880	0,1083	76	126	176	5,16
4	0,2496	0,1994	0,2089	64	114	164	8,571
5	0,4569	0,2179	0,0068	52	102	152	9,43

Table 5.8

№	Gradients			Control			*Det (M)*
	V_1	V_2	V_3	k_2	k_3	k_4	
1	-0,0031	-0,0059	0,0117	150	50	180	0,221
2	0,6150	-0,0700	0,0865	162	62	168	1,789
3	0,2020	0,1567	0,0485	156	74	156	4,354
4	0,1548	-0,0269	0,1320	150	68	144	4,425
5	0,0805	-0,2831	-0,0888	144	71	132	4,611
6	-0,2769	-0,2932	0,0875	138	74	138	5,450
7	-0,0070	-0,0331	0,0244	141	77	135	5,743

Table 5.9

№	Gradients			Control			*Det (M)*
	V_1	V_2	V_3	k_2	k_3	k_4	
1	-9,271	5,980	-5,511	80	50	150	4,679
2	-8,829	6,707	-7,953	82	48	152	5,443
3	-7,962	7,708	-11,26	84	46	154	6,366
4	-6,216	8,931	-14,81	86	44	156	7,389
5	-3,083	10,325	-19,29	88	42	158	8,440
6	2,022	11,855	-24,45	90	40	160	9,322
7	-0,104	11,730	-29,16	89	38	162	12,32
8	4,359	11,851	-33,95	90	36	164	13,88
9	1,379	9,683	-37,23	87	32	166	17,19
10	-7,088	5,565	-39,01	87	32	166	21,61
11	-2,009	2,709	-39,98	88	30	170	22,68
12	10,768	-0,216	-12,48	90	28	172	24,98

Table 5.10

№	Gradients			Control			*Det (M)*
	V_1	V_2	V_3	k_2	k_3	k_4	
1	14,919	11,328	-12,94	100	50	150	0,532
2	9,335	8,416	-10,144	98	48	152	1,120
3	8,479	9,184	-12,180	96	46	154	2,170
4	7,548	10,442	-15,618	94	44	156	3,834

5	5,552	11,480	-19,849	92	42	158	6,211
6	2,022	11,855	-24,444	90	40	160	9,322
7	-3,208	11,252	-28,978	88	38	162	13,09
8	0,564	11,197	-33,632	89	36	164	14,76
9	-2,855	9,977	-36,938	88	34	166	18,16
10	8,322	7,118	-39,551	89	32	168	19,48
11	-2,009	2,709	-39,983	88	30	170	22,68
12	4,487	-1,057	-39,169	89	28	172	23,22
13	2,132	3,100	-36,107	88	29	174	23,41

Table 5.11

№	Gradients			Control			*Det (M)*
	V_1	V_2	V_3	k_2	k_3	k_4	
1	-2,615	1,559	1,452	160	50	200	0,523
2	-3,439	1,646	1,514	162	48	198	0,714
3	-5,011	2,057	1,786	164	46	196	1,09
4	-7,485	2,685	2,114	166	44	194	1,78
5	-10,92	3,428	2,334	168	42	192	2,83
6	-15,25	4,128	2,207	170	40	190	4,35
7	-20,30	4,577	1,471	172	38	188	6,37
8	-25,30	4,539	-0,121	174	36	186	8,89
9	-28,94	4,253	2,037	176	34	187	10,23
10	-32,39	3,114	1,263	178	32	186	12,62
11	-34,37	0,866	-2,306	180	30	184	15,60
12	-34,01	-1,32	0,964	182	28	185	16,62
13	-29,58	1,925	0,164	184	29	184	16,84
14	-23,78	0,317	-5,117	186	28	182	17,47

Table 5.12

№	Gradients			Control			*Det (M)*
	V_1	V_2	V_3	k_2	k_3	k_4	
1	14,919	11,328	-12,940	100	50	150	0,536
2	4,114	5,283	-5,681	96	50	150	0,957

3	0,967	4,836	-4,988	92	50	150	1,593
4	-2,064	5,051	-5,025	88	50	150	2,435
5	-2,563	6,242	-7,423	88	46	150	3,721
6	-3,344	6,652	-10,304	88	42	150	5,559
7	4,499	5,500	-13,337	88	38	150	8,007
8	-6,071	2,331	-16,991	88	34	150	11,01
9	-8,036	-2,769	-18,621	88	30	150	14,39
10	-8,084	-2,548	-25,424	88	30	154	16,85
11	-7,642	-2,010	-31,784	88	30	158	19,08
12	-6,674	-1,082	-36,878	88	30	162	20,903
13	-4,927	0,382	-39,883	88	30	166	22,13

Table 5.13

The optimization of the trace of the information matrix

№	Gradients			Control			*Det (M)*
	V_1	V_2	V_3	k_2	k_3	k_4	
1	-0,132	0,332	0,139	100	50	150	39,34
2	-0,090	0,257	0,174	102	44	148	42,84
3	-0,047	0,176	0,203	103	39	145	45,44
4	-0,020	0,120	0,242	103	36	141	47,85
5	0,007	0,078	0,279	103	34	137	49,32
6	0,042	0,055	0,314	103	33	132	51,66
7	0,106	0,045	0,330	103	32	126	54,67
8	0,165	0,096	0,337	101	32	120	58,10
9	0,224	0,148	0,339	98	31	114	62,23
10	0,269	0,188	0,340	94	29	108	67,31
11	0,280	0,206	0,339	89	26	102	73,37
12	0,268	0,182	0,317	84	22	96	79,62
13	0,207	0,200	0,246	79	19	90	84,86
14	0,148	0,141	0,213	75	15	86	88,35
15	0,137	0,138	0,116	73	13	82	90,19
16	0,098	0,112	0,088	71	11	80	91,23
17	0,098	0,112	0,088	71	11	80	91,23

Table 5.14

№	Gradients			Control			*Det (M)*
	V_1	V_2	V_3	k_2	k_3	k_4	
1	0,018	-0,003	0,014	160	40	180	0,757
2	1,048	-0,006	0,019	162	38	182	1,102
3	0,089	-0,006	0,022	164	36	184	1,562
4	0,146	-0,004	0,016	166	34	186	2,132
5	0,218	-0,010	-0,004	168	32	188	2,769
6	0,267	-0,002	0,003	170	30	187	2,718

Table 5.15

1	0,041	0,002	-0,038	100	3	150	8,006
2	0,006	0,001	-0,005	102	32	148	9,123
3	0,006	0,001	-0,004	104	34	146	10,084
4	0,005	0,001	-0,002	106	36	144	10,809
5	0,003	0,001	-0,001	108	38	142	11,242
6	0,000	0,000	-0,001	110	40	140	11,357

The surface of the functional $J = \det M$, as a function of the other two control actions (Δ_1 and Δ_2 -are the displacement of the control pulses in time) is shown in the fig. 5.5 for

clarity at a fixed value of the third control action. The surface has the multimodal character: the largest maximum det$M$ =12 takes place at the origin of coordinates ($\Delta_1 = 0$ and $\Delta_2 = 2$), next the local maximum det$M$ =12 for $\Delta_1 = 4{,}5$ and $\Delta_2 = 2$, then we have the less extremum det$M$ =1 for $\Delta_1 = 7$, $\Delta_2 = 1$. As it follows from the fig. 5.5, at $\Delta_1(0) < 2.8$, the algorithm (4.66) leads to the first maximum; at $2.8 < \Delta_1(0) < 5.8$ it leads to the second one; at $\Delta_1(0) > 5.8$ it leads to the third maximum. Therefore, in general, the problem of optimization of the index (4.46) or D-criterion should be performed for different initial conditions.

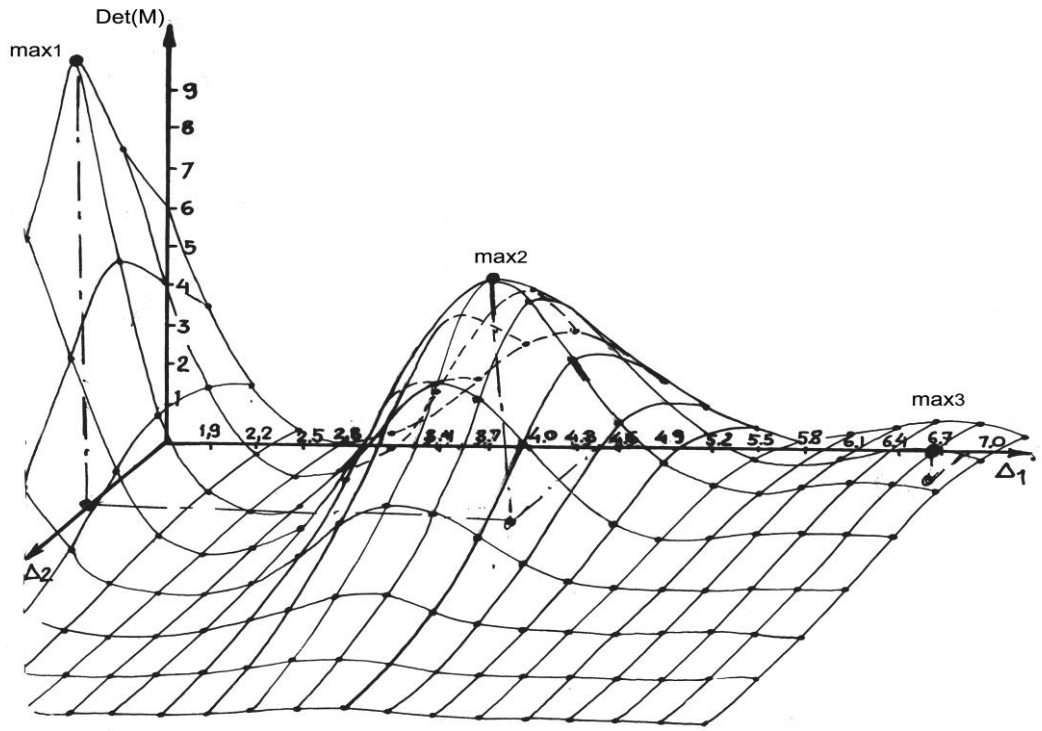


Fig. 5.5 The dependence det$M$ on Δ_1 and Δ_2 at $\Delta_3 = const$.

Chapter 6. The Unity and Difference of the Signal and Parametric Identification of Real Objects [26]

The emergence of the theory of identification as a mathematical formalization of cause-and-effect relationship in the objects of the real world has a long history. The peak of its development is the second half of the XXth century, the period of the emergence and rapid development of means of computerization and the automation of experimental researches. But now it cannot be said that this theory is finally formed and that its application to the real-world objects is always correct [26]. According to the basic categories of philosophy, real world is infinite dimensional, everything is interconnected directly or indirectly. Conventionally, it can be represented by the equation

$$\dot{X}_\infty = f_\infty\left(X_\infty\right), \qquad (6.1)$$

where X_∞ - is the infinity dimensional vector-function of the variables, $\dot{X}_\infty$ -is the speed of change X_∞ in accordance with a function of total interrelation f_∞. By reason of the total interrelation f_∞ is the component x_i of the infinite dimensional vector function, the system (6.1) is unstable (the matter and the motion are unified), and the total interrelation f_∞ (even if it is sustainable) is unknowable.

6.1. The Statement of the Problem

On restricting the space-time area of the change X_∞ by the small area, only the projection X_∞ to the finite number n of coordinates is considered [2]:

$$\dot{X} = f\left(X\right), \qquad (6.2)$$

where $X\left(t\right)$ - is n-dimensional vector-function of time, $\dot{X}$ -is its derivative in time *t*.

We proceed to the linear stationary model

$$\Delta\dot{X} = A\Delta X + B\Delta U, \qquad (6.3)$$

with the certain degree of accuracy ε, continuing to narrow the area,

where the significant variables of the real object, marked in the model (6.3), are divided into causal U and consequential X and taken in the deviations ΔX, ΔU from any center (X_0, U_0) of the region G [35].

The model as an operator that connects the conditionally accepted inputs ΔU and outputs ΔY of the object

$$\Delta Y = W \cdot \Delta U, \qquad (6.4)$$

where ΔY, ΔU and W may be the functions of time or Laplace's complex variable.

The unaccounted subset $(X_\infty - X)$ of variables of the real world determines the proximity of the models (6.3) (6.4). The error ε goes to zero only at narrowing the region G to the point. But the ratio of "noise $(X_\infty - X)$- signal $\Delta Y(t)$" infinitely increases. Thus, the model (6.2) (6.3) (6.4) and the methods of their identification cannot to be accurate: the fast (relatively X, U) variable components of the rejected subset, perceived as a random process $N_1(t)$ influence on the small samples; the boundary theorem of Chebyshev does not work for the large samples because of the influence of the slow variable components $N_2(t)$ of this subset, that contribute the non-stationarity into the average characteristics of the random process (the process is not ergodic).

Then, from the variety of similar models the best one will be the model by means of which the main objective,for which the model is determined, is achieved [9]. For example, these objectives may be the optimal control, the forecast of the behavior or the control of specific physical parameters of the object. According to the main objectives we distinguish two fundamentally different approaches to the problem of identification.

The first approach is the signal identification, if for a given set of input signals $U(t)$ it is necessary to choose the mapping that is random by its structure $U(t)$ in order that any rate of the error ε is less than the desired Δ:

$$\|\varepsilon\| < \Delta. \qquad (6.5)$$

Thus, the signal $Y(t)$ is approximated with an accuracy to Δ in the basis of signals $U(t);\ i=\overline{1,n}$ that are converted by the operators W_i. The second approach is the parametric identification, if at a set of signals ΔX or ΔY it is necessary to define (for the given point (X_0,U_0,t_0) of these region) the structure and/or parameters of the matrices A, B of the model (6.3) or the operator W of the model (6.4), which would correspond to the essential correlation of the variables of the real object. The signal identification must be used for the goal orientation of the models to the tasks of control and forecast, the parametric identification is used for the diagnosis and monitoring of specific, but not directly measurable, parameters of the real object.

6.2. The Strategy of the Signal Identification and the Invariance of the Adaptive Control (an Example of the First Approach)

Let us represent the model of the real object in the limited region as

$$\Delta Y = W\cdot\Delta U + W_1\cdot\Delta F + W_2\cdot\Delta N, \qquad (6.6)$$

where ΔF - is the vector of controlled perturbations, ΔN -is the vector of uncontrolled perturbations, W, W_1, W_2 - are corresponding operators.

It is necessary to construct the invariant optimal (in terms of the functional that is quadratic by ε and ΔU) regulator:

$$\Delta U(t) = W_p(\varepsilon(t,\beta)), \qquad (6.7)$$

where β is the unknown vector of parameters of the operators W and W_1 of the model (6.6).

The error in (6.7) is equal to

$$\varepsilon(t) = \Delta Y^*(t) - \Delta Y(t), \qquad (6.8)$$

where $\Delta Y^*(t)$ - is the desired optimal trajectory of the motion of the aircraft (AC).

The operator W_p of regulator is linear in the absence of restrictions

$$\Delta U(t) = W_{p1} \cdot \varepsilon(t) + W_{p2} \cdot \Delta F(t). \qquad (6.9)$$

Substituting the control (6.9) into the model (6.6), we obtain the expression for the closed system:

$$\Delta Y(t) =$$

$$= \left(W \cdot W_{p1} + I\right)^{-1} \cdot \left[W \cdot W_{p1} \cdot \Delta Y^{*}(t) + \left(W \cdot W_{p2} + W_1\right) \cdot \Delta F(t) + W_2 \cdot \Delta N(t)\right]. \qquad (6.10)$$

The condition of the invariance of its access to the controlled perturbation $\Delta F(t)$:

$$W_{p2} = W^{-1} \cdot W_1, \qquad (6.11)$$

where W and W_1 are the unknown operators of the model (6.6) of the object.

In order to indentify these operators W and W_1, let us consider the model:

$$\Delta \hat{Y} = \hat{W}\left(\hat{\beta}\right) \cdot \Delta U + \hat{W}_1\left(\hat{\beta}\right) \cdot \Delta F. \qquad (6.12)$$

The vector $\hat{\beta}$ of parameters of the model (6.12) is estimated from the conditions of the minimum of some index I of the proximity $\Delta Y(t)$ and $\Delta \hat{Y}(t)$:

$$I\left(\hat{\beta}\right) = \left\| \Delta Y(t) - \Delta \hat{Y}(\hat{\beta}, t) \right\|. \qquad (6.13)$$

In this case we have the **signal** identification: the faster is determined the vector $\hat{\beta}$ of parameter of the operator $\hat{W}$ i $\hat{W}_1$ from the conditions of the minimum I and the wider is the basis approximating functions ΔY, the closer are $\Delta \hat{Y}$ to ΔY. Thus, the impact of the uncontrolled slow perturbations $N_2(t)$ is indirectly compensated (due to the operational adjustment $\hat{\beta}$).

In the asymptotic behavior the system has the invariance [9] to change the parameters of the object, and controlled ΔF and uncontrolled low frequency perturbations. The high frequency component N_2 of the perturbations is usually smoothed by the nature of the object, and influences on the minimal time of the identification of the model. The complexity of the model's structure (6.12) can adapt to the pace of nonstationarity of

characteristics of the random process $N(t)$. It is appropriate to apply the orthogonal basis or the nonius approach [35] in order to adapt the base dimensionality to the non-stationarity.

The simple nonstationary model, where $\hat{\beta}(t)$ changes in the rate of processes, providing the proximity $\Delta\hat{Y}$ to ΔY of the object (the parametric feedback) and the hypothetical complex stationary model (6.1) are theoretically extreme in the series of complexity. There are the quasi-stationary models (6.12), the complications of which should help to improve the accuracy in terms of the index (6.13) and quasi-stationarity of the vector $\hat{\beta}$ of their parameters and, consequently, the degrees of optimality of the controlled object. In this case the account of the physical processes in the object is not required. The estimate $\hat{\beta}$ may not have the physical meaning. The strict convexity and the unimodality of the index (6.13), as a function, is not required either.

6.3. The Strategy of Parametric Identification (an Example of the Second Approach the Evaluation of the Aerodynamic Coefficients $\hat{\beta}$ of the Aircraft)

When selecting the structure of non-linearity in the model (6.12), the matrices A, B in the model (6.13) or the operator W in the model (6.14), it is necessary to take into account the "physics" of processes in AC. The models (6.13) and (6.14) are the linearization of the model (6.12). The linearization is allowed through the smoothness of nonlinearity f (in nature, due to the power of systems, the ideal jumps are absent).

It is necessary to ensure the strict convexity of the index (6.13) and, if it is possible, the autonomy for the unaccounted subsets $(X_\infty - X)$ of the variables by special planning of the nature experiment in order to obtain the unambiguous and objective estimate $\hat{\beta}$ of the physical parameters β of the object, for example, AC. It is also desirable that the estimate $\hat{\beta}$ linearly goes into the expression of the mistake (6.18) and n of its components

are the coefficients at the linearly independent functions of the sensitivity of error (6.8) on $\hat{\beta}_i$.

For this purpose we take the minimum of the left and the right sides of the incoherence of each row of the equation (6.13) instead of the minimization of the index (6.13) as the difference of outputs of the object and the model. Then the estimation $\hat{\beta}$ is reduced to the one-step solution of the minimization problem of the strictly convex quadratic indicators. The estimation is single, and at the relevant approaches [35], it is statistically unbiased and effective.

However, there is its displacement due to the proximity of the models (6.12), (6.13), (6.14). The proximity tends to zero if the region G of the change of variables collapses to the point $\left(X_0, U_0, t_0\right)$. With the decrease of ΔX, ΔU, the ratio of "noise – signal $\Delta X\left(t\right)$" increases. This leads to the loss of effectiveness of the estimate $\hat{\beta}$. The method that allows getting the accurate and unbiased estimate at the point $\left(X_0, U_0, t_0\right)$ is proposed in [37].

In order to do this it is required to accomplish the sequence of similar (but of diverse deviations' amplitude ΔX) active experiments at the object, each of which provides the convexity of the index (6.13) for the linear basis of the model. The sufficiently effective estimates, shifted as a result of ignoring the nonlinearity, are found. The unbiased estimate is determined using the regression dependence, that is constructed for each $\hat{\beta}_i$ at a set of amplitudes $\left\|\Delta X\right\|$ and taken at the point where $\left\|\Delta X\right\|$ is zero.

Test example. The accurate nonlinear model:

$$y(k) = \sum_{j=1}^{3} x_j(k) + \sum_{j,q=1, j\geq q}^{3} x_j(k) x_q(k), \qquad (6.14)$$

with single coefficients for four samples of different amplitude is approximated by its linear part for those signals

$$x_1(k) = x_{max}(l)\sin\left(\pi\frac{k-1}{M-1}\right),\ x_2(k) = x_{max}(l)\sin\left(2\pi\frac{k-1}{M-1}\right),$$

$$x_3(k) = x_{max}(l)\cos\left(2\pi\frac{k-1}{M-1}\right),\ k=\overline{1,M},\ l=\overline{1,4}. \qquad (6.15)$$

The estimates $\hat{\beta}_j$, $(j=1,2,3)$ are counted by the least square method (LSM). Here the index (6.13) of the error of approximation by the linear model $y(k)=\hat{\beta}_1 x_1(k)+\hat{\beta}_2 x_2(k)+\hat{\beta}_3 x_3(k)$ was by 2 orders less than the magnitude of the similar norm $\|\Delta y\|$. Thus, the approximation problem is solved on rather high quality level in the sense of (6.13). But the coefficients $\hat{\beta}_i$ are significantly shifted.

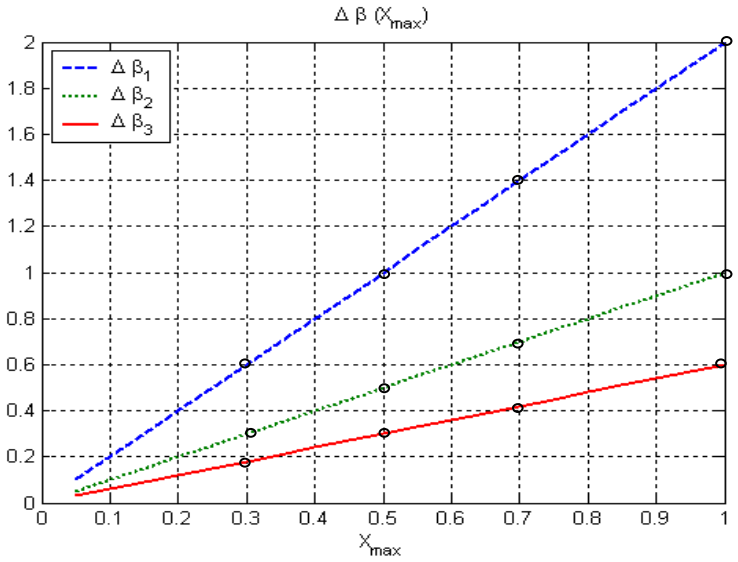


Fig. 6.1. The dependence $\Delta\beta(x_{max})$.

The linear regression dependences of the shifts $\Delta\beta_j$ (j = 1,2,3) from $x_{max}(l)$, (l = 1,2,3,4), agree at zero amplitude x_{max} to zero; the estimates $\hat{\beta}_j$ coincide with true $\beta_j = 1$, respectively.

If we take the complete model (6.14), then, with Tikhonov's regularization (the index (6.16) of the LSM-estimation of coefficients of the complete model (6.14)) we obtain the near-zero index (6.13), because the convexity of the index (6.13) is not strict, as a function $\hat{\beta}$.

The regularized index

$$I(\hat{\beta},\alpha) = \left\| \Delta Y(y) - \Delta \hat{Y}(\hat{\beta},t) \right\| + \alpha \left\| \hat{\beta} - \beta_{apr} \right\|, \qquad (6.16)$$

where α is the parameter of regularization, β_{apr} - is the vector of parameters, given apriori.

Here the near-zero meaning (6.13) is achieved at a variety of meanings β_{apr}. The estimate $\hat{\beta}$ is close to β_{apr}, and not to the actual unit.

To be more illustrative, we consider the two-dimensional problem when the object and the model are isomorphic

$$y(k) = \beta_1 x_1(k) + \beta_2 x_2(k), \qquad (6.17)$$

Here $x_1(k)$ and $x_2(k)$ are strongly correlated. Therefore, the functional (6.13) (the dashed lines in fig. 6.2) as a function $\hat{\beta}$, is strongly convex for the first and the third quadrants of the plane (β_1, β_2) and slightly curved (the "furrow") for the second and the fourth quadrants.

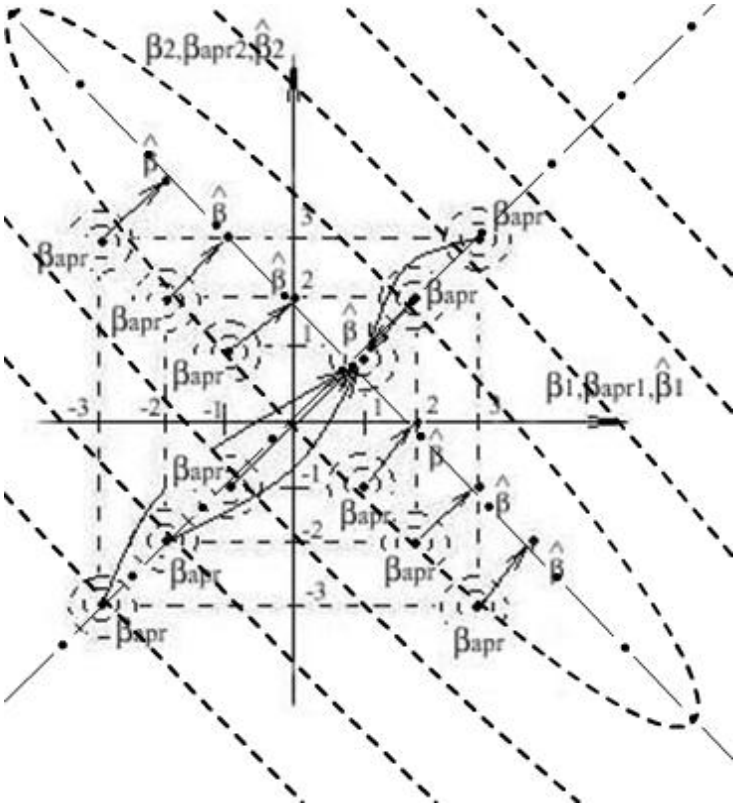


Fig. 6.2. The exact value β and its estimate $\hat{\beta}$, obtained from the condition of

minimum of the functional (6.16) for the various aprior meanings β_{apr}.

The regularizing one-in-ten addition $\alpha\left\|\hat{\beta}-\beta_{apr}\right\|$ of the functional (6.16) (the circles in the fig.6.2) is strongly convex and, consequently, makes the functional (6.16) convex.

The process of minimization of the functional (6.16) is shown by the arrows, the aprior estimate $\hat{\beta}$ is obtained from the aprior estimate β_{apr} as result of the process of minimization.

Table 6.3.

β_{apr}	$\hat{\beta}_1$	β_{apr}	$\hat{\beta}_2$	$\frac{\left\|\Delta Y(t)-\Delta\hat{Y}(t\right.}{\left\|\Delta Y(t)\right\|}$
1	1	1	1	0
-1	0.904	-1	0.904	0.0091
-2	0.857	-2	0.857	0.0091
-3	0.809	-3	0.809	0.036
2	1.048	2	1.048	0.0023
3	1.095	3	1.095	0.0091
1	1.9	-1	0.0015	0.0045
2	2.85	-2	-0.95	0.011
3	3.8	-3	-1.9	0.0226
-1	0.0015	1	1.9	0.0045
-2	-0.95	2	2.85	0.011
-3	-1.9	3	3.8	0.0226

As it follows from the table 6.3 and the fig. 6.2, the estimates $\hat{\beta}$ coincide with the region that is close to the actual value, when β_{apr} is in the first and the third quadrants (closer or farther on β depending on the distance β_{apr} from the β). If β_{apr} is in the second or the fourth quadrants, the estimates, as a result of minimization of the regularized

functional (6.61), are not far away from the aprior ones and find the point of compromise between the increasing addition $\alpha\left\|\hat{\beta}-\beta_{apr}\right\|$ and the decreasing value of the norm $\left\|\Delta Y(t)-\Delta\hat{Y}(\beta,t)\right\|$.

At the same time the ratio of this norm to the similar norm $\left\|\Delta Y(t)\right\|$ does not exceed 2.26%. Thus, the quality of the signal identification is high, while the parametric identification is not correct: $\hat{\beta}_2$ goes away from $\beta_{apr\,2}$ in the opposite side from $\hat{\beta}_2$ in the second quadrant, similarly, in the fourth quadrant $\hat{\beta}_1$. In general, for n variables, the function $I(\beta)$ (6.13) can have several "furrows" of minimal or even zero slope, which complicates the task much.

The real example. The seven modes of the change of the handlebar of height, the angle of attack $\alpha(t)$ and the angular velocity $\omega_{Z_1}(t)$ of the aircraft M-17 in the short-periodic longitudinal motion are represented in the fig.6.3. It ensures the non-interaction concerning the unaccounted multitude of variables (the lateral-directional motion is absent, the velocity, the height, the configuration of the aircraft and other variables are practically constant).

The complete model of the dependence $\dot{\omega}_{Z_1}(t)$ on $\delta_H(t)$, $\alpha(t)$ and $\omega_{Z_1}(t)$ is similar to the model (6.14). The estimates of the coefficients $\hat{\beta}_j$, $j=1,2,3.$, shifted due to the approximation of the models, were determined in each from these seven modes. The specific physical parameter was calculated using them: the distance between the center of mass and the aerodynamic focus of the plane, normalized by the mean aerodynamic chord of the wing, viz the reserve $\hat{\sigma}_n$ of the aperiodic stability at the vertical overload [39].

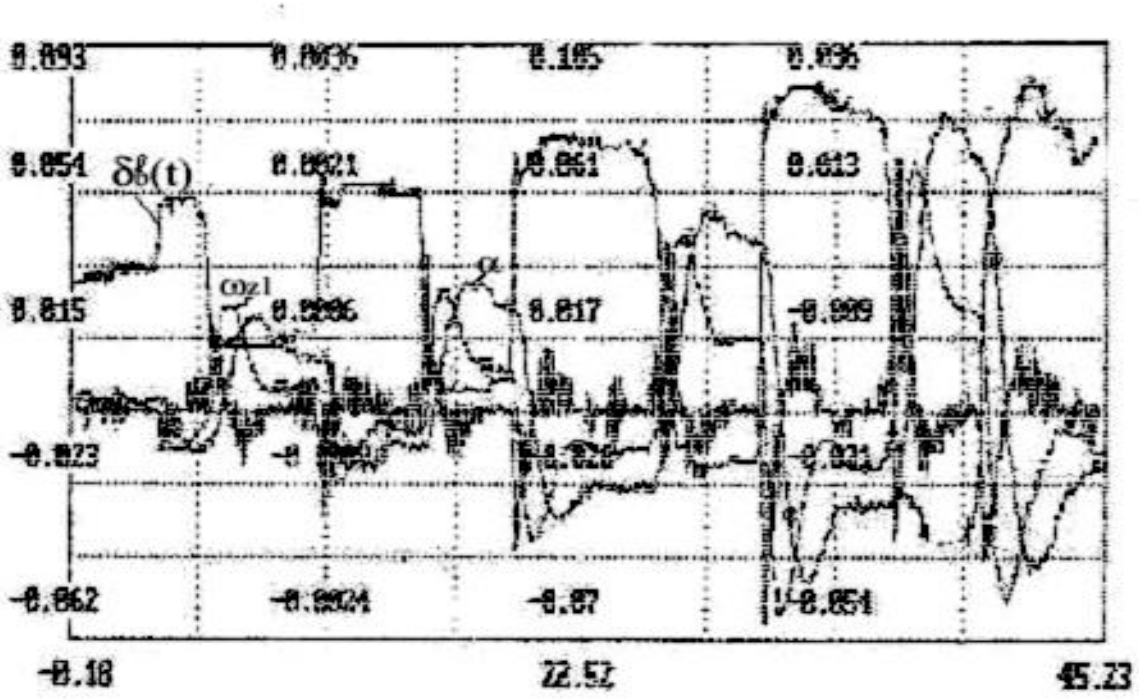


Fig. 6.3. The oscillograms of the modes of the change of handlebar of height, the angle of attack and the angular velocity.

Next the reserve was approximated by the linear dependence in the function $\|\Delta\alpha\|$ (the fig. 6.4):

$$\hat{\sigma}\left(\|\Delta\alpha\|\right) = 0{,}22 - 0{,}075\|\Delta\alpha\|. \qquad (6.18)$$

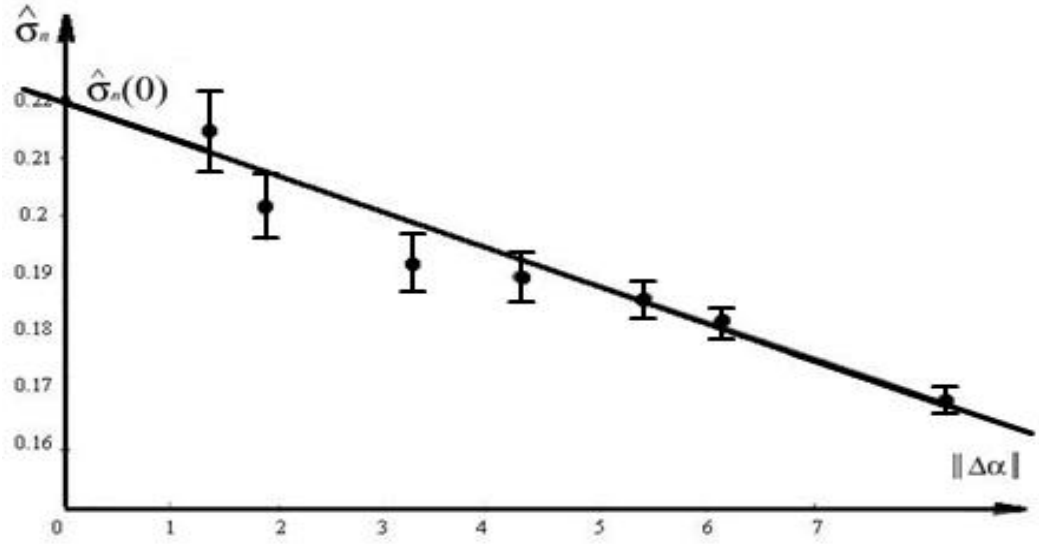


Fig. 6.4. The dependence of the estimate $\hat{\sigma}_n$ on $\|\alpha\|$.

The forecasting in $\|\Delta\alpha\| = 0$ value $\hat{\sigma}_n(0)$ is in the region of the real value σ_n. The unbiased estimate is obtained by the linear approximation of the dependence (6.18) and the calculation of its value at the point where the deviation is zero. The simple averaging of the results will give the significantly shifted estimate (the underestimate) $\hat{\sigma}_n = 0{,}188$. The further clarification $\hat{\sigma}_n$ can be achieved by approximating the estimates $\hat{\sigma}_n(0)$ by regression smooth dependence on the other flight parameters (the velocity, the altitude, etc.).

The linear regression dependence of the aerodynamic coefficients (ADC) on the initial angle of attack α and its derivative, which determines the pace of implementation of dynamic modes in the short-periodic longitudinal motion of the aircraft, are given in the table 6.4. The appropriate ADC are given in the first column of the table, next a_1, a_2, a_3 - are the coefficients of regression, next there is the mean square error (ASE) of approximation of the appropriate ADC, next the mean value ADC and its mean square deviation. As we can see, taking into account only α and $\dot{\alpha}$ allowed increasing the accuracy at the average of 2 times.

Table 6.4

The dependence of the estimates of ADC on α and $\dot{\alpha}$.

ADC	a_1	a_2	a_3	ASD apr.	Average	ASD aver.
$m_z^{\overline{\omega}_z}$	-8.54	-0.2289	-0.0112	0.01	-10.61	0.02
m_z^{α}	-0.0092	0	0	0.007	-0.0091	0.007
$m_z^{\delta_в}$	-0.0161	-0.0012	0	0.003	-0.02	0.009
$m_z^{C_y}$	-0.0873	-0.0029	0	0.004	-0.0928	0.006
C_y^{α}	0.1068	-0.003	0	0.001	0.0979	0.003
$C_y^{\delta_в}$	-0.01	0.0012	0.0001	0.431	0.0028	0.923
σ_n	-0.2301	0.0071	0.0001	0.001	-0.2119	0.002

The data, summarized for different types of the aircrafts, supporting the effectiveness of clarification, in particular, the stability margin σ_n by approximation by linear

regressive dependence on various flight parameters, are given in the table 6.5. The dimensionality was changed from 2 to 6 and the number of modes was changed from 15 to 190 for different aircrafts.

Table 6.5

The comparison of the simple averaging and the regressive approximation σ_n from the parameters of flight

№	Type of the aircraft	AQD %		Dimensionality ΔX	Number of modes
		Models	Average		
1	АН-72	5	102	6	190
2	ИЛ-86	7	31	2	25
3	Ту-154	4	13	4	70
4	Миг-29	7	50	4	50
5	М-17	0.5	1.5	2	15

The conclusions. In order to set the problem of identification correctly, we should clearly distinguish the signal and parametric approaches. Their generality is in the minimization of an error (6.18); their difference is in the models (abstract and "physically" adequate) and in the requirements to the functional (6.13), as a function of the estimate $\hat{\beta}$ (nonstrict and strict convexity respectively). Unfortunately, the signal identification is sometimes used in practice of flight test of the aircraft for the estimation of parameters, by putting the coefficients that are not objectively accurate apriori (using the calculation or the results of insufflation in the wind tunnel) in the model (6.13) and then adjusting them from the condition of minimum of the functional (6.16). The apparent adequacy of the model is achieved here: the error (6.18) is rather small, the estimates $\hat{\beta}$ are close to the aprior ones. But the latter may significantly differ from the actual physical parameters that could then lead to a decrease of safety of flights due to the incorrect evaluation of the aerodynamic coefficients of the aircraft which determine their stability and control.

Chapter 7. The Integration of Methods and Models [29]

The stability of dynamic systems is one of the most important characteristics, especially if they are non-stationary, not fully known and prone to random noises in the measurement channels of the state variables. These systems include the aircraft, in particular, with its stability in the longitudinal short-period motion [41]. In terms of the variables (the angle of attack, the angular velocity and the handlebar of height) this motion is described by the system of equations of the first order

$$\left.\begin{aligned}\dot{x}_1 &= \beta_{11}x_1 + \beta_{12}x_2 + \beta_{13}u,\\ \dot{x}_2 &= \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}u\end{aligned}\right\}, \qquad (7.1)$$

for the small deviations of variables of the balancing mode (the constant altitude and the speed in the vertical plane), where $\dot{x}_1$, $\dot{x}_2$ – are the derivatives of time from x_1 i x_2; β_{ij}, ($i=1,2;\quad j=1,2,3$) – are the aerodynamic coefficients.

The transfer functions that reflect the input signal u into the output x_{1M} or x_{2M}, for example u to x_{2M}:

$$W_u^{x_2}(p) = \frac{\beta_{23}p - (\beta_{23}\beta_{11} - \beta_{13}\beta_{22})}{p^2 - (\beta_{11} + \beta_{22})p + (\beta_{11}\beta_{22} - \beta_{12}\beta_{21})}. \qquad (7.2)$$

are the equivalents to the system (2.7).

The coefficients $(\beta_{11} + \beta_{22}) = -a_1$ and $\langle(\beta_{11}\beta_{22} - \beta_{12}\beta_{21}) = a_0\rangle$ of the denominator determine the oscillatory and aperiodic stability of the reserve respectively. Multiplying these coefficients by $J_{z_1}/qSb_a\beta_{11}$ (where J_{z_1} - is the moment of inertia, q - is the dynamic pressure, S - the area of the wing, b_a - is the mean aerodynamic chord of the wing), we obtain the appropriate reserve of stability in the fraction $\langle(\beta_{11}\beta_{22} - \beta_{12}\beta_{21}) = a_0\rangle$ from b_a. The coefficients β_{13},β_{23} determine the effectiveness of the handlebar of height, β_{12},β_{22} - are the damped forces and moments. The evaluation of the coefficients β_{ij} of the aircraft (AC), which determine the stability and control of the aircraft, is an actual task of the flight tests (FT).

7.1. The Formulation of the Problem

The estimates $\hat{\beta}_{ij}\beta_{ij}$ cannot be obtained with the relatively high accuracy from the measurements $\hat{x}_i, \hat{u}$ that are noisy by the interference, because of the limited cost of the nature experiment, the transience of time of the experiment and the nonlinearity of the range of changes of variables. For these reasons, the practice of FT is limited by rather approximate estimates of the stability and controllability AC [41]. We can increase the accuracy of the evaluation of the aerodynamic coefficients (ADC) without increasing the time, using the statistical modeling of the interferences.

In order to do this, it is enough to perform the filtering of noises and obtain the approximate implementation of the interference as a difference of the filtered and output signals, taking into account the information on the spaced spectra of signals and interferences (interferences are more high frequency). Then we should determine the statistical characteristics of these implementations and generate m statistically similar implementations of interferences for each variable x_i, u. Adding these implementations up with the relevant filtered signals we get m pseudosamples of data of FT. Having m statistically identical samples, the *p* identification methods and the *q* models of the aircraft, we can use their redundancy in order to obtain more accurate estimates of ADC [29].

7.2. The Methodology of the Research

With regard to the problem of determination of the ADC AC in the longitudinal short-periodic motion, let us consider three models $(q = 3)$: the models (7.1), (7.2) and the model, that leaks out from the first two:

$$\ddot{x}_2 = (\beta_{11} + \beta_{22})\dot{x}_2 - (\beta_{11}\beta_{22} - \beta_{12}\beta_{21})x_2 + \beta_{23}\dot{u} - \\ -(\beta_{23}\beta_{11} - \beta_{13}\beta_{22})u = -a_1\dot{x}_2 - a_0 x_2 + a_2\dot{u} + a_3 u. \quad (7.3)$$

The number p of methods for determining the ADC is defined by the number of functionals of proximity in the space L_2 [6] of variables AC and its models (7.1), (7.2), (7.3). The optimal estimates $\hat{\beta}$ of the ADC are defined from the condition of the minimum of these functionals:

$$\left\{\beta_{ij}^*\right\} = \underset{\beta_{ij}}{\operatorname{argmin}} \left\| \dot{\hat{x}}_i - \sum_{j=1}^{2} \beta_{ij}\hat{x}_j - \beta_{i3}\hat{U} \right\|^2, \quad (7.4)$$

where $i = 1,2;\ j = 1,2;\ \hat{x}_i,\ \dot{\hat{x}}_j,\ \hat{U}$ – are the variables, smoothed by filter.

$$\left\{a_i^*\right\} = \underset{a_i}{\operatorname{argmin}} \left\| x_2 - x_{2M} \right\|^2, \quad (7.5)$$

$$\left\{a_i^*\right\} = \underset{a_i}{\operatorname{argmin}} \left\| \ddot{\hat{x}}_2 + a_1\dot{\hat{x}}_2 + a_0\hat{x}_2 - a_2\dot{\hat{U}} - a_3 U \right\|^2. \quad (7.6)$$

In order to avoid the methodological error in the linear equation (7.1), (7.3) AC, all variables are smoothed by the same filter. The other methods of identification that give the unbiased evaluation ADC under noisy conditions can be applied [35].

Next for each from the algorithms (7.3), (7.4), (7.5) at a set of *m* implementations, for each implementation we define the optimal values $\left\{\beta_{ij}^*\right\}$, $\left\{a_i^*\right\}$, their average values for *m* implementations $\left\{\overline{\beta}_{ij}^*\right\}$, $\left\{\overline{a}_i^*\right\}$, the estimates of their own $\hat{\sigma}^2_{\beta_{ij}^*(k)}$, $\hat{\sigma}^2_{a_i^*(k)}$ and mutual $\hat{\sigma}^2_{\beta_{ij}^*(k)\beta_{ij}^*(l)}$, $\hat{\sigma}^2_{a_i^*(k)a_i^*(l)}$ variances, where $k = 1,2,3;\ p = 3;\ l = 1,2,3;\ k \neq l$. Let us denote the elements of sets $\left\{\beta_{ij}\right\}$, $\left\{a_i\right\}$ by means of $\left\{\alpha_i\right\}$ for brevity sake. Then we search for the best estimate α_i^* at a set of three methods–models in the form

$$\alpha_i^* = \sum_{k=1}^{3} C_k \alpha_i(k), \quad \sum_{k=1}^{3} C_k = 1. \quad (7.7)$$

The coefficients C_k are determined from the condition:

$$C_1 = 1 - C_2 - C_3; \qquad \frac{\partial \hat{\sigma}^2_{\alpha_i^*}}{\partial C_k} = 0, \quad k = 2,3, \tag{7.8}$$

where

$$\partial \hat{\sigma}^2_{\alpha_i^*} = \frac{1}{m-1} \sum_{j=1}^{m} \left(\sum_{k=1}^{3} C_k \left(\alpha_i(k,j) - \bar{\alpha}_i(k,j) \right) \right)^2 =$$

$$= \sum_{k=1}^{3} C_k^2 \hat{\sigma}^2_{\alpha_i(k)} + \sum_{\substack{k,l=1 \\ k \neq l}}^{3} C_k C_l \hat{\sigma}^2_{\alpha_i(k)\alpha_i(l)} = C^T A C.$$

For brevity sake let us denote $\hat{\sigma}^2_{\alpha_i(k)} = \hat{\sigma}^2_k$, $\hat{\sigma}^2_{\alpha_i(k)\alpha_i(l)} = \hat{\sigma}^2_{kl}$, then $C^T = \left[C_2, C_3 \right]$

$$A = \begin{bmatrix} \left(\hat{\sigma}^2_1 + \hat{\sigma}^2_2 - 2\hat{\sigma}^2_{12} \right) & \left(\hat{\sigma}^2_1 + \hat{\sigma}^2_{23} - \hat{\sigma}^2_{13} - \hat{\sigma}^2_{12} \right) \\ \left(\hat{\sigma}^2_1 + \hat{\sigma}^2_{23} - \hat{\sigma}^2_{13} - \hat{\sigma}^2_{12} \right) & \left(\hat{\sigma}^2_1 + \hat{\sigma}^2_3 - 2\hat{\sigma}^2_{13} \right) \end{bmatrix}.$$

Now the condition (7.8) is equivalent to the system

$$AC = B, \tag{7.9}$$

where

$$B = \begin{bmatrix} \hat{\sigma}^2_1 - \hat{\sigma}^2_{12} \\ \hat{\sigma}^2_1 - \hat{\sigma}^2_{13} \end{bmatrix}, \; C = \begin{bmatrix} C_2 \\ C_3 \end{bmatrix}.$$

The solution $C = A^{-1}B$ of the system (7.9) is single if matrix *A* is positively defined. The less correlated are errors in the evaluation of α_i by different methods (7.7) (7.8) (7.9), the closer is the matrix *A* to the diagonal one. Since the norms of proximity (7.4) (7.5) (7.6) are taken in the space of variables x, $\dot{x}$, $\ddot{x}$, it is possible to expect the weak correlation in errors of estimates α_i by different methods–models (7.4), (7.5), (7.6).

Substituting the optimal values C_2, C_3, and also $C_1 = 1 - C_2 - C_3$, found from the system (7.9), in the equation (7.7), we obtain the optimally weighted estimate α_i^* of the coefficient α_i (under good conditioning of the matrix *A* and a small error of the estimates $\hat{\sigma}_i^2$, $\hat{\sigma}_{ij}^2$ of the variances σ_i^2, σ_{ij}^2). In other case in order to guarantee the non

deterioration of the result of weighing of estimates α_i^* we should use the minimax approach.

If we assume that the errors of the estimates $\alpha_i(k)$ are distributed under the normal law, the estimates $\hat{\sigma}_i^2$, $\hat{\sigma}_{ij}^2$, calculated by the *m* generated sample of data, have errors with χ^2- distribution. The true values σ_i^2, σ_{ij}^2 for the given level of authenticity are in the range $(1\pm\gamma)^{-1}\hat{\sigma}_i^2$, $(1\pm\gamma)^{-1}\hat{\sigma}_{ij}^2$ or approximately $(1\pm\gamma)^{-1}\sigma_i^2$, $(1\pm\gamma)^{-1}\sigma_{ij}^2$, where γ is a fractile of distribution, $\gamma << 1$.

Let $\sigma_i^2 = \hat{\sigma}_i^2(1+\gamma)$ for $i \in I = \overline{1,n_1}$, $\sigma_j^2 = \hat{\sigma}_j^2(1+\gamma)$ for $j \in J = \overline{n_1+1,n}$. Then $\sigma_{ij}^2 = \hat{\sigma}_{ij}^2(1+\gamma)$ for $i,j \in I$; $\sigma_{ij}^2 = \hat{\sigma}_{ij}^2(1-\gamma)$ for $i,j \in J$; $\sigma_{ij}^2 = \hat{\sigma}_{ij}^2\sqrt{(1-\gamma^2)}$ for $i \in I, j \in J$ or $i \in J, j \in I$. The evaluation of the variance of the weighted estimate α_r^* *r*-th ADC:

$$\hat{\sigma}^2_{\alpha_r^*} = \sum_{i\in I} C_i^2\hat{\sigma}_i^2(1+\gamma) + \sum_{j\in J} C_i^2\hat{\sigma}_i^2(1-\gamma) + \sum_{\substack{i\neq j\\ i,j\in I}} C_iC_j\hat{\sigma}_{ij}^2(1+\gamma) +$$
$$+ \sum_{\substack{i\neq j\\ i,j\in J}} C_iC_j\hat{\sigma}_{ij}^2(1-\gamma) + \sum_{\substack{i\in I\\ j\in J}} C_iC_j\hat{\sigma}_{ij}^2\sqrt{(1-\gamma^2)} + \sum_{\substack{i\in J\\ j\in I}} C_iC_j\hat{\sigma}_{ij}^2\sqrt{(1-\gamma^2)}. \qquad (7.10)$$

The expression (7.10) is taken *L* times for all possible combinations $\hat{\sigma}_{ij}^2(1\pm\gamma)$, $i = \overline{1,3}$. For each of *j*-th combination from the condition (7.8) we obtain the system that is similar to (7.9), and solving it, we find the optimal ξ- th vector $C^*(\xi)$

$$C^*(\xi) = \left[C_1^*(\xi), C_2^*(\xi), C_3^*(\xi)\right]^T,\ \xi = \overline{1,L} \qquad (7.11)$$

and calculate $\sigma^2_{\alpha_r^*}\left(S, C^*(\xi)\right)$, $\xi = \overline{1,L}$, $S = \overline{1,L}$.

The minimax estimate C^* is determined from the condition

$$C^* = \underset{\xi\in\{1,L\}, S\in\{1,L\}}{\operatorname{argminmax}} \sigma^2_{\alpha_r^*}\left(S, C^*(\xi)\right). \qquad (7.12)$$

That is, each vector (7.11) is substituted in each case of the variance (7.10) and we accept as C^* that one, the greatest value of which is the lowest (7.10) in all variants among the greatest values of other vectors (7.11).

The example. For clarity we consider the two-dimensional case ($p = 2$). Let $\sigma_1^2 = 1$, $\sigma_2^2 = 9$; $\sigma_{12}^2 = 2{,}8$, $\gamma = 0{,}1$. The possible combinations are represented in the table 7.1.

Table 7.1.

The multipliers of possible deviations of the estimates of variances.

№ of variant	Multipliers for		
	σ_1^2	σ_2^2	σ_{12}^2
1	$1+\gamma$	$1+\gamma$	$1+\gamma$
2	$1-\gamma$	$1-\gamma$	$1-\gamma$
3	$1+\gamma$	$1-\gamma$	$\sqrt{\left(1-\gamma^2\right)}$
4	$1-\gamma$	$1+\gamma$	$\sqrt{\left(1-\gamma^2\right)}$

The estimate (7.7):

$$\alpha_i^* = C\alpha_i(1) + (1-C)\alpha_i(2). \qquad (7.13)$$

The variance (7.10):

$$\sigma_{\alpha_r^*}^2 = \left(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}^2\right)C^2 + 2\left(\sigma_{12}^2 - \sigma_1^2\right)C + \sigma_2^2. \qquad (7.14)$$

Substituting four variants of their evaluation from the table 7.1 instead of variances and executing the operations (7.8) (7.9), we obtain the set (7.11) for 4 variants (7.14):

$$C^*(1) = C^*(2) = \frac{\hat{\sigma}_1^2 - \hat{\sigma}_{12}^2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2 - 2\hat{\sigma}_{12}^2} = 1{,}41;$$

$$C^*(3)=\frac{\hat{\sigma}_2^2(1-\gamma)-\hat{\sigma}_{12}^2\sqrt{1-\gamma^2}}{\hat{\sigma}_1^2(1+\gamma)+\hat{\sigma}_2^2(1-\gamma)-2\hat{\sigma}_{12}^2\sqrt{1-\gamma^2}}=1{,}465;$$

$$C^*(4)=\frac{\hat{\sigma}_2^2(1+\gamma)-\hat{\sigma}_{12}^2\sqrt{1-\gamma^2}}{\hat{\sigma}_1^2(1-\gamma)+\hat{\sigma}_2^2(1+\gamma)-2\hat{\sigma}_{12}^2\sqrt{1-\gamma^2}}=1{,}361.$$

The value $\sigma^2_{\alpha_r^*}\left(S,C^*(\xi)\right)$ is represented in the table 7.2.

Table 7.2

The optimal values of variances.

$\sigma^2_{\alpha_r^*}\left(S,C^*(\xi)\right)$				
№ of variant	1	2	3	4
1	0,290	0,290	0,304	0,300
2	0,236	0,236	0,249	0,245
3	0,327	0,327	0.316	0,354
4	0,232	0,232	0,275	0,218

If $\sigma_1^2=1;\quad \sigma_2^2=9;\quad \sigma_{12}^2=2{,}8;\quad \gamma=0{,}1,$ then $C^*(1)=C^*(2)=0{,}756;$ $C^*(3)=0{,}737;\ C^*(4)=0{,}737$. The table 7.3 is similar to the table 7.2.

Table 7.3.

The optimal values of variances.

$\sigma^2_{\alpha_r^*}\left(S,C^*(\xi)\right)$				
S	1	2	3	4

ξ				
1	0,081	0,081	0,089	0,087
2	0,067	0,067	0,073	0,071
3	0,082	0,082	0,078	0,099
4	0,076	0,076	0,094	0,070

Consequently $\min\limits_{\xi\in\{1,4\},S\in\{1,4\}}\max \hat{\sigma}^2_{\alpha_r^*} = \hat{\sigma}^2_{\alpha_r^*}\left(S, C^*(\xi)\right) = 0{,}082$, $C^* = 0{,}756$, $\left(1-C^*\right) = 0{,}244$. If we assume that $\sigma_1^2 \le \sigma_2^2$, for different values of the correlation ratio $r_{12} = \dfrac{\sigma_{12}^2}{\sqrt{\sigma_1^2\sigma_2^2}}$ and the interrelation $F_{12} = \dfrac{\sigma_2}{\sigma_1}$, we obtain the family of graphs $C^*(r_{12}, F)$ (the fig. 7.1).

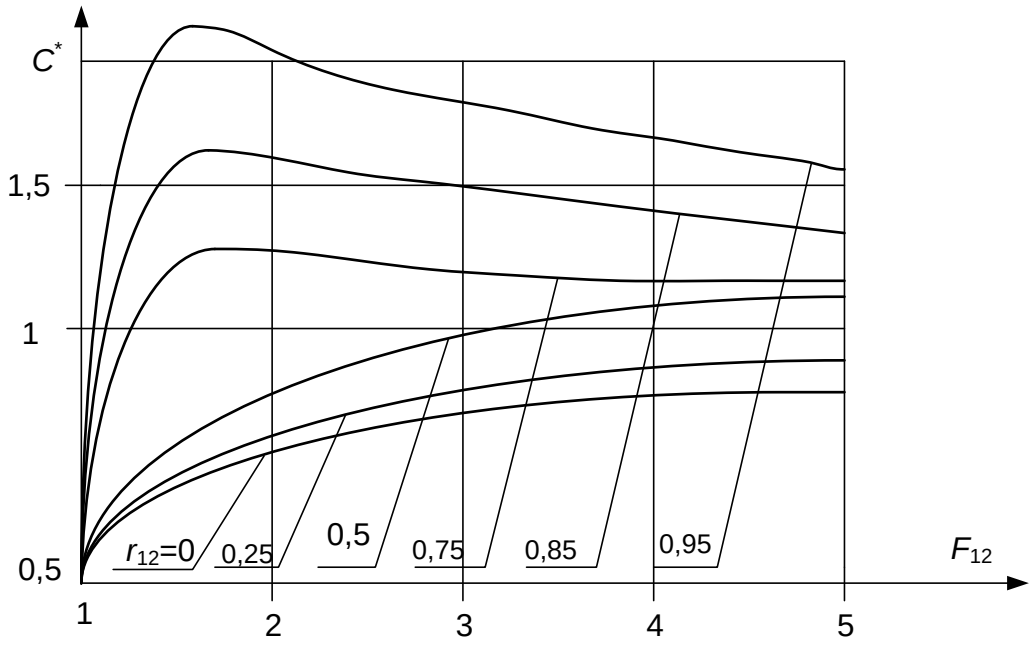


Fig. 7.1. The dependence of the optimal weight C^* on the degree of correlation and noisiness of estimates

For $F_{12} = 1$, $C^* = 0{,}5$ for any r_{12}; at $F_{12} \to \infty$, $C^* \to 1$, $\left(1-C^*\right) \to 0$, that means that the inefficient method is eliminated. It is characteristic that for the same

$\sigma_1^2, \sigma_2^2, \left|\sigma_{12}^2\right|$ the value $\hat{\sigma}^2_{\alpha_r^*}$ depends on the sign σ_{12}^2 (for the positive one it is 0,316, for the negative one it is 0,082). Thus, the weighing two methods with the negative σ_{12}^2 provides the better result. In this example for $\sigma_{12}^2 = 0$ the variance $\hat{\sigma}^2_{\alpha_r^*} = 0{,}9$.

It is much greater than the minimax values of variances at $\sigma_{12}^2 = \pm 2{,}8$ (0,316 and 0,082 respectively). The mutual correlation allows increasing the accuracy of the minimax estimate.

Thus, the application of several isomorphic models of the object of identification and several different methods of parametric estimation together using the technology of statistical modeling of interferences we can significantly improve the accuracy of the estimates of parameters ADC of the aircraft without increasing the length of the data sample of the nature tests. Moreover, so far as there is the nonstationarity of the object of identification, the samples cannot be arbitrarily large. In the [40] it is considered how to ensure the unbiased estimation by ignoring really existing nonlinearity.

Chapter 8. The Multilevel Systems with the Identifier of the Controlled Object

8.1. The System Approach

Let us return to the content of the section 1.1. The system of identification is a set $\{\Sigma, \beta, Opt, \alpha, Y, I, T\}$ at which the relaxation (optimization) process $\{Q_k\}$ is realized relative to *F*, where $Q_k \in \{\{\Sigma, \beta, Opt, \alpha\}, F \in \{Y, I\}\}$ is such that $F_{k+1} \le F_k, k = 1,2,...$

The content of the components of the system are as follows:

$\{\Sigma\}$ – is a set of models' structures;

$\{\beta\}$ –a set of vectors of the models' parameters;

$\{Opt\}$ – a set of methods of the parameters β estimation;

$\{\alpha\}$ – a set of parameters of the methods Opt;

$\{Y\}$ – the functional of the identification quality; the elements (Σ,β) of the model and (Opt,α) of the identification method are found under condition of the extremum $\{Y\}$;

$\{I\}$ - the functional (index) of quality of the goal achievement, for which the problem of identification is solved.

There are such classes of the identification systems (IS), depending on the elements that are in Q_k i F_k: the single-level IS in a narrow sense (sis), the single-level IS in a broad sense (SIS), the single-level adaptive in a narrow sense (sais) and in a broad sense (SAIS), the two-level adaptive in a narrow sense (tmais) and in a broad sense (TMAIS), the three-level (ThMAIS).

The structure of the set Q_k, that optimize the indices F_k for the appropriate identification systems are represented in the table 8.1.

Table 8.1

The classification of the identification systems.

№	Class of the system	$\{Q_k\}$				$\{F_k\}$		
		β	Σ	α	Opt	Y	I	Λ
1	sis	$\{\beta_k\}$	$\{\Sigma_k\}$	α	Opt	Y	0	0
2	SIS	$\{\beta_k\}$	Σ	α	Opt	Y	0	0
3	sais	$\{\beta_k\}$	$\{\Sigma_k\}$	$\{\alpha_k\}$	Opt	Y	0	0
4	SAIS	$\{\beta_k\}$	Σ	$\{\alpha_k\}$	Opt	Y	0	0
5	tmais	$\{\beta_k\}$	$\{\Sigma_k\}$	$\{\alpha_k\}$	$\{Opt_k\}$	$\{Y_k\}$	0	0
6	TMAIS	$\{\beta_k\}$	$\{\Sigma_k\}$	$\{\alpha_k\}$	$\{Opt_k\}$	$\{Y_k\}$	0	0
7	ThMAIS	$\{\beta_k\}$	$\{\Sigma_k\}$	$\{\alpha_k\}$	$\{Opt_k\}$	$\{Y_k\}$	$\{I_k\}$	Λ

The identification task is in the optimization of the internal criterion *J* in the single level IS. That is the identification problem is not enclosed by the main (external) criterion (goal). In the two level IS there is the process of finding the best set of those and other elements by the main criterion *I*. The criterion *I* is subordinate to the goal of higher level, which is defined by the criterion Λ in the three level system. Let us represent the content of the best IS (the 1st row in the table 8.1) and the most perfect IS (the 7th row in the table 8.1).

The first row: the single level identification system in a narrow sense (sis) is the relaxation process $\{\beta_k\}$ [16] relative to *J* if Σ, Opt, α is constant and the set I is empty. For example, the vector β of the equation of regression $y = x\beta$, where y - is a vector of measurements of the dependent variable, x - is the matrix of independent variables, is determined by the least square method. Here the functional $J = E^T E, E = y - x\beta$, the structure Σ of the model is given, the parameter α and the criteria I and Λ are absent.

The seventh row: the three level multi adaptive IS (ThMAIS) is the relaxation process $\{\Sigma_k, \beta_k, Opt_k, \alpha_k, Y_k, I_k, T_k\}$ relative to the index of the third level.

This process is the composition of three processes:

a) $\{\beta_{gkn}\}, n = 1,2,...$ (*g* and *k* is constant) relative to J_{gk} to the stationary point $\beta_{gk} = \text{argextr} J_{gk} \{\beta_{gkn}, \Sigma_{gk}, Opt_{gk}, \alpha_{gk}\} = \text{arg} J^*_{gk}$;

b) $J_{gk}\{\beta_{gkn}, \Sigma_{gk}, Opt_{gk}, \alpha_{gk}\}, k = 1,2,...$ (*g* is constant) relative to J_g to the stationary point $I^*_g = I_g(\beta_g, \Sigma_g, Opt_g, \alpha_g) = \text{argextr} I_g(Y^*_{gk}) = \text{arg} I^*_g$;

c) $\{I^k_g(Y_g(\beta_k, \Sigma_g, Opt_g, \alpha_g))\}, g = 1,2,...$ relative to the index Λ of the third level of the system to the stationary point $I^* = \text{argextr}\Lambda(I^*_g) = \text{arg}\Lambda$.

For example, there is the task of improving of the automated control of the technological process: Λ - is the production quality; $\{I_g\} = \{I_1, I_2\}$, where I_1 - is the quality index of stabilization of technological parameters (variables X of the process), I_2 - is the products quality index, $\{J_{gk}\}$ - is the set of indices of the identification quality of the dynamic models, $\{\Sigma_{g1}, \beta_{g1}\}$ - the mapping of the control variables U in the state transition $X\ (k=1)$ and the regression models; $\{\Sigma_{g2}, \beta_{g2}\}$ - is the mapping of the variables X in the index $I_2\ (k=2)$. At the first stage of automation ($I_n = I_1$) on J_{1k} the local models $\{\Sigma_{1k}, \beta_{1k}\}$ of the object are determined by the methods $\{Opt_{1k}, \alpha_{1k}\}$.

The local regulators, stabilizing the variables X close to the set values x_0 of the extremum I_1 conditions, are synthesized. According to the items a) and b) the ThMAIS determination, the element $\{\beta_1^*, \Sigma_1^*, Opt_1^*, \alpha_1^*\}$, optimal on I_1, is defined, next we construct the model $\{\Sigma_2, \beta_2\}$ of dependence I_2 on X after having stabilized x around x_0 by planning the experiment in deviations $\Delta x = x - x_0$.

The setpoints x_0 of the operating modes are refined with the help of this model: $x_0 \to x_0^*$. Then the optimal (on I_2) element $\{\Sigma_{2^*}, \beta_2^*, Opt_2^*, \alpha_2^*\}$ is determined as a result of the relaxation process according to the items a), b). The transition from Σ_1 to Σ_2 leads to the change of the indicator I_1 to I_2, in accordance with the item c), thereby optimizing the index Λ of the third level system, for example, the production efficiency.

8.2. The Reconciliation of the Quality Indicators

The convergence of the relaxation processes of optimization in two- and three-level identification systems can be significantly improved by conducting the specifically

planned active experiment, which would ensure the similarity of variations of the functionals of the related levels (J and I or I and Λ) in the space of the varying parameters β.

Obviously, if I as a quality indicator of the solution of the main problem at the object with the help of the model, is very sensitive to the errors in the evaluation of the i-th component β_i of the vector β of the model parameters and little sensitive to the j-th (β_j), it is desirable to have the evaluation β_i more accurate than β_j. The extreme values I^* and J^* must meet the same value β^*. Considering that the first variations of J and I in the extreme point β^* is zero, we obtain the variation of the second order:

$$\delta I(\delta\beta) \cong \delta\beta^T \cdot \frac{\partial^2 I}{\partial\beta \cdot \partial\beta^T}\Big|_{\beta^*} \cdot \delta\beta, \tag{8.1}$$

$$\delta J(\delta\beta) \cong \delta\beta^T \cdot \frac{\partial^2 J}{\partial\beta \cdot \partial\beta^T}\Big|_{\beta^*} \cdot \delta\beta. \tag{8.2}$$

The conditions of similarity of the variation (8.1), (8.2):

$$\frac{\partial^2 I}{\partial\beta \cdot \partial\beta^T} = K \cdot \frac{\partial^2 J}{\partial\beta \cdot \partial\beta^T}. \tag{8.3}$$

The ellipsoids of scattering $\delta\beta$ will be similar to the fixed values δI, δJ, that is the surface of the equal value of variations δI, δJ. The coefficient K in (8.3) depends on the informativeness of the experiment. For the ergodic process, the greater is the time of the experiment, the more accurate is the estimates β and, consequently, the less is K.

Let us denote

$$\frac{\partial^2 J}{\partial\beta \cdot \partial\beta^T}\Big|_{\beta^*} = a_{ij}, \ (i, j = 1,2,..n).$$

The coefficients a_{ij} are functionals of the vectors of testing influences $\delta U(t)$ and the parameters α, which are optimized under the following condition:

$$(\delta U(t), \alpha^*) = \arg\min_{dU,\alpha} \left\| [\Delta a_{ij}] \right\|, \tag{8.4}$$

where $\Delta a_{ij} = K \cdot a_{ij}(\delta U(t),\alpha) - a_{ij}^{*}$.

For example, let the model of dynamic of object be the weighted sum of operators $W_i(p,\alpha)$, reflecting the input testing influence $\delta U(t)$ into the output signal $y(t)$:

$$\hat{y}(t) = \sum_{i=1}^{n} \beta_i \cdot W_i(p,\alpha_i) \cdot \delta U(t) = \sum_{i=1}^{n} \beta_i x_j(t). \qquad (8.5)$$

where $p = \frac{d}{dt}$, $\delta U(t) = \sum_{k=1}^{m} \gamma_k \varphi_k(t)$, (8.6)

$\varphi_k(t)$- is the system of ortogonal functions.

If $J = \int_0^T \varepsilon^2(t)dt$, where $\varepsilon(t) = y(t) - \hat{y}(t)$, then

$$\frac{\partial^2 J}{\partial \beta \cdot \partial \beta^T} = [a_{ij}], \quad i,j = \overline{1,n} ,$$

where $a_{ij} = \int_0^T [W_i(p,\alpha_i) \cdot \delta U(t)] \cdot [W_j(p,\alpha_j) \cdot \delta U(t)]dt$.

Or taking into account (8.5) and (8.6)

$$a_{ij} = \int_0^T \left(\sum_{k=1}^{m} \gamma_k x_{ik}(t) \right) \cdot \left(\sum_{k=1}^{m} \gamma_k x_{jk}(t) \right) dt =$$

$$= \sum_{k=1}^{m} \sum_{q=1}^{m} \gamma_k \gamma_q \int_0^T x_{ik}(t) \cdot x_{jk}(t)dt = \sum_{k=1}^{m} \sum_{q=1}^{m} \gamma_k \gamma_q \cdot X_{ij}^{kq} , \qquad (8.7)$$

where $X_{ij}^{kq} = \int_0^T x_{ij}(t) \cdot x_{ik}(t)dt$ - are calculated separately.

The expression (8.4) is defined with the accuracy to the coefficients γ_k, γ_q, selected under condition of the minimum of the norm $\left\| [\Delta a_{ij}(\gamma_k,\gamma_k)] \right\|$, $i,j = \overline{1,n}$ with restrictions in order that δU (8.6) does not extend out of the zone of permissible values, as a function γ_k. The greater is the time *T*, the less is *K* in (8.3), the more accurate is the estimate $\hat{\beta}$ and, consequently, the less is the variation (8.1) of the main index.

Let us show the conditions of conformity of the functionals *I* i *J* and hence the conformity of their variations (8.1), (8.2) for specific cases.

The functional *I* is the unimodal function of the scalar β. Its allowable variation $\Delta I = I - I^* = C$ is composed by the deviation of the estimate $\hat{\beta}$ from β, $I^* = \inf_{\beta} I(\beta)$, β is a scalar. The evaluation $\hat{\beta}$ with the certain probability is in the range $[\beta_1, \beta_2]$ of length *L*. *L* is the distance between the points of the intersection $I(\beta)$ and $I = I^* + C$. Then there are two obvious statements.

The statement 1. From all the uncertainty intervals of length *L* of the evenly distributed evaluation $\hat{\beta}$ the $\min\limits_{K} \max\limits_{\beta_1 \beta_2} \{I(\beta_1), I(\beta_2)\}$ is achieved under condition that $K = K^*$, where $K = (\beta^* - \beta_1)/(\beta_2 - \beta^*)$, $K^* = (\beta^* - \hat{\beta}_1)/(\hat{\beta}_2 - \beta^*)$.

Indeed, any shift of the interval $[\beta_1, \beta_2]$ increases $\max\limits_{\beta_1 \beta_2} \{I(\beta_1), I(\beta_2)\}$.

The statement 2. The value $K = K^*$ ensures the minimum of losses, i.e.

$$\min_{\beta_1} \int_{\beta_1}^{L+\hat{\beta}_1} I(\beta) d\beta. \qquad (8.8)$$

Genuinely, any shift of the interval $[\beta_1, \beta_2]$ relative to $[\hat{\beta}_1, \hat{\beta}_2]$ increases the area, viz the integral $\int_{\beta_1}^{L+\hat{\beta}_1} I(\beta) d\beta$.

The example: $I(\beta) = a_2 \beta^2 + a_1 \beta, \quad a_2 > 0$. Let us find β^*:

$$\frac{\partial I(\beta)}{\partial \beta} = 0 = 2a_2 \beta^* + a_1, \ \beta^* = -\frac{a_1}{2a_2}.$$

Solving the equation $a_2 \beta^2 + a_1 \beta - C = 0$ for the variation *C*, we find $\hat{\beta}_1$, $\hat{\beta}_2$:

$$\hat{\beta}_{1,2} = -\left(a_1 \pm \sqrt{a_1^2 + 4a_2 C}\right) / 2a_2. \qquad (8.9)$$

Therefore, we find the dependence C on L - $C = (a_2^2 L^2 - a_1^2)/4a_2$ and substitute it in the expression (8.9), considering that $\hat{\beta}_2 - \hat{\beta}_1 = L$:

$$\hat{\beta}_{1,2} = -\frac{a_1}{2a_2} \pm L/2 = \beta^* \pm L/2.$$

It follows from the statement 1 that

$$K^* = (L/2)/(L/2) = 1. \qquad (8.10)$$

Using the statement 2, we show that (2.8.10) is optimal

$$\int_{\beta_1}^{L+\beta_1} (a_2\beta^2 + a_1\beta)d\beta = Q(\beta_1) = a_2[(L+\beta_1)^3 - \beta_1^3]/3 + a_1[(L+\beta_1)^2 - \beta_1^2]/2.$$

$$\frac{dQ}{d\beta_1} = a_2(L^2 + 2L\beta_1) + a_1 L \ , \ \frac{d^2Q}{d\beta_1^2} = 2La_2 > 0.$$

From the condition $\frac{dQ}{d\beta_1} = 0$ we find $\hat{\beta}_2 = -\frac{a_1}{2a_2} - L/2$.

Then $\hat{\beta}_2 = \hat{\beta}_1 + L = -\frac{a_1}{2a_2} + L/2$, $K^* = (L/2)/(L/2) = 1$, viz we obtain the ratio (8.10).

Let us consider the two-dimensional case: the vector β has two components β_1, β_2. We find the projection $\Delta I = C$ on the plane (β_1, β_2) under the same assumptions about $I(\beta)$ and $I(\beta)$. Let the area of this figure be equal to *S*. Let it be the ellipse *F* with parameters *a* and *b* and the area $S = \pi ab$. Let the evaluation $\hat{\beta}$ with a certain probability be in the ellipse of scattering with its parameters *p* and *q* of the same area πab. Then we have the following statement.

The statement 3. From all the ellipses of scattering *F* of the area πab of the evenly distributed evaluation $\hat{\beta}$ the $\min\limits_{p,q} \max\limits_{\beta \hat{I} F} \{I(p)\}$ is achieved under condition that $p = a$, $q = b$, i.e. $F = F^*$.

The proof of the statement is similar to the proof of the statement 1: for any point *N* of the arbitrary ellipse, which lies outside the ellipse–projection F^* the value $I(N)$ is larger than *C*, therefore the ellipse should not have the points *N*, that lie outside the ellipse-projection F^* for which $\min\limits_{p,q}\max\limits_{\beta\in F}\{I(\beta)\}$ is achieved

The statement 4. The ellipse-projection F^* provides the minimum of the loss function, viz $\min\limits_F \iint\limits_F I(\beta)d\beta_1 d\beta_2 = \iint\limits_F I(\beta)d\beta_1 d\beta_2$. (8.11)

The proof of the statement is similar to the proof of the statement 2: the loss function for two-dimensional β is the volume of the body, bounded at the top by the surface $I(\beta)$, at the bottom – by the ellipse on the plane $[\beta_1,\beta_2]$. The ellipse *F* is the generator of cylindrical body. Comparing two volumes *V* of cylindrical bodies, the generators of which are the ellipses F and F^*, it is easy to see that

$$\begin{aligned} &V(F) = V_0 + S_1 + I(M_1), \\ &V(F^*) = V_0 + S_1^{`} + I(M_2), \\ &S_1^{`} = S - S_0, \\ &\qquad S_1^{`} = S_0 - S_0 = 0, \\ &\qquad I(M_1) > I(M_2), \end{aligned} \qquad (8.12)$$

$S_0^{`} = F \cap F^*$ - is the generator of cylindrical body, which is the common part of both volumes; M_1 *i* M_2 – are the inner points, such that $M_2 \in F^*$, $M_1 \notin F^*$, $M_1 \in F$, $M_2 \notin F$. Then $V(F) > V(F^*)$.

The illustrative example.

$I(\beta) = (\beta_1^2 + \beta_2^2)/R$. It is clear, that the minimum $I(\beta)$ will be for $\beta = 0$. Let us intersect the paraboloid $I(\beta)$ by the plane $I = R$. The circumference $\beta_1^2 + \beta_2^2 = R$ with the radius R will be the intersection. The same circumference with the area πR^2 will be its projection on the plane $(\beta_1^2 + \beta_2^2)$.

Let us take the random ellipse F $\frac{\beta_1^2}{a^2}+\frac{\beta_2^2}{b^2}=1$ on the plane $\left(\beta_1^2+\beta_2^2\right)$ with the area πab. Let us show that the minimum loss function is achieved if $a=b=R$. We calculate the integral:

$$J(a)=\frac{4}{R}\int_a^b d\beta_1\int_{\beta_2}\left(\beta_1^2+\beta_2^2\right)d\beta_2=\frac{\pi R^2}{16}\left(a^2+\frac{R^4}{a^2}\right),$$

where $\beta_2=(R^2/a^2)\sqrt{a^2-\beta_1^2}$.

The condition of the minimum J on a: $\frac{dJ}{da}=\frac{\pi R^2}{16}\left(2a-2\frac{R^4}{a^3}\right)=0$. From here $a=R$. From the condition $\pi ab=\pi R^2$ we obtain $b=R^2/a=R$. So $a=b=R$ which was to be proved.

The statements 3 and 4 can be generalized in the case of n-dimensional vector β and n-dimensional ellipsoids of scattering that correspond to it.

The statement 5. From all the ellipsoids of scattering of the given volume $S^{(1)}$ the evenly or symmetrically distributed evaluation $\hat{\beta}$, the minimum of the loss function and also $\{I(\beta)\}$ are achieved if $F=F^{(1)}$, where $F^{(1)}$ is the ellipsoid with volume $S^{(1)}$, similar to the ellipse–projection F^*.

The overall conclusion. For the consistency of the quality functionals of the multilevel system it is preferably to maintain their similarities in the space of the estimated parameters β. Then, under the boundedness condition of the accuracy of calculations of functionals, that ellipsoids of scattering of the estimates $\hat{\beta}$ will be similar too, viz the errors $\delta\beta_i$ $(i=\overline{1,n})$ of the estimates $\hat{\beta}$ of component of the vector β are distributed among themselves in order that the variation of the main index $\delta I(\delta\beta)$ in the two-level system or $\delta\Lambda(\delta\beta)$ in the three-level system will be minimum.

8.3. The Recursive Bayesian Estimate of the Vector β

In probability sense, Bayes method that comes out of the equation for the appropriate densities of distribution, gives the most complete and accurate solution in order to obtain the estimate $\hat{\beta}$ of the parameters β of the object under examination:

$$P(y/\beta)\cdot P(\beta) = P(y,\beta) = P(\beta/y)\cdot P(y), \qquad (8.13)$$

$P(y/\beta)$, $P(\beta)$, $P(\beta/y)$, $P(y)$ - are densities of probability measurements *Y*, priori values β, posteriori values β (that depends on $P(\beta)$ and *Y*), measurements of *Y* for all possible β.

From the expression (8.13) we determine the a posteriori density of probability of the unknown vector β:

$$P(\beta/y) = \frac{P(y/\beta)\cdot P(\beta)}{P(y)}. \qquad (8.14)$$

The loss function C which is goaloriented to the main index *I*, is given:

$$C(\delta\beta) = \delta\beta^T \frac{\partial^2 I}{\partial\beta\partial\beta^T}\Big|_{\beta}\cdot\delta\beta\ , \qquad (8.15)$$

where *I* – the main index of the system quality.

We write the expression for the average-risk *R* under condition of statistically representative sample $Y=[y(i),\ldots,y(m)]$, viz the vector *Y* must be considered as a deterministic one, then

$$R = \int c(\delta\beta)\cdot P(\beta/y)d^n\beta, \qquad (8.16)$$

where $P(\beta/y)$ is determined by the expression (8.14), in which a priori density $P(\beta)$ is the regularizing multiplier for $P(\beta/y)$. In most practical problems $P(\beta)$ and $P(y/\beta)$ can be considered as normally distributed:

$$P(\beta) = C_1 exp\left(-\frac{1}{2}(\beta - \beta_0)^T P_0^{-1}(\beta - \beta_0) \right)$$
$$P(y/\beta) = C_2 exp\left(-\frac{1}{2}(Y - X\beta)^T Q^{-1}(Y - X\beta) \right), \quad (8.17)$$

where C_1, C_2 – are the normalizing multipliers, that provide the condition $\int_{-\infty}^{\infty} P(\beta)d\beta = \int_{-\infty}^{\infty} P(Y/\beta)dY = 1$, саме $C_1 = (2\pi)^{\frac{n}{2}} |P_0|^{-\frac{1}{2}}, C_2 = (2\pi)^{-\frac{m}{2}} \cdot |Q|^{-\frac{1}{2}}$;

$$Q = M\{\varepsilon\varepsilon^T\} = \begin{bmatrix} M\{\varepsilon^2(1)\}...M\{\{\varepsilon(1)\varepsilon(M)\} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ M\{\varepsilon(M)\varepsilon(1)\}...M\{\varepsilon^2(M)\} \end{bmatrix},$$

$Y - X\beta = \varepsilon$ – the discrepancy between *Y* and its model $X\beta$.

The density of distribution $P(y)$ (taking into account that $Y - X\beta = \varepsilon$) will also be normal with the expectation $x\beta_0$ and the dispersion matrix $P = xP_0x^T + Q$:

$$P(y) = C_3 exp\left(-\frac{1}{2}(Y - X\beta_0)^T p^{-1}(Y - X\beta_0) \right), \quad (8.18)$$

$C_3 = (2\pi)^{-\frac{m}{2}} |P|^{-\frac{1}{2}} = (2\pi)^{-\frac{m}{2}} [xP_0x^T + Q]^{-\frac{1}{2}}$.

Then the posterior density of the estimate β distribution $P(\beta/y)$ takes the form

$$P(\beta/y) = C_4 \exp\left\{ \begin{array}{l} -\frac{1}{2}(\beta - \beta_0)^T P_0^{-1}(\beta - \beta_0) + \\ +(Y - X\beta)^T Q^{-1}(Y - Y\beta_0)^T p^{-2}(Y - X\beta_0) \end{array} \right\}, \quad (8.19)$$

C_n –is the normalizing constant.

After the appropriate transformations, the expression (8.19) takes the form:

$$P(\beta/y) = C_4 \exp\left\{ -\frac{1}{2}(\beta - \hat{\beta})^T p^{-1}(\beta - \hat{\beta}) \right\}. \quad (8.20)$$

Hence the estimate $\hat{\beta}$ and its covariance matrix *P* are equal to [35]:

$$\hat{\beta} = \beta_0 + PX^T Q^{-1}(Y - X\beta_0),$$
$$P = (P_0^{-1} + X^T Q^{-1} X)^{-1}. \qquad (8.21)$$

For the symmetric loss function (8.15) minimum average risk (8.16) is achieved under condition $\beta = \hat{\beta}$, viz the estimate (8.21) is Bayesian.

On the other hand, if β_0 and P_0 are the estimates, obtained in the previous (*m*-the) step, $\hat{\beta}$ and *P* in *(m-1)*-th step (by adding *(m+1)*-th measurement), then the formula (8.21) implements the recursive method of Bayesian estimation of the vector β and its covariance matrix *P*.

Taking into account the symmetry of the matrices P_0 and Q, the expression (8.21) can be simplified for P:

$$P = P_0 - P_0 X^T [Q + XP_0 X^T]^{-2} XP_0. \qquad (8.22)$$

If $\varepsilon(k)$ is nonstationary "white noise", that $Q = \sigma_k^2 \cdot I$, $k = 1,2,...,m,m+1,...,$ and some vectors are replaced by the scalars $Y = y(m+1), X^T(m+1) = [x,(m+1),...,x_n(m+1)], Q^{-1}, (m+1) = 1/\sigma_{m+1}^2.$

Under these conditions, the method of recursive Bayesian estimation β degenerates into the weighted recursive least squares method:

$$\hat{\beta}(m+1) = \hat{\beta}(m) + x(m)\frac{1}{\sigma_m^2}[y(m) - X^T(m)\beta(m)],$$
$$P(m+1) = P(m) - P(m)X(m)[X^T(m)P(m) + \sigma_m^2]^{-1} X^T(m)P(m); \qquad (8.23)$$

and under condition of the stationary white noise ($\sigma^2 = const$) the expression (8.23) is simplified to the ordinary recursive least squares method:

$$\hat{\beta}(m+1) = \hat{\beta}(m) + x(m)[y(m) - X^T(m)\beta(m)]$$
$$P(m+1) = P(m) - P(m)X(m)[X^T(m)P(m) + 1]^{-1} X^T(m)P(m). \qquad (8.24)$$

If at the time of calculation there is the whole data sample $y(k)$, $X(k)$, $k = \overline{1,m}$, the estimates (8.23), (8.24) are calculated applying the ordinary least squares method:

$$\hat{\beta} = (X^T X)^{-1} X^T Y,$$
$$\operatorname{cov}\hat{\beta} = \sigma^2 (X^T X)^{-1}. \qquad (8.25)$$

The series of methods of the unbiased evaluation β, when Y and X are measured with the errors are considered in [9]. Obviously, the wider is the set of elements $\{J, \sum, \beta, opt, \alpha\}$ and more perfect is the algorithm for searching the element optimal on I, the better is solved the optimization problem of the main index I.

8.4. The Identification and Optimization of the Technological Process Parameters [30]

In order that Ukrainian products come to the world market and be competitive to the foreign standards, it is necessary not only automate the process of its production, but implement the most effective control principles such as adaptation of control systems to the controlled object, based on the correct methods of its identification; adaptation of the process of identification to the non-stationarity of stochastic processes that happens in the controlled object; adaptation of the reference models; optimization of setpoints and the process of stabilization of the appropriate variables of the object.

Let us consider the application of these principles by the example of the automated technological process control system of drawing the quartz tube of the specified diameter d_{int} and wall thickness δ_w from the glass block with external D_{ext} and internal D_{int} diameters (fig. 8.1).

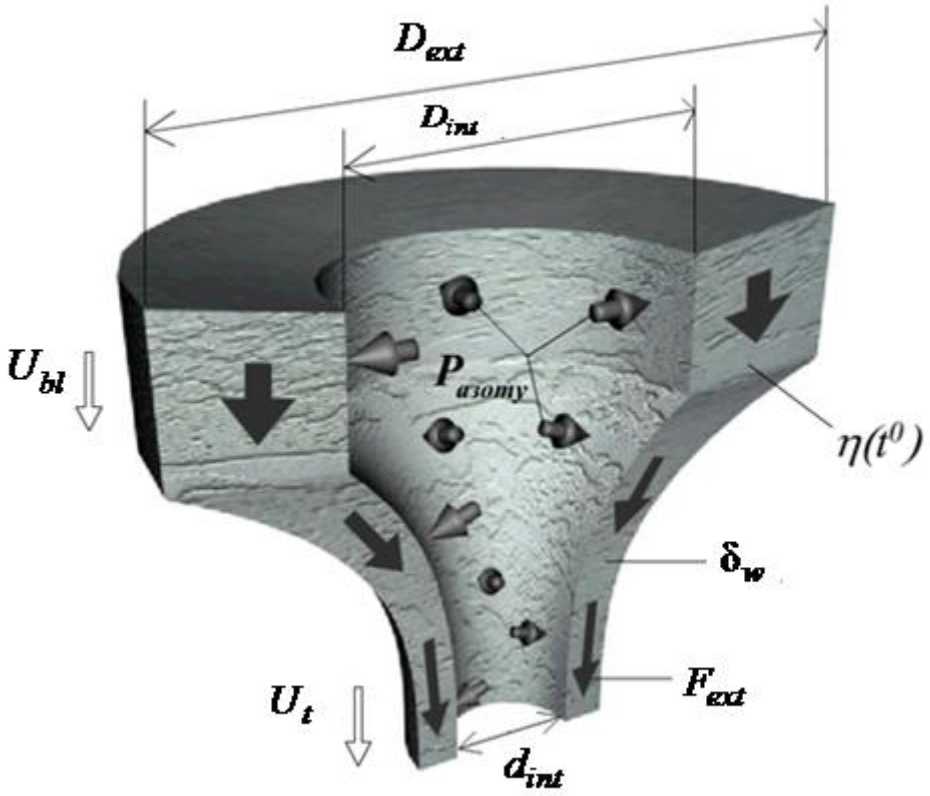


Fig. 8.1. The area of the block where the tube is formed.

The block is warmed up to the temperature t^0 of softening, it is spun out into the tube under effort F_{ext} and pressure P_{nt} from the inside. If the speed of the block's feeding is U_{bl}, and the speed of drawing the tube is U_t, then if the mass is unchangeable, we obtain the relation:

$$\frac{D_{ext}^2 - D_{int}^2}{d_{ext}^2 - d_{int}^2} = \frac{U_t}{U_{bl}} \tag{8.26}$$

or using the wall thickness δ_w

$$\delta_w \left(2d_{int} + \delta_w\right) = \frac{U_{bl}}{U_t}\left(D_{ext}^2 - D_{int}^2\right). \tag{8.27}$$

The equation (8.27) indicates the relation d_{int} and δ_w. Therefore, the controlled object should be considered as a multidimensional one with cross-connections: in order to ensure the stability d_{int} and δ_w, it is necessary to change the pressure P_{nt} and the speed

U_{bl} of drawing the tube simultaneously. Moreover it is required to stabilize the speed of the block feeding, its temperature t^0, the viscosity η of the heated glass mass or the drawing force (the moment $M(\eta)$).

The physical and chemical instability of the block acts as a stochastic perturbation, thus, the automatic control system aims to compensate its effect.

8.4.1. The Mathematical Model of the System in the Mode "Working"

The nonstationary nonlinear stochastic process that takes place in the object, can be represented by the stationary vector-matrix differential operator for the bounded deviations from the nominal mode and the interval of time:

$$L(t)\cdot X_B(t) = \tilde{L}(t)\cdot U(t) \qquad (8.28)$$

or by the matrix transfer function $W(s)$ if using the Laplace transform

$$X_O(s) = W(s)\cdot U(s), \qquad (8.29)$$

where X_O is the vector–function of the output values of the object, namely δ_w, d_{int}, U_{bl}, t^0, $M(\eta)$; $U(s)$, the input values: U_t, P_{nt}, U_{bl}^*, the current I of the block heater.

In the fig. 8.2 it is shown the structure which has the cross-connection of the first and the second channels (the transfer functions W_{12} and W_{21}). The impact of other channels on these ones are taken into account by the parametric perturbations ξ acting on W_{ij}, $i, j = 1,2$, in addition there are the signal perturbations caused by the instability δU_{bl}^* of speed U_{bl} and the influence t^0 on $M(\eta)$.

For their compensation the system has the PI-regulators on separate channels and the diagonalizer $W_g(s)$ is additionally enabled for decoupling the first and the second channels. Then, provided that

$$W_g(s)\cdot W(s)=\operatorname{diag}W(s), \qquad (8.30)$$

the operators W_{ij}^{g} of the diagonalizer are determined:

$$W_{12}^{g}(s)=-W_{12}\cdot W_{11}^{-1}(s), \qquad W_{21}^{g}(s)=-W_{21}\cdot W_{22}^{-1}(s). \qquad (8.31)$$

The dynamics of each channel is close to dynamics of the reference model (the fig. 8.3) by setting the parameters and k_I of the PI-regulator. In order to do this the appropriate channel of the object of inertial channel of the first order is approximated (the fig. 8.4) using the principle of minimal complexity. The coefficients of the PI-regulator are determined under condition of the equivalence of the reference model (the fig. 8.3) and the automation control system of i-th channel.

$$k_n=k_{em}\cdot k_{ii}^{-1}\cdot\tau_{ii};\;\; k_I=k_{em}\cdot k_{kk}^{-1}\;. \qquad (8.32)$$

The block $I(\varepsilon)$ controls the quality of the output product (the deviations of diameter and thickness of the tube) and if the quality decreases, the system goes into the mode “Training” (the fig. 8.5).

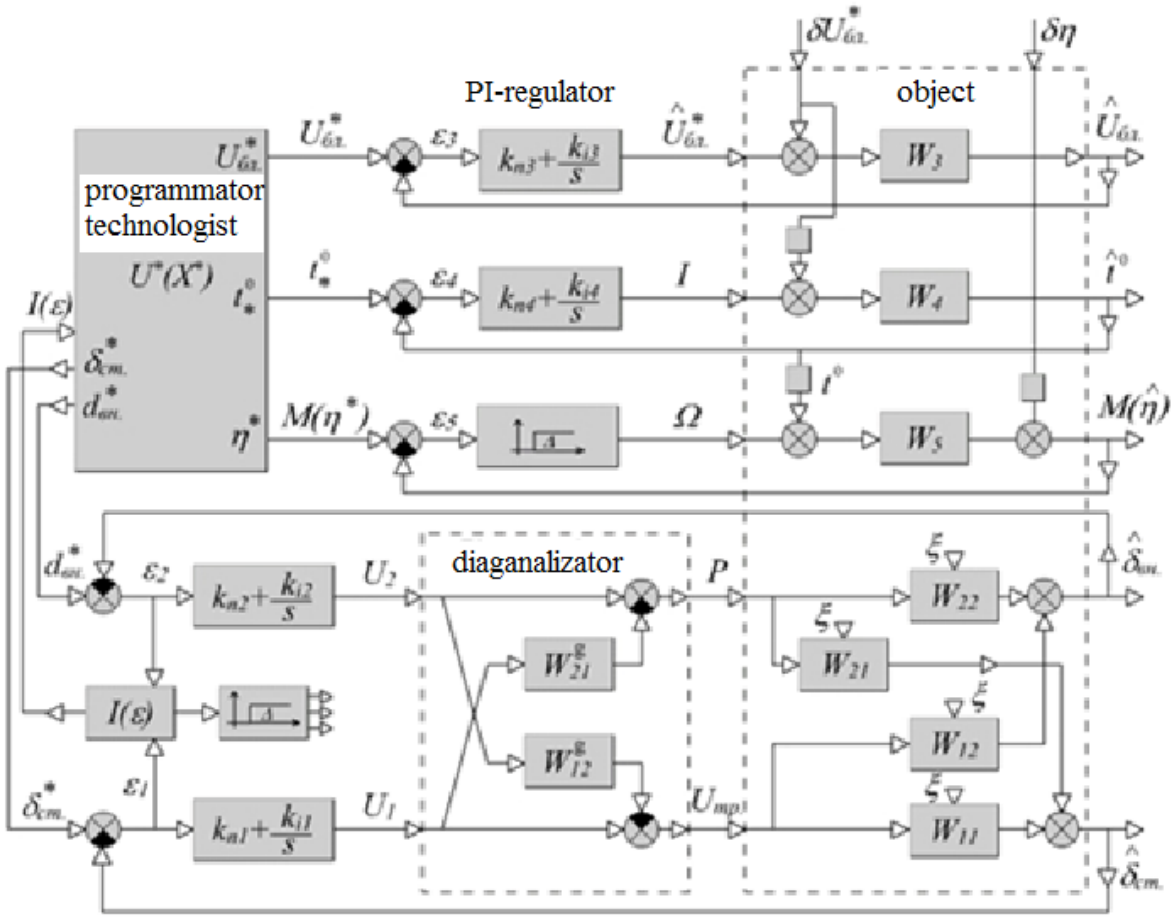


Fig. 8.2. The block diagram of the control system in the mode “Working”.

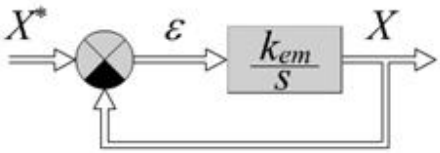


Fig. 8.3. The reference model.

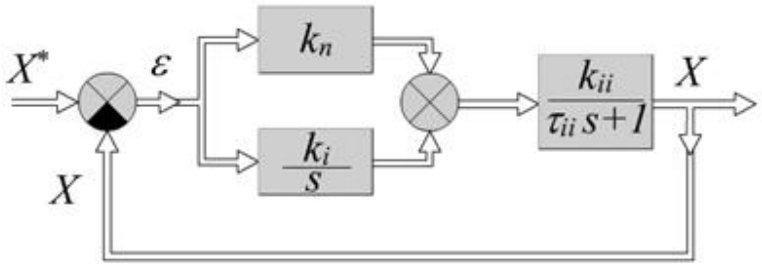


Fig. 8.4. The automation control system with the PI-regulator.

8.4.2. The Mathematical Model of the System in the Mode “Training”

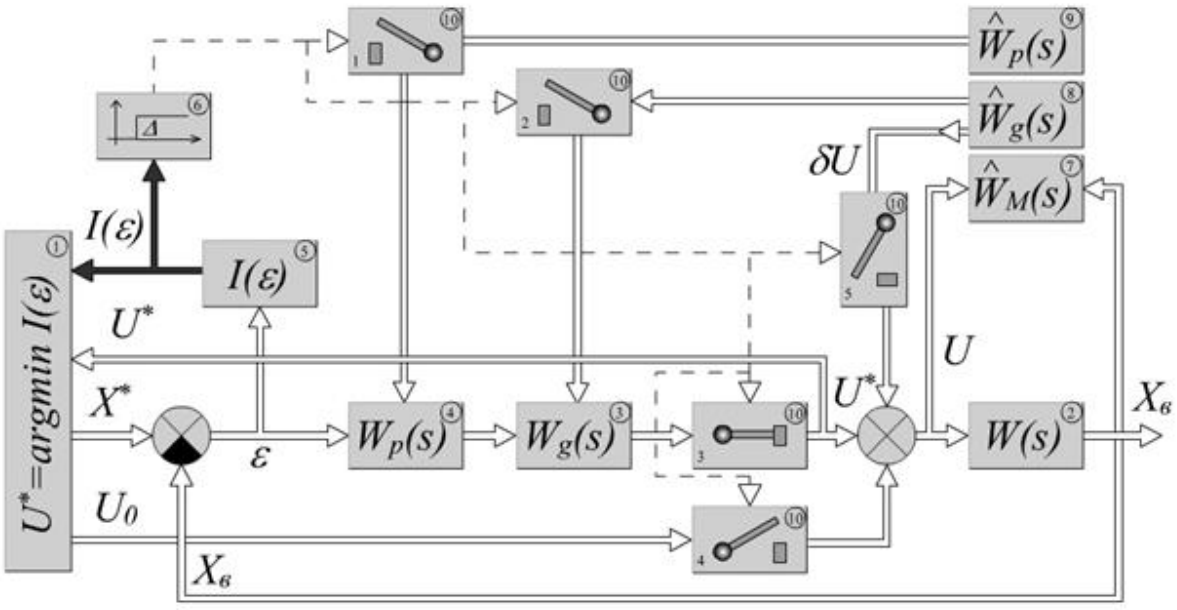


Fig. 8.5. The block diagram of the system.

Switching from the mode “Working” to the mode “Training” is done by closing the key 3 and opening other keys (fig. 8.5). Then the object is subjected by the program’s influence U_0 and by test influence δU. The object’s model $W_M(s)$ is defined in the block 7 and, in accordance with its parameters, the blocks 8, 9 adjust the diagonalizer $W_g(s)$ by the algorithm (6) and PI-regulators $W_p(s)$ - by the algorithm (7).

The mode “Training” always takes place at the beginning of the process of drawing the tube, and it also can occur while drawing, if $I(\varepsilon) > \Delta$.

As an example, let us consider the identification process of direct W_{11}, W_{22} and cross W_{12}, W_{21}, operators of the object at the beginning of tube drawing process. Depending on the sign deviations $d_{вн.}$ and $\delta_{ст.}$ from the nominal values ranging from plus or minus, the block ID feeds mutually independent sequence of steps ΔP and ΔU_t, the amplitude of which gradually decreases (fig. 8.6).

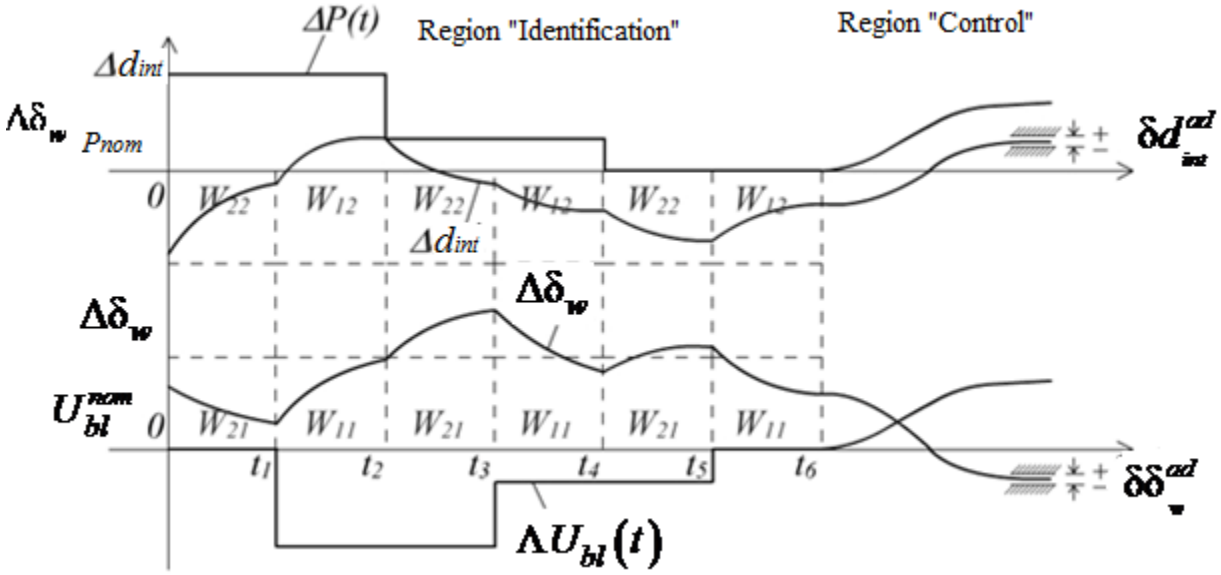


Fig. 8.6 The graphs of transients on d_{int} and δ_w.

Within each step the transition process $\Delta d_{int}(t)$ and $\Delta\delta_w(t)$ nearly ends and is described by the exponent. Then the estimate of the coefficient $\hat{k}_{ij}$, $i=1,2$ is defined as the ratio of the increment of the output variable of the j-th channel to the increment of the appropriate input ΔU_i; the evaluation of the time constant $\hat{\tau}_{ij}$ is determined by the robust Tukey's algorithm as a median of series $\tau_{ij}(k)$, ordered by value:

$$\hat{\tau}_{ij} = \mathrm{Me}\left\{\tau_{ij}(k)\right\}, \qquad (8.33)$$

where

$$\tau_{ij}(k) = -\frac{t_k}{ln\left|\frac{\Delta x_j(t_k)}{\Delta x_j(t_n)} - 1\right|},$$

$k=\overline{0,n}$ - is the discrete time of the particular area $[t_l, t_{l+1}]$, $l=\overline{0,5}$ (see the fig. 8.6).

The further refinement of the estimates $\hat{k}_{ij}$, $\hat{\tau}_{ij}$ is achieved by their linear approximation as functions of amplitude of the testing signal:

$$\hat{k}_{ij}\left(\Delta P,\Delta U_t\right)=k_{ij}^{*}+a_1^{ij}\cdot\Delta P+a_2^{ij}\Delta U_t$$
$$\hat{\tau}_{ij}\left(\Delta P,\Delta\delta_t\right)=\tau_{ij}^{*}+b_1^{ij}\cdot\Delta P+b_2^{ij}\Delta U_t \quad , \qquad (8.34)$$

k_{ij}^{*} and τ_{ij}^{*} will be the sought ones. Having defined the object parameters on all channels, the settings of diagonalizer and regulators are adjusted, the keys 1,2,4,5 are opened and the key 3 connects ACS to the object, viz the system goes into mode "Working".

8.4.3. The Working of the System in the Mode of Nonius Refinement of the Object's Model and Control Algorithm

If the functional $I\left(\varepsilon\right)$ (fig. 8.5) in the mode "Working" is significantly smaller than Δ and the control actions vary slightly, then the information about U^{*}, which specifies the modes' flow chart viz the ratio between the nominal values of the input and output variables of the object, comes to the subsystem "Technologist". Then at these values U^{*}, the commutator 3 (fig. 8.5) is opened again, and the others are closed and the system switches to the mode of nonius refinement of the object's model.

In order to do this, the mutually correlated and autononcorrelated pseudorandom binary sequences of testing signals are fed from the block 8 (fig. 8.5) to the appropriate channels (fig. 8.7) and the parameters k_{ij}, τ_{ij} of basic models are specified by the least squares method for the smoothed data; the approximation error of object by basic models is defined, next the parameters of refining operators of the nonius models are evaluated applying the same method, if the basic operator is already known. For example, the inert differential operator of the channel of temperature stabilization:

$$W_y\left(s\right)=\frac{T_4 s+1}{\tau_4 s+1}. \qquad (8.35)$$

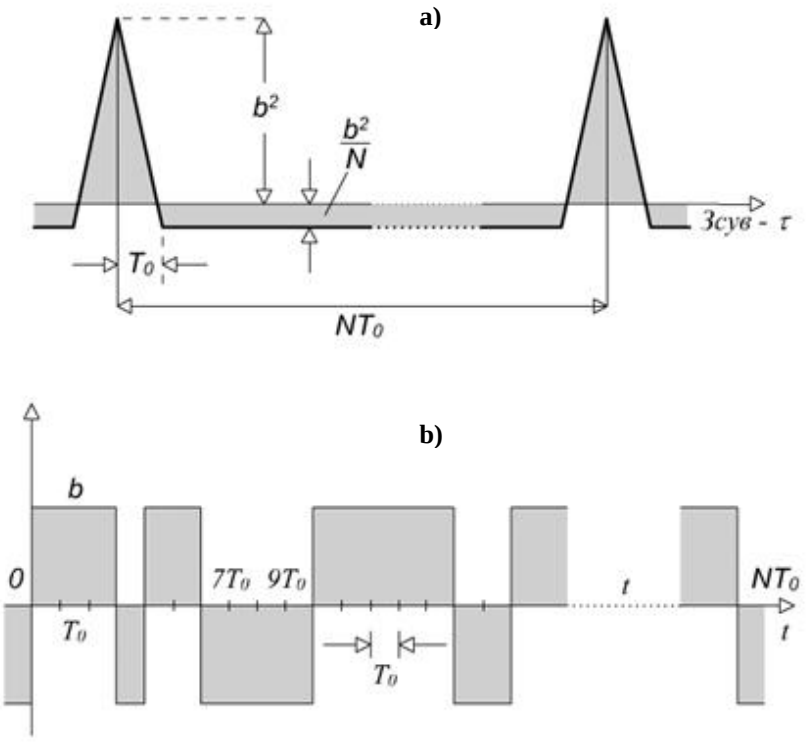


Fig. 8.7. The correlation function (a) of the signal (b).

Then the nonius model is

$$W_{M4} = \frac{k_4^*}{\tau_4^* s + 1} \cdot \frac{T_4 s + 1}{\tau_4 s + 1}. \qquad (8.36)$$

In order that dynamics correspond to the reference one, the control algorithm is corrected by connecting to the PI-regulator of compensating operator $W_y(s)$ of the level $W_y^{-1}(s)$ (fig. 8.8).

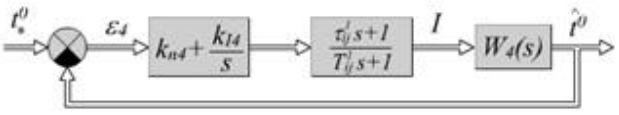


Fig. 8.8. The nonius correction of the stabilization t^0 channel.

In the process of the nonius identification the index $I(\varepsilon)$ value is controlled and, if it is close to Δ, the system switches to "Working" with the model that is succeeded to identify. Thus, the adaptability to non-stationary perturbations is achieved: the more

stationary is the process, the more accurate is the model and the less is the value $I(\varepsilon)$. But even under condition of non-stationarity the system succeeds to construct the simple base model and provide stability and quality of the control process.

Besides, if $I(\varepsilon) << \Delta$, the system can reduce the inertia of the reference models and, consequently, the channels of the stabilization of process parameters.

8.5. The Identification and Optimization of the Process of Self-Study with the Electronic Simulator

8.5.1. The Introduction

The process of self-study occupies an important place in the electronic teaching resources of the discipline (ETRD) [27, 31, 32]. An electronic simulator is one of means of self-study. The electronic simulator should provide the optimal process of transmission of information to the person who studies in order to approach the level of direct communication “teacher – student”.

The optimality of the process of the information transfer is understood as a minimization of the loss of information when transferring it from the “teacher” to the learner (pupil, student, etc.) and the assimilation of this information as well. Such electronic simulator as part of the ETRD and also the teacher and the learner create the two-level automated control system of learning process. The teacher (the upper level) defines the way (count) of passing the required sections of the discipline, the weight of these sections in the overall assessment of the level of the learner’s knowledge.

The electronic trainer and the learner create the lower level of the system where the learner is the controlled object and the simulator is the regulator of the learning process of this object. The teacher and the learner can plan the overall pace and the duration of studying the discipline and the desired rating (score) of learning. The simulator must provide the control influence that is adaptive (preferably optimal) to the student’s parameters. The controlling variables of the system are:

1. the average rate of information (time interval between classes);

2. the trajectory of moving forward from section to section and, if necessary, returning back to course modules that have been already passed, based on the information of the current student testing;

3. the volume of motivational information.

The returning to the previous modules as a controlling influence is required at the low index of current control (i.e., understanding) of knowledge; the motivation is required as the impact on the quality of learning.

Thus, having the sufficient level of formalization and capabilities to control the the current parameters of the model of the object of training, the learning process on an electronic simulator can to be submitted (with the appropriate degree of approximation) as the automatic control system (ACS) with the identifier and the synthesizer of the optimal mode.

8.5.2. The Description of the Elements of the System of Automatic Control of Training

The teaching material of the relevant subject is divided into thematically coherent blocks, that are of the same volume, the classes' duration is 75 ± 15 minutes.

The material of each n-th block is divided into informative $I(n)$ and motivational $M(n)$ components of the subject matter. There are several levels of incentives depending on the student's success rate $k(n)$. For n-th block the rate is set based on the results of current monitoring of individual microblock within the limits of n-th and final monitoring for the whole n-th block.

The control in the microblocks is exercised by representing the material in the interview mode: 30 seconds of language and appropriate on-screen information, the question and 3 answers: surface, normal, and profound. The student, opting for one of them, gains a certain number of points, and the trainer, having received this information by

the feedback channel, corrects the further way of giving teaching material: the skip to the next microblock or return to the microblocks which have been already passed, and which are logically associated with the current one.

On completion the process of submitting the information $I(n)$ of the n-th block, the final control is made for the n-th block by the index $K(n)$, the decision is accepted depending on the actual and the desired level (specified by the program path $KV^{*}(n)$ of the accumulation of rating):

If $K(n) \geq K_{min}(n)$, then, after having the desired pause, there is the transition to ($n+1$)-th block; if $K(n) < K_{min}(n)$, then there is the repeat of n-th block without pause with the increase of the motivational component $M(n)$.

At the beginning of each n-th class, the student undergoes the control of residual knowledge at the $(n-1)$-th block-lesson and gets his rating $KV(n)$, depending on the level, then he goes to the assimilation of n-th block in accordance with $KV(n)$ the level of motivation or he returns to re-studying the $(n-1)$-th block. The total rating $\sum_{n} KV(n)$ is considered as a function of time t and adjusted by the automation control system by changing the pauses T_n between the blocks-lessons. The sequence of actions at one lesson is represented in the fig. 8.9.

Input control	Micromodules of the n-th block (n,i) , $i = \overline{1,m}$										Output control
$KV(n\text{-}1)$	$M(n,1)$	$I(n,1)$	$K(n,1)$		$M(n,i)$	$I(n,i)$	$K(n,i)$		$M(n,m)$	$I(n,m)$	$K(n,m)$	$KV(n)$

Fig. 8.9. The components of the n-th lesson.

For the successful operation of the automatic control system it is required to build the structure of the learner's model as the controlled object and to define the current values of the parameters:

$$K(n) = K_0 + K_M \cdot M(n) + K_r \cdot R(n), \tag{8.37}$$

$$KV(n) = K(n) \cdot exp\left[-T_n / \tau_n \right] + KV_r \cdot R(n), \tag{8.38}$$

where K_0, K_M і K_r – are coefficients of the linear stochastic model of the dependence $K(n)$ on the level of motivation $M(n)$ and incidental $R(n)$, $R(n)$ - is the Gaussian noise with the unit variance, τ_n -is the time constant of exponential forgetfulness of information of the $(n-1)$-th lesson at the n-th lesson:

$$\tau_n = \frac{t_n - t_{n-1}}{ln\left(K(n) / KV(n) \right)}. \tag{8.39}$$

The exponential process of knowledge accumulation within one lesson is given by the linear model (8.37) due to the short time of one session; the process of information forgetfulness in a pause T_n between classes is taken as exponential.

The desired optimal trajectory $\sum_{j=1}^{n} KV^*(t_j)$, $n = \overline{1, n_K}$ of knowledge accumulation at the time t_j is given on the basis of psychological aspects of training by a teacher or a student, depending on the planned level of knowledge and time reserves.

The task of the simulator as a system of automatic process control is to provide the closeness of the real trajectory $\sum_{j=1}^{n} KV(t_j)$ to the desired $\sum_{j=1}^{n} KV^*(t_j)$ that minimizes the functional error $\varepsilon(n)$:

$$\varepsilon(n) = KV^*(n) - KV(n), \tag{8.40}$$

$n = \overline{1,n_k}$, where n_k - is the number of finitesimal lesson. This problem will be solved using the proportional (k_P) and integral (k_I) regulator [20] of the interval ΔT_n between n-th and m-th lessons:

$$\Delta T_n = \frac{\Delta T_{min} + \Delta T_{max}}{2}\left[\frac{1}{2} + K_P \frac{\varepsilon(n)}{\varepsilon_{max}} + K_I \frac{\sum_{j=1}^{n}\varepsilon(j)}{\varepsilon_{max}}\right], \qquad (8.41)$$

where ΔT_{min}, ΔT_{max} – are restrictions on the smallest and the largest interval. The model of discrete ACS with frequency-pulse modulation, identification and adaptation to the controlled object corresponds to the process of training on a simulator (fig. 8.10):

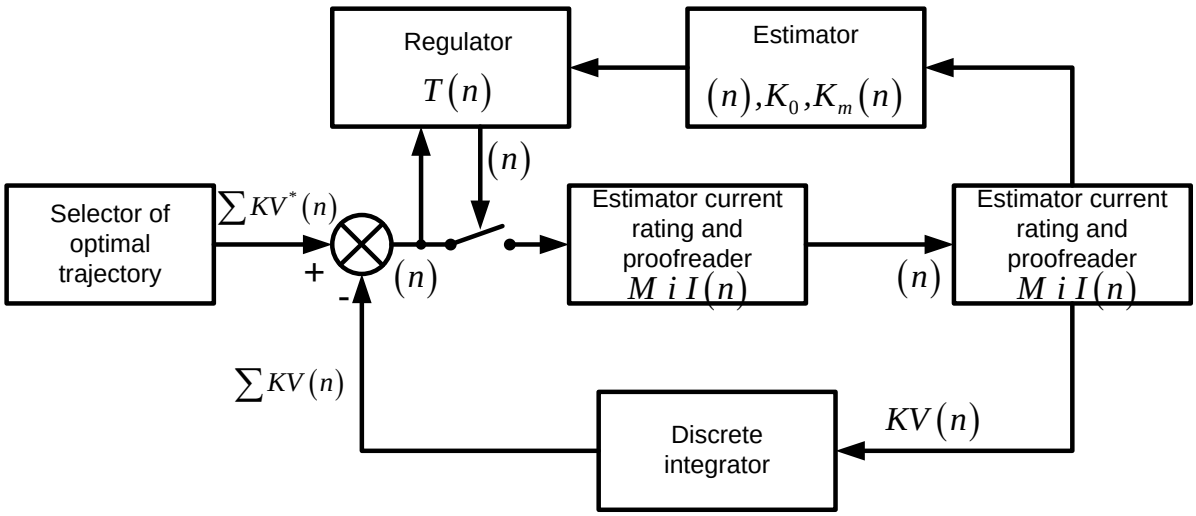


Fig. 8.10. The functional diagram of the ACS of training.

The control of motivation $M(n)$ is exercised during the n-th lesson depending on the quality $K(n,i)$ of learning of the i-th lesson. The increase of $M(n,i)$ leads to (at a fixed time $T \pm \Delta T$ of the n-th lesson) reduction of the number of *m* microblocks in the n-th block. Then the regulator $\Delta T(\varepsilon,\tau,n)$ reduces the break between classes in order that $\sum KV(n)$ approaches to $\sum KV*(n)$.

8.5.3. The Numerical Modeling of the Process

The simulation of automatic control of training (fig. 8.10) is carried out in such modes using MATLAB: $K_r = KV_r = 0$), $\Delta T(n) = const$ is determined (the perfectly concentrated student), viz the system (fig. 8.10) is disconnected and as a result the trajectory $\sum KV(n)$ is slightly ahead of the optimal $\sum KV^*(n)$ (fig. 8.11); there is the same student, but ACS is locked and has the K_P - regulator (fig. 8.12); there is the same system with proportional (K_P) and integral (K_I) regulating law $\Delta T(n)$ from $\varepsilon(n)$ (fig. 8.13).

As we see in the case (fig. 8.12) the mean square error of tracking the desired trajectory of training is minimal; if the random component is not zero $KV_r = 0{,}2$ in the model (8.38) (the non-ideal student), then the error for the system (fig. 8.13) and the number of lessons slightly increase (the error is from 1,62 to 3,67; the number of lessons is from 39 to 58).

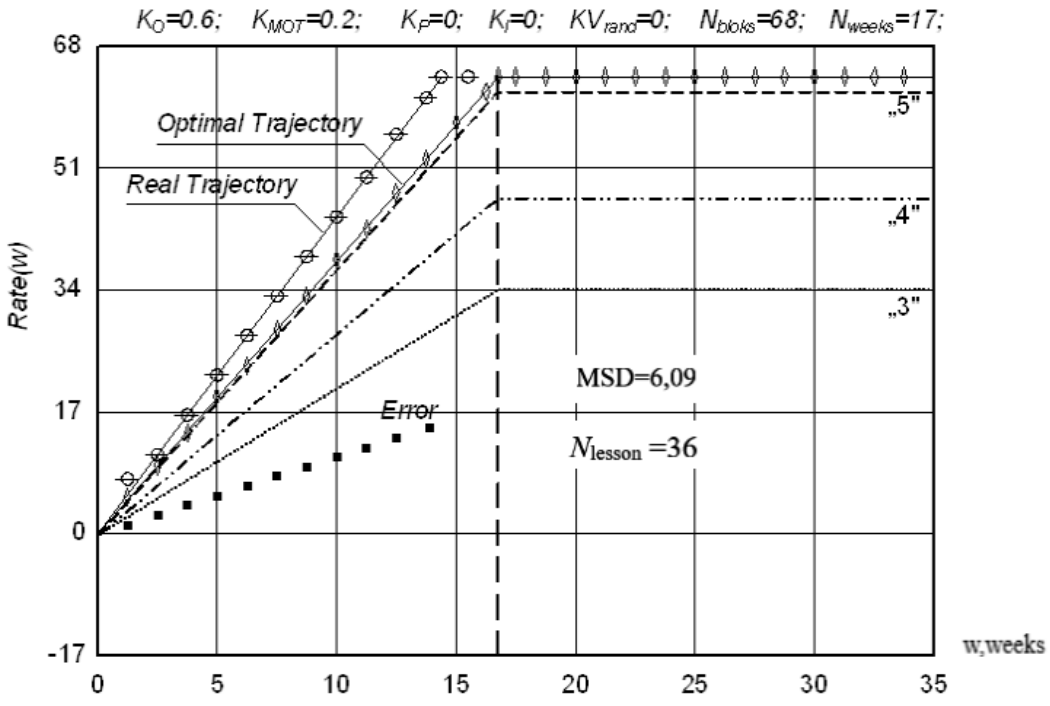


Fig. 8.11. The open-loop control system of the ideal student.

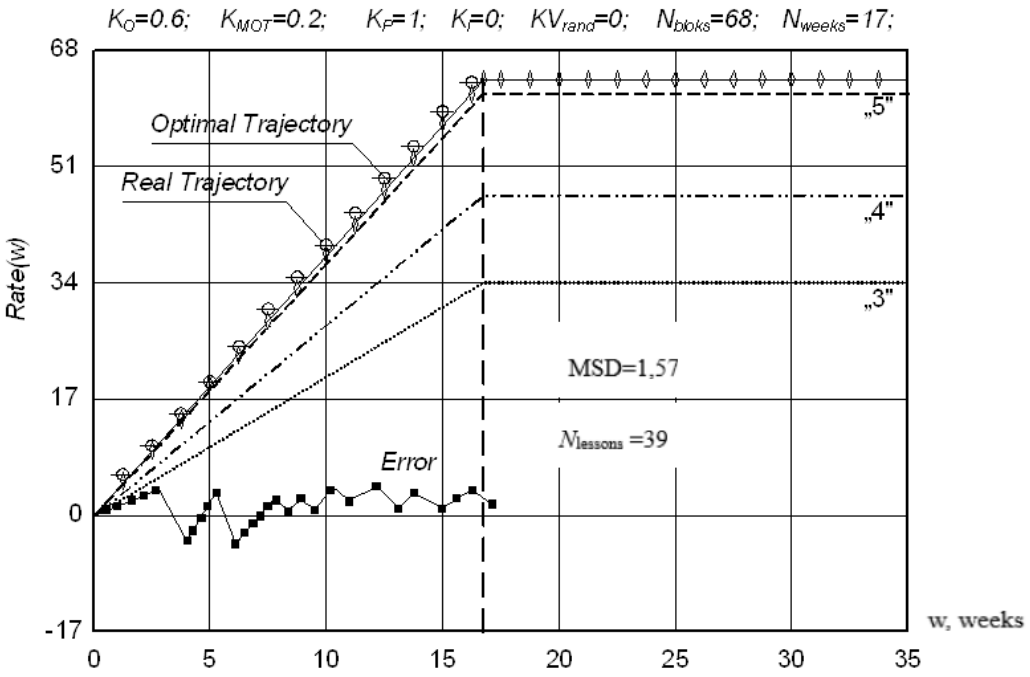


Fig. 8.12. The closed loop automatic control system with the proportional regulator (the ideal student).

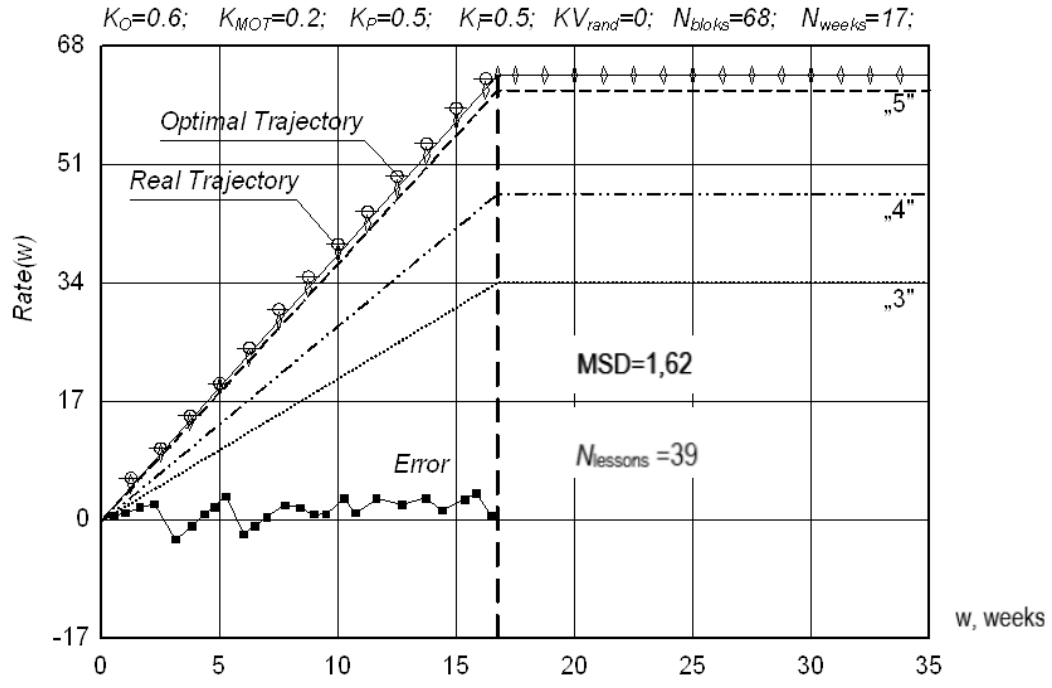


Fig. 8.13. The closed loop automatic control system with the integro-proportional regulator (the ideal student).

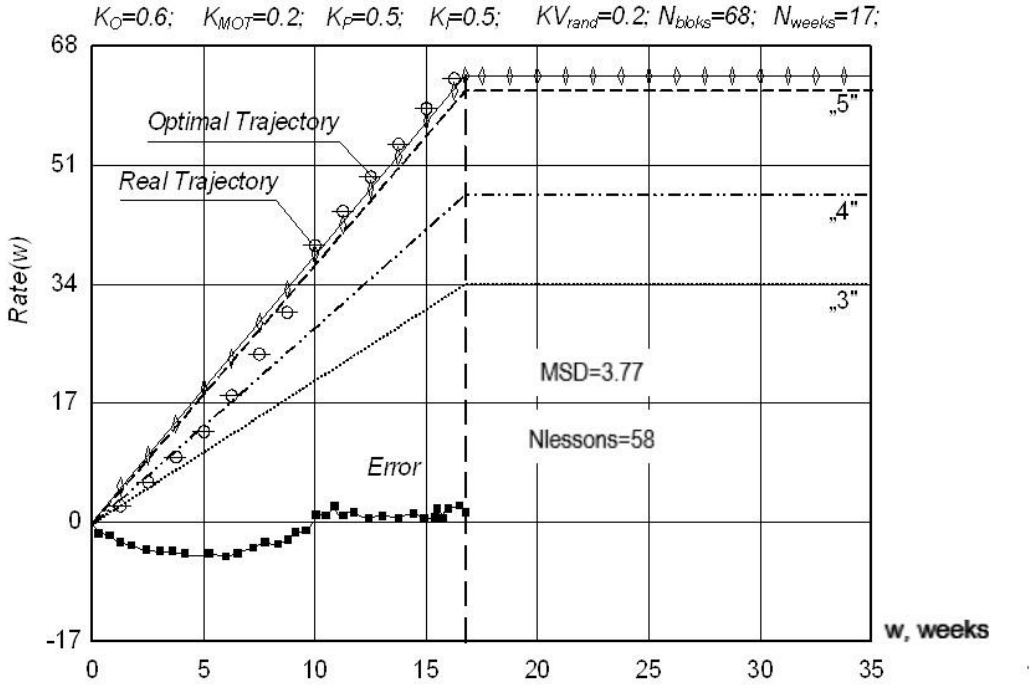


Fig. 8.14. The closed loop automatic control system with the integro-proportional regulator (the real student).

Thus, the standard proportional and proportionally integral control laws provide tracking the optimal trajectory with the error of 1.57 - 3.77 units with total maximum rating of 68 units, i.e. with the error 2 – 6 %.

For the further improvement of the process of training the real student $\left(KV_r = 0{,}2\right)$, let us consider the possibility of constructing the optimal adaptive (to the student) control law instead of the standard proportional and integral one. Let the total current rating of the student, after executing the input control of residual knowledge of the $\left(n-1\right)$-th lesson $\left(KV\left(n-1\right)\right)$, be $\sum_n KV\left(n\right)$ at the end of the *n*-th lesson (fig. 8.9, the time t_{n2}). The total of estimates of the current control on *m* micro blocks of the *n*-th lesson is $K\left(n\right)$.

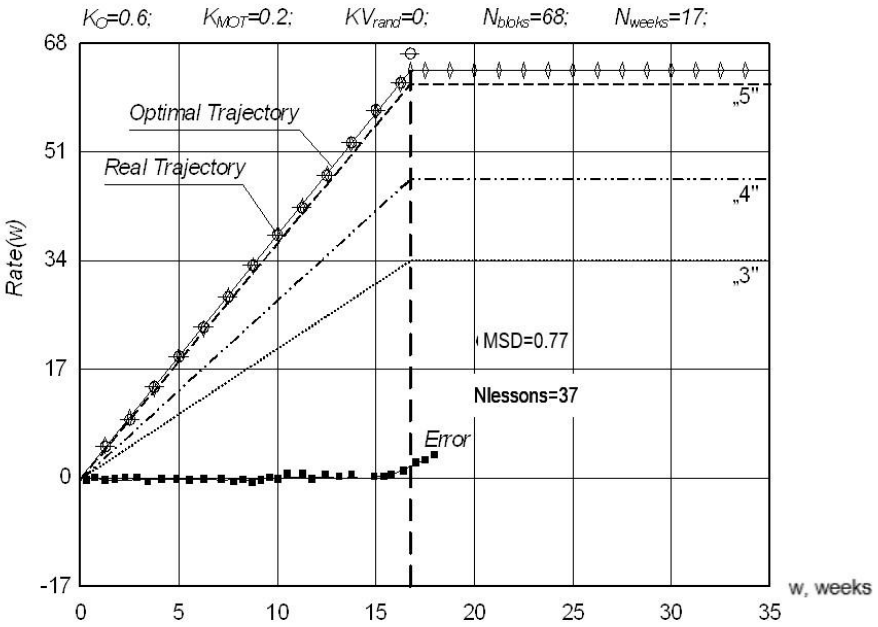


Fig. 8.15. The optimal system of automatic control of training process (the ideal student).

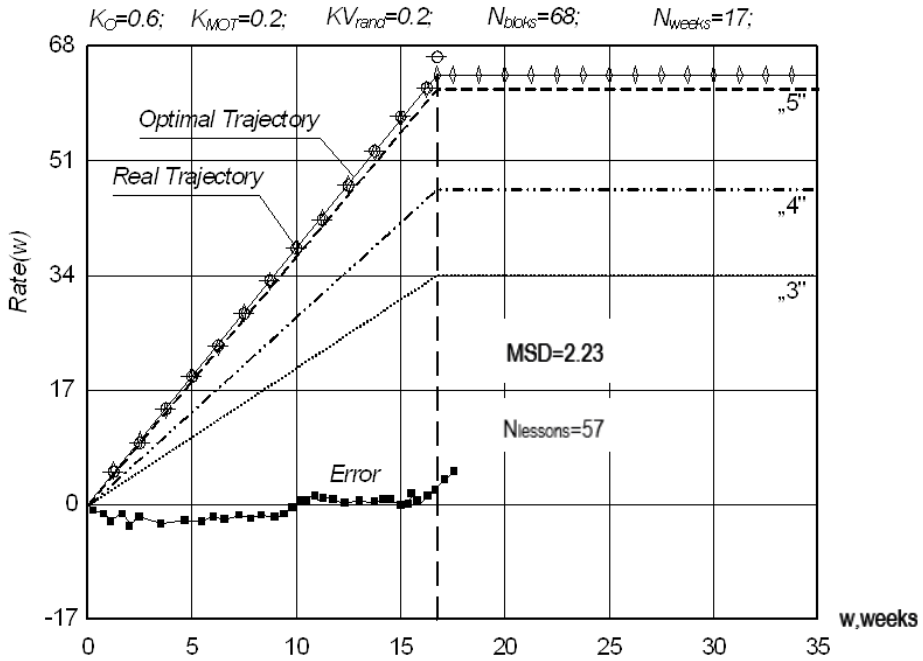


Fig. 8.16. The optimal system of automatic control of training process (the real student).

The optimal (desired) value of the total rating is known at the time t_{n1}, this is $\sum_n KV^*(n)$. Taking into account that the level of student's knowledge exponentially decreases (the time constant τ_n) at the beginning of the $(n+1)$ lesson, then for the moment $t_{n+1,1}$, the forecasting rating of student is determined by the expression:

$$\sum_n KV(n+1,\, t_{n+1,1}) = \sum_n KV(n,\, t_{n,2}) + K(n,\, t_{n,2}) \cdot \exp\left[\frac{t_{n+1,1} - t_{n,2}}{\tau_n}\right]. \quad (8.42)$$

Having equated the expression (8.42) to the optimal rating $\sum_n KV^*(n+1)$, we obtain the expression for determining the optimal time $t_{n+1,1}$ to start the $(n+1)$-th lesson:

$$\sum_n KV(n,\, t_{n,2}) + K(n,\, t_{n,2}) \cdot \exp\left[\frac{t_{n+1,1} - t_{n,2}}{\tau_n}\right] = \sum KV^*(n+1). \quad (8.43)$$

From here the optimal time $t_{n+1,1}$ to start the $(n+1)$-th lesson is

$$t_{n+1,1} = t_{n,2} + \tau_n \, ln \frac{\sum KV^*(n+1) - \sum KV(n, t_{n,2})}{K(n, t_{n,2})} \quad (8.44)$$

As we can see, the time constant τ_n and the indicators of total $\sum KV(n, t_{n,2})$ and current $\sum KV(n, t_{n,2})$ ratings have influence upon the $t_{n+1,1}$. On the right side of the expression (8.44) all components are known, with the exception of the time constant τ_n. It can vary from lesson to lesson depending on the state and degree of student's motivation. Therefore, in order to solve the equation (8.44) it is required to set τ_n. The easiest way to do this is to assume that $\tau_n = \tau_{n-1}$, where τ_{n-1} is found from the previous lesson from the expression (8.42), if taking *n* instead of $(n+1)$ and $(n-1)$, instead of *n:*

$$\sum_n KV(n,\, t_{n,1}) = \sum_n KV(n-1,\, t_{n\text{-}1,2}) + K(n-1,\, t_{n\text{-}1,2}) \cdot \exp\left[\frac{t_{n,1} - t_{n\text{-}1,2}}{\tau_{n-1}}\right]; \quad (8.45)$$

from here

$$\tau_{n-1} = \frac{\left(t_{n,1} - t_{n-1,2}\right)}{ln\frac{\sum_n KV\left(n,t_{n,1}\right) - \sum_n KV\left(n-1,t_{n-1,2}\right)}{k\left(n-1,t_{n-1,2}\right)}}. \qquad (8.46)$$

The more accurate forecasting value τ_n is determined (if n>3) by the linear approximation $\tau(t)$ and by the forecast to the interval $\left(t_{n+1,1} - t_{n,2}\right)$:

$$\tau_{n+1} = \tau_n + \left[t_{n+1,1} - t_{n,2}\right] \cdot \frac{\tau_n - \tau_{n-1}}{t_{n,1} - t_{n-1,2}}. \qquad (8.47)$$

The computer modeling of the optimal ACS, as a mean of ETRD, is conducted under the same conditions as the ACS with the proportionally integral control law: $k_0 = 0,6$; $K_M = 0,2$; the number of blocks $N = 68$, the number of weeks – 17, $K_r = 0,2$ (the real student). The result of the simulation of the optimal mode of training is shown in the fig. 8.15, 8.16. Comparing the ACS with the regulator (5) (fig. 8.15) and the optimal regulator (8.44) with the additional restrictions on a break T_n between lessons $\left(maxT_n > T_n > minT_n\right)$, due to the fact that the expression (8.44) is adjusted by the restrictions on the length of the pause, we can conclude that at nearly the same number of lessons (58 and 57) under condition of the optimal regulator, the trajectory of accumulation of student's rating is much closer to the desired one (the mean-square error is 2,23).

Thus, we have the possibility to automate and optimize the process of training under condition of its appropriate formalization, based on a systematic approach with the appropriate degree of adequacy.

We believe that the quality of education process will increase significantly and reach the level of direct communication between the teacher and the student, if the material is submitted by micro-blocks (the student is not distracted for 30 seconds) with the questions at the end of each micro-block and three correct answers of different degree of understanding the question. The student chooses the best one for him, the teacher-trainer changes the contents of the next micro block depending on the answers selected by

student. Such dialogue with the initial and current motivation and control at the end of the lesson can dramatically improve the training process.

References

1. Анализ и статистическая динамика систем автоматичного управления/ Под ред. Н.Д. Егунова. – М.: Изд-во МГУ им. Н.С. Баумана, 2000, 748 с.
2. Галай В.М., Сільвестров А.М. Нонісна адаптивна система стабілізації стохастичних технологічних процесів. В сб. «Системні дослідження та інформаційні технології», К.: НТУУ «КПІ», ІПСА, 2003, №1, с. 135-142.
3. Гроп Д. Методы идентификации систем. М.: Мир, 1979, 684с.
4. Ейкхофф П. Основы идентификации систем управления. - –.: Мир, 1975, 464с.
5. Івахненко А.Г. Долгострочное прогнозирование и управление сложными системами. – К.: Техніка, 1975. – 312 с.
6. Казаков И.Е. Статистическая теория систем управления в пространстве состояний.– М.: Наука, 1975. – 432 с.
7. Калман Р., Фалби П., Арбиб М. Очерки по математической теории систем. М.: Мир, 1971, 400 с.
8. Кашин Г.М., Федоренко Г.И. Автоматическое управление продольным движением упругого самолета. М.: Машиностроение, 1974, 312с.
9. Костюк В.И., Сильвестров А.Н. Принцип многократной адаптивности// достижения и перспективы развития тех. Кибернетики. К.: Изд-во РДИП, 1975. т.1, с.65-69.
10. Костюк В.И., Сильвестров А.Н. Асимптотическая инвариантность многократно адаптивных систем// Теория инвариантности и ее применение. – К.: Наукова думка, 1979, с. 155-167.
11. Корн Г. Справочник по математике / Г. Корн, Т. Корн М.: Наука, 1970, – 832 с.
12. Крамер Г. Математические методы статистики. – М. Мир, 1975, 648с.
13. Красовский А.А.. Буков В.Н., Шендрик В.С. Универсальные алгоритмы оптимального управления непрерывными процессами. М.: Наука, 1977, 270с.
14. Круг Г.К., Сосулин Ю.А., Фатуев В.А. Планирование эксперимента в задачах идентификации и экстраполяции.М.:Наука 1977, 207с.
15. Ланкастер П. Теория матриц. М.: Наука, 1978. – 280 с.
16. Любич Ю.И., Майстровский Г.Д. "Общая теория релаксаційних процессов.- Успехи математических наук, 1970, №1, с.57-112.
17. Льюнг Л. Идентификация систем: Теория для пользователя. Пер. с англ./ Под ред. Я.З.Цыпкина. - –.: Наука, 1991. -432 с.
18. Маркова Е.В., Лысенков В.Н. Комбинированные задача в планах многофакторного эксперимента. М.: Наука, 1977, 345с.

19. Маркус М., Минк Х. Обзор по теории матриц и матричных неравенств. М: Наука, 1979, 232с.

20. Математические основы теории автоматического управления/ Под ред. Б.К. Чемоданова. М.: высшая школа, 1977, т.1, 366с.

21. Методи класичної та сучасної ТАУ: Підручник в 3-х т./Під ред. М.Д. Егупова.- М.: Вид-во МГТУ ім.. И.П. Баумана, 2000. – 736с.

22. Поляк Б.Т.,Цыпкин Я.З. Помехоустойчивая идентификация // Идентификация и оценка параметров системы // Труды IV Симпозиума ИФАК, препринты ИПУ.М.1976.4.1.с.190-213.

23. Пухов Г.Е., Хатиашвили Ц.С. Модели технологических процессов. Киев: Техника, 1974.- 200 с.

24. Райбман Н.С., Чадеев В.М. Основы управления технологическими процессами. М.: Наука, 1980, 374с.

25. Рао С.Р. Линейные статистические методы и их использование. – М.: Наука, 1968, 548с.

26. Самсонов В.В., Сильвестров А.Н., Зименков Д.К. Две стратегии идентификации реальных объектов // Сучасні методи, інформаційне, програмне та технічне забезпечення систем управління організаційно-технологічними комплексами. – К.: НУХТ, 2009. – с.93-94.

27. Самсонов В.В., Сільвестров А.М., Скринник О.В. Системний підхід до розробки моделі керування процесом навчання// Наукові праці Національного університету харчових технологій К.: НУХТ.- 2010 р., № 32, с. 104-106.

28. Самсонов В.В., Сільвестров А.М. Ідентифікація моделі прогнозу прихованих закономірностей// Тези доповідей “Міжнародна науково-практична конф. “Новітні технології, обладн., безпека та якість харчових продуктів: сьогодення та перспективи” К.: НУХТ. - 2010, с.59-60.

29. Самсонов В. В., Сильвестров А.Н. Комплексування методів ідентифікації// Збірник наукових праць «Проблеми транспорту». – К.: НТУ, 2010, с. 155-161.

30. Самсонов В.В., Сільвестров А.М. Адаптивна система стабілізації параметрів технологічного процесу// Наукові праці Національного університету харчових технологій К.: НУХТ.- 2011 р., № 37, 38 с. 168-173.

31. Самсонов В.В., Поворознюк Н.И., Сильвестров А.М. Моделирование процесса самостоятельного обучения с электронным тренажером// Інформаційні технології в освіті. Вип.1. Херсон: ХДУ, 2008. с. 126-133.

32. Самсонов В.В., Сільвестров А.М., Костіков М.П. Алгоритм адаптивного навчання в системі електронних навчально-методичних ресурсів дисципліни// Тези доповідей Міжнар. наук.-практ. конф. «Інформаційні технології в освіті, науці і техніці» - Черкаси: ЧДТУ, 2012. – т.2. – с. 85-86.

33. Себер Дж. Линейный регрессионный анализ. -М.: Мир, 1980, 456с.

34. Сейдж Э.П., Мелса Дж. Идентификация систем управления. М.: Наука,1974, 248с.

35. Сильвестров А.Н., Чинаев П.И. Идентификация и оптимизация автоматических систем. М.: Энергоатомиздат, 1983, 200 с.

36. Сильвестров А.Н., Самсонов В.В. Адаптация системы управления технологическим процессом// XXVI междун. наук. конф. “Математические методы в технике и технологиях (ММТТ-24)”, сб. трудов, т.6. – К.: НТТУ “КПИ”, 2011. – с.17-21.

37. Сильвестров А.М., Панченко О.М. Многократно адаптивные системы идентификации. К.: "Т«хника",»1983, 110 с.

38. Синеглазов В.М, Сильвестров А.Н. Идентификация статических и динамических характеристик реальных объектов//Електроніка та системи управління. НАУ. 2005. №4(6), с. 80-87.

39. Сільвестров А.М., Резніченко В.М. Спосіб вимірювання аеродинамічних коефіцієнтів повздовжнього руху літаків. Патент N48566A, Бюл. N8 від 15.08.2002р.

40. Сільвестров А.М., Синєглазов В.М. Ідентифікація статистичних і динамічних характеристик реальних об’єктів// В.сб. «Електроніка та системи управління». К.: НАУ, 2006, №4.

41. Снешко Ю. И. Устойчивость и управляемость самолета. М.: Машиностроение, 1987, – 300 с.

42. Современные методы идентификации систем/ Под ред. П. Ейкхоффа – М.: Мир, 1983 – 440 с.

43. Солодовников В.В., Матвеев П.С. Расчет оптимальных систем автоматического управления при наличии помех. М.: Машиностроение, 1973, 240с.

44. Фельдбаум А.А. Основы теории оптимальных автоматических систем. М.: Физматгиз. 1963, 552с.

45. Wiener N. Extrapolation, interpolation and smoothing of stationary time series//John Wiley, 1949. – 63р.

46. Петров Ю.П., Сизиков В.С. Корректные, некорректные и промежуточные задачи с приложениями. Санк - Петербург, Политехника, 2003. – 261 с.

Content

The Introduction……..3
Chapter 1. The System Approach to Identification……11
1.1.The Definition of the System……11
1.2. The Multi-Level Decomposition of Systems……20
1.3. The Principle of Criteria's Consistency……31
1.4. The Composition of the Subsystems of Different Levels into the Single System……44
Chapter 2. The Basic Model of the Real Processes and its Derivative Models……..49
2.1. The Basic Mathematical Models of Real Processes……49
2.2. The Discrete Time Mathematical Models that Correspond to the Basic Continuous Model Exactly or Approximately (2.2)……54
2.3. The Determining the Structure of the Nonlinear Operator f of the Model (2.1) by Composing the Local Models (2.2) into the Global one……..57
2.4. The Modeling and Forecasting of the Solution (2.11) of the System (2.1) if the Information on the Mapping $\mathbf{T}\times\mathbf{T}\times\mathbf{X}\times\mathbf{\Omega}\to\mathbf{X}$ is Missed……63
2.5. The Modeling, Forecasting and Diagnostics of the Solution (2.11) of the System (2.1) if the Information on the Mapping $T\times T\times X\times\Omega\to X$ is Available…….71
Chapter 3. The Analysis of the Simplified Methods of the Confluent Analysis…….78
3.1. The Least Square Method's Analysis if the Input and Output Signals are Noisy……78
3.2. The Generalized LSM and its Practical Implementation …….81
3.3. The Analysis of the Integral-Correlation Criterion and Method of its Minimization (the Integrated LSM) ……..85
3.3.1 The Method's Main Point……...85
3.3.2. The Analysis of the Method Components……88
3.3.3. The Definition of the Maximum Shift……..90
3.3.4. The Displacement Estimation of the Vector's $\hat{\beta}$ Estimate……93
3.3.5. The Variance Estimation of the Vector's $\hat{\beta}$ Estimate……...94
3.3.6. The Recursive Form of Calculations……97
3.4. The Method of the Auxiliary Variable (MAV)…… 100
3.4.1. The Method's Main Point ……..100

3.4.2. The Shift of the Estimates (3.72)……101
3.4.3. The Covariance of the Estimates (3.72) ……102
Chapter 4. The Increasing of Data Informativeness and, as a Consequence, of the Accuracy of the Estimates of Parameters of the Examined Objects……103
4.1. The Increasing of Informativeness of Data Samples in Terms of the Passive Experiment……103
4.2. The Improving of the Convergence……110
4.3. The Increasing of the Informativeness of Data Samples in Terms of the Active Experiment ……116
4.3.1. The Theoretical Argumentation of the Expediency of the Coordinatewise Quasi-diagonalization……...116
4.3.2. The Synthesis of the Testing Input Actions on the Examined Object……122
4.4. The Method of the Separate Estimation of Static Nonlinear and Dynamic Linear Components of the Hammerstein's Model ……..129
Chapter 5. Constructing of the Models of the Deterministic Processes, the Measurements of which are Perturbated by Random Interferences……..146
5.1. Modeling, Identification and Forecasting of Economy Performance by Methods of the Simplified Confluent Analysis……..146
5.2. The Numerical Modeling of the Task of Active Identification of the Parameters of Discrete Dynamical Systems……155
Chapter 6. The Unity and Difference of the Signal and Parametric Identification of Real Objects [26]……...165
6.1. The Statement of the Problem……166
6.2. The Strategy of the Signal Identification and the Invariance of the Adaptive Control (an Example of the First Approach)……. 168
6.3. The Strategy of Parametric Identification (an Example of the Second Approach the Evaluation of the Aerodynamic Coefficients $\hat{\beta}$ of the Aircraft)……...170
Chapter 7. The Integration of Methods and Models [29]……..178
7.1. The Formulation of the Problem……179
7.2. The Methodology of the Research ……180

Chapter 8. The Multilevel Systems with the Identificator of the Controlled Object……187
8.1. The System Approach ……187
8.2. The Reconciliation of the Quality Indicators……..190
8.3. The Recursive Bayesian Estimate of the Vector β……196
8.4. The Identification and Optimization of the Technological Process Parameters [30].200
8.4.1. The Mathematical Model of the System in the Mode "Working"…….. 201
8.4.2. The Mathematical Model of the System in the Mode "Training"……...204
8.4.3. The Working of the System in the Mode of Nonius Refinement of the Object's Model and Control Algorithm ……..206
8.5. The Identification and Optimization of the Process of Self-Study with the Electronic Simulator……208
8.5.1. The Introduction……..208
8.5.2. The Description of the Elements of the System of Automatic Control of Training ……...208
8.5.3. The Numerical Modeling of the Process…….213
References……..220