IFAC

INTERNATIONAL FEDERATION
OF AUTOMATIC CONTROL

WARSZAWA 1969

Translations of papers from Russian into English

Fourth Congress of the International
Federation of Automatic Control
Warszawa 16—21 June 1969

Vol. III

TECHNICAL
SESSIONS
48 — 70

NOT

Organized by
Naczelna Organizacja Techniczna w Polsce

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

Translations of papers from Russian into English

Vol. III

TECHNICAL SESSIONS 48 – 70

FOURTH CONGRESS OF THE INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL
WARSZAWA 16 – 21 JUNE 1969

Organized by
Naczelna Organizacja Techniczna w Polsce

Contents

48.1

SYSTEMS WITH VARIABLE STRUCTURE IN MULTIDIMENSIONAL PLANTS IDENTIFICATION AND CONTROL

S.V.Yemel'anov, N.Ye.Kostyleva, V.I.Utkin
Institute of Automation and Telemechanics
Moscow, USSR

In discussing the various problems of synthesis in a class of systems with variable structure it is generally assumed [1-6] that there is one quantity to be controlled and one control action in the system (or, in multivariable systems the number of controlled variables is equal to that of control actions) while the control action itself is formed of the error magnitude and its derivatives. In practice this approach may prove rather hard to apply due to the difficulties in obtaining the derivatives. At the same time in many cases we can measure both the magnitude of the error and certain coordinates that characterize the state of the system. In forming the control function it would be wise to use that possibility. Another feature, often inherent in automatic control systems, is that control actions can be applied to different points of the plant; in other words, control is a vectorial quantity.

In this connection it is interesting to discuss the possibilities of control in a class of systems with variable structure for a general case where the control function is formed by using information in the shape of the system coordinates which generally are not the system error or its derivatives, while the control function itself is a vectorial quantity. The motion of such a system without external disturbances is described by the differential equations

$$\frac{dx}{dt} = Ax + Bu, \tag{1}$$

$x=(x_1,...,x_n)$ is an n-dimensional vector which characterizes the state of the system,
$u=(u_1,...,u_m)$ is an m-dimensional control vector
A is an $n \cdot n$-dimensional matrix with constant elements a_{ij}
$(i,j=1,...,n)$, B is an $n \times m$-dimensional matrix with constant elements $b_{ij}(i=1,...,n;j=1,...,m)$. For control u it is assumed that its any component is a sum of actions by certain coordinates of the system; the coefficients of actions are piecewise-constant functions and they change with the state of the system. It is in this sense that the structure of the system is variable.

Assume that the system coordinates space is divided by a certain set of hyperplanes into regions; within each of these the system has a certain linear structure. The structure changes at the boundaries of the regions; as a result the control function may become discontinuous. If certain conditions are met the sliding mode appears in the system and the trajectory of the describing point belongs to the boundary of a discontinuity[7]. If the dynamic properties of sliding mode satisfy certain requirements to a control system, then it is practical to choose a control so as in any point of discontinuity boundaries the conditions for the existence of a sliding mode are met. After the describing point has hit the discontinuity boundary the motion in a sliding mode will appear and will not cease.[x] This synthesis technique was used in Refs[1-6] for the case where control is a scalar function of coordinates and its derivatives, while sliding mode appears in a certain hyperplane in the space of these coordinates.

We will apply this approach to systems with variable structure of a more general form of eq. (1). with the assumption that each component of the control vector is an action by the controlled variable (let this be the coordinate x_1) with a discontinuously changing coefficient of the action. We will show that for a structure change law to form in this case, data on the plant parameters is to be available. If they are not known in advance, then we have to identify these parameters; the ways to do this

[x] The describing point hitting the discontinuity boundary deserves special study and will not be discussed in this paper.

will be examined by the methods of the variable structure systems theory. In conclusion we will describe a procedure for identification of a second order dynamic system.

1. Selection of Control

Let us examine a case where control is a vectorial quantity and the motion of the system is described by eq.(1). Assume that a change in the structure of the system and the resulting discontinuous change in control u will occur if in the space $(x_1,\dots, x_n)$ the describing point will hit a certain hyperplane S given by

$$s = (c, x) = 0, \tag{1.1}$$

where $c = (c_1, \dots, c_n)$, $c_1, \dots, c_n$ - const, $c_n = 1$.

It has been already noted that in such a dynamic system a sliding mode is possible that would require the following inequality[2] in the neighbourhood of the hyperplane S

$$s\dot{s} < 0 \qquad \left(\dot{s} = \frac{ds}{dt}\right). \tag{1.2}$$

Geometrically this equation means that in the neighbourhood of a discontinuity of S the phase trajectories of both structures are directed towards each other and the describing point after having hit S continues its motion in the sliding mode along the trajectories that belong to that hyperplane. By Ref.[7] during that kind of motion the vector of phase velocity is directed along the hyperplane S and therefore in sliding mode the following relations are true

$$s = 0, \qquad \dot{s} = 0. \tag{1.3}$$

Let us find what differential equations describe the various kinds of sliding modes that can occur in the class of systems under consideration if each component $u_1, \dots, u_m$ of the vector u has discontinuities when the describing point hits the appropriate hyperplane in the space $(x_1, \dots, x_n)$.

Let the scalar function u_m have discontinuities on a certain hyperplane S^1 described in the space $(x_1, \dots, x_n)$ by

$$s^1 = (c^1, x) = 0 \quad , \quad \text{where the vector } c^1 = (c_1^1, \dots, c_n^1), \quad c_1^1, \dots, c_n^1 = \text{const}, \quad c_n^1 = 1. \tag{1.4}$$

When eq. (1.2) is true, there is a sliding mode on the hyperplane S^1. Let us obtain equations for the motion of the describing point along the trajectories that belong to S^1.

It has been already noted that in sliding mode eqs (1.3) are true. From the conditions $s^1 = 0$, $\dot{s}^1 = 0$ and eq.(1) for motion we find

$$x_n = -\sum_{i=1}^{n-1} c_i^1 x_i ,$$
$$u_m = -\frac{1}{(b_m, c^1)} \sum_{j=1}^{m-1} (b_j, c^1) u_j - \frac{1}{(b_m, c^1)} \sum_{j=1}^{n-1} \left[(a_j, c^1) - c_j^1 (a_n, c^1) \right] x_j . \tag{1.5}$$

where $a_j \, (j = 1, \dots, n)$, $b_j \, (j = 1, \dots, m)$ are the columns of the matrices A and B, $(b_m, c^1) \neq 0$. By substituting the values for x_n and u_m obtained above into (1) we will determine the differential equation of the $n-1$-th order for $x_1, \dots, x_{n-1}$ with $m-1$-dimensional control that would describe the motion of the system in sliding mode

$$\frac{dx^1}{dt} = A^1 x^1 + B^1 u^1, \tag{1.6}$$

where the vector $x^1 = (x_1, \dots, x_{n-1})$, the vector $u^1 = (u_1, \dots, u_{m-1})$, A^1 and B are $(n-1) \times (n-1)$ and $(n-1) \times (m-1)$ - dimensional matrices respectively with the elements $a_{ij}^1 \, (i, j = 1, \dots, n-1)$ and $b_{ij}^1 \, (i = 1, \dots, n-1; j = 1, \dots, m-1$

$$a_{ij}^1 = a_{ij} - a_{in} c_j^1 - \frac{b_{im}}{(b_m, c^1)} \left[(a_j, c^1) - c_j^1 (a_n, c^1) \right],$$
$$b_{ij}^1 = b_{ij} - \frac{b_{im}}{(b_m, c^1)} (b_j, c^1). \tag{1.7}$$

A solution to eq. (1.6) is a mapping of the solution to the initial system in sliding mode on the subspace ($x_1, \dots, x_{n-1}$). In that subspace we can also select a hyperplane S^2 given by the equation $s^2 = (c^2, x^1) = 0$, $c^2 = (c_1^2, \dots, c_{n-1}^2)$, $c_{n-1}^2 = 1$ and with the aid of the discontinuous control u_{m-1} ensure the appearance of a sliding mode S^2. Evidently the motion in the sliding mode lies in the linear subspace of the $n-2$-th order which is determined in the space ($x_1, \ldots, x_n$) by the intersection of S^1 and S^2. That motion is described by differential equations of the $(n-2)$-th order for $x^2 = (x_1, \ldots, x_{n-2})$ with an $(m-2)$-th control $u^2 = (u_1, \ldots, u_{m-2}$

By iterating these operations we will obtain at the k-th step a motion in sliding condition of the n-k-th order over the intersection of hyperplanes $S^1, \ldots, S^k$ given by

$$s^i = (c^i, x^{i-1}) = 0, \tag{1.8}$$

where the vector $c^i = (c_1^i, \ldots, c_{n-i+1}^i)$, $c_{n-i+1}^i = 1$, the vector $x^i = (x_1, \ldots, x_{n-}$
$i = 1, \ldots, k, \quad 1 \leq k \leq m, \; x^0 = x.$

The equation of this motion for coordinates $x_1, \ldots, x_{n-k}$ is given by

$$\frac{dx^k}{dt} = A^k x^k + B^k u^k, \tag{1.9}$$

where A^k and B^k are matrices with dimensionalities $(n-k) \times (n-k)$ and $(n-k) \times (m-k)$ respectively, with elements a_{ij}^k and b_{ij}^k, the vector $u^k = (u_1, \ldots, u_{m-k})$.

The elements a_{ij}^k and b_{ij}^k of the matrices A^k and B^k are described by the recurrent relations

$$a_{ij}^k = a_{ij}^{k-1} - a_{i\,n-k+1}^{k-1} c_j^k - \frac{b_{i\,m-k+1}^{k-1}}{(b_{m-k+1}^{k-1}, c^k)} \left[(a_j^{k-1}, c^k) - c_j^k (a_{n-k+1}^{k-1}, c^k) \right],$$

$$b_{ij}^k = b_{ij}^{k-1} - \frac{b_{i\,m-k+1}^{k-1}}{(b_{m-k+1}^{k-1}, c^k)} (b_j^{k-1}, c^k), \tag{1.10}$$

$$k = 1, \ldots, m, \qquad a_{ij}^0 = a_{ij}, \; b_{ij}^0 = b_{ij}.$$

The upper index for a and b shows what matrix this quantity belongs to as an element, the double lower index signifies the numbers of the row and the column, the single lower index shows that this element is a column-vector and denotes its number. It is assumed that in (1.10) $(b^{k-1}_{m-k+1}, c^k) \neq 0$ for any k.

For the case where $k=m$ we obtain a linear homogeneous differential equation of the $(n-m)$-th order

$$\frac{dx^m}{dt} = A^m x^m \tag{1.11}$$

As follows from (1.10), (1.11), when sliding modes appear at all hyperplanes $S^1, \dots, S^m$ the motion in the system is described by an equation with dimensionality lower than in the initial equation and depends on all the coefficients c^k_τ $(k=1,\dots,m;\ \tau=1,\dots,n-k+1)$.

If by the appropriate choice of coefficients c^k_τ the motion has the desired dynamic properties, then it would be wise to make each hyperplane S^k in the space $(x_1,\dots,x_{n-k+1})$ a hyperplane of sliding i.e. conditions (1.2) are true for any point of the hyperplane. Then as a result of successive decreasing the order of the motion differential equation in sliding mode, the control process will be described by equation (1.11). Let us try to solve the problem stated in the class of systems with variable structure.

Choose the scalar function u_{m-k+1} such that for a system described by the equation

$$\frac{dx^{k-1}}{dt} = A^{k-1} x^{k-1} + B^{k-1} u^{k-1}, \tag{1.12}$$

the conditions for existence of a sliding mode in eq. (1.2) are met in any point of the hyperplane S^k given in the space $(x_1,\dots,x_{n-k+1})$ by the equation

$$s^k = (c^k, x^{k-1}) = 0. \tag{1.13}$$

We will assume that each of the components of the control vector is the action by a certain coordinate, e.g. x_1, with the discontinuous coefficient

$$u_i = -\psi^i x_1, \qquad \psi^i = \begin{cases} \alpha^i & \text{at } s^{m-i+1} x_1 > 0 \\ \beta^i & \text{at } s^{m-i+1} x_1 < 0 \end{cases} \quad i=1,\dots,m, \tag{1.14}$$

where α^i, β^i - const

(note that a similar procedure was discussed in Ref.[2]).

Let us find the magnitude of $\dot{s}^k$ on the hyperplane S^k from (1.12-1.14)

$$\dot{s}^k = \sum_{j=2}^{n-k}\left[(a_j^{k-1}, c^k) - c_j^k(a_{n-k+1}^{k-1}, c^k)\right]x_j - \left[(a_1^{k-1}, c^k) - \right. \tag{1.15}$$

$$\left. - c_1^k(a_{n-k+1}^{k-1}, c^k) - \sum_{j=1}^{m-k}(b_j^{k-1}, c^k)\psi^j - (b_{m-k+1}^{k-1}, c^k)\psi^{m-k+1}\right]x_1.$$

Note that by eq.(1.14) the quantity ψ^{m-k+1} changes discontinuously on the hyperplane S^k. From (1.12) and (1.15) we obtain the necessary and sufficient conditions for the existence of sliding mode in any point of the hyperplane

$$(b_{m-k+1}^{k-1}, c^k)\alpha^{m-k+1} > (a_1^{k-1}, c^k) - c_1^k(a_{n-k+1}^{k-1}, c^k) - \min_{\psi^1,\dots,\psi^{m-k}} \sum_{j=1}^{m-k}(b_j^{k-1}, c^k)\psi^j,$$

$$(1.16)$$

$$(b_{m-k+1}^{k-1}, c^k)\beta^{m-k+1} < (a_1^{k-1}, c^k) - c_1^k(a_{n-k+1}^{k-1}, c^k) - \max_{\psi^1,\dots,\psi^{m-k}} \sum_{j=1}^{m-k}(b_j^{k-1}, c^k)\psi^j,$$

$$\frac{(a_j^{k-1}, c^k)}{c_j^k} = (a_{n-k+1}^{k-1}, c^k), \quad j = 2,\dots,n-k.$$

If these conditions are met for $k = 1,\dots,m$, then each of the hyperplanes $S_*^1,\dots,S^m$ will be a hyperplane of sliding.

Remark. In synthesis of control the choice of the coefficients α^k and β^k should start with α^1 and β^1 because by (1.16) all values of these coefficients that follow depend on the preceding ones.

For the control law of eq. (1.14) all C_r^k should satisfy the second group of conditions in (1.16). The algorithm described is suitable if with these constraints the motinn of the system in sliding mode described by eq.(1.11) can have the required dynamic properties.

2. Techniques of Systems with Variable Structures in the Problem of Identifying a Linear Plant

For the above control algorithm to be implemented data on the parameters a_{ij} of the controlled plant is to be available. In a number of cases, however, the exact values of these constant parameters may be unknown in advance; e.g. they can vary from a plant to a plant or the case may be that the controller is intended for different types of plants. Finally, the parameters the plant may be found to change but so slowly that within one process this change can be neglected. In all these cases we have the problem of identification which reduces to distribution of parameters in a linear plant of a known structure. We will discuss possible solutions to this problem by methods of systems with variable structure.

Let us take first this auxiliary problem. Let a certain dynamic first-order element be described by the equation

$$\frac{dx}{dt} = (a, f(t)), \tag{2.1}$$

where the vectors $a = (a_1, \dots, a_n)$, $f(t) = (f_1(t), \dots, f_n(t))$, a_i - $const$, $f_i(t)$ - are arbitrary linearly independent time functions, $(a, f(t))$ is a scalar product. The output quantity x and the functions $f_i(t)$ are assumed known. We have to find the unknown coefficients a_i .

To solve this problem let us construct a model with variable structure described by the equation

$$\frac{dy}{dt} = u, \qquad \text{control } u = (\psi, f), \tag{2.2}$$

where the vector $\psi = (\psi_1, \dots, \psi_n)$,

$$\psi_i = \begin{cases} \alpha_i & \text{at } f_i s > 0 \\ \beta_i & \text{at } f_i s < 0, \end{cases} \tag{2.3}$$

$$\alpha_i,\ \beta_i - const, \qquad \beta_i \leq a_i \leq \alpha_i, \tag{2.4}$$

(the feasible range of values of coefficients a_i will be assumed known).

$$s = x - y. \tag{2.5}$$

From eqs.(2.2)-(2.5) follows that the structure of the model varies with the error between the output values of the element to be identified x and of the model y .

To study the possible kinds of motion in this system of (2.1), (2.2) and (2.5) let us find the quantity $\dot{s}$

$$\dot{s} = ((a-\psi), f). \tag{2.6}$$

By eqs. (2.3) - (2.6) the functions s and $\dot{s}$ differ in signs. Therefore the describing point on the plain (x, y) always hits the line S $(s=0)$ and then moves along S with zero error in sliding mode.

When a system operates in sliding mode the control changes at an infinitely high frequency. By Filippovs'[7] difinition $\dot{s}=0$in a sliding mode and

$$u = (a, f). \tag{2.7}$$

The control u "makes" the describing point move along the line S ; by (2.7) depends on the parameters a_i . However, the control action is implemented without changing these parameters by the sliding condition, and by (2.7) is a continuous function. This happens because the Fillipov's definition averages at every instant of time the control action that varies at an infinitely high frequency. Let us further define the average value of each ψ_i . To do this the law of eq.(2.3) for variation of ψ_i is more conveniently represented in the form

$$\psi_i = \frac{\alpha_i + \beta_i}{2} + \frac{\alpha_i - \beta_i}{2} \operatorname{sign} f_i s. \tag{2.8}$$

Let, in a small interval of time Δt the motion take place in the following way: the interval Δt_1corresponds to the structure $s>0$, the interval Δt_2 to the structures $s<0$ while $\Delta t_1 + \Delta t_2 = \Delta t$.

Then

$$\psi_{icp} = \gamma\left(\frac{\alpha_i+\beta_i}{2} + \frac{\alpha_i-\beta_i}{2}\,\mathrm{sign}\, f_i\right) + \qquad (2.9)$$

$$+ (1-\gamma)\left(\frac{\alpha_i+\beta_i}{2} - \frac{\alpha_i-\beta_i}{2}\,\mathrm{sign}\, f_i\right),$$

where $\gamma = \frac{\Delta t_1}{\Delta t}$.

From the condition $\dot{s}=0$ we will find the value of γ .

$$\gamma = \frac{\sum_{i=1}^{n}\left(-a_i - \frac{\alpha_i+\beta_i}{2} + \frac{\alpha_i-\beta_i}{2}\,\mathrm{sign}\, f_i\right) f_i}{\sum_{i=1}^{n}\left[\left(\alpha_i-\beta_i\right)\mathrm{sign}\, f_i\right] f_i} . \qquad (2.10)$$

Evidently γ =const if the ratio of linear forms coefficients is constant, i.e.

$$\frac{-a_i - \frac{\alpha_i+\beta_i}{2} + \frac{\alpha_i-\beta_i}{2}\,\mathrm{sign}\, f_i}{(\alpha_i-\beta_i)\,\mathrm{sign}\, f_i} = \frac{-a_n - \frac{\alpha_n+\beta_n}{2} + \frac{\alpha_n-\beta_n}{2}\,\mathrm{sign}\, f_n}{(\alpha_n-\beta_n)\,\mathrm{sign}\, f_n} . \qquad (2.11)$$

From (2.11) we obtain the value of β_i at which the magnitude of γ does not change in time

$$\beta_i = \beta_{i0} = \frac{\left(-a_i - \frac{\alpha_i}{2}\right)(\alpha_n-\beta_n)\,\mathrm{sign}\, f_n - \alpha_i\left(-a_n - \frac{\alpha_n+\beta_n}{2}\right)\mathrm{sign}\, f_i}{\frac{\alpha_n-\beta_n}{2}\,\mathrm{sign}\, f_n - \left(-a_n - \frac{\alpha_n+\beta_n}{2}\right)\mathrm{sign}\, f_i} . \qquad (2.12)$$

If condition (2.12) is valid for all β_i , then γ =const and by (2.9) ψ_{icp} =const. By the definition of sliding mode $\dot{s}$ =0, therefore due to linear independence of the functions f_i from (2.6) follows

$$\psi_{cp} = a \qquad \text{or} \qquad \psi_{icp} = a_i . \qquad (2.13)$$

Thus if γ =const, then by measuring the average value of ψ_{icp}

we will obtain the value of α_i[x]. Hence the idea: it after the quantity s has become zero with the coefficient β_n assumed constant we can change each of the fixed linear structures of the model by changing $\beta_1, \ldots, \beta_{n-1}$ so as to maintain the constant value of γ. A possible approach to this problem can be as follows. Since in the desired point $\gamma^{(i)} = 0,\ i = 1, \ldots, n-1$ we will obtain a set of equations with n-1 unknown $\beta_1, \ldots, \beta_{n-1}$. This problem can be solved e.g. by minimizing the function $\sum_{i=1}^{n-1} |\gamma^{(i)}|$ by one of the known hill-climbing techniques. If for the desired values of β_i there is just one extremum, the parameters α_i will be found at any initial errors between the values α_i and $\psi_{i\,ср}$; if there are several extrema, the feasible initial errors have constraints. The search for the required values of coefficients β_i will be discussed in more detail for a system with two unknown parameters.

A technique of finding the parameters of a controlled plant follows from the above. Indeed, since the plant to be identified is described by the equation

$$\frac{dx}{dt} = A x, \tag{2.14}$$

where the n-dimensional vector x and the $n \times n$ - dimensional matrix A are given by eq. (1), let us construct a model with a variable structure whose motion is described by the equation

[x] To measure γ and $\psi_{i\,ср}$ one can use a relay element whose input and output are related as

$$u_{вых} = \begin{cases} A_1 \text{ at } & u_{вх} > 0 \\ A_2 \text{ at } & u_{вх} < 0, \quad A_1, A_2 - const. \end{cases}$$

The average valueof the output quantity in sliding mode equals

$$u_{вых\,ср} = A_1 \gamma + A_2 (1 - \gamma).$$

Hence to find γ we have to assume $u_{вх}$ =sign s, $A_1 = 1, A_2 = 0$ and for $\psi_{i\,ср}$ $u_{вх} = \text{sign}\, f_i s, A_1 = \alpha_i,\ A_2 = \beta_i$! The average value can be obtained if the output quantity of a relay element is filtered through an inertial unit with a small time constant but sufficient to filter off the high frequency constituent.

$$\frac{dy}{dt} = u, \qquad u = \Psi x, \tag{2.15}$$

where Ψ is an $n\times n$ dimensional matrix with the elements Ψ_{ij} $(i,j=1,\dots,$

$$\Psi_{ij} = \begin{cases} \alpha_{ij} & \text{at } s_i x_j > 0 \\ \beta_{ij} & \text{at } s_i x_j < 0, \end{cases} \quad i,j=1,\dots,n, \tag{1.16}$$

$$\alpha_{ij}, \beta_{ij} - const, \quad \beta_{ij} \le a_{ij} \le \alpha_{ij}, \tag{2.17}$$

$$s_i = x_i - y_i. \tag{2.18}$$

In further discussion the components $x_1,\dots,x_n$ of the vector x which characterizes the state of the system are assumed to be linearly independent time functions. (If for instance, the initial conditions are such that these quantities are linearly dependendent time functions, the plant parameters cannot be found unambignously by the coordinates $x_1,\dots,x_n$).

Then as follows from the above auxiliary problem, if fixed values of β_{ij} can be found such that the quantities γ_i which characterize the motion in sliding mode are constant, then

$$\Psi_{cp} = A \qquad \text{or} \qquad \Psi_{ij\,cp} = a_{ij}. \tag{2.19}$$

Thus, the parameters of the plant to be identified can be found unambignously by the characteristics of sliding mode and the model when the model parameters are appropriately adjusted. With information on the coefficients a_{ij} on hand we can implement control by the algorithms of (1.14), (1.16) obtained above.

Let us illustrate this method of parameters identification using as an example a second-order system described by the equations

$$\left.\begin{aligned}\frac{dx_1}{dt} &= x_2\\ \frac{dx_2}{dt} &= a_2 x_2 + a_1 x_1\end{aligned}\right\} \qquad (2.20)$$

where a_1, a_2 are constant parameters to be found.

Generally, in a model with variable structure four coefficients have to be changed discontinuously. But since one of the equations which describe the plant to be identified is known, the model is chosen as a first-order element

$$\frac{dy}{dt} = \psi_1 x_1 + \psi_2 x_2, \qquad (2.21)$$

where ψ_1 and ψ_2 are found by (2.16).

With the generality of the presentation preserved, we can discuss only the process of identification at $x_1 > 0$ and $x_2 > 0$. Then by (2.10) for $n=2$ we have

$$\gamma = \frac{(a_1-\beta_1)x_1 + (a_2-\beta_2)x_2}{(\alpha_1-\beta_1)x_1 + (\alpha_2-\beta_2)x_2}. \qquad (2.22)$$

The required value of $\mu_1 = \beta_{10}$ at which γ is a constant quantity and $\psi_{1cp} = a_1$, $\psi_{2cp} = a_2$ is found from (2.12)

$$\beta_{10} = \frac{a_1\alpha_2 - a_1\beta_2 - \alpha_1 a_2 + \alpha_1\beta_2}{\alpha_2 - a_2}. \qquad (2.23)$$

Assuming β_1 =const and β_2=const, from (2.22) we obtain the magnitude of $\dot{\gamma}$

$$\dot{\gamma} = \frac{[-(\alpha_1-\beta_1)(a_2-\beta_2) + (\alpha_2-\beta_2)(a_1-\beta_1)][x_2^2 + a_1 x_1^2 + a_2 x_1 x_2]}{[(\alpha_1-\beta_1)x_1 + (\alpha_2-\beta_2)x_2]^2} \qquad (2.24)$$

A few words have to be said on this relation. If the quadratic form $P(x) = x_2^2 + a_1 x_1^2 + a_2 x_1 x_2 = 0$, then the identification process is not feasible. Let us explain this phenomenon for the case of constant a_1 and a_2. In case of real roots of the characteristic equation for the plant to be identified its output quantity is a sum of two exponents; if one of the addends is not present, $P(x) \equiv 0$. In this case no information can be obtained on the plant parameters by the quantity x_1 since there is an infinite number of second-order linear systems with this particular solution. For any other initial conditions as well as for complex roots of the characteristic equation the quantity $P(x)=0$ only at $x_1 = x_2 = 0$.

The sign of the quadratic form will be further assumed definite. The dependence $\dot{\gamma}(\beta_1)$ is represented in Fig. 1.

We will describe a possible procedure of a search for the quantity β_{10} if after the appearance of sliding mode the quantity β_1 satisfies the inequality

$$\beta_{1э} < \beta_1 < \alpha_1, \tag{2.25}$$

where $\beta_{1э}$ is the extremum point of the function $\dot{\gamma}(\beta_1)$.

The function $\dot{\gamma}(\beta_1)$ on the interval ($\beta_{1э}$, α_1) can be represented as

$$\dot{\gamma}(\beta_1) = f(\beta_1)(\beta_1 - \beta_{10}), \tag{2.26}$$

where the function $f(\beta_1)$ meets the relations

$$0 < |f(\beta_1)| \leq A, \tag{2.27}$$

A - is a constant quantity independent from x_1, x_2,

$$sign\, f(\beta_1) = sign \frac{\partial \dot{\gamma}}{\partial \beta_1}. \tag{2.28}$$

Let us change β_1 discretely by the procedure

$$\beta_1(t+1) = \beta_1(t) - K \frac{\Delta \dot{\gamma}}{\Delta \beta_1}(t)\, \dot{\gamma}(t). \tag{2.29}$$

From (2.26) and (2.29) we have

$$\beta_1(t+1) = \cdot\beta_1(t) - K \frac{\Delta \dot{\gamma}}{\Delta \beta_1}(t)\, f[\beta_1(t)]\, [\beta_1(t) - \beta_{10}]. \tag{2.30}$$

By (2.27), (2.28) there is always .: such K that $\lim_{t \to \infty} \beta \cdot)=$ $= \beta_{10}$, which solves the problem.

The above procedure was obtained under the assumption that the function $Y(\beta_1)$ does not change in time.The curve of Fig.2 will actually be shifted due to changes in x_1 and x_2 .Therefore the step in (2.30) should be chosen such that during the search for the required value of β_1 this change can be neglected.

REFERENCES

I. Емельянов С.В. - О высококачественном управлении некоторыми нелинейными объектами с переменными параметрами. Изв. АН СССР, Энергетика и Автоматика, № 4, I962.

2. Емельянов С.В.,Таран В.А. - Об одном классе систем автоматического регулирования с переменной структурой. Изв. АН СССР, Энергетика и Автоматика, № 3, I962.

3. Емельянов С.В., Уткин В.И. - Применение систем автоматического регулирования с переменной структурой для управления объектами, параметры которых изменяются в широких пределах. ДАН СССР, т.I52, I963.

4. Уткин В.И. - Квазиинвариантное управление вынужденным движением линейных объектов с переменными параметрами. Изв. АН СССР, Техническая кибернетика, № 5, I966.

5. Петров Б.Н.,Емельянов С.В., Костылева Н.Е. - Об управлении линейными объектами с переменными параметрами, ДАН СССР, т. I55, № I, I964.

6. Петров Б.Н., Емельянов С.В., Гриценко М.Б. - Автономность в многосвязанных системах автоматического управления с переменной структурой. ДАН СССР, т. I69, № I, I966.

7. Филиппов А.Ф. - Дифференциальные уравнения с разрывной правой частью. Математический сборник, т. 5I, № I, I960.

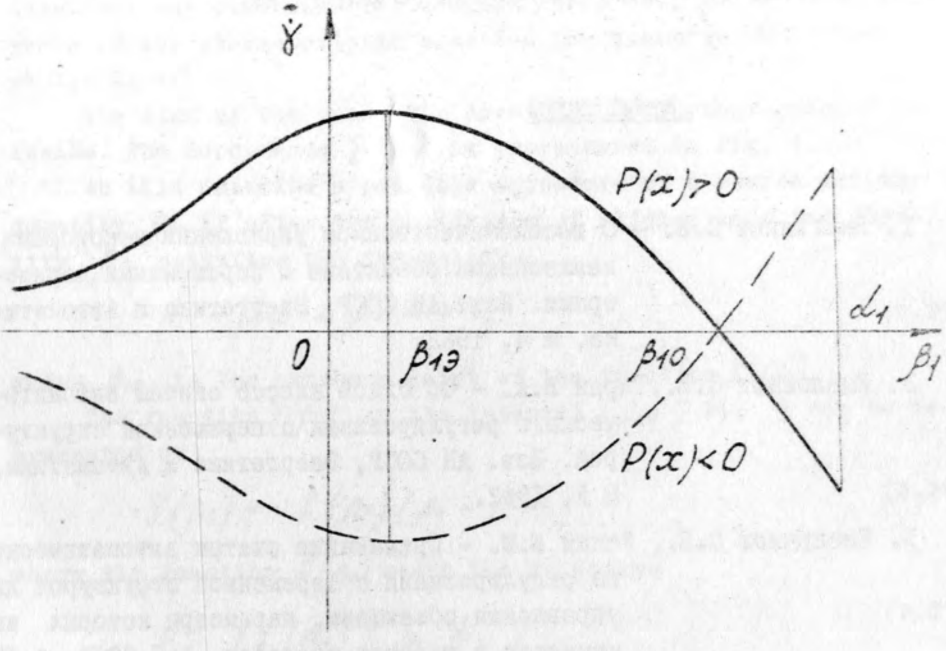


Fig 1

48.4

ON PROBLEM OF INVARIANT SYSTEMS SYNTHESIS

V.V. Velichenko

Moscow Phisico-Technical Institute
Moscow
USSR

The problem of invariant systems synthesis consist in development of the methods of constructing such systems whose definite control criterions don't depend on external disturbances.

This problem is very important and attracts attention of many investigators. The invariance theory of linear stationary systems and realization methods for non-disturbable technical devices was developed by G.V. Scipanov, N.N. Luzin, V.S. Kulebakin, B.N. Petrov and other authors. Survey of the problem and results achieved in this region are cited in[1] and [2]. Some results, tied with investigations of invariance conditions in linear non-stationary and nonlinear systems are stated in papers[2-6].

Control criterion whose independence on disturbances are ensured when we design invariant systems, are the functionals depending on the behavior of these systems. The said abobe determines variational character of the invariace problem[7] and openes the possibilities to investigate the problem by mathematical tools of modern theory of optimal processes[8,9]. On the grounds of variational approach to the invariance problem L.I. Rozonoer[7] gave the complete solution of the problem of necessary and sufficient conditions of the invariance in linear stationary systems and of perfect invariance in nonlinear systems.

Development of the variational approach in direction of investigations of large variations of the functional gives the possibility to achieve some new results in the problem of invariance in nonlinear systems. In the present paper we formulated the necessary and at the same time sufficient conditions of the invariance with respect to disturbances and with respect to initial conditions for systems of sufficiently general nature. On this ground we solved a problem of the synthesis of correcting circuit which ensures the invariance of the given nonlinear object.

1. Formulation of the problem

Synthesis of invariant system is realized on the grounds of invariance conditions which we formulate for the following problem.

Consider a system which is discribed by an n-dimensional system of equations

$$\dot{x}(t) = f(x,u,t). \tag{1}$$

Here x denotes a vector of phase coordinates and u is a vector of external disturbances. We shall study the functional

$$\mathcal{J}(x,u) = \Phi[x(T),T] \tag{2}$$

as the control criterion, which we define as the function $\varphi(x,t)$ depending on the coordinates of the system (1) and the time in prescribed instant T . Instant T may be fixed or be defined by means of any conditions on the variables of the system (1). In the general case it may be threated as being defined by the condition that the trajectory $x(t), t$ of the system (1) reaches the hypersurface M defined by the equality

$$M(x,t) = 0. \tag{3}$$

In the general case nonlinearity of the problem does not permit us to threat disturbances $u(t)$ as quite arbitrary. In any case, we must restrict the class of disturbances by thouse vector functions $u(t)$ which don't let the point $\{x,u,t\}$ out of the region G in which right-hand parts of the system (1) are defined. All the following construction we shall perform for preescribed region A of $(n+1)$-dimensional space $X \times t$ and regard such piecewise-continuous vector functions $u(t)$ to be admissible disturbances, that the trajectories $x(t), t$, corresponding to them and initiating in A , are all in A and .every point $\{x(t), u(t), t\}$ belongs to G . The functions $f(x,u,t)$, $\varphi(x,t)$

and $M(x,t)$ are assumed to be continuous together with their first and second-order partial derivatives with respect to all of theirs arguments.

The system (1) is said to be Φ -invariant on M with respect to u , if on its trajectories, initiating in points $\{x,\tau\}\in A$, the value of the functional (2) does not depend on disturbance $u(t)$. Let us formulate the problem on conditions, which ensure invariance of the system (1) in the above sense. In such a form, which is a generalization of the weak invariance problem[7], a wide class of problems encountered in ingineering can be formulated. In this form the problems in which the control criterion is determined in another, for example in integral, form can be formulated as well.

2. Necessary and sufficient condition of invariance

Support field. Let us consider a vector function $u(t)$ to be the control, the value of which at any instant we may choose at will. Let us choose some support control $\tilde{u}(t)$ and consider all the trajectories $\tilde{x}(t), t$, initiating in points $\{x,\tau\}\in A$ and corresponding to the control $u(t)=\tilde{u}(t)$. The trajectory $\tilde{x}(t), t$ reaches M at the instant $t=\tilde{T}$. Along this trajectory we define a vector function $\tilde{p}(t)$ by the equation

$$\dot{\tilde{p}} = -grad_x H(x,\tilde{p},\tilde{u},t); \quad x(t)\equiv\tilde{x}(t), \; H\equiv(p, f(x,u,t)) \tag{4}$$

with the boundary condition

$$\tilde{p}(\tilde{T}) = \left[-grad_x \Phi(x,t) + \left(\frac{d\Phi(x,t)}{dt}\Big/\frac{dM(x,t)}{dt}\right) grad_x M(x,t)\right]_{t=\tilde{T},\; x=\tilde{x}(\tilde{T})} .$$

Equation (4) puts the vector $\tilde{p}=\tilde{p}(x,\tau)$ in correspondence to every point $\{x,\tau\}\in A$. The vector field $\tilde{p}(x,\tau)$ is said to be the support field.

Support function. The trajectory $\tilde{x}(t), t$ passing the point $\{x,\tau\}$ puts the value of the functional (2) in unique correspondence to this point, which makes a value of the functional (2) a function of the point $\{x,\tau\}$. Call thus defined function

$$\tilde{V}(x,\tau) \equiv \Phi[\tilde{x}(\tilde{T}),\tilde{T}] \tag{5}$$

a support function.

Explicit formula for change of functional. With the help of the support function $\tilde{V}(x,\tau)$ a difference between the values of the functional (2) for trajectories $\hat{x}(t),t$ and $\tilde{x}(t),t$, passing from one and the same point $\{x^\circ,t^\circ\}\in A$ and corresponding to the controls $\hat{u}(t)$ and $\tilde{u}(t)$ may be written as

$$\Delta J = \Phi[\hat{x}(\hat{T}),\hat{T}] - \Phi[\tilde{x}(\tilde{T}),\tilde{T}] = \tilde{V}[\hat{x}(\hat{T}),\hat{T}] - \tilde{V}(x^\circ,t^\circ) = \int_{t^\circ}^{\hat{T}} \frac{d\tilde{V}[\hat{x}(t),t]}{dt}\,dt\,.$$

Now applying an approximate formula for small change of the functional (2)[10] to calculation of the derivative of $\tilde{V}(x,\tau)$ with respect to time along the trajectory $\hat{x}(t),t$ we obtain

$$\Delta J = -\int_{t^\circ}^{\hat{T}} \Big[H[\hat{x},\tilde{p}(\hat{x},t),\hat{u},t] - H[\hat{x},\tilde{p}(\hat{x},t),\tilde{u},t]\Big]\,dt \tag{6}$$

Utilization of the explicit formula (6) makes it possible to formulate the necessary and at the same time sufficient condition of invariance of the system (1).

Theorem 1.(Independence principle). A necessary and sufficient condition for the system (1) to be Φ -invariant on M with respect to u is the independence of the function

$$H[x,\tilde{p}(x,\tau),u,\tau] \equiv (\tilde{p}(x,\tau), f(x,u,\tau)),$$

corresponding to any support control $\tilde{u}(t)$, from u .

From the formulation of the theorem follows that if its condition is valid at least for one support control, it is valid also for any other. This follows from the property of the support field vectors and Hamiltonian $H[x,\tilde{p}(x,\tau),\tilde{u},\tau] = \tilde{H}(x,\tau)$ of the invariant system to be invariant with respect to $\tilde{u}(t)$. This means that the values of $\tilde{p}(x,\tau)$ and $\tilde{H}(x,\tau)$ are determined only by the point $\{x,\tau\}$ and don't depend on the control function $\tilde{u}(t)$, which we used in their calculations. Namely owing to this remarkable property the synthesis problem, very difficult for the general case of optimal systems construction,becomes very simple for case of invariant systems construction. This very property permits us as well to utilize as a support function the function, depending not only on time, but also on coordinates of the system. Such support functions are handly in practice.

In the terms of the support function, which proves to be connected with the support field and Hamiltonian by means the relations

$$\tilde{p}(x,\tau) = -grad_x \tilde{V}(x,\tau), \quad \tilde{H}(x,\tau) = \frac{\partial \tilde{V}(x,\tau)}{\partial \tau},$$

the necessary and sufficient condition of invariance may be formulated the following way.

Theorem 2. In order that the system (1) be Φ -invariant on M with respect to u, it is necessary and sufficient that the support function, corresponding to any support control $\tilde{u}(t)$, satisfy the partial differential equation

$$\frac{\partial \tilde{V}}{\partial \tau} = -\left(grad_x \tilde{V}, f(x,u,t)\right)$$

and right-hand part of this equation do not depend on u.

Note that according to definition (5) the function $\tilde{V}$ sutisfies the boundary condition $\tilde{V}(x,\tau) = \Phi(x,\tau)$ on M.

Conditions of the theorems 1 and 2 are close to wordings of necessary[8,9] and sufficient[11,12] conditions of optimality in the Pontrjagin maximum principle form. This analogy turns out to be very deep. Actually, the invariant functional for any control, and for the support control $\tilde{u}(t)$ too (the trajectories $\tilde{x}(t), t$ corresponding to $\tilde{u}(t)$ satisfy all conditions of regular synthesis[11,12]) achieves its maximal and at the same time its minimal values. On this reason for wide class of problems the conditions of invariance must differ from corresponding optimality conditions only by such property, that Hamiltonian, achieving on control $\tilde{u}(t)$ its minimal and at the same time its maximal values, does not depend on u .

3. Structure of region of invariance and invariance conditions with respect to initial conditions

The support function defines the structure of the phase space of the invariant system. The region of invariance A exfoliates into invariant smooth manifolds of dimensionality ℓ , with $\ell \leq n$, in the $(n+1)$ -dimensional space $X \times t$ .These manifolds are level surfaces of the function $\tilde{V}(x,\tau)$. For any disturbances $u(t)$ the trajectories of the invariant system, initiating at same point $\{x,\tau\} \in A$, belong to one and the same invariant manifold. Thus as in case of linear systems with constant coefficients, nonlinear invariant system proves to be not completely controllable: disturbance does not let the phase point out of the invariant manifold and can only change the location of the trajectory on this manifold.

Invariance with respect to initial conditions. If a set $L \subset A$ is given, we say the system (1) to be Φ -invariant with respect to initial conditions on L , if the values of the functional (2) don't depend on disturbance $u(t)$ and for all trajectories, initiating on L , are equal. The structure of the region of invariance shows, that for the system (1) to be invariant with respect to initial conditions on L it is necessary and sufficient that set L should belong to a level surface of the function $\tilde{V}(x,\tau)$. The $(n+1)$ -dimensional vector $\tilde{P}(x,\tau) = \{-\tilde{p}(x,\tau), \tilde{H}(x,\tau)\}$ is normal to this surface. Thus, if L is a smooth manifold and $\ell(x,\tau)$ is any tangent vector to L in the point $\{x,\tau\}$, the following assertions is true.

Theorem 3. In order that the system (1) be Φ-invariant with respect to initial conditions on L , it is necessary and sufficient that the condition of theorem 1 be satisfied and in any poir $\{x,\tau\} \in L$

$$(\tilde{P}(x,\tau), \ell(x,\tau)) = 0.$$

4. Invariant systems synthesis

The problem of the invariant systems synthesis we shall consider as a problem of finding a corrective circuit which ensures invariance of the system with a given invariable part. A correct formulation of the problem is such:

Let the control system is described by a system of equation

$$\dot{x}(t) = f(x, u, v, t). \tag{7}$$

As before u is a disturbance and a scalar control parameter v characterizes a changeable part of the system. It is necessary to choose such a corrective function

$$v = v(x, u, t), \tag{8}$$

that the system

$$\dot{x}(t) = f[x, u, v(x,u,t), t] \tag{9}$$

be Φ-invariant on M with respect to u . We suppose that for $u \equiv 0$ the behavior of the system is as desired if $v \equiv 0$.

It may happen that in the region of the varying of the variables x, u and t , defined by conditions of the problem, the continious corrective function does not exist. This makes us take into consideration piecewise-continious corrective functions $v(x,u,t)$. Then the right-hand part of the system (9) turns to be discontinious, but this does not result in appearing any essential peculiarities, because the properties of the invariant system are defined by the support field but system (9) is continious when the support control $\tilde{u}(t) \equiv 0$ is used. Let us build the support field $\tilde{p}(x,\tau)$ corresponding to $\tilde{u}(t) \equiv 0$ and write the defining equation

$$H[x, \tilde{p}(x,t), u, v, t] = H[x, \tilde{p}(x,t), 0, 0, t]$$

which defines v as a function of the variables x, u and t .

Theorem 4. In order that the function $v(x,u,t)$ be corrective, it is necessary and sufficient that it satisfies the defining equation.

This statement shows that the corrective function must be built as a solution (or from solutions) of the defining equation. Thus the synthesis of the invariant system is reduced to the prognosis of its nondisturbed motion[+].

Complete solution of the problem of synthesis is guaranteed if we know vector function $\tilde{p}(x,\tau)$ determined by nondisturbed systems (9) and (4), with the values of this vector function to be calculated only for points $\{x,\tau\}$ of the disturbed trajectory being realized.

With the help of the support function the defining equation may be written as follows

$$-(\mathrm{grad}_x \tilde{V}(x,t), f(x,u,v,t)) = \frac{\partial \tilde{V}(x,t)}{\partial t} = -(\mathrm{grad}_x \tilde{V}(x,t), f(x,0,0,t)).$$

But the problem of the function $\tilde{V}(x,t)$ determination is equivalent to the problem of integration of the nondisturbed system (9). Than, if we can find the complete solution of this system, the invariant system synthesis problem may be solved in final form. Let

$$\tilde{x}(t) = \tilde{\varphi}(x,\tau,t)$$

[+] It is necessary to note that the solution of the optimum systems synthesis problem requires that the boundary problem for conjugate variables be solved.

is the complete solution of the system (9) for $u \equiv 0$, $v \equiv 0$, where point $\{x,\tau\}$ plays a role of initial conditions. If we substitute $\tilde{x}(t)$ into (3) and solve the obtained equation with respect to $\tilde{T}$, we obtain $\tilde{T} = \tilde{T}(x,\tau)$ and consequently,

$$\tilde{x}(\tilde{T}) = \tilde{\varphi}\left[x,\tau,\tilde{T}(x,\tau)\right] = \varphi(x,\tau).$$

Than the support function is determined by the expression

$$\tilde{V}(x,\tau) = \Phi[\tilde{x}(\tilde{T}),\tilde{T}] = \Phi[\varphi(x,\tau),\tilde{T}(x,\tau)].$$

Conditions of the corrective function existence coincide with the conditions of the existence of the defining equation solution. As a simple sufficient condition of existence of the solution of this equation we may indicate an Φ - controllability condition of the system (1) with respect to v . This condition consist in fulfilment of unequality

$$\frac{\partial}{\partial v} H[x,\tilde{p}(x,t),u,v,t] \neq 0$$

for all the values of the variables x, t, u and v under consideration. In the case when in the given region of the values of the variables x, t and u, or in any part of this region, a solution of the defining equation does not exist, or any solution of this equation does not satisfy constructive restricrions on the control parameter v, imposed by the conditions of the problem, the realization of the invariant synthesis with the help of corrective function is impossible. However this does not mean to ensure the invariance of the system (7) is impossible and shows only that the control law for the parameter v must be looked for in another form.

It follows from the explicit formula (6) for change of the functional, that the necessary and sufficient condition for the system (7) to be invariant is that equality

$$\Delta J = -\int_{t^\circ}^{T}\Big[H[x,\tilde{p}(x,t),u,v,t] - H[x,\tilde{p}(x,t),0,0,t]\Big]dt = 0 \qquad (10)$$

is fulfilled along any disturbed trajectory $x(t),t$ initiating in the point $\{x^\circ, t^\circ\}$. This equality must be considered as a specific equation with respect to v , solution of which must be looked for in accordance with available information about the disturbance and the state of the system. For example, if we are given the constrain $v \in W$ the simplest method to ensure the condition (10) to be fulfilled consists in addition to the control law, found from the defining equation, the conditions

$$H[x,\tilde{p}(x,t),u,v,t] = \sup_{w \in W} H[x,\tilde{p}(x,t),u,w,t] \quad \text{if } \Delta J(t) > 0$$

and

$$H[x,\tilde{p}(x,t),u,v,t] = \inf_{w \in W} H[x,\tilde{p}(x,t),u,w,t] \quad \text{if } \Delta J(t) < 0$$

where $\Delta J(t)$ is a value of the integral in expression (10) when upper limit is t .

All above results are easily generalized to the case of invariance of $m \geqslant 2$ functionals in the system (1). To have the possibility of the synthesis we must consider the control parameter v as an m -dimensional vector, that permits us to construct m independent corrective functions.

In the case when the right-hand sides of the system (1) (of (7)) are discontinuous above results will be true, if consider that the vector function $\tilde{p}(t)$ in (4), when the trajectory $\tilde{x}(t),t$ passes through the surface of discontinuity, satisfies the jump conditions[13]. It is interesting to note, that if at least one support control exosts for which right-hand sides of system (1) are continuous along the trajectories $\tilde{x}(t),t$, the support field $\tilde{p}(x,t)$ will be continuous and in the case of the systems with the discontinuous right-hand parts.

5. Examples

Corrective function may be built in an explicit form in any problem for which the complete solution of nondisturbed motion can be found. As an examples of such problems the corrective function synthesis problems for the systems of the form

$$\dot{x}(t) = Ax + g(x,u,v,t), \text{ where } g(x,0,0,t)=0,$$

may serve, as well as the problems, connected with the control of a material point motion in the central field of the gravity forces, and other problems.

Example 1. Consider the system

$$\dot{x}_1 = (1+u)x_2, \qquad \dot{x}_2 = (1+v)x_1,$$

trajectory of which must terminate on the surface determined by the equation

$$M(x_1, x_2, t) \equiv x_1 + x_2 - 1 = 0.$$

The only corrective functions, which guarantee the invariance of this system with respect to u for the functionals

$$\Phi \equiv x_1^2(T) + x_2^2(T), \quad \Phi \equiv x_1(T) - x_2(T) + T \quad \text{and} \quad \Phi \equiv T$$

will be, respectively, functions

$$v = u, \quad v = \frac{2x_1x_2(x_1+x_2) - x_2}{2x_1x_2(x_1+x_2) + x_1}\,u \quad \text{and} \quad v = -\frac{x_2}{x_1}\,u.$$

Example 2. In problem

$$\dot{x}_1 = x_2\left[1 + u(1-v^2)\right], \qquad \dot{x}_2 = x_1 + v(x_1^2 - u^2);$$

$$M \equiv x_1 + x_2 - 1 = 0; \qquad \Phi \equiv x_1^2(T) + x_2^2(T)$$

we have

$$\tilde{V} = 0.5\left[1+(x_1^2-x_2^2)^2\right]; \quad \tilde{p}_1 = -2x_1(x_1^2-x_2^2), \; \tilde{p}_2 = 2x_2(x_1^2-x_2^2)$$

and defining equation is written in the form

$$\upsilon^2 x_1 u + \upsilon(x_1^2-u^2) - x_1 u = 0.$$

An infinite number of discontinious corrective functions may be constructed from two distinct solutions of this equation. If we restrict the choice of υ by additional demand that $|\upsilon|$ is to be minimal, we have

$$\upsilon = \frac{u}{x_1} \text{ for } |x_1| > |u| \, , \quad \upsilon = -\frac{x_1}{u} \text{ for } |x_1| \leq |u| \text{ and } u \neq 0,$$

$$\upsilon = 0 \quad \text{for } u = 0 \, .$$

In this problem corrective function, which is continious in all space $X \times U \times t$ does not exist.

BIBLIOGRAPHY

1. Кулебакин В.С. Теория инвариантности автоматически регулируемых и управляемых систем. Труды I Конгресса ИФАК. Изд-во АН СССР, Москва, 1961, т. I, стр. 247-258.
2. Петров Б.Н. Принцип инвариантности и условия его применения при расчете линейных и нелинейных систем. Труды I Конгресса ИФАК. Изд-во АН СССР, Москва, 1961, т. I, стр. 259-275.
3. Фельдбаум А.А., Шрейдер Ю.А. Теория компенсации. В книге А.А. Фельдбаума "Вычислительные устройства в автоматических системах". Физматгиз, Москва, 1959, стр. 492-512.
4. Беля К.К. Инвариантность системы регулирования по отношению к измерению параметров системы. Труды I Конгресса ИФАК. Изд-во АН СССР, Москва, 1961, т.I, стр. 282-289.

5. Кухтенко А.И. Критерии абсолютной инвариантности для систем регулирования с переменными параметрами. Известия АН СССР, отделение технических наук. Энергетика и автоматика. № 2, стр. I06-II3, I96I.

6. Петров Б.Н., Уланов Г.М., Емельянов С.В. Инвариантность и оптимизация в системах автоматического управления с жесткой и переменной структурой. Труды П Конгресса ИФАК. Наука, Москва, I965, т. I, стр. 2I4-229.

7. Розоноэр Л.И. Вариационный подход к проблеме инвариантности систем автоматического управления. Автоматика и телемеханика, т. XXIУ, I963, I-№ 6, стр. 744-756; П-№ 7, стр. 86I-870.

8. Понтрягин Л.С., Болтянский В.Г., Гамкрелидзе Р.В., Мищенко Е.Ф. Математическая теория оптимальных процессов. Физматгиз, Москва, I96I.

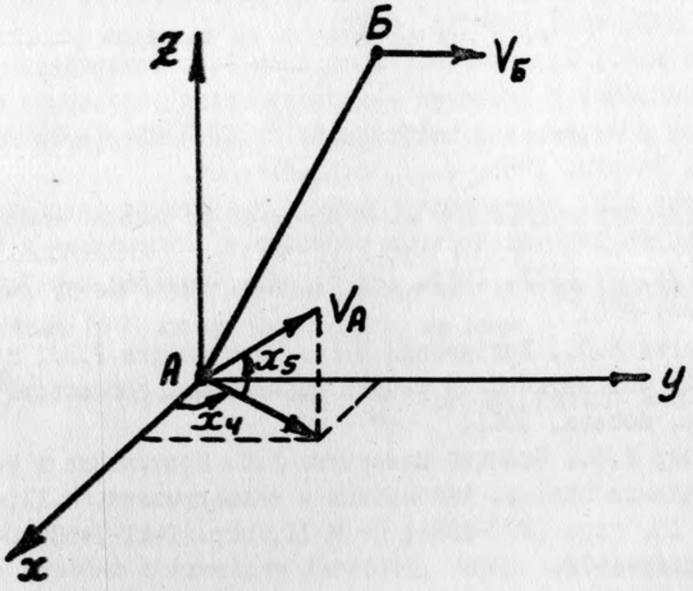
9. Розоноэр Л.И., Принцип максимума Л.С. Понтрягина в теории оптимальных систем. Автоматика и телемеханика, т.XX, I959, I - № IO, стр. I320-I334; П- № II, стр. I44I-I45B; Ш- № I2, стр. I56I-I578.

IO. Величенко В.В. Численный метод решения задач оптимального управления. Журнал вычислительной математики и математической физики, т.6, I966, № 4, стр. 635-647.

II. Болтянский В.Г. Математические методы оптимального управления. Наука, Москва, I966.

I2. Величенко В.В. К достаточным условиям оптимальности в принципе максимума. Доклады АН СССР, т.I82, № 4, I968, стр. 747-749.

I3. Величенко В.В. О задачах оптимального управления для уравнений с разрывными правыми частями. Автоматика и телемеханика, I966, № 7, стр. 20-30.


Z
Б
V_Б
V_A
x_5
A
y
x_4
x

49.1

On Controllability in a Pursuit Problem

Babakov N.A., Kim D.P.
Institute of Automation and Telemechanics

Moscow
USSR

1. Consider point A pursueing point B under the following conditions. The pursued point B moves at constant velocity along a straight line. The pursueing point A has the velocity constant in magnitude. Control actions of point A are angular velocities bounded in magnitude. Velocity V_A of the pursueing point A is less than velocity $V_{Б}$ of pursued point B ($V_A \leq V_{Б}$). This condition means that point A, depending on the initial conditions, may or may not hit point B or reach an E-vicinity of the latter.

Hence it is required to find a condition which must be met by point A in order to hit point B when pursueing it. The condition defects a controllability region within the space of the pursuit initial conditions and it will be termed as the condition for physical possibility of pursuit, or shorter, the controllability condition. Therefore the problem under study may also be formulated as search for a controllability region within the space of the pursuit initial conditions, and we shall further call it the controllability problem.

The two-dimensional controllability problem, the plane pursuit, was discussed elsewhere [1]. This paper deals with the three-dimensional problem.

2. Problem Stated. If coordinate system is chosen appropriately, that is axis AY is directed as vector $\vec{V}_{Б}$ of velocity of point B is, and at an initial time instant point A is localised at the origin of the coordinate system, then motion equations for points A and B will be of the form:

$$\begin{aligned}
\dot{x}_A &= V_A \cos x_5 \cos x_4 \\
\dot{y}_A &= V_A \cos x_5 \sin x_4 \\
\dot{z}_A &= V_A \sin x_5 \\
\dot{x}_4 &= u_1 \\
\dot{x}_5 &= u_2
\end{aligned} \qquad (2.1)$$

$$\dot{x}_{\text{Б}} = 0, \; \dot{y}_{\text{Б}} = V_{\text{Б}}, \; \dot{z}_{\text{Б}} = 0 \qquad (2.2)$$

Control actions of point A fulfil the condition

$$\Omega : \; |u_1| \le \omega_1 , \; |u_2| \le \omega_2 \qquad (2.3)$$

At an initial time instant of the pursuit

$$\begin{aligned}
& x_A(t_0) = y_A(t_0) = z(t_0) = 0 \\
& x_4(t_0) = x_4^\circ , \; x_5(t_0) = x_5^\circ \\
& x_{\text{Б}}(t_0) = x_1^\circ , \; y_{\text{Б}}(t_0) = x_2^\circ , \; z_{\text{Б}}(t_0) = x_3^\circ
\end{aligned} \qquad (2.4)$$

Introduce the notation

$$x_1 = x_{\text{Б}} - x_A , \; x_2 = y_{\text{Б}} - y_A , \; x_3 = z_{\text{Б}} - z_A \qquad (2.5)$$

Then eqs. (2.1) and (2.2) give the set

$$\begin{aligned}
\dot{x}_1 &= -V_A \cos x_5 \cos x_4 \\
\dot{x}_2 &= V_{\text{Б}} - V_A \cos x_5 \sin x_4 \\
\dot{x}_3 &= -V_A \sin x_5 \\
\dot{x}_4 &= u_1 \\
\dot{x}_5 &= u_2
\end{aligned} \qquad (2.6)$$

which describes relative motion of points A and B, and consequently, point A as pursueing point B. The pursuit initial conditions are as follows (cf. eq (2.4))

$$x_1(t_0)=x_1^\circ,\ x_2(t_0)=x_2^\circ,\ x_3(t_0)=x_3^\circ,\ x_4(t_0)=x_4^\circ,\ x_5(t_0)=x_5^\circ \qquad (2.7)$$

Now, the controllability problem will be formulated in the following manner. Consider a pursuit described with set (2.6) and relation (2.3) together with a condition

$$V_A \leq V_B \qquad (2.8)$$

For this pursuit find the controllability region within space X° of initial conditions $x^\circ = (x_1^\circ, x_2^\circ, x_3^\circ, x_4^\circ, x_5^\circ \quad)$.

One must note that the problem statement and the solution procedure do not change to any substantial degree, if, first, the motion of points A and B is non-uniform, the velocity magnitudes of points A and B are assumed to have a time dependence, and if, second, the pursueing point as well as the pursued one are assumed to be complex dynamic systems and equations are influenced by the equations for these systems.

3. To begin with, consider the two-dimensional controllability problem when point A and B always move on the plane (y , z). The pursuit equations may now be obtained in (2.6) $x_1 \equiv 0$, $x_4 = \frac{\pi}{2}$

$$\begin{aligned} \dot{x}_2 &= V_B - V_A \cos x_5 \\ \dot{x}_3 &= -V_A \sin x_5 \\ \dot{x}_5 &= u_2 \end{aligned} \qquad (3.1)$$

Set Ω_1 of possible values of control action u_2 is described with the inequality $|u_2| \leq \omega_2$.

Denote t_1 a time instant corresponding to point A as crossing a trajectory of pursuied point B for the first time:

$$x_3(t_1) = 0 \qquad (3.2)$$

and $x_3(t) \neq 0$ at all $t \in (t_0, t_1)$. Quantities t_1 and $x_2(t_1)$ - depend on the initial condition $\bar{x}^0 = (x_2^0, x_3^0, x_5^0)$ of the pursuit and on control action U_2. If state $\bar{x}^0$ is controllable, there exists such a control action $u_2 \in \Omega_1$ that x_2 is less than or equal to zero, $x_2(t_1) \leq 0$, at time instant t_1. Actually, if all feasible control actions give $x_2(t_1) > 0$, then whatever control action $u_2(t) \in \Omega_1$ is used point A will be behind point B on the trajectory of the latter at time instant t_1 and it will not be able to reach point B by virtue of condition (2.8).

Thus, a necessary condition of controllability of state $\bar{x}^0$ will be

$$\min_{u_2 \in \Omega_1} x_2(t_1, \bar{x}^0) \leq 0 \qquad (3.3)$$

If set Ω_1 coincides with the entire axis ($\omega_2 = \infty$), condition (3.3) is sufficient as well. But if set Ω_1 of permissible control is bounded, then states $\bar{x}^0$ are possible such that whatever control action $u_2 \in \Omega_1$ is used point A will cross the trajectory of point B being in front of it for the first time, $x_2(t_1) < 0$, and behind it for the second time, $x_2(t_2) > 0$, t_2 - time instant corresponding to point A as crossing the target trajectory for the second time. Here point A cannot reach point B, though condition (3.3) holds. The following assumtion is valid. For state $\bar{x}^0$ to be controllable, it is necessary and sufficient that condition (3.3) be fulfilled together with at least one of the following conditions:

$$\max_{u_2 \in \Omega_1} x_2(t_1, \bar{x}^0) \geq 0 \qquad (3.4)$$

$$\min_{u_2 \in \Omega_1} x_2(t_2, \bar{x}^0) \leq 0 \qquad (3.5)$$

<u>Necessity.</u> The necessity of condition (3.3) has been already shown. Necessity of inequalties (3.4) or (3.5) is proved as follows: if these inequalities are both invalid,

then point A, condition (3.3) being satisfied at any $u_2 \in \Omega_1$ will be in front of point B at $t = t_1$ and behind it at $t = t_2$ It has already been pointed out, however, that this leads to non-controllability of state $\bar{x}^\circ$.

Sufficiency. If condition (3.3) holds, then for a given state $\bar{x}^\circ$ there exists control action such that point A will cross the trajectory of point B in front of it at time instant t_1 whereas if condition (3.4) holds, then control action exists such that point A will cross the trajectory of point B behind it at time instant t_1 . It is clear that the two conditions holding simultaneously, a control action must exist such that point A hits point B, $x_2(t_1) = 0$, at time instant t_1 . In other words, state $\bar{x}^\circ$ is here controllable. Controllability of state $\bar{x}^\circ$ will now be demostrated under conditions (3.3) and (3.5). Here point A moves from the trajectory of point B for a stretch of time belonging to interval (t_1 , t_2). Therefore, one can make point A hit point B at time instant $t = t_2$ by varying the stretch with the aid of the appropriate choice of control action. Q.E.D.

One can also see that, the plane pursuit being considered, the controllability problem resolves in determining the minimum and maximum of functional $x_2(t_1, \bar{x}^\circ)$, and the minimum of functional $x_2(t_2, \bar{x}^\circ)$. But to find the minimum (maximum) of functional $x_2(t_2, \bar{x}^\circ)$ one has to solve the Mayer variational problem which is easy to formulate[1]. To find the minimum of functional $x_2(t_2, \bar{x}^\circ)$ one has to solve the three-point variational problem where the values of function $\bar{x}(t) = \{x_2(t), x_3(t), x_5(t)\}$ are given that correspond to three values of time, t_0 , t_1 , and t_2 . Paper[2] describes a procedure to solve the multipoint variational problem and, in particular, the three-point one.

4. Let us now consider the general problem of the three-dimensional pursuit. Here necessity and sufficiency cannot be formulated as above. For example, the fact that the conditions

$$\min_{u \in \Omega} x_2(t_1, x^\circ) < 0 , \quad \max_{u \in \Omega} x_2(t_1, x^\circ) > 0$$

are met simultaneously does not necessary imply controllability of state x° . Here $x^\circ=(x_1^\circ, x_2^\circ, x_3^\circ, x_4^\circ, x_5^\circ)$ is an initial of the pursuit, t_1 means, as in 3.above, a time instant corresponding to point A when this for the first time crosses a trajectory of point B when pursuing it

$$x_1(t_1)=0,\ x_3(t_1)=0 \qquad (4.1)$$

and at least one of these inequalities does not hold at any $t_\circ \in (t_\circ, t_1)$.

But the same procedure as developed in 3. above allows to show that necessary conditions of controllability of state x° are the inequalities

$$\min_{u\in\Omega} x_2(t_1, x^\circ) \leq 0 \qquad (4.2)$$

A variational problem which has to be solved in order to find the minimum of functional $x_2(t_1, x^\circ)$ may be stated as follows. Consider continuous and piecewise-differetiable functions $x(t)=\{x_1(t), \cdots, x_5(t)\}$ for which set (2.6) is satisfied within interval $(t_\circ, t_1)$ while conditions (2.7) and (4.1) are at the ends, and consider piecewise-continuous control actions $u(t)\in\Omega$; among the functions and the control actions find such $x(t)$ and $u(t)$ that the functional

$$S=x_2(t_1, x^\circ) \qquad (4.2)$$

takes its minimum value.

We shall solve the problem using L.S.Pontryagin's principle of maximum.

Construct Hamiltonian function:

$$H=-\psi_1 V_A \cos x_5 \cos x_4+\psi_2(V_{\text{Б}}-V_A \cos x_5 \sin x_4)- \qquad (4.3)$$
$$-\psi_3 V_A \sin x_5+\psi_4 u_1+\psi_5 u_2$$

Write the "conjugated" equations:

$$\begin{aligned} &\dot{\psi}_1 = \dot{\psi}_2 = \dot{\psi}_3 = 0 \\ &\dot{\psi}_4 = -\psi_1 V_A \cos x_5 \sin x_4 + \psi_2 V_A \cos x_5 \cos x_4 \\ &\dot{\psi}_5 = -\psi_1 V_A \sin x_5 \cos x_4 - \psi_2 V_A \sin x_5 \sin x_4 + \psi_3 V_A \cos x_5 \end{aligned} \qquad (4.4)$$

Variables ψ_2 , ψ_4 , ψ_5 and function H must satisfy the following conditions [3] at a final time instant t_1 :

$$\psi_2(t_1) = -1, \; \psi_4(t_1) = \psi_5(t_1) = 0 \qquad (4.5)$$

$$H(t_1) = \left[-\psi_1 V_A \cos x_5 \cos x_4 - V_B + V_A \cos x_5 \sin x_4 - \psi_3 V_A \sin x_5\right]_{t=t_1} = 0 \qquad (4.6)$$

From (4.4) and (4.5) we obtain

$$\begin{aligned} &\psi_1 = C_1 \,, \; \psi_2 = C_2 \,, \; \psi_3 = C_3 \\ &\dot{\psi}_4 = -C_1 V_A \cos x_5 \sin x_4 - V_A \cos x_5 \cos x_4 \\ &\dot{\psi}_5 = -C_1 V_A \sin x_5 \cos x_4 + V_A \sin x_5 \sin x_4 + C_3 V_A \cos x_5 \end{aligned} \qquad (4.7)$$

As function H does not depend explicitly on time, it is constant and vanishes by virtue of condition (4.6):

$$\begin{aligned} -C_1 V_A \cos x_5 \cos x_4 - V_B + V_A \cos x_5 \sin x_4 - \\ - C_3 V_A \sin x_5 + \psi_4 u_1 + \psi_5 u_2 = 0 \end{aligned} \qquad (4.8)$$

The principle of maximum gives for the optimal control action at $\psi_4 \neq 0$ and $\psi_5 \neq 0$

$$u_1 = \omega_1 \operatorname{sgn} \psi_4 , \; u_2 = \omega_2 \operatorname{sgn} \psi_5 \qquad (4.9)$$

If $\psi_4 = 0$ and $\psi_5 = 0$, then function H does not depend explicitly on control action $u = (u_1, u_2)$, and the principle of maximum does not allow to obtain an expression for optimal control action. Therefore in this case we will try another approach.

Let $\Psi_4 = 0$ at a certain interval. Then [illegible]rivative of function ψ_4 vanishes also at this interva[illegible]h the possible exception of its ends, $\dot{\psi}_4 = 0$.

Transform $\dot{\psi}_4$ in this equality by means of expression (4.4); we obtain

$$\cos x_5 (C_1 \sin x_4 + \cos x_4) = 0 \qquad (4.10)$$

By using geometric considerations together with the results obtained in paper [1] one can show that equality $\cos x_5 = 0$ is out of the question within a certain non-degerate interval at optimum control action. Hence from (4.10) it follows

$$C_1 \sin x_4 + \cos x_4 = 0$$

hence the expression for x_4 is easily obtained

$$x_4 = k\pi - \operatorname{arc\,sin} \frac{1}{\sqrt{1+C_1^2}} \qquad (4.11)$$

Thus, if function Ψ_4 is identically zero at a certain interval, parameter u_1 of the optimal control action is expressed by (4.12) at this interval with the possible exception of the boundary point of the latter.

$$u_1 = \dot{x}_4 = 0 \qquad (4.12)$$

Suppose function Ψ_4 vanishes at a finite number of isolated points of interval $[t_0, t_1]$. Then the set of isolated or boundary points of the interval, the set which makes function Ψ_4 vanish, will be finite because permissible control action is assumed to be piecewise-continuous and to contain the finite number of discontinuity points. Therefore, the value of parameter u_1 of optimum control action may be assumed to vanish at these points without violating the optimality condition. So we have at the whole interval $[t_0, t_1]$ for u_1

$$u_1 = \omega_1 \, sgn\, \psi_4(t) \tag{4.13}$$

where

$$Sgn\, \psi_4(t) = \begin{cases} 1, & \psi_4(t) > 0 \\ 0, & \psi_4(t) = 0 \\ -1, & \psi_4(t) < 0 \end{cases}$$

Now let function ψ_5 vanish at a certain interval. Then derivative $\dot{\psi}_5$ vanishes at the interval with the possible exception of its boundary points:

$$-c_1 \sin x_5 \cos x_4 + \sin x_5 \sin x_4 + c_3 \cos x_5 = 0 \tag{4.14}$$

By differentiating this identity we obtain after easy transformations

$$u_2 = \frac{(c_1 \sin x_4 + \cos x_4)\, u_1}{ctg\, x_5 (c_1 \cos x_4 - \sin x_4) + c_3}, \tag{4.15}$$

where u_1 is a parameter of optimum control action, the parameter given by eq. (4.13). Formula (4.15) holds at such points of the interval with $\dot{\psi}_5 \equiv 0$ where functions x_4 and x_5 are differentiable. One may believe, however, that formula (4.15) is true over the entre interval where $\dot{\psi}_5 \equiv 0$ without violating the generality, because there is only a finite number of the points where functions x_4 and x_5 are non-differentiable. Let $u_2 = 0$ at the isolated and boundary points of the interval where $\psi_5 = 0$. Then parameter u_2 of the optimal control action is given by formula (4.15) at non-degenerate intervals where derivative $\dot{\psi}_5$ is indentically zero, or by formula (4.9) at other points of interval $[t_0, t_1]$.

To summarize, we have found the optimal control action $u(t) = \{u_1(t), u_2(t)\}$. To find now the controllability condition one has to solve set (2.6) together with set (4.4) under initial conditions (2.7), (4.1), (4.5) and (4.6)

Figure. $A\ (x_A, y_A, z_A)$ are coordinates of the pursueing point B; $Б\ (x_Б, y_Б, z_Б)$ are coordinates of the pursued point B; x_4 is the angle between AX axis and a projection of vector $\vec{V}_A$ of the point A velocity onto plane (x, y); x_5 is the angle between plane (x, y) and the velocity vector $\vec{V}_A$.

Reference

I. Бабаков Н.А., Ким Д.П. Об области управляемости и оптимальных траекториях сближения двух космических аппаратов. Доклад на симпозиуме ИФАК по автоматическому управлению в космосе, воде и под землей. Вена, I967.

2. Троицкий В.А. Вариационные методы решения задач оптимизации процессов управления. Труды Всесоюзного совещания по автоматике. Оптимальные системы. Статистические методы, "Наука", I967г.

3. Розоноэр Л.И. Принцип максимума Л.С.Понтрягина в теории оптимальных систем. I, П. Автоматика и телемеханика, I959, т.XX, №№ IO, II.

49.2

DYNAMICS OF THE TETHERED ASTRONAUT MOVING TOWARD THE SPACECRAFT AND AN APPROACH TO SYNTHESIS OF SPACECRAFT CONTROL BASED ON A THEORY OF THE VARIABLE-STRUCTURE SYSTEMS.

Soshnikov V.N., Ulanov, G.M.

Moscow
USSR

INTRODUCTION

A flexible tetherline can serve as a means of retrieving an astronaut. The use of the technique,causes the difficulties which have been treated in works[1,2]. Wrapping of the tetherline around the spacecraft rotation rates and excessive impact velocities of the astronaut are the main difficulties. Because of these effectsthe astronaut's retrieval at arbitrary initial condition cannot be provided.

One purpose of this work is to select a region of the initial conditions in which the astronaut's retrieval can be provided under given restrictions. The solution of this problem permits to estimate the practical feasibility of an uncontrolled tetherline system. Another purpose of this study is to dev#lop a perspective synthesis technique for attitude spacecraft control which would permit to eliminate some system properties causing the difficulties of the practical system uses. These problems are solved for a case of plane motion at the constant reel-in rate.

EQUATIONS OF MOTION

A mathematical model is developed here under the following assumptions.

1. The spacecraft and the astronaut are solid bodies.
2. The tetherline is a non-stationary linear constraint.
3. The external disturbances including those from the gravity gradient are negligible.

At thése assumptions the equations of motion for the center of mass and the rotational equations are mutually independent.

The dynamical model of rotational motion and the generalized coordinate system are shown in Fig. 1.

Fig.1 shows:

1 and 3 - the spacecraft and the astronaut with masses m_1 and m_3, respectively, and moments of inertia J_1 and J_3, respectively, about the axis going through the centres of masses O_1, O_3 and normal to the plane of the figure.

$X'O'Y'$ - the inertial coordinate system with origin at the center of system mass.

r_1 - the distance from the center of mass of the spacecraft to the point of the line output from the spacecraft.

r_2 - the tetherline lenth varying according to:

$$r_2 = r_{20} + \dot{r}_{20} t$$

r_3 - the distance from the center of mass of the astronaut to the attachment point on the surface of the space-craft.

$\varphi_1, \varphi_2, \varphi_3$ - generalized coordinates in the $X'O'Y'$ axis system.

ℓ_4 - the distance between the center of mass of the space-craft and the center of mass of the astronaut.

The second-kind Lagrangian equations of angular motion written with respect to the second derivatives are:

$$\ddot{\varphi}_1 = -\frac{F_2 r_1 \sin(\varphi_1-\varphi_2)}{J_1}$$

$$\ddot{\varphi}_2 = \frac{1}{r_2}\Big\{\frac{F_2[r_1^2\cos(\varphi_1-\varphi_2)\sin(\varphi_1-\varphi_2)]}{J_1} - \frac{F_2[r_3^2\cos(\varphi_2-\varphi_3)\sin(\varphi_2-\varphi_3)]}{J_3} - $$

$$- 2\dot{r}_{20}\dot{\varphi}_2 + r_1\sin(\varphi_1-\varphi_2)\dot{\varphi}_1^2 + r_3\sin(\varphi_2-\varphi_3)\dot{\varphi}_3^2\Big\} \qquad (1)$$

$$\ddot{\varphi}_3 = \frac{F_2 r_3 \sin(\varphi_2-\varphi_3)}{J_3}$$

where a force F_2 of line tension is :

$$F_2 = \frac{K[r_1\cos(\varphi_1-\varphi_2)\dot{\varphi}_1^2 + r_2\dot{\varphi}_2^2 + r_3\cos(\varphi_2-\varphi_3)\dot{\varphi}_3^2]}{1 + \frac{Kr_1^2\sin^2(\varphi_1-\varphi_2)}{J_1} + \frac{Kr_3^2\sin^2(\varphi_2-\varphi_3)}{J_3}}; \quad K = \frac{m_1 m_3}{m_1+m_3} \qquad (2)$$

As the generalized coordinate is cyclic (or ignored) the order of Eqs.(1) can be reduced by two. The conventient reduetion procedure is gained invølving Routh variables[3] such as:

$t, \alpha_1, \alpha_2, \varphi_1, \dot{\alpha}_1, \dot{\alpha}_2, p_1$

where $p_1 = \frac{\partial L}{\partial \dot{\varphi}_1} \quad \alpha_1 = \varphi_1 - \varphi_2 \quad \alpha_2 = \varphi_3 - \varphi_2$ (3)

With the help of the Routh function $R(t, \alpha_1, \alpha_2, \varphi_1, \dot{\alpha}_1, \dot{\alpha}_2, p_1)$ independent of φ_1 in this case the system (1) can be replaced an equivalent system as :

$$\frac{d}{dt}\left(\frac{\partial R}{\partial \dot{\alpha}_1}\right) - \frac{\partial R}{\partial \alpha_1} = 0$$

$$\frac{d}{dt}\left(\frac{\partial R}{\partial \dot{\alpha}_2}\right) - \frac{\partial R}{\partial \alpha_2} = 0$$

$$\frac{d\varphi_1}{dt} = \frac{\partial R}{\partial p_1} \qquad (4)$$

$$\frac{dp_1}{dt} = -\frac{\partial R}{\partial \varphi_1}$$

Since the Routh function is independent of φ_1 , we have $p_1 = p_{10} =$ const and the problem comes to integration of the ferst two Eqs.(4). These equations written with respect to the second derivatives are :

$$\ddot{\alpha}_1 = \bar{c}_{11} f_{11} + c_{12}\dot{\alpha}_1^2 + c_{13}\dot{\alpha}_2^2 + c_{14}\dot{\alpha}_1\dot{\alpha}_2 + c_{15}\dot{\alpha}_1 + c_{16}\dot{\alpha}_2$$
$$\ddot{\alpha}_2 = \bar{c}'_{11} f'_{11} + c'_{12}\dot{\alpha}_1^2 + c'_{13}\dot{\alpha}_2^2 + c'_{14}\dot{\alpha}_1\dot{\alpha}_2 + c'_{15}\dot{\alpha}_1 + c_{16}\dot{\alpha}_2 \qquad (5)$$

where $\bar{c}_{11}, \bar{c}'_{11}$ are functions of α_1, α_2 , and c_{ik}, c'_{ik}, f_{11}, f'_{11} are functions of α_1, α_2 and t. After Eqs.(5) have been integrated, the expressions for an angular position of the spacecraft in the inertial space are obtained. They are :

$$\dot{\varphi}_1 = \frac{\partial R}{\partial p_1} = \frac{p_{10} + \dot{\alpha}_1\left[K(r_2 b_1 + r_2^2 + 2b_2 r_2 + b_1 b_2 + a_1 a_2) + J_3 + K r_3^2\right] - \dot{\alpha}_2\left[K(b_2 r_2 + b_1 b_2 + a_1 a_2) + J_3 + K r_3^2\right] + K(a_1 - a_2)\dot{r}_2}{J_1 + J_3 + K l_t^2} \qquad (6)$$

$$\varphi_1 = \int \dot{\varphi}_1(t)\,dt + C \qquad (7)$$

where $a_1 = r_1 \sin\alpha_1 \quad a_2 = r_3 \sin\alpha_2 \quad b_1 = r_1\cos\alpha_1 \quad b_2 = r_3 \cos\alpha_2$

$$l_t^2 = r_1^2 + r_2^2 + 2b_1 r_2 + r_3^2 + 2b_1 b_2 + 2a_1 a_2 + 2b_2 r_2 \qquad (8)$$

It can be shown that the value of p_1 is angular momentum and an integral $p_1 = p_{10}$ signifies the conservation of angular momentum. With all these states in mind one can conclude that Eqs.(5), (6) and (7) describe the system motion completely.

SYSTEM STATIONARY MOTIONS

The equilibrium states of the system (5) or the points α_1^*, α_2^* for which $\dot{\alpha}_1 \equiv \dot{\alpha}_2 \equiv \ddot{\alpha}_1 \equiv \ddot{\alpha}_2 \equiv 0$, can be defined as roots of algebraic equations :

$$\bar{c}_{11}(\alpha_1, \alpha_2) = 0$$
$$\bar{c}'_{11}(\alpha_1, \alpha_2) = 0 \qquad (9)$$

The solutions α_1^* and α_2^* may be found by using the following values of a_1^* and a_2^* :

$$a_1^* = r_1 \sin\alpha_1^* \quad a_2^* = r_3 \sin\alpha_2^* \qquad (10)$$

The values of a_1^* and a_2^* which are the solutions of Eqs.(9) may be desined by equations :

$$a_{11}^* = -\frac{P_{10} + K a_{21}^* \dot{r}_{20}}{K \dot{r}_{20}} \qquad (11)$$

$$a_{12,13}^* = \frac{P_{10} J_1}{(J_1+J_3) 2K\dot{r}_{20}} \pm P_{10}\sqrt{\left[\frac{J_1}{(J_1+J_3)2K\dot{r}_{20}}\right]^2 - \frac{2J_1^2}{K(J_1+J_3)P_{10}^2}}\ ;\ a_{22}^* = \frac{J_3}{J_1}a_{12}^*\ ;\ a_{23}^* = \frac{J_3}{J_1}a_{13}^* \qquad (12)$$

where the second subscripts of a_1^* and a_2^* relate to various solution forms. It is easy to show that Eq.(11) correspond to an infinite set of system ststionary motions when the nonrotational spacecraft and the nonrotational astronaut are freely closing at a range rate $\dot{r}_{20}$ with their constant position with respect to the line at its zero tention. Two stationary motions corresponding to Eqs.(12) are some ones at which the system untwists at an increasing velocity with respect to the inertial space while the spacecraft and the astronaut attitudes are fixed with respect to the line.

It is an interesting peculiarity that the system equilibrium ststes and ststionary motions depend not only on the system parameters but on the initial conditions as well. It is the latter which define a value of angular momentum P_{10}. The real equilibruim values of α_1^* via the system parameters and angular momentum as well as signs of the member $C_{11} = \overline{C}_{11} f_{11}$ in the right side of Eqs.(5) it is convenient to represent by means of a bifurcation diagram (Fig.2)

Fig.2 shows :

$\underline{\overline{II}}$ and $\underline{\overline{II}}'$ - curves defined by an equation :

$$P_{10} = \pm\left[\dot{r}_{20} K \frac{J_1+J_3}{J_1} r_1 \sin\alpha_1^* + \frac{2\dot{r}_{20} J_1}{r_1 \sin\alpha_1^*}\right]$$

$\overline{\overline{II}}$ - curves inclosed between the curves $\underline{\overline{\overline{II}}}'$ and $\underline{\overline{\overline{II}}}''$, and defined by an equation :

$$P_{10} = -K\left(a_{11}^* + a_{21}^*\right)\dot{r}_{20} \qquad (14)$$

and transferring to $\underline{\overline{\overline{II}}}'$ and $\overline{\overline{II}}''$ respectively at $a_{21}^* = +r_3$ and $a_{21}^* = -r_3$.

The function α_1^* of P_{10} takes the form of $\underline{\overline{II}}$ or $\underline{\overline{II}}'$ which depends on the fact whether the value of

$$a_{1m}^* = J_1\sqrt{\frac{2}{K(J_1+J_3)}} \qquad (15)$$

is less or greater than τ_1 .

The values of α_1^* are defined by equations (Fig.2) :

$$\alpha_{12}^* = \arcsin \frac{|a_{12}^*|}{\tau_1} \qquad \alpha_{13}^* = \arcsin \frac{|a_{13}^*|}{\tau_1}$$
$$\alpha_{12}^{*\prime} = \pi - \arcsin \frac{|a_{12}^*|}{\tau_1} \qquad \alpha_{13}^{*\prime} = \pi - \arcsin \frac{|a_{13}^*|}{\tau_1} \tag{16}$$
$$\alpha_{1m}^* = \arcsin \frac{|a_{1m}^*|}{\tau_1} \qquad \alpha_{1m}^{*\prime} = \pi - \arcsin \frac{|a_{1m}^*|}{\tau_1}$$

The values of parameters P_{10}' , $\bar{P}_{10}$ and P_{10}^{min} , at which the number of equilibruim ststes is changing, are bifurcational and defined by equatios :

$$P_{10}' = |\dot{\tau}_{20}| \left[\frac{2J_1}{\tau_1} + \frac{K(J_1+J_3)\tau_1}{J_1} \right] ; \quad P_{10}^{min} = 2\sqrt{2}\, |\dot{\tau}_{20}| \sqrt{K(J_1+J_3)} \tag{17}$$

$$\bar{P}_{10} = K(\tau_1 + \tau_3)|\dot{\tau}_{20}|$$

The equilibruim values α_2^* depend on the system parameters and P_{10} in the same manner as shown in Fig.2 with the only distinction that a scale of the α_2^* - axis and the equations of $\overline{II}$ and $\overline{II}'$ must be changed to :

$$a_1^* = \frac{J_1}{J_3} a_2^*$$

Now the bifurcation parameter P_{10}' is defined by:

$$P_{10}' = |\dot{\tau}_{20}| \left[\frac{2J_3}{\tau_3} + \frac{K(J_1+J_3)}{J_3} \tau_3 \right] \tag{18}$$

Since the values of α_1^* and α_2^* differed by 2kπ (k=1,2...) relate to the same system configuration ($\alpha_1, \dot{\alpha}_1$ and $\alpha_2, \dot{\alpha}_2$ are cylindrical subspaces) , all possible equilibruim states are exhausted by the consideration of equilibruim values on an interval ($-\pi$, π). For the bifurcation diagram in the form $\overline{II}'$ and at $P_1 > P_{10}'$, $P_{10} < P_{10}'$ the major types of spacecraft-tetherline motions are shown in Fig.3.

APPROXIMATE EQUATIONS OF SPACECRAFT- TETHERLINE RELATIVE MOTION

For arbitrary geometric and inertial characteristics of the closing object an analysis of stationary motion was performed. A formulae spacecraft - line relative motion is derived taking into account that the following inequalities.

$$m_1 \gg m_2 \qquad \tau_1 \gg \dot{\tau}_2 \qquad \tau_1 \gg l_2 \tag{19}$$

are valid for the astronaut's retrieval system. In this case a structure of the phase space for a system model with an astronaut as a point mass and a structure of the phase subspace $\alpha_1 \dot{\alpha}_1$ for a considered model are practically the same as it follows from the equatuions of equilibruim states α_1^* and from the values of bifurcation parameters. For aur simplified model the equations of spacecraft-line relative motion linearized in the vicinity of equilibruim states $\alpha_{13}^*, \alpha_{13}^{*\prime}$ are:

$$\bar{\alpha}_1'' + A_1(\tau_2)\bar{\alpha}_1' + B_1(\tau_2)\bar{\alpha}_1 = 0 \qquad (20)$$

and linearized in the vicinity of states $\alpha_{11}^*, \alpha_{11}^{*\prime}$ are:

$$\bar{\alpha}_1'' + A_1(\tau_2)\bar{\alpha}_1' + \bar{B}_1(\tau_2)\bar{\alpha}_1 = 0 \qquad (21)$$

where $\bar{\alpha}_1 = \alpha_1 - \alpha_1^*$ and $\bar{\alpha}_1'$ denotes differentiations with respect to a new independent variable τ_2 which is a linear function of t. The values of $A_1(\tau_2)$, $B_1(\tau_2)$ and $\bar{B}_1(\tau_2)$ are estimated for any equilibruim state using equations:

$$A_1(\tau_2) = \frac{2(J_1 + \kappa\tau_1^2 + \kappa b_1^* \tau_2)}{\tau_2(J_1 + \kappa l_t^{*2})} \qquad (22)$$

$$B_1(\tau_2) = \frac{2b_1^*(2J_1 - \kappa a_1^{*2})}{\tau_2(J_1 + \kappa l_t^{*2})\, a_1^{*2}} \qquad (23)$$

$$\bar{B}_1(\tau_2) = -\frac{2\kappa b_1^*}{\tau_2(J_1 + \kappa l_t^{*2})} \qquad (24)$$

$$l_t^{*2} = \tau_1^2 + 2b_1^*\tau_2 + \tau_2^2 \qquad (25)$$

$$b_1^* = \tau_1 \cos\alpha_1^* \qquad (26)$$

and a_1^* is estimated using one of Eqs.(10).

Eqs.(20) and (21) can be classified that is they are linear differential equations of Fuks class[4] with four singular points. The integration og these is a problem the solution of which is not yet available. Let usconsider the characteristics of Eqs.(20) and (21) to compare them to ones of nonlinear system solutions (5) near the equilibruim states. With theorems used in[5] one may ascertain that the solutions of Eqs.(20) and (21) cannot be oscillatory if correspond to equilibruim points $\alpha_{12}^{*\prime}$, α_{13}^{*} , α_{11}^{*} and are oscillatory in a quite broad interval of argument variation if correspond to equilibruim points α_{12}^{*} , $\alpha_{13}^{*\prime}$, $\alpha_{11}^{*\prime}$. Near the coefficients of a singular point =0 the zeros of oscillatory solutions of Eqs.(20) and (21) have no accumulation points and the derivative of a general solution goes to infinity as it can be found by integration of Eqs.(20) and (21) near the singular point on the basis of Frobenius method. One can see for instance work [5] or [6] . By simulation of the nonlinear system (5) it can be ascertained that properties of system solutions for spacecraft-line relative motion are in a good agreement with the characteristics of linear Eqs.(20) and (21) established above.

Using Sonin-Polia theorem[5] it can be proved that the sequence of extremuma for the oscillatory solutions of Eqs.(20), (21) is decreasing on an interval (τ_{20} , $\tau_{21}^{\varphi_1}$) of τ_2 variation and is increasing on an interval ($\tau_{21}^{\varphi_1}$, 0). When conditions (19) are fulfiled the value of $\tau_{21}^{\varphi_1}$ is positive and defined by:

$$\tau_{21}^{\varphi_1} = \sqrt{\frac{J_1 + K\tau_1^2}{K}} \tag{27}$$

The exact solution I and II of Eqs. (20) and (5) describing spacecraft-line relative motion near the equilibruim point α_{12}^{*} are shown in Fig.4. It can be seen that the solutions of linear Eqs.(20) possess the properties of an associated solution o a nonlinear sxstem (5) near α_{12}^{*} and are in a satisfactory quantity agreement with it. Also, it may be observed that while returning the solution deviation from the associated equilibruim state is substantial only quite near the point τ_2=0. Utilizing the above information the linear Eqs.(20) and (21) can be considered acceptable to describe spacecraft-linm relative motion.

AN ASYMPTOTIC REPRESENTATION FOR THE SOLUTION OF THE EQUATION OF SPACECRAFT-TETHERLINE RELATIVE MOTION

Accepting the authority of[6], obtain solutions of Eqs.(20) and (21) in an asymptotic representation. To bring Eqs.(20) and (21) into accord with desired form introduce a large parameter T by changing a scale of an independent variable in the following way:

$$\tau = \frac{\tau_2}{T} \tag{28}$$

For the sake of simplicity let the formula

$$\mathcal{L}_1 = \exp\left(-\frac{1}{2}\int A_1(\tau_2)\,d\tau_2\right)\beta \tag{29}$$

be substituted for an unknown function thus letting Eqs.(20) and (21) be reduced to a canonical form as:

$$\beta'' + I(\tau_2)\,\beta = 0 \tag{30}$$

where $\bar{I}(\tau_2)$ takes values of

$$I_1(\tau_2) = B_1(\tau_2) - \frac{1}{4}A_1^2(\tau_2) - \frac{1}{2}A_1'(\tau_2) \tag{31}$$

or

$$\bar{I}_1(\tau_2) = \bar{B}_1(\tau_2) - \frac{1}{4}A_1^2(\tau_2) - \frac{1}{2}A_1'(\tau_2) \tag{32}$$

according to coefficients of Eqs.(20) and (21). Linear independent solutions β_1 and β_2 are searched as series in terms of negative powers of the large parameter. At some unimportant limitations (for our case) the series are asymptotical and defined by:

$$\beta_1 = e^{\int \lambda_1(\tau) T d\tau}\left[C_0(\tau) + C_1(\tau)\frac{1}{T} + \dots C_K(\tau)\frac{1}{T^K} + \dots\right]$$

$$\beta_2 = e^{\int \lambda_2(\tau) T d\tau}\left[\bar{C}_0(\tau) + \bar{C}_1(\tau)\frac{1}{T} + \dots \bar{C}_K(\tau)\frac{1}{T^K} + \dots\right] \tag{33}$$

where $C_K(\tau)$, $\bar{C}_K(\tau)$ are unknown functions derived from the following recurrent relationships:

$$\begin{aligned} &2\lambda_1\dot{C}_0 + \dot{\lambda}_1 C_2 = 0 && 2\lambda_2\dot{\bar{C}}_0 + \dot{\lambda}_2\bar{C}_0 = 0 \\ &\ddot{C}_0 + 2\lambda_1\dot{C}_1 + \dot{\lambda}_1 C_1 = 0 \quad (34) && \ddot{\bar{C}}_0 + 2\lambda_2\dot{\bar{C}}_1 + \dot{\lambda}_2\bar{C}_1 = 0 \quad (35) \\ &- - - - - - - && - - - - - - \\ &\ddot{C}_K + 2\lambda_1\dot{C}_{K+1} + \dot{\lambda}_1 C_{K+1} = 0 && \ddot{\bar{C}}_K + 2\lambda_2\dot{\bar{C}}_{K+1} + \dot{\lambda}_2\bar{C}_{K+1} = 0 \end{aligned}$$

and

$$\lambda_{1,2}(\tau) = \pm i\sqrt{I(\tau)} \qquad (36)$$

With the possible exception of a close neighbourhood of τ_2 = 0 it may be established that for our case the following inequalities:

$$C_0(\tau) \gg \frac{C_1(\tau)}{T} \gg \frac{C_k(\tau)}{T^k}\dots;\quad \bar{C}_0(\tau) \gg \frac{\bar{C}_1(\tau)}{T} \gg \dots \frac{\bar{C}_k(\tau)}{T^k} \gg \dots \qquad (37)$$

Therefore an asymptotic representation for the solution of Eqs.(20) and (21) will be sufficient ,namely:

$$\alpha_1 = e^{-\frac{1}{2}\int A_1(\tau_2)d\tau_2}\, I^{-\frac{1}{4}}(\tau_2)\left[D_1 e^{i\int\sqrt{I(\tau_2)}d\tau_2} + D_2 e^{-i\int\sqrt{I(\tau_2)}d\tau_2}\right] \qquad (38)$$

It can be shown that properties of the solution of Eq.(38) are the same as those of (20) and (21) which have been ascertained earlier. Exactness of the solution has been estimated quantitatively as shown in Fig.5 where the solution in the asymptotic representation as well as its envelope are plotted with a continuous line whereas the exact solution of the linear equtaion with a dot line.

AN ADMISSIBLE REGION OF INITIAL CONDITIONS FOR SPACECRAFT-TETHERLINE RELATIVE MOTION

Using Eqs.(11), (12), (16) and (38) spacecraft-line relative motion is given as a known function of time, system parameters and initial conditions. Some restrictions put on an angle α_1 between the spacecraft and the linę, on an angular velocity $\dot{\varphi}_1$ of the spacecraft with respect to the inertial space, on a force F_2 or line tension on a velocity value V at the impact instant are given by the following inequalities:

$$|\alpha_1(t)| \le \alpha_m$$
$$|\dot{\varphi}_1(t)| \le \dot{\varphi}_m$$
$$|F_2(t)| = F_2(t) \le F_m$$
$$\sqrt{[\tau_2(T)\dot{\alpha}_1(T)]^2 + \dot{\tau}_{20}^2} \le V_m \qquad (39)$$

where T being an instant of mooring does not coincide with τ_2 = 0 and is chosen from design considerations. The inequalities (39) define a certain region with a bound S in space of initial conditions and system parameters. In is called the region of admissible conditions or system parameters.

In this study a projection of the region bounds σ on a plane of initial astronaut's distances - astronaut's tangential velocities with coordinates r_{20}, $V_{\tau} = r_{20}\dot{\alpha}_{13}$ is determined. The study results shown in Fig.6 and Fig.7 are obtained for the following system parameters and accepted constrains:

$$J_1 = 700\ \text{кGm}\,\text{sec}^2 \quad K = 20\ \frac{\text{кG}\,\text{sec}^2}{m} \quad \varphi_1 = \varphi_2 \quad \alpha_m = \frac{\pi}{2}$$

at the fixed values of $\bar{r}_m$, $\dot{\varphi}_m$ and V_m.

Fig.6 shows:

I - between these curves the first of inequalities (39) is held at $\dot{\varphi}_{10} = 0$, $\dot{\varphi}_{20} \neq 0$.

II - above this curve the first of inequalities (39) is held at $\dot{\varphi}_{10} = \dot{\varphi}_{20}$.

$\dot{\varphi}_m$, $\bar{r}_m$ - below these curves the second and the third of inequalities (39) are held at given constraints.

In Fig.6 it is a shadowed region where the first three of inequalities (39) are satisfied.

Fig.7 shows:

$P_{10} < P_{10}^{min}$ - a region where angular momentum is less than P_{10}^{min}

$V_m^{(1)}$, $V_m^{(1,5)}$, $V_m^{(2)}$, $\bar{V}_m^{(1)}$, $\bar{V}_m^{(1,5)}$, $\bar{V}_m^{(2)}$

to the of these curves the last inequalities (39) is held; the first three curves are plotted for $\dot{\varphi}_{10} = 0$, $\dot{\varphi}_{20} \neq 0$ and the next three for $\dot{\varphi}_{10} = \dot{\varphi}_{20}$.

From Fig.6 and 7 it can be seen that for each initial distance an astronaut's tangential velocity has to be bounded from above and below if a successful retrieval is assumed. Thus the use of an uncontrolled tetherline system is unlikly to be acceptable.

AN APPROACH TO SYNTHESIS OF CONTROL SYSTEM FOR SPACECRAFT-TETHERLINE RELATIVE MOTION

In this study a possible control system is considered to be an attitude system to control rotational motion of the spacecraft with the use of reaction jets with bang-bang characteristics. In the case of the accepted operational system and at con conditions (19) it can be shown that spacecraft-line relative conditions (19) it can be snown that spacecraft-line relative motion is described by:

$$\ddot{\alpha}_1 + \frac{J_1 + \kappa(r_1^2 + b_1 r_2)}{J_1 + \kappa \bar{l}_+^2}\left[\frac{2\dot{r}_{20}}{r_2}\dot{\alpha}_1 + \frac{\kappa a_1(b_1 + r_2)}{J_1 + \kappa a_1^2}\dot{\alpha}_1^2\right] + \frac{[P_1 + \kappa a_1 \dot{r}_{20}]}{r_2(J_1 + \kappa a_1^2)(J_1 + \kappa \bar{l}_+^2)} \cdot$$

$$\cdot [P_1 a_1 - \kappa a_1^2 \dot{r}_{20} - 2\dot{r}_{20} J_1] = \frac{M \operatorname{sign} \delta_1 (b_1 + r_2)}{r_2(J_1 + \kappa a_1^2)}; \quad P_1 = P_{10} + \int_0^t M \operatorname{sign} \delta_1 \, dt \qquad (40)$$

where $\bar{l}_+^2 = r_1^2 + 2 b_1 r_2 + r_2^2$, $M \operatorname{sign} \delta_1$ - a control moment applied to the spacecraft, and δ_1 - a control signal.

When Eqs.(40) are integrated an angular velocity and a rotational angle of the spasecraft are defined with respect to the inertial space, namely

$$\dot{\varphi}_1 = \frac{P_{10} + \kappa a_1 \dot{r}_{20} - \kappa \dot{\alpha}_1 r_2 (b_1 + r_2) + \int_0^t M \operatorname{sign} \delta_1 \, dt}{J_1 + \kappa \bar{l}_+^2}; \quad \varphi_1 = \int \dot{\varphi}_1(t)\, dt + C \qquad (41)$$

A parameter value $M \operatorname{sign} \delta_1$(from Eq.(40} is changing by jumps when a certain switching surface $\delta_1(\alpha_1, \dot{\alpha}_1) = 0$ (42) is transfered by an image point in the phase space.

The systems described by such equations and called the variable-structure systems have a number of interesting peculiarities in which the component structures[7] described in this case by nonlinear, nonautonomous equations depending on history of motion are lacking in general. The most important peculiarity is as follows: at some given conditions stable motion of the system along the switching surface in so-called "slippy" regime is possible. Then the system dynamical characteristics are defined in general by a form of the switching surface and weakly depend on the plant parameters. For our problem the provision of controlled motion of such parameters as angular momentum and a line length give us the possibility of getting a system which works in a great range of initial conditions without excessive angular velocities at the end of the retrieval and hence without excessive impact velocities. With all these states it is possible to formulate a synthesis problem for control of spacecraft-tetherline relative motion as a problem of a switching surface choice so that:

1. The surface (42) would have such a "slippy" region $\mathcal{S}$ that while moving in this region the system come to and remain in given neighbourhood of a point $\alpha_1 = \dot{\alpha}_1 = 0$ in the phase space.

2. Phase trajectories of the system would come into the re-

gion Ω at arbitrary initial conditions.

3. The surface Ω would correspond to some quality conditions.

In this work a specific case is considered when switching surfaces are chosen as straight lines:

$$K_{\dot\alpha}\dot\alpha_1 + K_\alpha \alpha = 0 \tag{43}$$

where an angle value α is defined by

$$\alpha = \left.\begin{matrix} \alpha_1 - \alpha^0 \\ 0 \\ \alpha_1 + \alpha^0 \end{matrix}\right\} \begin{matrix} \alpha_1 > \alpha^0 \\ -\alpha^0 \leq \alpha_1 \leq \alpha^0 \\ \alpha_1 < -\alpha^0 \end{matrix} \tag{44}$$

The reasonable choice of gains K_α, $K_{\dot\alpha}$, α^0, M succeeds in providing controlled relative motion weakly dependent of initial conditions and of the line length thus eliminating wrap of the line around the spacecraft as well as excessive impact velocities.

Fig.8 shows how an angle α_1 between the spacecraft and the line is varying when system parameters are of values:

$K_\alpha = 1 \quad K_{\dot\alpha} = 2\,\text{sec} \quad \tau_{20} = 25\,\text{m} \quad M = 10\,\text{кгm} \quad \alpha^0 = 0{,}5^\circ \quad \tau_2(T) = 0{,}5\,\text{m}$

The comparison of controlled and uncontrolled motion characteristics permits us to conclude that the considered approach to synthesis of a control system for spacecraft-line relative motion is fruitful.

REFERENCES

1. CR Poli and EP Hanavan. A Three-Mass Retrieval Study for the Gemini. Tethered Astronaut, The Journal of the Astronautical Sciences, vol.XIII, No.2, March-April 1966.

2. CR Poli. A Study of Retrieval for Tethered Astronauts. SEG-TDR-65-30. Syst. Egn. Group, Research and Technology Division, Wright-Patterson [illegible] Force Base, Ohio.

3. A.I. Lurie. Analytical Mechanics. [A.I. Lurie. Analytycheskaya Mekhanika.] ~~In Russian~~. Fizmathgiz. Moscow, 1961. In Russian.

4. V.V. Golubev. Lectures on Analytical Theory of Differential Equations. [V.V. Golubev. Lektsü po Analyticheskoi Teorii Differentsialnykh Uravnenii]. GITTL, 1950. In Russian.

5. F. Tricomi. Differential Equations. [F. Tricomi. Differentsialnye Uravnenya]. IIL, Moscow, 1962. Russian Translation.

6. E.A. Koddington and N. Levinson. Theory of Ordinary Differential Equations. [E.A. Koddington i N. Levinson. Teorya Obyknowvennikh Differentsialnykh Uravnenii]. IIL, Moscow, 1958. Russian Translation.

7. S.V. Emelyanov. Automatic Control Systems with Variable Structure. [S.V. Emelyanov. Systemy Avtomaticheskogo Upravleniya s Peremenoi Strukturoi]. Izdatelstvo "Nauka", Moscow, 1967. In Russian.

ABSTRACT

DYNAMICS OF THE TETHERED ASTRONAUT MOVING TOWARD THE SPACECRAFT AND AN APPROACH TO SYNTESIS OF SPACECRAFT BASED ON A THEORY OF THE VARIABLE-STRUCTURE SYSTEMS.

Soshnikov V.N., Ulanov G.M.
Moscow
USSR

In this paper analysis aspects of an uncontrolled system are considered, the necessity of control to provide a retrieval of given quality is proved and an approach to the development of a control system based on properties of the variable-structure systems is suggested.

For an accepted model of the system the possible equilibruim states are studied, their dependence on system parameters and initial conditions is considered, bifurcation values of the parameters are determined.

A linearized equation representing a linear differential equation with varying coefficients of Fuks class is derived near equilibruim states of practical interest. Properties of the derived equation are investigated with the use of asymptotic representations near singular points and of Sonin-Polya theorem qualitatively on the whole of time interval. By digital computer simulation the properties coincidence for linear, nonlinear equations as well as asymptotic formulae is established.

In this paper a control system for the retrieval is defined by an equation which permits to consider the sustem as one of the variable-structure systems (VSS). A possible approach to synthesis of such system is shown.

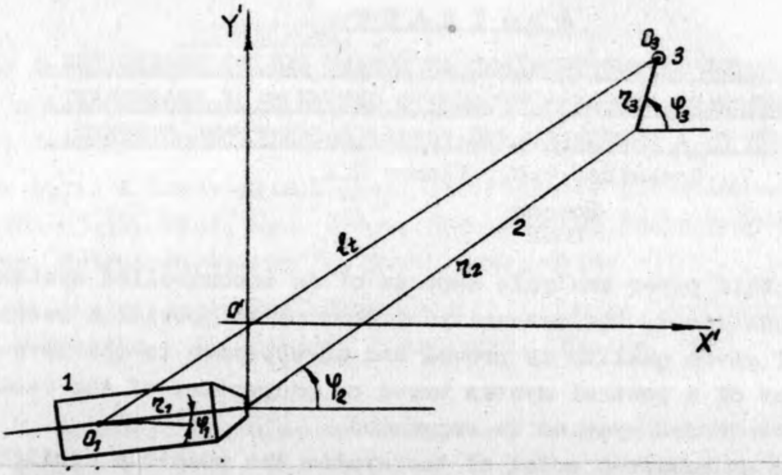


Fig.1. System geometry and generalized coordinates.

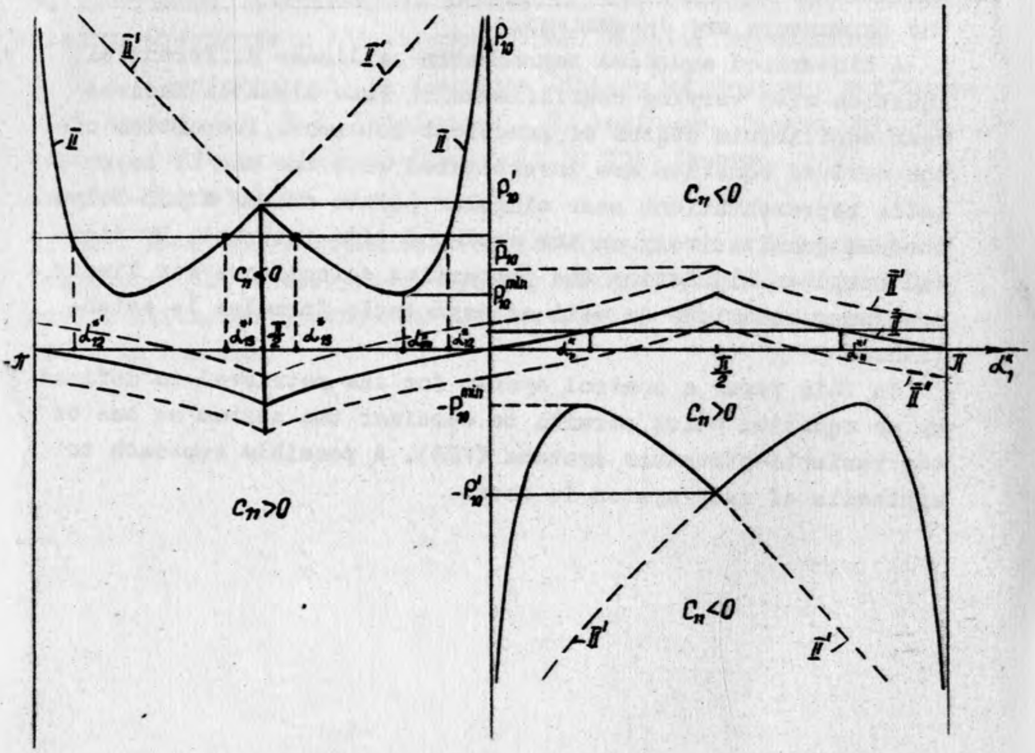


Fig.2. Bifurcation diagram.

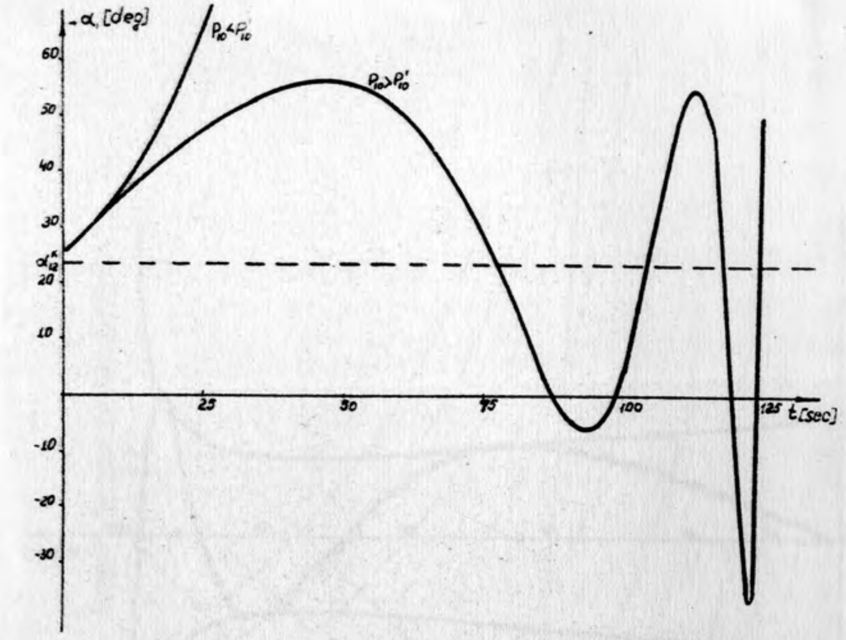


Fig.3. Major types of spacecraft-tetherline relative motions.

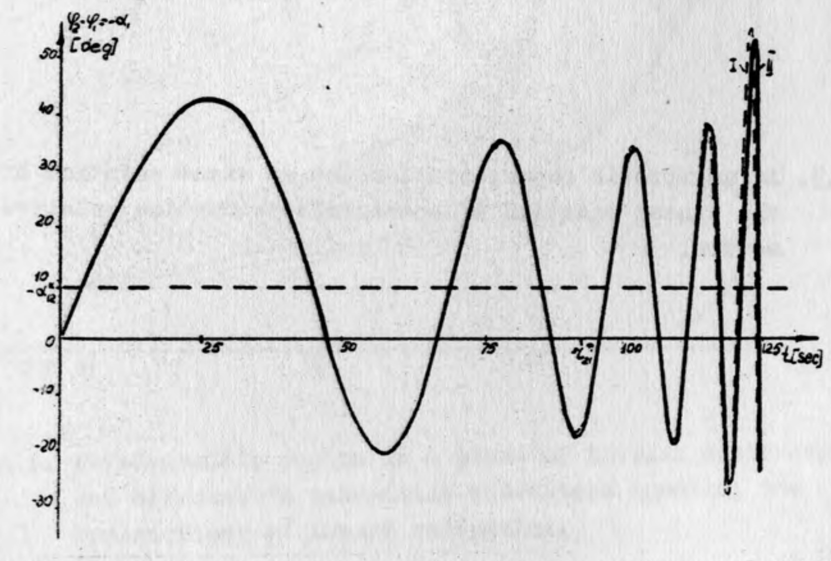


Fig.4. Oscillatory solutions of nonlinear systems and approximate linear systems for spacecraft-tetherline relative motion.

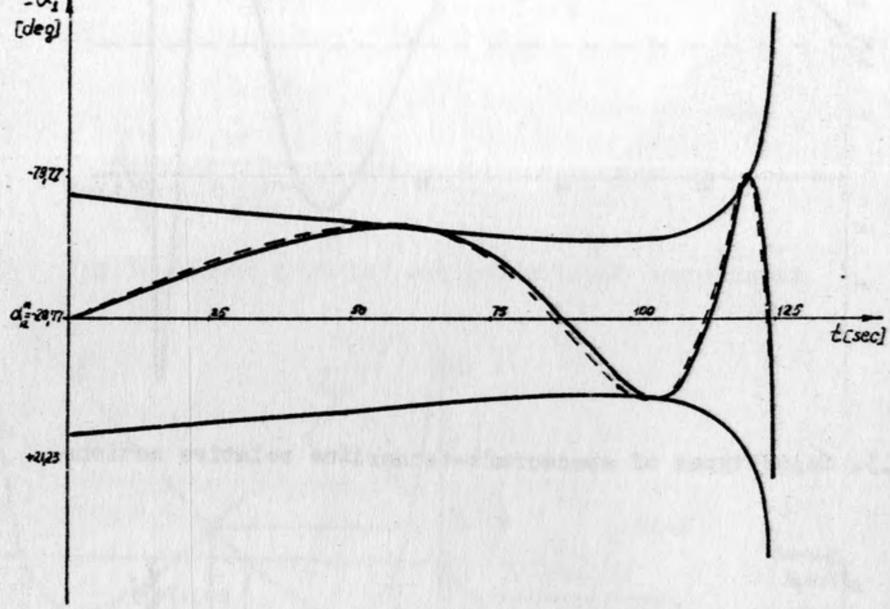


Fig.5. An asymptotic representation and an exact solution of the linear equation of spacecraft-tetherline relative motion.

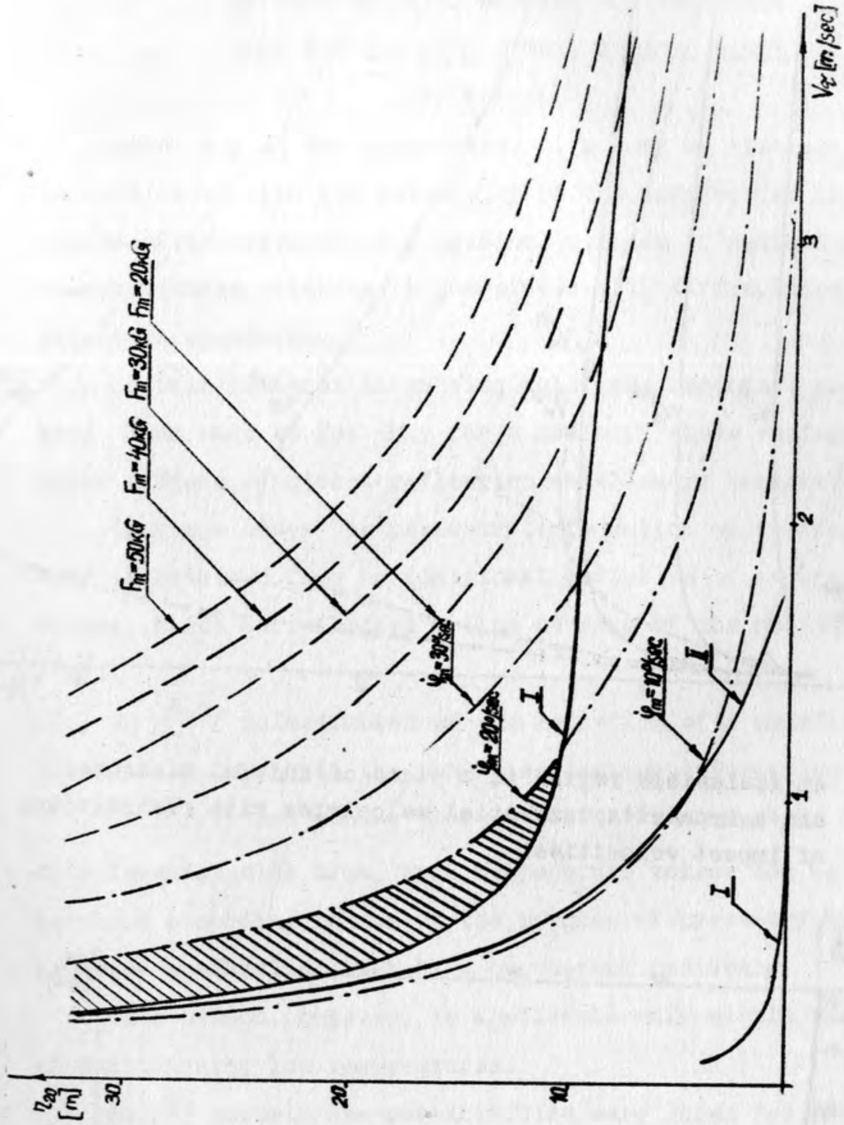


Fig.6. An admissible region in a plane of initial distances and astronaut's tangential velocities ignoring the restrictions of impact velocities.

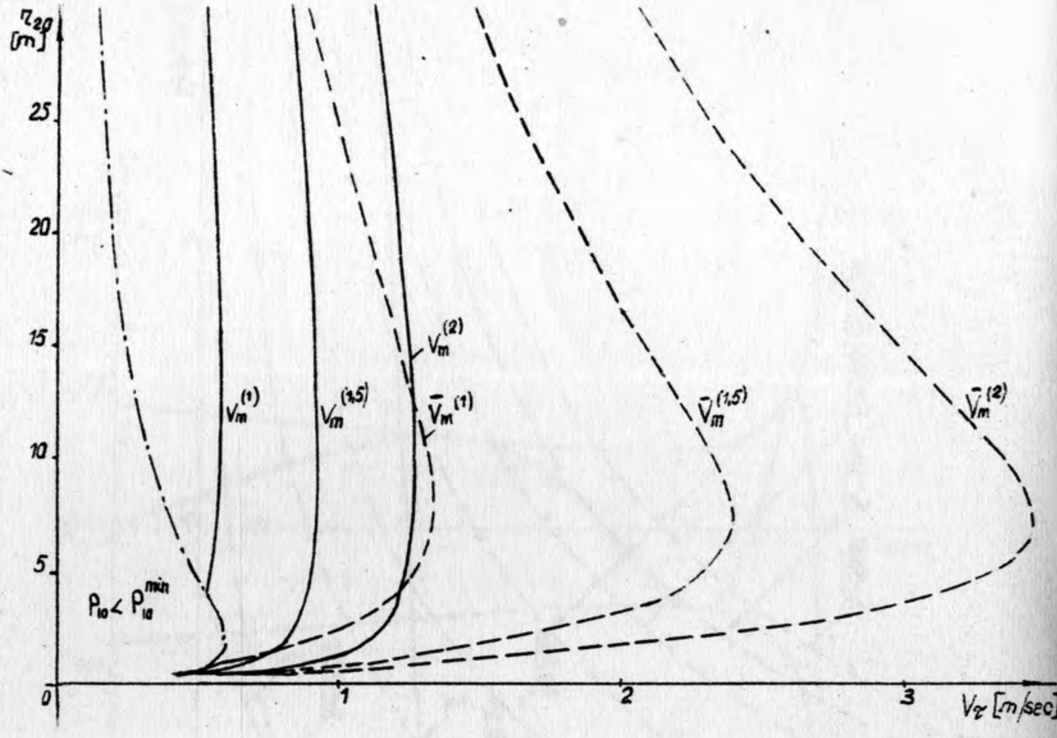


Fig.7. An admissible region in a plane of initial distances and astronaut's tangential velocities with restrictions of impact velocities.

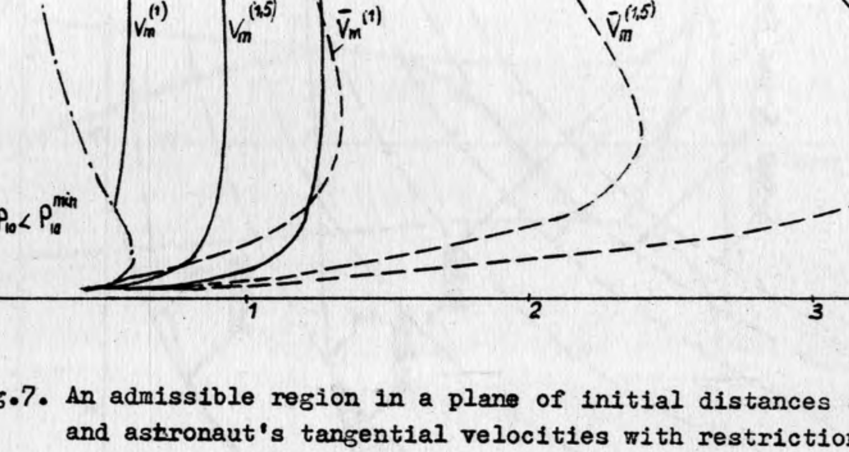


Fig.8. Spacecraft-tetherline relative motion for a particular type of a spacecraft control system

50.5

ON SELF-ADAPTIVE SYSTEMS FOR MEASURING REAL TEMPERATURES WITHIN OPTICAL RANGE

D.Ya.Svet

Measuring of the temperature of a body by its radiation in such cases when the emissivity of the body varies in the course of measurement is a cardinal problem of radiation pyrometry, whose solution is associated with difficulties of principal character.

Certain advances in solving this very important problem have been made so far only for a radiator whose surface features diffuse or direct reflection which obeys Lambert's law.

In these cases the necessary information on the emissivity is obtained from an additional radiation of an external source, which is reflected by the surface of the radiator /1,2/.

In /3,4/ polarization of the radiation of a metallic mirror is used to obtain the otherwise lacking information.

In /5/ it was shown, that within the limits where Drude's formula holds true, real temperature values can be determined according to one of the methods of pyrometry by the relative spectral distribution of thermal radiation.

This method, however, is applicable only within the range of sufficiently low temperatures.

In /6/ certain new possibilities were shown for measuring real temperatures under the conditions of varying emissivity, based on obtaining additional information derived directly from the flux of polychromatic radiation /7/ on the basis of a certain new form of Wien-Planck's law for the spectral concentration of the radiant emittance found by us /8/.

In these papers we have shown, that though real temperature and emissivity values cannot be determined directly by the values of intrinsic radiation, the opinion commonly shared by the specialists in optical pyrometry and astrophysical measurements as to the impossibility of evaluating the influence of emissivity on the results of measuring the intrinsic thermal radiation characterized by Wien-Planck's equation independently of the temperature, proves to be not always correct.

Let us consider a certain form of spectral distribution $W_i(\lambda_i,T)$ which is obtained from the spectral concentration of the radiant emittance of a full radiator $b_0(\lambda_i,T)$ within the limits where Wien's approximation holds true, if the values of each i-th spectral component of the isotherm at a temperature T are raised to the power with the exponent equal to its wavelength λ_i

$$W_0(\lambda_i,T) = b_0^{\lambda_i}(\lambda_i,T) = C_1^{\lambda_i}\lambda_i^{-5\lambda_i}\exp\left(-\frac{C_2}{T}\right)$$

Here constants C_1 = 37413.10^4 W/cm^{-2}.mu^4 and C_2 = 14388 mu °K.

An important feature of the obtained spectrum $W_0(\lambda_i,T)$ is, that its maximum does not shift with temperature. The wavelength at which the function $W_o(\lambda_i;T)$ has a maximum is λ_m = 1,905 mu. This results in that the relative spectral density distribution $R_o(\lambda_i,\lambda_j,T)=W_o(\lambda_i,T)/W_o(\lambda_j,T)$ is independent of temperature. For a full radiator

$$R_o(\lambda_i,\lambda_j,T)=C_1^{(\lambda_i-\lambda_j)}\left(\frac{\lambda_j^{\lambda_j}}{\lambda_i^{\lambda_i}}\right)^5$$

For any values of i and j

$$\partial R_o(\lambda_i,\lambda_j,T)/\partial T=0$$

For real bodies the distribution considered above will have the form

$$W(\lambda_i,T)=\alpha^{\lambda_i}(\lambda_i)\,b_o^{\lambda_i}(\lambda_i,T)=\alpha^{\lambda_i}(\lambda_i)\,W_o(\lambda_i,T)$$

where $\alpha=\varepsilon(\lambda)\tau(\lambda)$ is a certain function which characterizes the spectral distribution of the emissivity factor $\varepsilon(\lambda)$ of the radiator and of the transmission factor $\tau(\lambda)$ of the medium, including the elements of the optical system.

It is obvious, that the relation of any pair of components (i-th and j-th) of the distribution $R(\lambda_i,\lambda_j,T)$ will also be independent of the temperature x/

$$R(\lambda_i,\lambda_j)=\frac{\alpha^{\lambda_i}(\lambda_i)}{\alpha^{\lambda_j}(\lambda_j)}$$

The fact, that the equation $R(\lambda_i,\lambda_j)=R_o(\lambda_i,\lambda_j)$ holds true with $\tau(\lambda_i)=\tau(\lambda_j)=1$ is a specific criterion of the "absolute blackness" of the radiation $[\alpha(\lambda_i)=\alpha(\lambda_j)=1]$, except those cases when

$$\alpha^{\lambda_i}(\lambda_i)=\alpha^{\lambda_j}(\lambda_j),\ \text{but}\ \alpha(\lambda_i)\neq\alpha(\lambda_j)$$

x/ For $\partial\alpha(\lambda)/\partial T=0$

For three and more spectral components (i, j, K...) whose frequencies $\nu = C_0/\lambda$ satisfy the condition $\sum_i^n \nu_i = \sum_j^m \nu_j$ (it always being, that $i \neq j$) relative distribution can also be obtained, the value of which, within the limits where Wien's approximation holds true, is determined only by the parameter $\alpha(\lambda)$ and does not depend on temperature: $\angle(\lambda_i,\lambda_j)$.

These distributions will be generally described by the equations of the form[x/]

$$\angle(\lambda_i,\lambda_j)=\frac{\prod_i^n b(\lambda_i,T)}{\prod_j^m b(\lambda_j,T)}=\frac{C_1^P\prod_i^n\lambda_i^{-5}\alpha(\lambda_i)}{C_1^G\prod_j^m\lambda_j^{-5}\alpha(\lambda_j)}$$

where P and G are the amounts of spectral components in the numerator and denominator,respectively.

For grey radiation

$$\angle_0(\lambda_i,\lambda_j)=\frac{\prod_i^n\lambda_i^{-5}}{\prod_j^m\lambda_j^{-5}}C_1^{(P-G)}$$

In the particular case of three components whose frequencies satisfy the condition $\nu_1=\nu_2+\nu_3$ for a full radiator we can write

[x/] It is evident,that a third kind of relative distributions can also be created (satisfying the condition of being temperature independent) which are based on combined relationships of R and $\angle$ type.

$$L_0(\lambda_1,\lambda_2,\lambda_3)=\left(\frac{\lambda_1}{\lambda_2\lambda_3}\right)^5 C_1=(\lambda_2+\lambda_3)^{-5}C_1$$

For a real body this relative spectral distribution will have the form

$$L(\lambda_1,\lambda_2,\lambda_3)=\frac{\alpha(\lambda_2)\alpha(\lambda_3)}{\alpha(\lambda_1)}L_0(\lambda_1,\lambda_2,\lambda_3)$$

Similarly, for four components whose frequencies satisfy the condition $\nu_1+\nu_2=\nu_3+\nu_4$

$$L_0(\lambda_1,\dots,\lambda_4)=\left(\frac{\lambda_3\lambda_4}{\lambda_1\lambda_2}\right)^5;\ L(\lambda_1,\dots\lambda_4)=\frac{\alpha(\lambda_1)\alpha(\lambda_2)}{\alpha(\lambda_3)\alpha(\lambda_4)}L_0(\lambda_1,\dots\lambda_4)$$

The condition $L(\lambda_i,\lambda_j\dots)=L_0(\lambda_i,\lambda_j\dots)$ being satisfied for the case when the number of spectral components is the same in the numerator and denominator of the expressions for relative spectral distributions of L type is a criterion, that the radiation $\partial\alpha(\lambda)/\partial\lambda=0$ is grey.An exception should be made for such particular cases,when with $\partial\alpha(\lambda)/\partial\lambda\neq 0$ there holds the equality of products of the form

$$\alpha(\lambda_1)\alpha(\lambda_2)=\alpha(\lambda_3)\alpha(\lambda_4)$$

The method considered above for deriving the information on the character of $\alpha(\lambda)$ from the intrinsic radiation is effectively employed, particularly, in a pyrometer for measuring the real temperature of a radiator,the emissivity of whose surface varies in the course of measurements.

The system developed at the Institute of Metallurgy of the Academy of Sciences of the USSR for measuring real temperatures by radiation is self-adaptive.

From the engineering standpoint this system is realized in a pyrometer which is provided with a device that produces signals determined by the value of the relation of R_i or L_i type. Depending on the value of this signal which is obtained simultaneously with a signal indicative of the temperature that is determined by the relative spectral distribution of radiation, a certain correction value ΔT_i is automatically introduced into the pyrometer readings.

The accordance between this correction value and the signals determined by the values of the functions $R(\lambda_i,\lambda_j)$ or $L(\lambda_i,\dots \lambda_j)$ is attained by pre-adaptation of the system.

Practically this adaptation process consists in certain simple pre-scaling operations. Usually it proves sufficient to effect the latter for several extreme values of the function $\varepsilon(\lambda)$.

Thus, for example, for an automatic control of the real temperature of molten steel whose emissivity varies in the course of measurements due to the appearance or disappearance of oxide film, with the aggregate metering error [x/] not

[x/] The aggregate error should be understood here as the sum of errors due to the instrument and variation of spectral emissivity.

worse than ±1%, it is sufficient to effect the adaptation-scaling process for three points, The same adaptation cycle proved to be satisfactory for most various chemical compositions of oxide films, this being of great interest and practical importance.

This phenomenon is in good agreement with the results of investigations of the emissivity of oxides in the visible and near infrared regions of the spectrum,which are presented, for example, in /9/.

To explain the above-stated, let us consider the relation between the temperature correction in terms of the temperature reciprocal ΔT^{-1} and the distribution function $R(\lambda_i, \lambda_j)$. Let us denote by $\varepsilon_M(\lambda_i)$ and $\varepsilon_M(\lambda_j)$ the spectral emissivity values of a pure metal surface for the effective wavelengths λ_i and λ_j, and by $\varepsilon_f(\lambda_i)$ and $\varepsilon_f(\lambda_j)$ the corresponding values of spectral emissivity of the oxide film.

Due to the dispersion phenomenon, for metals $\varepsilon_M(\lambda_i) > \varepsilon_M(\lambda_j)$ if $\lambda_i < \lambda_j$.

It is known also, that due to dispersion,including the abnormal one, $\varepsilon_M(\lambda_i) < \varepsilon_f(\lambda_j)$.

Let M denote that portion of the metal surface which is free from the film, and N stand for that covered with the oxide film. For the "normalized" surface $M + N = 1$.

When measuring temperature by the relation between two emissivities with effective wavelengths λ_i and λ_j ,the correction value determined by spectral emissivity in terms of temperature reciprocals can be written as

$$\Delta T^{-1} = \frac{\Lambda}{C_2} \ln \frac{\varepsilon_M(\lambda_i) + N[\varepsilon_f(\lambda_i) - \varepsilon_M(\lambda_i)]}{\varepsilon_M(\lambda_j) + N[\varepsilon_f(\lambda_j) - \varepsilon_M(\lambda_j)]}$$

where $\Lambda = \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1}$ is the value of the equivalent wavelength.

For pure metal surface

$$\Delta T_1^{-1} = \frac{\Lambda}{C_2} \ln \frac{\varepsilon_M(\lambda_i)}{\varepsilon_M(\lambda_j)}.$$

Accordingly, for a fully oxidized surface $\Delta T_2^{-1} = 0$, since the film radiation is grey. For the case when a portion, say, one half of the metal mirror is coated with an oxide film ($M=N=0{,}5$).

$$\Delta T_3^{-1} = \frac{\Lambda}{C_2} \ln \frac{\varepsilon_M(\lambda_i) + \varepsilon_f(\lambda_i)}{\varepsilon_M(\lambda_j) + \varepsilon_f(\lambda_j)}$$

The values of the function $R(\lambda_i, \lambda_j)$, corresponding to these correction values will be:

$$\delta_{T_1} = K \frac{\varepsilon_M^{\lambda_i}(\lambda_i)}{\varepsilon_M^{\lambda_j}(\lambda_j)}, \quad \delta_{T_2} = K \varepsilon_f^{\lambda_i - \lambda_j} \quad \text{and} \quad \delta_{T_3} = K \frac{[\varepsilon_M(\lambda_i) + \varepsilon_f(\lambda_i)]^{\lambda_i}}{[\varepsilon_M(\lambda_j) + \varepsilon_f(\lambda_j)]^{\lambda_j}}$$

where K = const.

The non-linear relationship between the function $R(\lambda_i, \lambda_j)$ and the emissivity cannot cause ambiguity of the result, this possibility being automatically excluded due to the above-mentioned dispersion phenomenon which is characterized by the function $\varepsilon(\lambda)$ that decreases in the visible and near infrared regions of the spectrum for all metals.

This system, with a somewhat different adaptation algorithm, makes it possible to measure the values of real temperature of the surface featuring constant emissivity,but with a variable transmission factor of the medium.

Examples of such a problem in metallurgy are encountered when the peep hole is polluted with metal vapors in the course of vacuum melting, and for instance, in the sun temperature measurements, with the appearance of atmospheric haze, etc.

No less interesting results are obtained when implementing a self-adaptive system on the basis on the spectral density distribution of the type $L(\lambda_i, \dots, \lambda_j)$.

Introduction of redundancy by using more than two spectral components makes possible significant improvements in the stability and accuracy of measurements in the presence of various irregular outside effects.

The development of self-adaptive pyrometric systems on the basis of R and L distributions makes it possible to attain successful solution of certain problems associated with the plasma diagnosis, surface flaw detection, etc.

The key diagram of an automatic self-adaptive pyrometer is shown in Fig. 1.

Here S stands for the surface of a radiator with varying emissivity;

O- stands for the condenser and sighting optical system of the pyrometer;

N stands for a monochromatizing device;

P are radiation receivers;

A is an amplifying device;

P is a preliminary converter;

C.U. is a correlation unit;

L is a logometering device;

R is a measuring and recording output instrument.

The system operates as follows.

From the radiation emitted by S on the monochromatizing device M the radiation to be investigated is condensed with the help of O. The receiver or receivers F convert spectral radiant fluxes at the output of M into electric signals which are amplified by A and converted into the required form by the preliminary converter P.

From the output of the latter the converted signals are fed to the logometering device L and to the correction unit C.U. The latter, depending on the pre-adaptation,produces the required value of the correction signal. This signal can be applied either directly to the measuring and recording insturment R, or first to one of the stages of the logometering system. On the basis of this diagram different versions of self-adaptive pyrometric systems have been created.

For illustration,shown in Fig. 2 are comparative data of different pyrometric systems operating with steel melts in an induction furnace equipped with a special device which enables the formation on the metal surface of an oxide film of the required chemical composition, or the provision of a pure metal mirror in the atmosphere of argon.

In the process of melting the film could either partially "break" or entirely disappear. This is well illustrated by the indications of luminous, total and colour radiation pyrometers.

The pyrometric system for measuring real temperatures developed at the Institute of Metallurgy of the Academy of Sciences of the USSR in all cases showed the same class of accuracy as the thermocouple pyrometer, that is, ±15°.

Under the same conditions the indications of the colour pyrometer varied by 50-60°, of the luminous pyrometer by about 100°, and of the total radiation pyrometer, by 160-200°.

It should be pointed out, that the conditions of the melting process in this experimental installation from the point of view of operation of the real temperature pyrometer were considerably more severe than those observed in real steel melting practice.

The measurements have shown, that under the conditions of open melts the results of employing the real temperature pyrometric system are still better.

REFERENCES

1. W.G.Fastie - J.Opt. Soc. 1951, 4, 872.

2. Д.Я.Свет.Авт.свид.№I555I от 20 XI I952 г.

3. C.Tingwaldt, U.Shley - Z.Instrum., 1961, 69, H.7, 204.

4. W.Pepperhoff, Arch. Eisenhuttenws., 1959, 30, 131.

5. Д.Я.Свет. "Металлургия СССР за 40 лет", изд-во Металлургиздат, I958 г.

6. Д.Я.Свет."Объективные методы высокотемпературной пирометрии при непрерывном спектре излучения,"Наука",М.,I968 г.

7. Д.Я.Свет. Д.А.Н. т.I70, №4, стр.825, I966 г.

8. Д.Я.Свет. Теплофизика №3, I967 г.

9. Д.Я.Свет."Температурное излучение металлов и некоторых веществ", I964 г., "Металлургиздат".

(D.Ya. Svet, Thermal Radiation of metals, semiconductors, ceramics, partly transparent bodies and films. Consultant Bureau, N.Y., 1965).

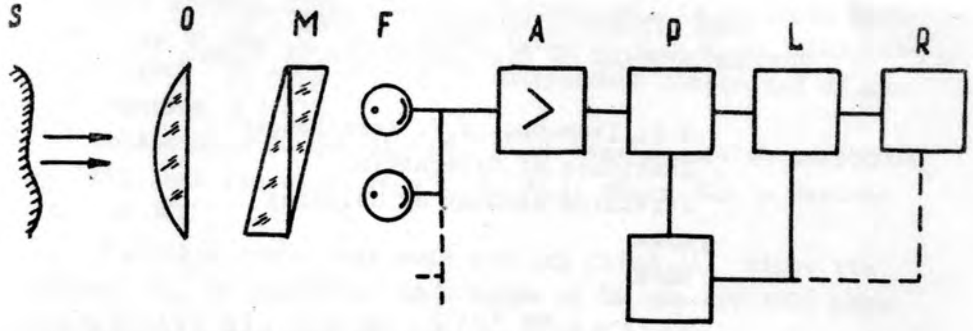

S
O
M
F
A
P
L
R
CИ

Fig. 1

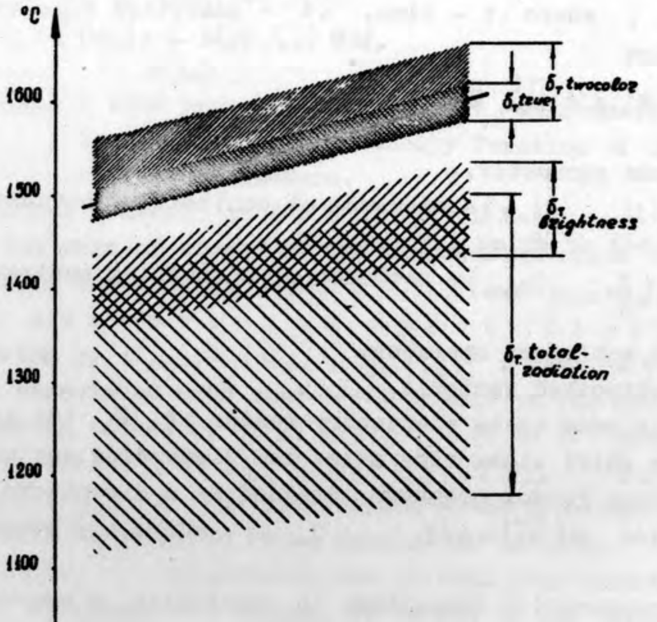

°C
1600
1500
1400
1300
1200
1100
1500
1550
1600
°C
δ_T twocolor
δ_T true
δ_T brightness
δ_T total-radiation

Fig. 2

53.1
SOME SYNTHESIS PROBLEMS OF ADAPTIVE CONTROL SYSTEMS OF STATIONARY RANDOM PLANTS [*]

V.I. Ivanenko, D.V. Karachenets
Institute of Cybernetics
Ukrainian Academy of Sciences
Kiev
USSR

1. The mathematical model of many controlled plants (CP), important in practice, can be represented in discrete time ($\kappa = \frac{t}{\Delta t}$, where t - time, Δt - quantized time interval) in the form

$$x_\kappa = G(\bar{U}^\ell_\kappa,\ \bar{Z}^m_\kappa,\ \theta), \qquad (1)$$

where:

θ - some parameter,

$\bar{U}^\ell_\kappa = (U_\kappa, \dots, U_{\kappa-\ell})$ - sequence of controlling actions magnitudes of which influence x_κ,

$\bar{Z}^m_\kappa = (Z_\kappa, \dots, Z_{\kappa-m})$ - sequence of disturbing actions influencing x_κ,

G - CP nonlinear operator,

x_κ- controlled variable.

CP is said to be stationary random if G (1) does not depend on shift along time axis, and disturbing action $\{Z\}$ is a stationary random sequence defined by n-dimensional density function $p(\bar{Z}^n_\kappa) = p(Z_\kappa, \dots, Z_{\kappa-n})$, where n - correlation interval.

Expression (1) describes, in particular, a number of industry processing plants under conditions of normal operation.[2,3] In such cases time of normal operation $N = \frac{T}{\Delta t} \gg max(\ell, m, n)$, and N may be assumed equalling infinity.

Consider the problem of determining an algorithm of controller (C) in closed automatic control system (CS) shown in Fig. 1.

It is supposed that C lags one time step, i.e., C selects U_κ value on the base of data obtained by measuring all measurable variables up to $(\kappa-1)$ -th time moment inclusive.

[*] Direct generalization of.

In a general case, control variable x_κ is measured with error h_κ so that C can observe variable $y_\kappa = x_\kappa + h_\kappa$ only. Sequence $\{h\}$ is supposed to be independent determined by density function $p(h_\kappa)$.

To estimate CS quality at κ-th time moment we introduce specific loss function $W_\kappa = W(x_\kappa, x^*_\kappa)$, where x^*_κ - desired value of x_κ.

W_κ is a random variable for any fixed $\bar{u}^\ell_\kappa$ since its argument x_κ is dependent on a value of CP unmeasurable parameters $\zeta_\kappa = (\bar{z}^m_\kappa, \theta)$, i.e. $W_\kappa = W(\bar{u}^\ell_\kappa, \bar{z}^m_\kappa, \theta, x^*_\kappa)$.

Therefore we estimate CS quality at κ-th moment by mean specific losses or specific risk:

$$R_\kappa = E\{W_\kappa\} = \int\limits_{\omega(\zeta_\kappa)} W_\kappa p(\zeta_\kappa)\, d\omega, \tag{2}$$

where: $p(\zeta_\kappa)$ - some probabilistic measure, unknown so far, in the form of the density function of CP unmeasurable parameters.

Introduce a notion of C observation space $\mathcal{U}$. In the case of closed CS the observation space is enclosed within coordinates $u_{\kappa-1}, u_{\kappa-2}, \ldots$ and $y_{\kappa-1}, y_{\kappa-2}, \ldots$. Introduce observation space dimension i denoting by $\alpha_{\kappa-i+\nu} = (u_{\kappa-i+\nu}, y_{\kappa-i+\nu})$ a pair of measured variables associated with some ν-th moment in the past, i.e., ν can assume values 0, 1, 2, ..., $i-1$. In $\mathcal{U}$ this determines vector $\bar{\alpha}^i_{\kappa-1} = (\bar{u}^i_{\kappa-1}, \bar{y}^i_{\kappa-1})$, $\bar{u}^i_{\kappa-1} = (u_{\kappa-1}, \ldots, u_{\kappa-i})$, $\bar{y}^i_{\kappa-1} = (y_{\kappa-1}, \ldots, y_{\kappa-i})$. The observation space constructed in this way will be referred to as initial or natural.

Let number i determining the initial observation space dimension, and being a C memory, be chosen by the designer. Then vector $\bar{\alpha}^i_{\kappa-1}$ will contain the whole information available for C to select u_κ control.

Consider a joint density function of unmeasurable parameters and of an observation sequence supposing that

$$p^i(\zeta_\kappa) = \int\limits_{\omega(\bar{\alpha}^i_{\kappa-1})} p(\zeta_\kappa, \alpha_{\kappa-1}, \alpha_{\kappa-2}, \ldots, \alpha_{\kappa-i})\, d\omega = p^i_\kappa \tag{3}$$

is the unconditional a posteriori density of unmeasurable parameters.

Determine $p(\zeta_\kappa)$ in (2) according to (3).

Substitute (3) in (2) and denote

$$R_\kappa^i = \int\limits_{\omega(\zeta_\kappa)} W_\kappa \cdot p^i(\zeta_\kappa)\, d\omega \,. \tag{4}$$

With an unbounded expansion of the observation space dimension we consider a specific risk limit. Using the Lebeg[7] theorem on transfer to the limit under the integral we have:

$$\lim_{i\to\infty} R_\kappa^i = \int\limits_{\omega(\zeta_\kappa)} W_\kappa \cdot \lim_{i\to\infty} p^i(\zeta_\kappa)\, d\omega \,. \tag{5}$$

Let the limit

$$\lim_{i\to\infty} p^i(\zeta_\kappa) = \lim_{i\to\infty} \int\limits_{\omega(\bar{\alpha}^i_{\kappa-1})} p\,(\zeta_\kappa, \bar{\alpha}^i_{\kappa-1})\, d\omega \tag{6}$$

exist and equal

$$p_\kappa^\infty = p^\infty(\zeta_\kappa) = \int\limits_{\omega(\alpha_{\kappa-1},\dots,\alpha_{\kappa-\infty})} p\,(\zeta_\kappa, \alpha_{\kappa-1}, \dots, \alpha_{\kappa-\infty})\, d\omega. \tag{7}$$

Let some distance between current unconditional a posteriori density p_κ^i and its limit value be termed $\pi_\kappa^i = \pi(p_\kappa^i, p_\kappa^\infty)$.

The data storing or adaptive process will be linked with vanishing distance π_κ^i for $\kappa \to \infty$. The specific risk corresponding to p_κ^∞ will be denoted by R_κ^∞ .

CS is said to be optimal under stationary working conditions corresponding to limit density function p_κ^∞ , if its quality estimation equals

$$\bar{\rho}^{\infty,0} = \inf_{\{U\}} \lim_{N\to\infty} \frac{1}{N} \sum_{\kappa=1}^{N} R_\kappa^\infty \,. \tag{8}$$

Some mathematical problems associated with (8) has been discussed in[4,5] .

Considering a change in probabilistic measure p_κ^i to be occuring at each time moment, i.e. $i \to \infty$ if $\kappa \to \infty$, we note that the data storing process conditions some transient in the system.

The CS quality in the transiton regime can be estimated by the functional

$$\tilde{\rho} = \sum_{\kappa=1}^{\infty'} (R_\kappa^i - \bar{\rho}^{\infty,0}), \quad i = \kappa. \tag{9}$$

The expression (9) can be used in certain cases for the synthesis of optimal adaptive systems, i.e. the systems with the optimal transient.

The problem of this type seems to be too complex for the present. For understanding the features of this problem it is necessary to consider particular cases, to investigate a problem for subclasses of stationary CP, to clear up the possibilities of an approximate solution.

2. Let us consider some of these questions conformably to CP, the mathematical model of which may be represented in the form (Fig. 2)

$$x_\kappa = \varphi\Big(\sum_{j=0}^{\ell} \alpha_j \cdot u_{\kappa-\ell+j},\, z_\kappa,\, \theta\Big). \tag{10}$$

The error h_κ of measuring the value x_κ is absent and $x_\kappa^* = \mathrm{const}$.

Suppose function φ in (10) to be monotonic with z_κ. We shall show that the following equality exists in this case

$$\lim_{\substack{\kappa\to\infty \\ i=\kappa}} \min_{u_\kappa} R_\kappa^i = \bar{\rho}^{\,\infty,0}. \tag{11}$$

This means that CS acted upon by control chosen by minimizing the specific risk with an unrestricted increase of a κ number in operation becomes optimal in a sense of (8) automatically, i.e. with no intervention from outside. Such a system may be related to adaptive systems.

We denote

$$\sum_{j=0}^{\ell} \alpha_j \cdot u_{r-\ell+j} = v_r. \tag{12}$$

The specific risk expression now will be written as

$$R_\kappa^i = \int_{\omega(\bar{a}_{\kappa-1}^i)} r_\kappa^i \cdot \rho(\bar{a}_{\kappa-1}^i)\, d\omega, \tag{13}$$

where r_κ^i - conditional specific risk equalling

$$r_\kappa^i = \int_{\omega(z_\kappa,\theta)} W(v_\kappa, z_\kappa, \theta) \cdot \rho(z_\kappa, \theta / \bar{a}_{\kappa-1}^i)\, d\omega. \tag{14}$$

The conditional a posteriori density of unmeasurable parameters can be represented as follows

$$\rho(z_\kappa, \theta / \bar{a}_{\kappa-1}^i) = \rho(\theta / \bar{a}_{\kappa-1}^i) \cdot \int_{\omega(\bar{z}_{\kappa-1}^n)} \rho(z_\kappa / \bar{z}_{\kappa-1}^n) \cdot \rho(\bar{z}_{\kappa-1}^n / \bar{a}_{\kappa-1}^i, \theta)\, d\omega. \tag{15}$$

We shall consider factors in right-hand part of (15). It is clear that conditional density function $\rho(z_\kappa / \bar{z}_{\kappa-1}^n)$ can be determined immediately by initial data containing joint density function $\rho(\bar{z}_\kappa^n)$.

To determine density function $\rho(\bar{z}^{n}_{k-1}/\bar{\alpha}^{i}_{k-1},\theta)$ we take the equation

$$x_{k-i+\nu}=\varphi(v_{k-i+\nu},z_{k-i+\nu},\theta),\ \nu=0,1,\dots,i-1. \tag{16}$$

Since function φ in (16) is monotonic relative to $z_{k-i+\nu}$ there corresponds to the pair of values $(x_{k-i+\nu},v_{k-i+\nu})=\alpha_{k-i+\nu}$ single value $z_{k-i+\nu}$ with constant θ . Vector $\bar{\alpha}^{n}_{k-1}$ is associated with single vector $\bar{z}^{n}_{k-1}$. Therefore we can state that it is enough to choose $i=n$, and

$$\rho(\bar{z}^{n}_{k-1}/\bar{\alpha}^{n}_{k-1},\theta)=\delta(\bar{z}^{n}_{k-1}-\bar{z}^{n}_{k-1}(\bar{\alpha}^{n}_{k-1},\theta)). \tag{17}$$

Finally, denoting $\rho(\theta/\bar{\alpha}^{i}_{k-1})=\hat{p}^{i}_{k}$ we shall discuss behaviour of the parameter a posteriori density with the k number increase. This is the very case with which the data storing process and the resulted transient in the system are connected. Functional (9) characterizing this process reflects the control duality in the A. Feldbaum's [6] sense. As was shown[6] there exist subclasses of the so called untransformed and transformed controlled plants. The transformed CP's are said to be such plants for which the choice of control does not influence the evolution of a posteriori density function of unmeasurable parameters. In this case problem (9) loses its sense.

A special criterion is necessary to classify CP's into untransformed and transformed.

The first step in this direction has been done in the mentioned work in which this classification was connected with the additive or non-additive occurence of an unknown parameter in the CP equation (10).

Another, more general, criterion of untransformability for CP of the form (10) can be suggested.

We write

$$\rho(\theta/\bar{\alpha}^{i}_{k-1})=\int_{\omega(\bar{\chi}^{i}_{k-1})}\rho(\theta/\bar{\alpha}^{i}_{k-1},\bar{\chi}^{i}_{k-1})\cdot\rho(\bar{\chi}^{i}_{k-1}/\bar{\alpha}^{i}_{k-1})\,d\omega \tag{18}$$

Introduce a one-to-one transformation of the observation space

$$\chi_{k-i+\nu}=\chi(\alpha_{k-i+\nu})=\chi(\varphi(v_{k-i+\nu},z_{k-i+\nu},\theta),v_{k-i+\nu}). \tag{19}$$

Now it can be shown that

$$\rho(\theta/\bar{\alpha}^{i}_{k-1})=\rho(\theta/\bar{\chi}^{i}_{k-1}) \tag{20}$$

where components of vector $\bar{\chi}^{i}_{\kappa-1}$ are found according to (19).

Now using (19) and (20) we can classify CP of the form (10) into untransformed and transformed with more confidence. Indeed, if there exists one-to-one transformation $\chi(\alpha)$ such that the CP equation may be permitted in the form [※)]

$$\chi_{\kappa-i+\nu} = \chi(x_{\kappa-i+\nu}, v_{\kappa-i+\nu}) = g(z_{\kappa-i+\nu}, \theta) \qquad (21)$$

then sequence $\{\chi\}$ is uncontrollable, and CP - transformed. It is impossible to influence the evolution of a posteriori density $\hat{p}^{i}_{\kappa}$ here because parameter θ is a parameter of some transformation g of uncontrollable random sequence $\{z\}$. The case when unknown parameter θ determines the disturbance density function is apparently a particular case of transformed CP. It follows here from the physical sense of the problem that a posteriori density function of parameter θ is not dependent on a choice of control. In compliance with these considerations the transformed CP may be represented as illustrated in Fig. 3.

In the case of the untransformed CP its equation cannot be represented in form (21) and we find that

$$\chi_{\kappa-i+\nu} = g(z_{\kappa-i+\nu}, \theta, v_{\kappa-i+\nu}). \qquad (22)$$

Now sequence $\{\chi\}$ proves to be controllable. Such CP's are conveniently depicted in the form of the block diagram shown in Fig. 4.

Approximate solution of (22) in form (21) can be utilized for estimation of "untransformability". In the case of a weak "untransformability" the effect of control influence on a posteriori density function $\hat{p}^{i}_{\kappa}$ may be ignored.
Employing (20) and (17) we substitute (15) to (14). Integrating over $\bar{z}^{n}_{\kappa-1}$ we have

$$r^{i}_{\kappa} = \int\limits_{\omega(z_{\kappa},\theta)} W_{\kappa} \cdot p(z_{\kappa}/\bar{z}^{n}_{\kappa-1}(\bar{\alpha}^{n}_{\kappa-1},\theta)) \cdot p(\theta/\bar{\chi}^{i}_{\kappa-1})\, d\omega \qquad (23)$$

Find limit r^{i}_{κ} for $i \to \infty$.

[※)] Determination of mathematical conditions of solvability (21) is an independent problem.

Using the Lebeg's theorem we obtain

$$r_k^{\infty} = \lim_{i\to\infty} r_k^i = \int\limits_{\omega(z_k,\theta)} W_k \cdot p\,(z_k/\bar{z}_{k-1}^{n}(\bar{u}_{k-1}^{n},\theta)) \cdot \lim_{i\to\infty} p(\theta/\bar{\chi}_{k-1}^{i})\,d\omega \qquad (24)$$

Supposing that control $\{v\}$ does not take on the values causing degeneracy of operator (1) we find $\hat{p}_k^i$ to be a martingal[8] and, consequently, there exists limit $\hat{p}_k^{\infty}$ equalling

$$\lim_{i\to\infty} \hat{p}_k^i = \delta\,(\theta - \theta^*) \qquad (25)$$

where θ^* – true value of unmeasurable parameter.

Substituting (25) in (24) and integrating over θ we calculate

$$r_k^{\infty} = r\,(v_k, \bar{u}_{k-1}^{n}, \theta^*) = \int\limits_{\omega(z_k)} W(v_k, z_k, \theta^*) \cdot p\,(z_k/\bar{z}_{k-1}^{n}(\bar{u}_{k-1}^{n},\theta^*))\,d\omega. \qquad (26)$$

Now we shall determine the specific risk. Evidently

$$R_k^{\infty} = \int\limits_{\omega(\bar{u}_{k-1}^{n})} r_k^{\infty} \cdot p(\bar{u}_{k-1}^{n})\,d\omega = R\,(v_k, \theta^*). \qquad (27)$$

Find the minimum of R_k^{∞} with respect to control v_k. As v_k in (27) is a parameter tnen, differentiating the integrand and taking into account that density function $p(\bar{u}_{k-1}^{n}) \neq 0$ we arrive at the conclusion that the minimum of R_k^{∞} with respect to v_k is determined by the minimum of r_k^{∞}.

Taking account of (12) we denote

$$\min_{u_k} r_k^{\infty} = r_k^{\infty,0}\,(\bar{u}_{k-1}^{\ell}, \bar{u}_{k-1}^{n}, \theta^*) \qquad (28)$$

It follows from (28) that $r_k^{\infty,0}$ depends on choice of $\bar{u}_{k-1}^{\ell}$. We shall show that

$$\min_{v_k} r_k^{\infty} = \min_{u_k} r_k^{\infty} \qquad (29)$$

if no restrictions are imposed on the choice of control action (or if restrictions are not reached), and loss function W_k is positive and differentiable with respect to x_k.

Setting derivative $\frac{\partial r_k^{\infty}}{\partial v_k}$ equal to zero we determine optimal value $v_k^0 = v(\bar{u}_{k-1}^{n}, \theta^*)$.

Taking into account (12) we have

$$v_k^0 = v\,(\bar{u}_{k-1}^{n}, \theta^*) = \alpha_0 \cdot u_k + \alpha_1 \cdot u_{k-1} + \ldots + \alpha_\ell \cdot u_{k-\ell}. \qquad (30)$$

Assume that $u_{k-1}, \ldots, u_{k-\ell}$ are chosen arbitrary. Then from (30) the optimal control u_k^0 can be found in the form $u_k^0 = u\,(\bar{u}_{k-1}^{\ell}, \bar{u}_{k-1}^{n}, \theta^*)$. Thus

$$\min_{v_\kappa} R_\kappa^\infty = \min_{u_\kappa} R_\kappa^\infty = R_\kappa^\infty (u_\kappa^\circ). \qquad (31)$$

In practice the unknown parameter data storing process (adaptation process) is understood to be converging to limit R_κ^∞ over a finite number of steps $N_1 \ll N$, followed by stationary working conditions of CS operation. Here control u_κ° is the optimal stationary control.

Ignoring losses in a transient which is caused by data storing over interval N_1 we can, omitting subscript κ , write that

$$\bar{\rho}^{\infty,0} = R^\infty (u^\circ). \qquad (32)$$

Now note that risks r_κ^i and R_κ^i converge their limit values uniformly with u_κ . This shows that

$$\min_{u_\kappa} \lim_{\substack{\kappa \to \infty \\ i = \kappa}} R_\kappa^i = \lim_{\substack{\kappa \to \infty \\ i = \kappa}} \min_{u_\kappa} R_\kappa^i . \qquad (33)$$

Proof (33) from the contrary is apparent. Equality (33) with account taken of (32) and (31) proves (11).

3. Consider a simple example.

The CP equation has the form $x_\kappa = u_\kappa \cdot \theta^* + z_\kappa$, where $\{z\}$ - independent sequence distributed with parameters m_z and σ_z according to Gaussian law. Choosing the loss function $W_\kappa = (\theta^* u_\kappa - 1)^2$, using the Bayes' formula for a posteriori density function of parameter θ and letting a priori distribution of θ be Gaussian, we minimize specific risk R_κ and find

$$u_\kappa^\circ = \frac{\frac{1}{\sigma_\theta^2} + \frac{1}{\sigma_z^2} \cdot \sum_{\nu=1}^{\kappa-1} u_\nu}{\frac{m_\theta}{\sigma_\theta^2} + \frac{1}{\sigma_\theta^2} \sum_{\nu=1}^{\kappa-1} u_\nu \cdot (x_\nu - m_z)} .$$

For $\kappa \to \infty$, majorizing the control by number $M < \infty$ we obtain

$$\lim_{\kappa \to \infty} u_\kappa^\circ = \frac{1}{\theta^* + \frac{\sum_{\nu=1}^{\kappa-1} u_\nu \cdot z_\nu^2}{\sum_{\nu=1}^{\kappa-1} u_\nu^2}} = \frac{1}{\theta^*} , \quad z_\nu^\circ = z_\nu - m_z$$

which is the expected result according to (25).

In many practically important cases disturbing sequence $\{Z\}$ is a result of quantization in time of random, continuous in a sense, function $Z(t)$. Using the method of transforming the observation space, described in[9], we can show that equality (11) holds approximately in the case when function φ in (10) is non-monotonic over Z_κ. Problems of such a type arise specifically in synthesizing systems of automatic optimization under conditions of a random wandering of the extremum[10]. Generalization of the result obtained in[9] for the case of unknown and unmeasurable parameter θ, provided function in (10) can be subjected to m-power series expansion over U_κ, Z_κ and θ, is simple but rather unwiedly. Therefore we shall explain the main idea of transformation by a particular example.

Suppose the CP equation has the form

$$x_\kappa = U_\kappa^2 + Z_\kappa^2 + \theta^2 + U_\kappa \cdot (Z_\kappa + \theta) + Z_\kappa \cdot \theta$$

Consider two adjacent moments r and $r-1$. Subtracting x_{r-1} from x_r after some transformations we obtain

$$\frac{x_r - x_{r-1}}{U_r - U_{r-1}} = (U_r + U_{r-1}) + Z_r + \theta + \left\{ \frac{(Z_r + Z_{r-1}) + U_{r-1} + \theta}{U_r - U_{r-1}} \right\} (Z_r - Z_{r-1}).$$

Denote $Z_r - Z_{r-1} = \Delta Z_r$; $\dfrac{x_r - x_{r-1}}{U_r - U_{r-1}} = \varkappa_r$;

and $\{\cdot\} = A$.

Assuming that owing to the continuity $E\{(\Delta Z_r)^2\} < \varepsilon_2$ if $\Delta t_1 < \varepsilon_1$, restricting difference $(U_r - U_{r-1})$ from bottom, and ignoring item $A_r \cdot \Delta Z_r$ we have

$$\varkappa_r \approx (U_r + U_{r-1}) + Z_r + \theta \quad \text{and} \quad \chi_r = \varkappa_r - (U_r + U_{r-1}) \approx Z_r + \theta .$$

Here $\varkappa_r$ is already monotonic function of Z_r and we can apply to it all inferences obtained above.

REFERENCES

1. В.И.Иваненко К синтезу замкнутых адаптивных систем управления стационарными объектами. Кибернетика, № 4, 1968 г.

2. В.И.Иваненко, Д.В.Караченец. Задачи статистического синтеза систем автоматической оптимизации массообмен-

ных установок. Труды IУ всесоюзного совещания по автоматическому управлению. Тбилиси, I968 г.

3. Л.С.Дубина, А.А.Снегур, В.М.Томашов. Субоптимальная адаптивная система управления реакционной колонной алкилирования. Труды симпозиума IФАК "Проблемы идентификации в системах регулирования", Прага, I967г.

4, А.Н.Ширяев Некоторые новые результаты в теории управляемых случайных процессов. **Transactions of the Fourth Prague Conf. on Information Theory, Decision Functions, Random Processes. Prague, 1967.**

5. D. Blackwell **Discounted Dynamic Programming, AMS, 36, 1 (1965).**

6. А.А.Фельдбаум Основы теории оптимальных автоматических систем. М., I966 г.

7. Г.Е.Шилов, Б.Л.Гуревич. Интеграл, мера и производная. М., I967 г.

8. Д.Л.Дуб Вероятностные процессы. М., I956 г.

9. В.И.Иваненко Метод преобразования пространства наблюдений при статистическом синтезе нелинейных замкнутых автоматических систем. Кибернетика, № 6, I967 г., № 2, I968 г.

I0. Л.Л.Вознюк, В.И.Иваненко. Анализ влияния шума на качество замкнутой системы автоматического управления. Кибернетика, № I, I969 г.

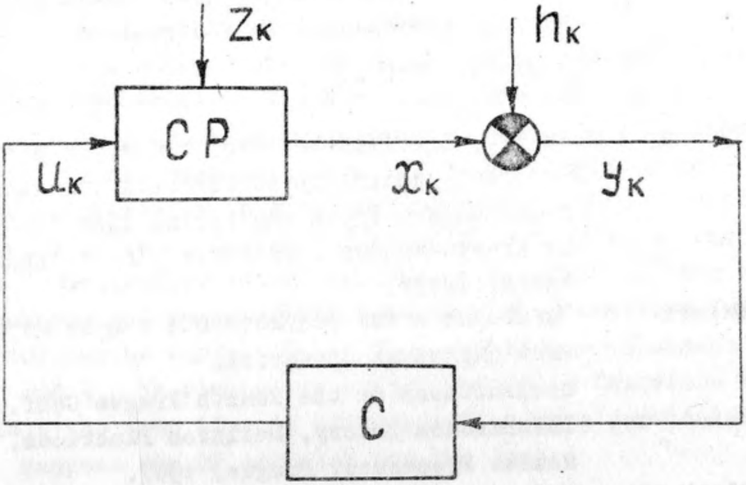


Figure 1

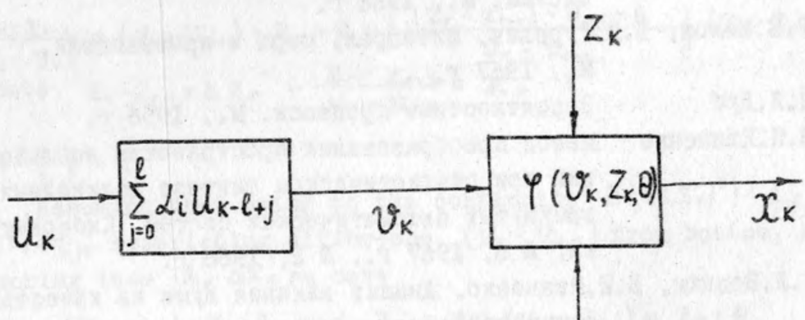

Z_K
U_K
$\sum_{j=0}^{\ell} \alpha_i U_{K-\ell+j}$
ϑ_K
$\varphi(\vartheta_K, Z_K, \theta)$
x_K
θ

Figure 2

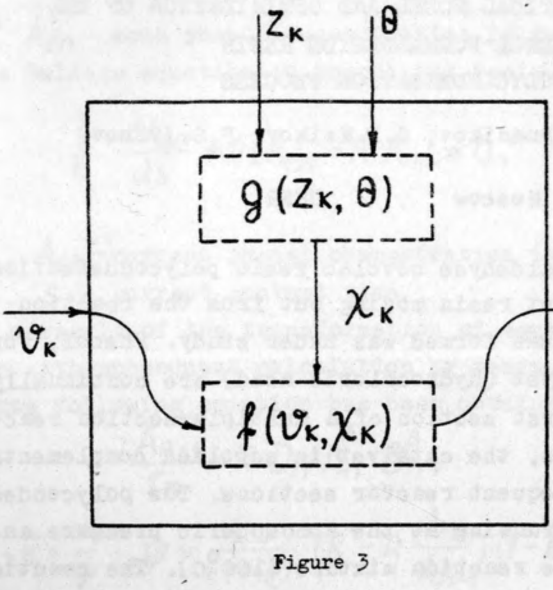


Figure 3

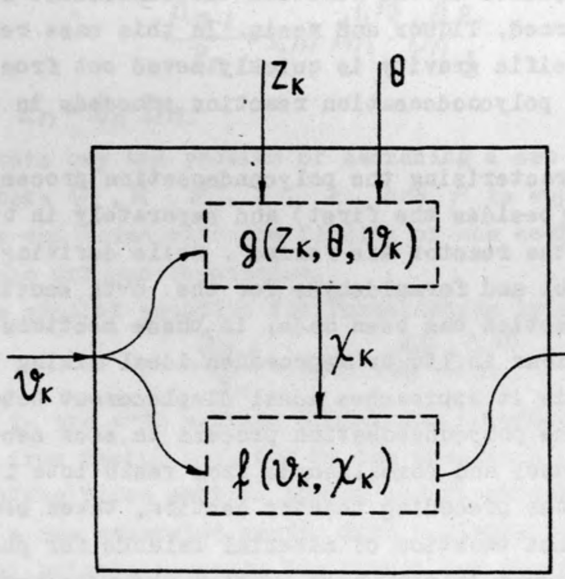


Figure 4

53.6
A MATHEMATICAL MODEL AND OPTIMIZATION OF THE PHENOL-FORMALDEHYDE RESIN POLYCONDENSATION PROCESS

E.G.Dudnikov, G.P.Maikov, P.S.Ivanov

Moscow USSR

The phenol-formaldehyde novolac resin polycondensation process based on fast resin moving out from the reaction zone as fast as it was formed was under study. Phenol, formaldehyde and catalyst (hydrochloric acid) are continually delivered to the first section of a multiple-section reactor (Fig.1). In addition, the catalyst is supplied complementary to each of the subsequent reactor sections. The polycondensation reaction is running at the atmospheric pressure and boiling point of the reaction mixture (100°C). The reaction mixture stirring is carried out at the expense of boiling only. In the consequence of the reaction two immiscible liquid phases are formed, liquor and resin. In this case resin having a large specific gravity is quickly moved out from the reaction zone. The polycondensation reaction proceeds in one phase (liquor).

Equations, characterizing the polycondensation process, in n-th section (besides the first) and separately in the first sections of the reactor are formed[1]. While deriving the equations for phenol and formaldehyde for the n-th section the following assumption has been made: in these sections hydrodynamic behaviour in liquor approaches ideal mixing behaviour and in resin it approaches ideal displacement behaviour. Along with the polycondensation process in each section an isolation of phenol and formaldehyde from resin into liquor, supplied by the preceding reactor section, takes place. With a glance to that equation of material balance for phenol the equation for liquor in the n-th section (besides the first) is of the form:

$$A_{n-1}(1-\xi)\cdot q - A_n(1-\xi)\cdot q + \eta(\bar{A}_{cn} - R\cdot A_n)\cdot V_n = \kappa\cdot D_n\cdot A_n^{\alpha}\cdot B_n^{\beta}\cdot(1-\rho)\cdot V_n, \tag{1}$$

where

$\bar{A}_{cn}$ - mean phenol concentration in resin.

The balance equation on phenol for resin is the following:

$$\frac{\xi}{\varepsilon}\cdot\frac{dA_{cn}}{ds}+\eta(A_{cn}-R\cdot A_n)=0, \qquad (2)$$

where

A_{cn} - current phenol concentration in resin,

S - current contact time.

As a result of the transformation of equations (1) and (2) and with subsequent calculation by means of digital computer the following equation has been obtained:

$$\frac{A_{n-1}}{\tau_n}=\bar{\kappa}\cdot D_n\cdot A_n^{\alpha}\cdot B_n^{\beta},$$

where

$$\tau_n=\frac{V_n}{q},\ \bar{\kappa}=\frac{\kappa}{\psi},\ \psi=\frac{\xi}{\varepsilon(1-\rho)}\cdot(R^*-R\frac{A_n}{A_{n-1}})\cdot[1-exp(-\eta\cdot\tau_n\cdot\frac{\varepsilon}{\xi})].$$

Aft substitution of coefficients obtained we have

$$\frac{A_{n-1}}{z_n}=1{,}14\cdot A_n^{1,16}\cdot B_n^{0,8}, \qquad (3)$$

where

$$z_n=\tau_n\cdot D_n.$$

In this way the problem of searching a set of the unknown parameters η , R , R^*, ξ , ε and ρ is substituted by the problem connected with the finding of one coefficient including these unknown quantities.

The similar equation for formaldehyde is of the form

$$\frac{B_{n-1}}{z_n}=1{,}69\cdot A_n^{0,4}\cdot B_n^{1,36}. \qquad (4)$$

If in the n-th section phenol and formaldehyde go into liquor from resin, supplied by the preceding ractor section, then in the first section phenol and formaldehyde are extracted by a new generated resin. With a glance to extraction the material balance equation for phenol for the first section has been formed. After the transformation and calculation of the unknown coefficients the equations for phenol are of the form:

$$\frac{A_0}{\tau_1}=\bar{\kappa}_1\cdot D_1\cdot A_1^{\alpha_1}\cdot B_1^{\beta_1},$$

where

$$\bar{K}_1 = \frac{K_1}{\psi_1}, \quad \psi_1 = 1 - \frac{A_1}{A_0}[1+(R^*-1)\cdot\xi].$$

After substitution of the obtained coefficients we have

$$\frac{A_0}{Z_1} = 41{,}7\cdot A_1^{0{,}7}\cdot B_1^{0{,}57}, \tag{5}$$

where

$$Z_1 = \tau_1\cdot D_1.$$

The similar equation for formaldehyde is of the form

$$\frac{B_0}{Z_1} = 5{,}8\cdot A_1^{0{,}42}\cdot B_1^{1{,}12}. \tag{6}$$

An important parameter, specific to novolac resin, is the degree of polycondensation which indirectly may be appraised by means of resin viscosity. The following relationship has been found

$$m_n = 3{,}02 + 1{,}18\cdot Z_n - 0{,}26\cdot A_n + 0{,}3\cdot B_n, \tag{7}$$

where

m_n - resin viscosity in the n-th section (viscosity of the 20%-solution of resin dried up in alcohol), cp.

Novolac resin viscosity is set within the range of 4-5 cp, and taking into account (7) we can write

$$0{,}98 \leqslant 1{,}18\cdot Z_n - 0{,}26\cdot A_n + 0{,}3\cdot B_n \leqslant 1{,}98. \tag{8}$$

The identity of equations (3-7) to experimental data has been checked by means of Fisher's criterion.

Equations obtained (3-7) enable the process of novolac resins to be optimized.

The sum of phenol and formaldehyde concentration in liquor at the outlet of the last reactor section $J = A_N + B_N$ has been taken for the optimality criterion. The optimization problem consisted in the criterion minimization. For the three-section reactor the following problem has been solved. On the basis of bond equations (3-6) and taking into account restrictions imposed on resin viscosity (8) it was necessary to find the mean time distribution of the residence and concentration of the catalyst over the sections, which minimizing the optimality criterion $J = A_3 + B_3$ under the desired initial phenol and formaldehyde concentration.

The feature of this problem is that at first the optimal distribution of Z is defined (product of the mean residence time and catalyst concentration), and then the optimal values τ and D are selected. In this case τ may be assumed to be constant for all sections, since it is more profitable to produce reactors with the same volume, when just the same value of the criterion optimality can be reached by the corresponding catalyst supply.

The stated problem has been solved by means of the dynamic programming method[2] . Let's reduce bond equations (3-6) to the form, necessary for solving the problem by means of the abovementioned method

$$A_n = 1,19 \cdot \frac{A_{n-1}^{1,08}}{Z_n^{0,34} \cdot B_{n-1}^{0,64}}, \tag{9}$$

$$B_n = 0,64 \cdot \frac{B_{n-1}^{0,92}}{Z_n^{0,6} \cdot A_{n-1}^{0,32}}, \tag{10}$$

$$A_1 = 0,003 \frac{A_0^{2,06}}{Z_1^{1,01} \cdot B_0^{1,05}}, \tag{11}$$

$$B_1 = 1,84 \frac{B_0^{1,29}}{Z_1^{0,52} \cdot A_0^{0,77}}. \tag{12}$$

On the basis of this equations, describing the system transfer from one state into another (from one section to another), the recurrent relations for the three-section reactor can be written:

For the third (the last one) section

$$f_1(A_2, B_2) = \min_{Z_3} J(A_3, B_3) =$$

$$= \min_{Z_3} \left(1,19 \cdot \frac{A_2^{1,08}}{Z_3^{0,34} \cdot B_2^{0,64}} + 0,64 \frac{B_2^{0,92}}{Z_3^{0,6} : A_2^{0,32}}\right).$$

For the second section

$$f_2(A_1,B_1)=\min_{z_2} f_1(A_2,B_2)=$$

$$=\min_{z_2} f_1\left(1{,}19\cdot\frac{A_1^{1{,}08}}{z_2^{0{,}34}\cdot B_1^{0{,}64}},\ 0{,}64\,\frac{B_1^{0{,}92}}{z_2^{0{,}6}\cdot A_1^{0{,}32}}\right).$$

For the first section of the reactor

$$f_3(A_0,B_0)=\min_{z_1} f_3(A_1,B_1)=$$

$$=\min_{z_1} f_2\left(0{,}003\,\frac{A_0^{2{,}06}}{z_1^{1{,}01}\cdot B_0^{1{,}05}},\ 1{,}84\,\frac{B_0^{1{,}29}}{z_1^{0{,}52}\cdot A_0^{0{,}77}}\right).$$

The calculation is beginning from the third (the last) section of the reactor with the subsequent using of the above recurrent relations. In the plane of variables A_2-B_2 a net is built. Different values z_3 are set and optimality criterion values for each node of this net are found, in this case a minimum of J and the appropriate z_3 are remembered. Unless a restriction imposed on viscosity for some node is met, this node is not considered. The calculation of the second section is similar, only in this case the minimal value of the optimality criterion for each node of the net in the plane A_1-B_1 is found, but now for the second and the first sections only. Having calculated the first section similarly, the net of the initial concentrations A_0-B_0 is obtained. Minimal values of the optimality criterion $J=A_3+B_3$ and an optimal sequence of z_1, z_2 and z_3 for each node of this net are found. The calculations were carried out by the digital computer "Ural-2". Values of the initial concentrations of phenol and formaldehyde lied within the limits:

$A_0 =$ 45 - 65 (weight %), $B_0 = 8 - 15$ (weight %).

The table lists the optimal sequence of Z_1 , Z_2 and Z_3 and the minimal value for A_0=55 (weight %) at the various initial concentrations B_0 .

Table

A_0	B_0	Z_1	Z_2	Z_3	J
55	8	1.4	1.4	1.6	1.15
"	10	1.2	1.2	1.5	1.29
"	11	1.2	1.2	1.2	1.84
"	12	1.0	1.1	1.2	2.00
"	15	0.8	0.8	1.1	2.59

SYMBOLS

A_0, B_0 - concentration of phenol and formaldehyde respectively in the starting mixture, weight %.

A, B, D - concentration of phenol, formaldehyde and catalyst in liquor, weight %.

q - total weight velocity of the mixture, kg/hour

ξ - coefficient characterizing a flow of liquid resin in a total flow.

ε - cross-sectional area of the liquid resin flow in the total cross sectional area of the reactor.

ρ - volume of the liquid resin in the total volume.

V - amount of the reaction mixture in one section, kg.

τ - average residence time, hour.

k - specific constant of the chemical reaction rate,

$$\frac{1}{\text{hour}\cdot(\text{weight }\%)^{\alpha+\beta}}$$

α, β - exponents for the concentrations of phenol and formaldehyde

η - effective constant of the phenol diffusion rate, weight % / hour.

R - equilibrium coefficient of the phenol distribution.

R^* - operating coefficient of the phenol distribution.

Index "n" means a number of a section for the appropriate symbol.

REFERENCES

I. Майков Г.П., Иванов П.С., Кабыкина Н.И.,Исследование процесса поликонденсации, сопровождающегося экстракцией, в многосекционном реакторе непрерывного действия. "Теоретические основы химической технологии", 2, № 2, I968.

2. Беллман Р. Процессы регулирования с адаптацией. Из-во "Наука", I968.

54.5

ALGORITHM OF OPTIMUM AIR EXCESS CONTROL IN STEAM BOILER FURNACES FIRED WITH SOLID FUEL

Kazimierz TARAMINA

Institut of Power Engineering Systems, Wrocław, Poland

1. Introduction

The purpose of air excess control in steam boilers fired with solid fuel is to reduce to minimum the sum of losses in flue gases and losses due to incomplete combustion in any working conditions.

Equation giving the condition of minimum sum of losses in flue gases and losses due to incomplete combustion was obtained as a result of differentiating adequately expanded formula for the sum of these losses with respect to gram-molecule content of free oxygen in dry flue gases at boiler outlet and equating this first derivative to zero, and next performing certain transformations. In this equation gram-molecule contents of dry or moist flue gases at boiler outlet appears, and their derivatives in respect to the gram-molecule content of free oxygen These derivatives, because of impossibility of arranging a measuring system for continuous and direct measurement of their value, are replaced by quotient of small increments the value of which are calculated on the basis of results of two consecutive measurements of each particular flue gas components. Such measurements are conducted at different values of air excess in the combustion chamber, other combustion factors being constant. Each derivative determined in this way is related to mean value of gram-molecule content of free oxygen in flue gases, calculated from two measurements of this parameter.

It is generally known that equation of the condition of minimum sum losses in flue gases and losses due to incomplete combustion, written in the form of small increments quotient is very sensitive to measuring errors. For this reason additional relations for measuring results coordination has been introduced to algorithm of optimum control of combustion process. The first value of the function determining minimum losses evalueted on the basis of results of first and second measurement differs from zero in majority of cases and is therefore insufficient for determination of the optimum air excess in flue gase or in the combustion chamber. Therefore the quantity of excess air must be changed in the known direction and when the combustion process reaches a steady state, a third measurement should be done. From the results of second and third measurement, a second function describing conditions for minimum losses should be calculated. Then, after coordination of results of all three measurements, gram-molecule content of free oxygen in flue gases should be calculated according to the formula obtained by equating to zero the linear function that passes through values of minimum losses condition and through two mean values of optimizated parameter.

After the optimum gram-molecule content of free oxygen in flue gases has been calculated, the quantity of air flowing into combustion chamber must be changed so, that the free oxygen content in flue gases becomes equal or very nearly equal to optimum quantity.

2. Expansion of formulae for minimum losses in flue gases and losses due to incomplete combustion

2.1. Initial formulae

In expanding the formula for minimum losses the following relationships were used:

A. General formula for the sum of losses in flue gases and losses due to incomplete combustion

$$q = q_2 + q_3 + q_4 = \frac{\left[100 \cdot \bar{c}_p (T_{sp} - T_o) + CO \cdot W_{CO}\right](1-x) + \frac{12 \cdot W_c}{22{,}71}(RO_2 + CO) \cdot x}{\frac{12}{22{,}71} \cdot \frac{Q_c - 2512(9h+w)}{c + 0{,}375 \cdot s}(RO_2 + CO)} \qquad (2.1),$$

where: $\bar{c}_p$ - mean specific heat of flue gases with moisture content $\left[\frac{kJ}{(1+[H_2O])m_n^3 \cdot deg}\right]$;

$[H_2O]$ - gram-molecule rate of moisture content in flue gases, in decimal fraction ;

$T_{sp}-T_o$ - temperature difference of flue gases at boiler outlet and air $[deg]$;

RO_2+CO - sum of gram-molecule content of carbon dioxid taking into account sulphur dioxide and carbon monoxide present in dry flue gases at boiler outlet $[\%]$;

W_{CO} - calorific value of carbon monoxide $\left[\frac{kJ}{m_n^3}\right]$;

W_c - calorific value of combustible components present in solid combustion products $\left[\frac{kJ}{kg}\right]$;

Q_c - calorific value of fuel $\left[\frac{kJ}{kg}\right]$;

c,h,n, o,s,w - gram-molecule content in fuel of: carbon, hydrogen, nitrogen, oxygen, combustible sulphur, and full moisture content, expressed in decimal fractions;

x - ratio of incomplete combustion;

22,71 - volume of kilo-molecule weight of gas at conventional temperature and pressure (273 °K and 1 bar) $[m_n^3]$;

12 - carbon atomic weight;

0,375 - atomic mass quotient of carbon and sulphur;

2512 - heat loss in evaporating 1 kg of full moisture content in fuel $\left[\frac{kJ}{kg}\right]$.

B. General formula for mean specific heat of flue gases with moisture content, multiplied by 100

$$100\cdot\bar{C}_p = N_2\cdot C_{pN_2} + RO_2\cdot C_{pCO_2} + O_2\cdot C_{pO_2} + CO\cdot C_{pCO} + H_2O\cdot C_{pH_2O} \quad \left[\frac{kJ}{(1+[H_2O])\,m_n^3\cdot deg}\right] \qquad (2.2),$$

where: N_2, O_2 - gram-molecule content of free nitrogen and oxygen in dry flue gases at boiler outlet [%];

H_2O - gram-molecule ratio of moisture content in flue gases [%];

C_{pN_2}, C_{pCO_2}, C_{pO_2}, C_{pH_2O} - mean specific heat of corresponding flue gases contents $\left[\frac{kJ}{m_n^3\cdot deg}\right]$.

C. Relationship[1] for incomplete combustion ratio to the chemical composition of fuel and dry flue gases and gram-molecule content of moisture in flue gases and in the air

$$1-x = \frac{n'_{H_2} + n'_{H_2O} + \frac{x_L}{100}\left(\frac{1}{2}n'_{H_2} - n'_{O_2} - n'_{N_2}\right)}{n'_C}\cdot\frac{RO_2+CO}{H_2O - X_L\left(1-\frac{CO}{200}\right)} = A\cdot\frac{RO_2+CO}{H_2O - X_L\left(1-\frac{CO}{200}\right)} \qquad (2.3),$$

where: n'_C, n'_{H_2}, n'_{H_2O} n'_{N_2}, n'_{O_2} - numbers of kilo-molecule of: carbon, hydrogen, full moisture content, nitrogen and oxygen in unit quantity of fuel $\left[\frac{kmol}{kg}\right]$

X_L - moisture content in air [%].

D. Relationship[1] of incomplete combustion ratio to the chemical composition of fuel and dry flue gases at boiler outlet

$$1-x=\frac{0{,}79\left(\frac{1}{2}n'_{H_2}-n'_{O_2}\right)+0{,}21\cdot n'_{N_2}}{n'_C}\cdot\frac{RO_2+CO}{N_2-79+0{,}395\cdot CO}=\frac{(L_R-1)(RO_2+CO)}{N_2-79+0{,}395\cdot CO}=\frac{(L_R-1)(RO_2+CO)}{21-RO_2-O_2-0{,}605\cdot CO} \quad (2.4),$$

where: $L_R=1+0{,}79\cdot 0{,}375\frac{8h-o+0{,}3n}{c+0{,}375\cdot s}$ - characteristic coefficient of fuel chemical composition.

From statistic data[5,6] it is apparent that values of this coefficient are almost constant and equal to:

a/ for hard coal from Upper Silesia coalfields

$$L_R=1{,}1034 \pm 0{,}0237$$

b/ for brown coal from Turoszów coaldfields

$$L_R=1{,}083 \pm 0{,}033$$

hence:

$$N_2=79+\frac{L_R-1}{1-x}(RO_2+CO)-0{,}395\cdot CO \quad [\%] \qquad (2.5).$$

E. Balance of dry flue gases at boiler outlet

$$N_2+RO_2+O_2+CO=100 \quad [\%] \qquad (2.6).$$

Substituting (2.5) into (2.6) we obtain a relationship

$$RO_2+CO=\frac{1-x}{L_R-x}(21-O_2+0{,}395\cdot CO) \qquad (2.7),$$

for checking the corectness of values indicated by systems measuring RO_2, O_2 and CO; this relationship, taking into account the fact, that quotient $\frac{1-x}{L_R-x}$ is only very slightly dependent on the incomplete combustion ratio, can be written in the form

$$L_R(RO_2+CO)=21-O_2+0{,}395\cdot CO \qquad (2.8).$$

The relationships (2.3), (2.5), and (2.7) were used for expanding the general formula (2.2) expressing mean specific heat of flue gases, namely:

a/ when determining composition of dry flue gases at boiler outlet

$$100\cdot\bar{C}_p = 100a - f\cdot CO + d\cdot x_L + \frac{RO_2+CO}{1-x}\,(b-e\cdot x) \quad \left[\frac{kJ}{(1+[H_2O])m_n^3\cdot deg}\right] \tag{2.9}$$

b/ when determining composition of flue gases with moisture content at boiler outlet

$$100\cdot\bar{C}_p = 100a - f'\cdot CO + \frac{RO_2+CO}{1-x}(p-e\cdot x) + d\cdot H_2O \quad \left[\frac{kJ}{(1+[H_2O])m_n^3\cdot deg}\right] \tag{2.10}$$

where: $a = 0{,}79C_{pN_2} + 0{,}21C_{pO_2} = 1{,}3075 \pm 0{,}003 \left[\frac{kJ}{m_n^3\cdot deg}\right]$ - for

$T_{sp} = 463 \pm 30$ °K and $T_o = 288 \pm 15$ °K ;

$b = C_{pCO_2} - C_{pN_2} - L_R(C_{pO_2} - C_{pN_2}) + A\cdot C_{pH_2O} = 1{,}278 \pm 0{,}277$ $\left[\frac{kJ}{m_n^3\cdot deg}\right]$ - for boilers fired with hard coal and $3{,}217 \pm 0{,}322$ for boilers fired with brown coal

$d = C_{pH_2O} = 1{,}522 \pm 0{,}0075 \left[\frac{kJ}{m_n^3\cdot deg}\right]$;

$e = C_{pCO_2} - C_{pO_2} = 0{,}456 \pm 0{,}031 \left[\frac{kJ}{m_n^3\cdot deg}\right]$;

$f = C_{pCO_2} - C_{pCO} - 0{,}395(C_{pO_2} - C_{pN_2}) + \frac{x_L}{200}\cdot C_{pH_2O} = 0{,}471 \pm$ $\pm 0{,}033 \left[\frac{kJ}{m_n^3\cdot deg}\right]$;

$f' = C_{pCO_2} - C_{pCO} - 0{,}395(C_{pO_2} - C_{pN_2}) = 0{,}456 \pm 0{,}033 \left[\frac{kJ}{m_n^3\cdot deg}\right]$;

$p = C_{pCO_2} - C_{pN_2} - L_R(C_{pO_2} - C_{pN_2}) = 0{,}452 \pm 0{,}031 \left[\frac{kJ}{m_n^3\cdot deg}\right]$.

Then the general formula (2.1) for the sum of losses in flue gases and losses due to incomplete combustion takes the form:

a/ when determining composition of dry flue gases at boiler outlet

$$q = \frac{\left\{\left[100a + d\cdot x_L - f\cdot CO + \frac{RO_2+CO}{1-x}(b-ex)\right](T_{sp}-T_o) + CO\cdot W_{CO}\right\}(1-x) + \frac{12\cdot W_c}{22{,}71}(RO_2+CO)\cdot x}{\frac{12}{22{,}71}\cdot\frac{Q_c - 2512(9h+w)}{c+0{,}375s}(RO_2+CO)} \tag{2.11}$$

b/ when determining composition at boiler outlet of flue gases with moisture content

$$q=\frac{\Big\{\Big[100a-f'\cdot CO+\frac{RO_2+CO}{1-x}(p-ex)+d\cdot H_2O\Big](T_{sp}-T_o)+CO\cdot W_{CO}\Big\}(1-x)+\frac{12\cdot W_c}{22{,}71}(RO_2+CO)\cdot x}{\frac{12}{22{,}71}\cdot\frac{Q_c-2512(9h+w)}{c+0{,}375\cdot s}(RO_2+CO)} \qquad (2.12).$$

2.2. General equation describing condition for minimum sum of losses in flue gases and losses due to incomplete combustion

General equation describing condition for minimum sum of losses has been obtained by differentiating (2.11) or (2.12) with respect to the free oxygen content expressed in gram-molecule in dry flue gases at boiler outlet and by equating the first derivative to zero and substituting for $\frac{1}{1-x}\cdot\frac{dx}{dO_2}$ an equivalent expression obtained as a result of differentiation expressions (2.7) and (2.8) with respect to O_2 and by dividing the derivatives by original expressions (2.7) and (2.8), namely:

$$\frac{RO_2+CO}{1-x}\cdot\frac{dx}{dO_2}=1-0{,}395\cdot\frac{dCO}{dO_2}-\frac{21-O_2+0{,}395\cdot CO}{21-O_2-RO_2-0{,}605\cdot CO}\cdot\left(\frac{dRO_2}{dO_2}+0{,}605\cdot\frac{dCO}{dO_2}+1\right) \qquad (2.13),$$

$$\frac{RO_2+CO}{1-x}\cdot\frac{dx}{dO_2}=\frac{RO_2+CO}{H_2O-X_L\left(1-\frac{CO}{200}\right)}\left[\frac{d(H_2O-X_L)}{dO_2}+\frac{X_L}{200}\cdot\frac{dCO}{dO_2}\right]-\left(\frac{dRO_2}{dO_2}+\frac{dCO}{dO_2}\right) \qquad (2.14).$$

Then the general expression for minimum losses will have the form:

a/ when determining composition of dry flue gases

$$\begin{aligned}\frac{dq}{dO_2}=&\Bigg\{\Big[\frac{12W_c}{22{,}71}-e(T_{sp}-T_o)\Big]+\Big[W_{CO}-\frac{0{,}395\cdot 12\cdot W_c}{22{,}71}-(f-0{,}395e)(T_{sp}-T_o)\Big]\frac{dCO}{dO_2}+\Big[100a+d\cdot X_L-f\cdot CO+\\&+\frac{RO_2+CO}{1-x}(b-ex)\Big]\frac{d(T_{sp}-T_o)}{dO_2}\Bigg\}(21-O_2-RO_2-0{,}605CO)-\Bigg\{\Big[\frac{12\cdot W_c}{22{,}71}-e(T_{sp}-T_o)\Big](21-O_2)-\Big[W_{CO}+\\&-\frac{0{,}395\cdot 12\cdot W_c}{22{,}71}-(f-0{,}395e)(T_{sp}-T_o)\Big]CO-(100a+d\cdot X_L)(T_{sp}-T_o)\Bigg\}\left(\frac{dRO_2}{dO_2}+0{,}605\frac{dCO}{dO_2}+1\right)=F(O_2)=0\\&\qquad (2.15)\end{aligned}$$

b/ when determining composition of flue gases with moisture content

$$\begin{aligned}\frac{dq}{dO_2}=&\Bigg\{\Big[\frac{12\cdot W_c}{22{,}71}-e(T_{sp}-T_o)\Big]RO_2+\Big[\frac{12\cdot W_c}{22{,}71}-W_{CO}+(f'-e)(T_{sp}-T_o)\Big]CO-(100a+d\cdot X_L)(T_{sp}-T_o)\Bigg\}\cdot\Big[\frac{d(H_2O-X_L)}{dO_2}+\\&+\frac{X_L}{200}\frac{dCO}{dO_2}\Big]-\Bigg\{\Big[\frac{12\cdot W_c}{22{,}71}-e(T_{sp}-T_o)\Big]\frac{dRO_2}{dO_2}+\Big[\frac{12\cdot W_c}{22{,}71}-W_{CO}+(f'-e)(T_{sp}-T_o)\Big]\frac{dCO}{dO_2}-(100a+p\cdot RO_2+d\cdot H_2O)\cdot\\&\cdot\frac{d(T_{sp}-T_o)}{dO_2}\Bigg\}\Big[H_2O-X_L\big(1-\frac{CO}{200}\big)\Big]=F(O_2)=0 \qquad (2.16).\end{aligned}$$

2.3. Simplified expression describing condition for minimum sum of losses in flue gases and losses due to incomplete combustion

From analysis of formulae (2.15) and (2.16) it becomes obvious , that values of certain components depend only very slightly on the changes of temperature of flue gases at boiler outlet $T_{sp} = 463 \pm 30\ {}^{o}K$, air temperature $T_o = 288 \pm 15\ {}^{o}K$ and on the temperature difference $T_{sp} - T_o = 180 \pm 40\ {}^{o}K$, namely:

a/ when burning hard coal

$$\frac{12 \cdot W_c}{22{,}71} - e(T_{sp} - T_o) = 17832 \pm 24 \left[\frac{kJ}{m_n^3}\right]; \quad W_{CO} - \frac{0{,}395 \cdot 12 \cdot W_c}{22{,}71} - (f - 0{,}395e)(T_{sp} - T_o) = 5517 \pm 15 \left[\frac{kJ}{m_n^3}\right];$$

$$\frac{12 \cdot W_c}{22{,}71} - W_{CO} + (f' - e)(T_{sp} - T_o) = 5274 \pm 1 \left[\frac{kJ}{m_n^3}\right]$$

b/ when burning brown coal

$$\frac{12 \cdot W_c}{22{,}71} - e(T_{sp} - T_o) = 17235 \pm 24 \left[\frac{kJ}{m_n^3}\right]; \quad W_{CO} - \frac{0{,}395 \cdot 12 \cdot W_c}{22{,}71} - (f - 0{,}395e)(T_{sp} - T_o) = 5753 \pm 15 \left[\frac{kJ}{m_n^3}\right];$$

$$\frac{12 \cdot W_c}{22{,}71} - W_{CO} + (f' - e)(T_{sp} - T_o) = 4677 \pm 1 \left[\frac{kJ}{m_n^3}\right]$$

Taking into account the fact, that component values $\left[100a + d \cdot X_L + - fCO + \frac{RO_2 + CO}{1-x}(b - ex)\right] \frac{d(T_{sp} - T_o)}{dO_2}$ and $\left[100a + p RO_2 + d H_2O\right] \frac{d(T_{sp} - T_o)}{dO_2}$ are appreciably smaller than the rest of the components in formulae (2.15) and (2.16) and additionaly that the values of derivatives $\frac{d(T_{sp} - T_o)}{dO_2}$ are very close to the value of $1 \left[\frac{deg}{\Delta O_2 = 1\%}\right]$ these components can be calculated with adequate accuracy for all practical purposes from simplified formulae:

A. for boilers fired with hard coal

a/ when determining composition of dry flue gases at boiler outlet

$$\left[100a + d \cdot X_L - fCO + \frac{RO_2 + CO}{1-x}(b - ex)\right] \frac{d(T_{sp} - T_o)}{dO_2} \simeq (133{,}8 + 1{,}3 \cdot RO_2) \frac{d(T_{sp} - T_o)}{dO_2} \qquad (2.17)$$

b/ when determining composition of flue gases with moisture content

$$\left[100a + p RO_2 + d H_2O\right] \frac{d(T_{sp} - T_o)}{dO_2} \simeq (130{,}7 + 0{,}452 \cdot RO_2 + 1{,}52 \cdot H_2O) \frac{d(T_{sp} - T_o)}{dO_2} \qquad (2.18)$$

For boilers fired with brown coal

a/ when determining composition of dry flue gases at boiler outlet

$$\left[100a+d\cdot X_L-f\cdot CO+\frac{RO_2+CO}{1-x}(b-ex)\right]\frac{d(T_{sp}-T_0)}{dO_2}\simeq(133{,}8+3{,}36\,RO_2)\frac{d(T_{sp}-T_0)}{dO_2} \tag{2.19}$$

b/ when determining composition of flue gases with moisture content

$$\left[100a+pRO_2+d\cdot H_2O\right]\frac{d(T_{sp}-T_0)}{dO_2}\simeq(130{,}7+0{,}453\,RO_2+1{,}52\cdot H_2O)\frac{d(T_{sp}-T_0)}{dO_2} \tag{2,20}$$

In accordance with the above simplifying assumptions formulae (2.15) and (2.16) that describe conditions for minimum losses take the form:

A. for boilers fired with hard coal:

a/ when determining composition of dry flue gases et boiler outlet

$$\frac{dq}{dO_2}=\left[17832+5517\cdot\frac{dCO}{dO_2}+(133{,}8+1{,}3RO_2)\frac{d(T_{sp}-T_0)}{dO_2}\right](21-O_2-RO_2-0{,}605\,CO)-\left[17832\,(21-O_2)+\right.$$
$$\left.-5517\cdot CO-133{,}8(T_{sp}-T_0)\right]\left(\frac{dRO_2}{dO_2}+0{,}605\,\frac{dCO}{dO_2}+1\right)=F(O_2)=0 \tag{2.21}$$

b/ when determining composition of flue gases with moisture content

$$\frac{dq}{dO_2}=\left[17832\cdot RO_2+5274\cdot CO-(130{,}7+1{,}52\cdot X_L)(T_{sp}-T_0)\right]\left[\frac{d(H_2O-X_L)}{dO_2}+\frac{X_L}{200}\,\frac{dCO}{dO_2}\right]-\left[17832\,\frac{dRO_2}{dO_2}+\right.$$
$$\left.+5274\cdot\frac{dCO}{dO_2}-(130{,}7+0{,}452RO_2+1{,}52\cdot H_2O)\frac{d(T_{sp}-T_0)}{dO_2}\right]\left[H_2O-X_L\left(1-\frac{CO}{200}\right)\right]\approx F(O_2)=0 \tag{2.22}$$

B. for boilers fired with brown coal;

a/ when determining composition of dry flue gases at boiler outlet

$$\frac{dq}{dO_2}\simeq\left[17235+5753\,\frac{dCO}{dO_2}+(133{,}8+3{,}36RO_2)\frac{d(T_{sp}-T_0)}{dO_2}\right](21-O_2-RO_2-0{,}605\cdot CO)-\left[17235(21-O_2)+\right.$$
$$\left.-5753\cdot CO-133{,}8\,(T_{sp}-T_0)\right]\left(\frac{dRO_2}{dO_2}+0{,}605\,\frac{dCO}{dO_2}+1\right)=F(O_2)=0 \tag{2.23}$$

b/ when determining composition of flue gases with moisture content

$$\frac{dq}{dO_2}=\left[17235\,RO_2+4677\cdot CO-(130{,}7+1{,}52\cdot X_L)(T_{sp}-T_0)\right]\left[\frac{d(H_2O-X_L)}{dO_2}+\frac{X_L}{200}\,\frac{dCO}{dO_2}\right]-\left[17235\,\frac{dRO_2}{dO_2}+\right.$$
$$\left.+4677\,\frac{dCO}{dO_2}-(130{,}7+0{,}453\,RO_2+1{,}52\cdot H_2O)\frac{d(T_{sp}-T_0)}{dO_2}\right]\left[H_2O-X_L\left(1-\frac{CO}{200}\right)\right]=F(O_2)=0 \quad (2.24).$$

From the above equations describing condition of minimum losses it follows that it is very sensitive to measuring errors , and therefore it cannot be used directly for optimization of combustion process without coordination of measuring results.

2.4. Equation in the form of small increments describing condition for minimum sum of losses in flue gases and losses due to incomplete combustion

Because it is impossible to arrange adequate measuring systems for continuous and direct determination of derivatives $\frac{dRO_2}{dO_2}$, $\frac{dCO}{dO_2}$, $\frac{d(H_2O-X_L)}{dO_2}$ and $\frac{d(T_{sp}-T_o)}{dO_2}$ the values of these derivatives are evaluated using the quotients of small increments:

$$\frac{dRO_2}{dO_2} \simeq \frac{(RO_2)_1-(RO_2)_2}{(O_2)_1-(O_2)_2}\,; \quad \frac{dCO}{dO_2} \simeq \frac{(CO)_1-(CO)_2}{(O_2)_1-(O_2)_2}\,; \quad \frac{d(H_2O-X_L)}{dO_2} \simeq \frac{(H_2O)_1-X_{L1}-(H_2O)_2+X_{L2}}{(O_2)_1-(O_2)_2}\,;$$

$$\frac{d(T_{sp}-T_o)}{dO_2} \simeq \frac{(T_{sp}-T_o)_1-(T_{sp}-T_o)_2}{(O_2)_1-(O_2)_2} \qquad (2.25).$$

where: $(RO_2)_1$, $(CO)_1$, $(H_2O)_1$, $X_{L\ 1}$, $(T_{sp}-T_o)_1$ - are values obtained during measurement made before change of air excess quantity in combustion chamber in stabilized firing conditions

$(RO_2)_2$, $(CO)_2$, $(H_2O)_2$, $X_{L\ 2}$, $(T_{sp}-T_o)_2$ - are values obtained during measurement after change of excess air quantity , when firing conditions have reached a stabilized condition.

To simplify the problem, these derivatives are related to the mean value of O_2 calculated from expression

$$O_2 = \frac{1}{2}\left[(O_2)_1+(O_2)_2\right] \qquad (2.26).$$

In a similar way the mean value of remaining parameters measured in two consecutive time intervals, should be calculated, namely:

$$RO_2 = \frac{1}{2}\left[(RO_2)_1+(RO_2)_2\right]; \quad CO = \frac{1}{2}\left[(CO)_1+(CO)_2\right]; \quad X_L = \frac{1}{2}\left(X_{L1}+X_{L2}\right); \quad H_2O = \frac{1}{2}\left[(H_2O)_1+(H_2O)_2\right];$$

$$T_{sp}-T_o = \frac{1}{2}\left[(T_{sp}-T_o)_1+(T_{sp}-T_o)_2\right] \qquad (2.27).$$

Substituting (2.25), (2.26) and (2.27) to (2.21), (2.22), (2.23), (2.24) , a expression in the form of small increments was obtained that describes the condition for minimum sum of losses to flue gases and losses due to incomplete combustion, namely:

A. for boilers fired with hard coal

a/ when determining composition of dry flue gases

$$F(O_2)_{1,2}=\left\{17832\left[21-(O_2)_2\right]-5517(CO)_2-133{,}8(T_{sp}-T_o)_2\right\}\frac{21-(O_2)_1-(RO_2)_1-0{,}605(CO)_1}{(O_2)_1-(O_2)_2}+$$

$$-\left\{17832\left[21-(O_2)_1\right]-5517(CO)_1-133{,}8(T_{sp}-T_o)_1\right\}\frac{21-(O_2)_2-(RO_2)_2-0{,}605(CO)_2}{(O_2)_1-(O_2)_2} \qquad (2.28)$$

b/ when determining composition of flue gases with moisture content

$$F(O_2)_{1,2}=\left\{17832(RO_2)_2+5274(CO)_2-\left[130{,}7+0{,}76(X_{L1}+X_{L2})\right](T_{sp}-T_o)_2\right\}\frac{(H_2O)_1-X_{L1}+\frac{(CO)_1}{400}(X_{L1}+X_{L2})}{(O_2)_1-(O_2)_2}+$$

$$-\left\{17832(RO_2)_1+5274(CO)_1-\left[130{,}7+0{,}76(X_{L1}+X_{L2})\right](T_{sp}-T_o)_1\right\}\frac{(H_2O)_2-X_{L2}+\frac{(CO)_2}{400}(X_{L1}+X_{L2})}{(O_2)_1-(O_2)_2} \qquad (2.29)$$

B. for boilers fired with brown coal

a/ when determining composition of dry flue gases

$$F(O_2)_{1,2}=\left\{17235\left[21-(O_2)_2\right]-5753(CO)_2-133{,}8(T_{sp}-T_o)_2\right\}\frac{21-(O_2)_1-(RO_2)_1-0{,}605(CO)_1}{(O_2)_1-(O_2)_2}+$$

$$-\left\{17235\left[21-(O_2)_1\right]-5753(CO)_1-133{,}8(T_{sp}-T_o)_1\right\}\frac{21-(O_2)_2-(RO_2)_2-0{,}605(CO)_2}{(O_2)_1-(O_2)_2} \qquad (2.30)$$

b/ when determining composition of flue gases with moisture content

$$F(O_2)_{1,2}=\left\{17235(RO_2)_2+4677(CO)_2-\left[130{,}7+0{,}76(X_{L1}+X_{L2})\right](T_{sp}-T_o)_2\right\}\frac{(H_2O)_1-X_{L1}+\frac{(CO)_1}{400}(X_{L1}+X_{L2})}{(O_2)_1-(O_2)_2}+$$

$$-\left\{17235(RO_2)_1+4677(CO)_1-\left[130{,}7+0{,}76(X_{L1}+X_{L2})\right](T_{sp}-T_o)_1\right\}\frac{(H_2O)_2-X_{L2}+\frac{(CO)_2}{400}(X_{L1}+X_{L2})}{(O_2)_1-(O_2)_2} \qquad (2.31)$$

Calculated in this way value of $F(O_2)_{1,2}$ after coordination of first and second measuring results, enables to determine only the deficiency or excess of air in boiler combustion chamber, and it is therefore inadequate for determination of the optimal value of O_2. It is therefore necessary to change the quantity of air excess, and get a third set of measurements after the combustion process becomes stabilized; on the basic of measuring results obtained in the second and third measurement

$F(O_2)_{2,3}$ must be calculated according to one of the formulae (2.28), (2.29), (2.30) or (2.31) changing only indexes 1 to 2 and 2 to 3.

After evaluation of $F(O_2)_{1\ 2}$ and $F(O_2)_{2,3}$ and coordination the results of all three measurements it is not difficult to calculate the optimum participation of oxygen in flue gases from the expression

$$(O_2)_{opt.} = \frac{[(O_2)_2+(O_2)_3]\cdot F(O_2)_{1,2} - [(O_2)_1+(O_2)_2]\,F(O_2)_{2,3}}{2[F(O_2)_{1,2} - F(O_2)_{2,3}]} \qquad (2.32),$$

which has been obtained by equating to zero the linear function passing through the coordinates of two points

$$\left\{F(O_2)_{1,2}\,,\tfrac{1}{2}[(O_2)_1+(O_2)_2]\right\};\ \left\{F(O_2)_{2,3}\,,\tfrac{1}{2}[(O_2)_2+(O_2)_3]\right\}.$$

Expression (2.32) for $(O_2)_{opt.}$ may be used directly only in the following cases:

a/ when $F(O_2)_{1,2} > 0$ and $(O_2)_1 > (O_2)_2$

b/ when $F(O_2)_{1,2} < 0$. and $(O_2)_1 < (O_2)_2$

In the remaining cases the numeration of measuring sequence must be interchanged, e.g second to first and conversly, that is:

$$(O_2)_{opt.} = \frac{[(O_2)_1+(O_2)_3]\cdot F(O_2)_{1,2} - [(O_2)_1+(O_2)_2]\,F(O_2)_{1,3}}{2\,[F(O_2)_{1,2} - F(O_2)_{1,3}]}$$

3. Coordination of parameter values measured in two and three successive time intervals

Coordination of measuring results has the purpose to minimize the sensitivity of expression (2.32) to measuring errors.

In the case, when composition of dry flue gases at boiler outlet is determined, there are three initial equations for coordination of measuring results, arising from expression (2.8), namely:

$$L_R(RO_2)_1+(L_R-0{,}395)(CO)_1+(O_2)_1-21=-w_1 \qquad (3.1),$$

$$L_R(RO_2)_2+(L_R-0{,}395)(CO)_2+(O_2)_2-21=-w_2 \qquad (3.2),$$

$$L_R(RO_2)_3+(L_R-0{,}395)(CO)_3+(O_2)_3-21=-w_3 \qquad (3.3),$$

where: L_R - the most probable mean value of this coefficient corresponding to given fuel;

w_1, w_2, w_3 - deviation from the basic equation (2.8) in three consecutive measuring intervals.

In the case when composition of flue gases with moisture content is being determined, results of three consecutive measurements of RO_2, O_2, CO, H_2O and X_L may be coordinated according to two initial equations arising from relationship

$$(21-O_2-RO_2-0{,}605\cdot CO)\left[\frac{d(H_2O-X_L)}{dO_2}+\frac{X_L}{200}\frac{dCO}{dO_2}\right]+\left(\frac{dRO_2}{dO_2}+0{,}605\frac{dCO}{dO_2}+1\right)\left[H_2O-X_L\left(1-\frac{CO}{200}\right)\right]=0 \quad (3,4),$$

which was obtained by comparison of formula (2.13) and (2.14), namely

$$\left[21-(O_2)_2-(RO_2)_2-0{,}605(CO)_2\right]\left[(H_2O)_1-X_{L1}+\frac{(CO)_1}{400}(X_{L1}+X_{L2})\right]-\left[21-(O_2)_1-(RO_2)_1-0{,}605(CO)_1\right]\left[(H_2O)_2-X_{L2}+\right.$$
$$\left.+\frac{(CO)_2}{400}(X_{L1}+X_{L2})\right]=-w_1 \qquad (3.5),$$

$$\left[21-(O_2)_3-(RO_2)_3-0{,}605(CO)_3\right]\left[(H_2O)_2-X_{L2}+\frac{(CO)_2}{400}(X_{L2}+X_{L3})\right]-\left[21-(O_2)_2-(RO_2)_2-0{,}605(CO)_2\right]\left[(H_2O)_3-X_{L3}+\right.$$
$$\left.+\frac{(CO)_3}{400}(X_{L2}+X_{L3})\right]=-w_2 \qquad (3.6).$$

The two next initial equations for coordination the results of three consecutive measurements of composition of flue gases with moisture content RO_2, O_2, CO, H_2O and X_L were obtained by substituting into (3.5) and (3.6), for $21-(O_2)_1-(RO_2)_1-0{,}605(CO)_1$ $21-(O_2)_2-(RO_2)_2-0{,}605(CO)_2$ and $21-(O_2)_3-(RO_2)_3-0{,}605(CO)_3$ equivalent expressions arising from relationships (3.1), (3.2), and (3.3), namely:

$$21-(O_2)_1-(RO_2)_1-0{,}605(CO)_1=(L_R-1)\left[(RO_2)_1+(CO)_1\right] \qquad (3.7)$$

and so on

Then the next two equations for coordination the results of three consecutive measurements of composition of flue gases with moisture content will take the form:

$$\left[(RO_2)_2+(CO)_2\right]\left[(H_2O)_1-X_{L1}+\frac{(CO)_1}{400}(X_{L1}+X_{L2})\right]-\left[(RO_2)_1+(CO)_1\right]\left[(H_2O)_2-X_{L2}+\right.$$

$$\left.+\frac{(CO)_2}{400}(X_{L1}+X_{L2})\right]=-W_3 \qquad (3.8),$$

$$\left[(RO_2)_3+(CO)_3\right]\left[(H_2O)_2-X_{L2}+\frac{(CO)_2}{400}(X_{L2}+X_{L3})\right]-\left[(RO_2)_2+(CO)_2\right]\left[(H_2O)_3-X_{L3}+\right.$$

$$\left.+\frac{(CO)_3}{400}(X_{L2}+X_{L3})\right]=-W_4 \qquad (3.9).$$

The initial relationships given above are still insufficient for calculation of coordinated values of three measuring results. Therefore additional equations are given which arise from the following relationships:

a/ from linearization of basic relationships for coordination of measuring results with regard to each particular measurement of given parameter

$$\sum_i a_{ki} \cdot v_i = w_k \qquad (3.10),$$

where: a_{ki} - partial derivatives of the k-th initial equation with respect to the i-th measurement of given parameter;

v_i - correction of the i-th measurement value of given parameter;

w_k - deviation of the k-th initial equation.

b/ from the condition of least squares

$$m_i^{-2} \cdot v_i = \sum_k a_{ki} \cdot k_k \qquad (3.11),$$

where: m_i - errors of readings determined by the meter accuracy and its highest scale reading

k_k - indetermined LAGRANGE coefficient for the k-th initial equation

From solution of equations (3.1), (3.2), (3.3), (3.10) and (3.11), or (3.5), (3.6), (3.8), (3.9), (3.10) and (3.11) we obtain correction factors and thus the coordinated values of each particular parameter measured in three seperate time intervals.

4. Conclusions

A. After coordination of measuring results, the reading errors are smaller than errors before coordination. Results of these calculations are listed in tables T.4.1 and T.4.2.

T.4.1 Comparing reading errors after coordination of three measuring results for each particular component of dry flue gases

Measured parameters	RO_2 [%]			O_2 [%]		CO [%]	L_R
Measuring sequence	1	2	3	1	2	1	
Flue gases composition	14,30	13,98	13,65	5,0	5,4	0,374	1,1
Meter accuracy	2,5	2,5	2,5	2,5	2,5	2,5	
m_i - meter error	0,5	0,5	0,5	0,25	0,25	0,012	0,026
$\bar{m}_i$-meter error after coordination	0,238	0,235	0,224	0,186	0,186	0,010	0,015

T.4.2 Comparing reading errors with errors after coordination of three measuring results for each particular component of flue gases with moisture content

Measured parameters	RO_2 [%]			O_2 [%]		CO [%]	H_2O [%]	X_L [%]
Measuring sequence	1	2	3	1	2	1	1	1
Flue gases composition	14,30	13,98	13,65	5,0	5,4	0,374	10,0	2
Meter accuracy	2,5	2,5	2,5	2,5	2,5	2,5	2,5	2,5
m_i - meter error	0,5	0,5	0,5	0,25	0,25	0,012	0,25	0,12
$\bar{m}_i$-meter error after coordination	0,238	0,235	0,222	0,151	0,151	0,009	0,133	0,085

B. Problem of combustion process optimization cannot be solved without coordination of measuring results, what is obvious from maximum error $(\bar{m}_{O_2})_{opt.}$ calculations before and after coordination of results of three measurements of dry flue gases composition T.4.3 .

T.4.3 Comparing errors $(\bar{m}_{O_2})_{opt.}$ before and after coordination of three measuring results of flue gases composition containing moisture

Range of change of O_2 in flue gases with respect to $(O_2)_{opt.} = 5,7$ [%]			Error $(\bar{m}_{O_2})_{opt.}$ – combustion process optimization method			
$(O_2)_1$	$(O_2)_2$	$(O_2)_3$	Meter accuracy 2,5		Meter accuracy 1,5	
			without	after	without	after
			coordination		coordination	
5,2	5,4	5,6	43,5	5,071	26,1	3,11
5,0	5,4	5,8	10,17	1,555	6,10	0,96
4,8	5,4	6,0	4,42	0,884	2,65	0,55
4,6	5,4	6,2	2,53	0,615	1,52	0,383

C. From calculations listed in table T.4.3 it is obvious that optimization problem of combustion process can be solved with adequate accuracy with an aid of measuring systems of accuracy 1,5 with changes of O_2 content in flue gases not smaller than 0,8 [%] in two consecutive measurements.

D. Maximum error calculation results of the combustion process of optimization method listed in table T.4.3 should be regarded as aproximate, because the calculation, owing to the lack at that time of specific measuring results of flue gases with moisture content, were made under assumption that the following relationships are known:

$$CO = 2\,\frac{6-O_2}{3-2O_2} + 0{,}22 \quad [\%]$$

$$x = 0{,}002(6-O_2)^2 + 0{,}005$$

$$RO_2 = (21-O_2+0{,}395\cdot CO)\frac{1-x}{L_R-x} - CO \quad [\%]$$

for $L_R = 1,1$

$$H_2O = A\,\frac{RO_2+CO}{1-x} + \left(1-\frac{CO}{200}\right)\cdot X_L \quad [\%]$$

for $A = 0,542$
and $X_L = 2$ %

$$T_{sp}-T_o = 180+0{,}9\cdot O_2 \quad [deg]$$

according to
measurements

Further deviations of initial equations for coordination of measuring results of composition of flue gases with moisture content were calculated according to law of errors transposition

$$w_k = \sqrt{\sum_i a_{ki}^2 \cdot m_i^2}$$

Hence it is to be expected that the optimization problem of combustion process according to the above method can be solved with sufficient accuracy for practical purposes, by using flue gases composition meters with class 2,5, because for calculation of $(O_2)_{opt.}$ and $(\overline{m}_{O_2})_{opt.}$ such measuring results can be choosen, that give appreciably smaller deviation from value w_k of initial equations for coordination of measuring results, calculated according to law of errors transposition.

Contents

In this article two algorithms of combustion process optimization are given. One of them is based on the coordinated results of three consecutive measurements of dry flue gases composition and the second is based on coordinated results of three consecutive measurements of composition of flue gases with moisture content.

These measurements are made at different values of air excess in combustion chamber, when the combustion process have reached a steady state condition.

Bibliography

1 St. Ochęduszko, J. Szargut - Metody wyznaczania stosunku niecałkowitego spalania. Gospodarka Cieplna - Energetyka Przemysłowa. Zeszyt 6. 1953r.

2 J. Szargut, Z. Kolenda - Uzgadnianie bilansów substancji i energii w procesach chemicznych. Pomiary Automatyka Kontrola. 1967r.

3 J. Szargut - Gospodarka cieplna w hutnictwie book in print

4 J. Szargut, Z. Kolenda - Theory of coordination material and energy balances in chemical process. Archiwum Hutnictwa (in print)

5 K. Taramina - Teoretyczne podstawy ciągłego pomiaru straty wylotowej wyraźnej w kotłach parowych (Doctor Degree Thasis) Wrocław 1962r.

6 K. Taramina - Nowe postacie wzorów na stratę wylotową wyraźną w kotłach parowych opalanych węglem brunatnym. Prace Instytutu Automatyki Systemów Energetycznych. Zeszyt 4 Wrocław 1965r.

7 K. Taramina - Teoretyczne podstawy optymalizacji procesu spalania w kotłach parowych opalanych węglem kamiennym i brunatnym. Prace Instytutu Automatyki Systemów Energetycznych. Zeszyt 7. Wrocław 1966r.

55.3

THE TRAFIC PLANNING ALGORITHMS OF THE PASSENGER AIRCRAFT AND THEIR OPERATIVE CORRECTION.

(automatic control of the tendentional system)

L.D.Atabegov, O.R.Frolov, H.B.Kordonsky, V.K.Linis,
Y.M.Paramonov

The civil aviation research and computing centre
Riga
USSR

GENERAL CONSIDERATION OF THE SYSTEM

Transport line of the USSR civil aviation provides communication between 150 large towns and represents graph,having about 2500 basic ribs.

Every flight between two peaks in spite of the number of the intermediate landing is named a flight.The flight is completely characterized by the set of $\{j_1, j_2, \dots j_k\}$ peak number in which terminal and intermediate take-offs (landings) are performed and by the time of the original take-off from the peak.

The flights with the identical set of $\{j_1. j_2, \dots j_k\}$ peaks are of the same name.The track connecting the $\{j_1, j_2, \dots j_k\}$ peaks is named an airline.The flights of the same name are performed on the given airline not less than once for the control period to be considered.Air passengers transportations are provided by setting the airline net (on the communication graph given),getting the flight number on every airline and at last by coordinating of all the flights into a single passenger traffic time-table.

An airline set with the indication of the flight number having the same name is called the traffic plan.

Demand for air-transportation during the given period of time is a random value.The chance of demand makes it necessary to correct the flight number depending on concrete situation.Thus arises the problem of control in the conditions of control in the conditions of the disturbing actions.

If the demand for air transportation changes it becomes necessary to change the traffic plan.However these changes have generally a partial character and represents addition or ellimination of some flights.It would be quite impossible to change the traffic plan completely as a current correction. It is connected with the fact that the traffic plan gives not only the passenger transportation conditions but functioning conditions of the whole civil aviation system,that is airfield loadings,airfield service work,additional passenger transportation service,distributed means,etc.

Formal description of the system.

The control object is an aviation communication graph,a set of airlines,a set of flights and a set of potentional aviation passengers (fig.1.).

The potentional aviation passengers are the persons who pretend to movement and who could make use of air transport under the suitable conditions.The considered object is a large system with a number of communications and parameters.The system parameters are the following: the air transportation demand between every two towns (more than 5000 substantional values),airlines (about 700),the number of flights of the same name on each airline (about 700).

The control effect iş in the flight number changing on every line or the lines (landing ways) type changing.

The control effect according to one parameter is performed with the receiving of one control pulse.For example,it is necessary to give about 1400 control pulses on all parameters for control effect.The demand change for transportation has systematic and random components.For example,season changes belong to systematic ones.The demand change in time results in time object state change.We shall call this change the displacement along the trajectory of the control system in the parameter space.

The system behaviour is given by the transitional function $P\{X(T)<M|\overline{X}(t), t\in(r,s)\}$ where $X(T)$ is the system state in the time moment T , $\overline{X}(t)$ is the system trajectory (enumeration of all the states) for the time (r,s) .As it is known the transitional function makes it

possible to predict the future behaviour of the system (here we mean probability prediction) with the known last behaviour.

The system behaviour estimation is performed on the basis of the functional,given displacement trajectory of the system. Such a functional can be profit or cost price of the transportation for the control time period considered.

The task of the control actions is in the trajectory optimization of the system displacement,that is in the choice of such trajectory which converts the functional into maximum (minimum).In the probability plan we mean either mathematical functional expectation or limit meaning endless-step control process.

In the considered type the control object (fig.I) is an example of the control random process.However here we have some features which effect radically the control method of this object and the control system construction.

As it was considered the control object is a large system. There is a great number of links in this system.It results in certain difficulties in the system control.

Let the control pulse number which is in the control object and needs the control effects (actions) be the ΔT period. The large system is called tendentional if the control action delay in relation to a corresponding pulse grows together with the control volume growth.Note that the control effect delay is characteristic for all the dynamic systems.Tendentional system feature is in the delay growth with the increasing number pulses,which are fed to the system.It's clear that there is a certain danger during the tendentional system control.The control effect will be performed when the inputs have radically changed compared with their state during control pulse feeding that is (i.e.) inputs/outputs misalignment will occur.

When the volume of control is small,for instance,when only one control pulse is in object,then the control effect comes quickly enough.

In conformity with the transport system it means the following.If a large number of pulses is fed into the object, that is if a large number of modifications in a number of flights and in the kind of airlines is demanded, it will mean the requirements of the passenger traffic time-table reorganization, the airfield service work reorganization and the means redistribution. The control effect realization will require much time. If only one pulse is fed, that is if one additional flight is introduced on Moscow - Leningrad airline, then this introduction can be performed during some hours.

Now we shall point the tendentional system control features.

1. The long-term planning necessity.

It is necessary to predict displacement of the system trajectory considering the control action delay presence for the successful control. It is also necessary to indicate the descrete time moments of the control actions. The long-term plan represents the control pulse set with a large control volume which is fed into the control object in advance considering its reaction delay.

2. The current control must have a small volume.

The current control is required for input and output mismatching correction, when the input changes are unforeseen (in prediction). The control must be small in volume as the system reaction on the current control must be quick.

3. The maximum (minimum) of the functional is out of reach when there is non-stationary random process of input change (the random component is considered). It is necessary to provide absolutely exact input change prevision to reach functional maximum (minimum) with the control effect delay presence such prevision is impossible with unstationary input change.

Tendentional object control system is constructed according to said features. This system includes the following blocks.

1. Input prediction block (prediction demand for air transportations)
2. Traffic plan calculations block (control pulse forestall set) considering the communications between flight number and the demand for transportations.

3.Choose block of the local all of the current control and pulse design of the current control.

Note that we do not consider such blocks which provide the pulse realization into the control effect.It depends on the fact that such "drive" blocks have no automatic character and "drive" is performed by maintenance personnel.

The above considered control blocks are electronic ones and performed automatically in

The control object pattern with no provision for blocks-"drives" is given in fig.2.

Every control block will be described in detail later.

Input prediction block.

As is is shown in fig.2.input prediction is based on traffic system prehistory.According to this we consider input behaviour extrapolation for the future.

The extrapolation has a grounded character if the input change process is a process with a strong afteraction.

The inputs represent the random demand for air transportations between towns.

Here there is a typical situation of the demand-supply, taking place in any kind of trade.

The error is well-known [4] when input is investigated separately with no provision for connection with output.Then the simple extrapolational formulae (polynomas),etc.are for input prediction,without the construction mathematical model of the demand-supply.

The erroneous approach is in misunderstanding that the offer change points form Markov's introduced circuit in the input change process (demand) and thus disturb after action.

As the direct calculations showed the extrapolational formula used leads to errors during prediction of the demand for air transportations.These errors are sometimes 1000 % of the unknown quantity.

Thus mathematical model is required for input extrapolation considering the connection between input and output.

It is characteristic that question is about the mathematical description of man's activity as a transportation subject.It is not necessary to speak that a great number of factors effects man's activity.But among them we can indicate predominant ones and effect of the rest we can take into account by the presentation of man's activity as a random event. If we choose a kind of transport and take a decision about the performing of a trip then money and time resources of the given passenger,cost and comfort of one or another kind of transport will become predominant factors.Here we mean that comfort is a total flight time,flight frequency and general comfort.The flight frequency (the flight number between A, B towns) (fig.2) is the control object output.

Thus organic connection is formed between input and output. Note that the transportation comfort increases with the growth of the flight frequency,as a passenger can easily choose departure and arrival time suitable for him.So the growth of flight number up to some limit leads to the increase of transportation demand.

The model construction task is in indicating of probability that a potential passenger will use air transport.

We shall point out,not going into details,that if communication between towns is provided with only a railway and aviation then in suggestion of logarithmic normal,distribution of passenger money resources (this kind of distribution was confirmed with the statistical inspection of population incomes of the socialist countries[7]).

Probability that the potential passenger agrees to perform a trip is $P^{+} = 1-(1-q)\,\Phi\left(\frac{\ln\delta_{*}-m}{\sigma}\right)$, $\delta_{*} = \min(\delta^{a}, \delta^{\text{ж}})$
If here $\delta^{a} > \delta^{\text{ж}}$ then the probability that the passenger will use air transport is

$$P^{a} = \frac{1-(1-q)\,\Phi\left(\frac{\ln\delta^{a}-m}{\sigma}\right)}{1-(1-q)\,\Phi\left(\frac{\ln\delta^{\text{ж}}-m}{\sigma}\right)}\left(1-\frac{k^{\text{ж}}}{k^{\text{ж}}+k^{a}}\exp\left(-\frac{k^{a}}{168}\left(\alpha\tau^{\text{ж}}-\tau^{a}\right)\right)\right)$$

if $\delta^{a} \geqslant \delta^{\text{ж}}$

then $$P^{a} = 1-\frac{1-(1-q)\,\Phi\left(\frac{\ln\delta^{\text{ж}}-m}{\sigma}\right)}{1-(1-q)\,\Phi\left(\frac{\ln\delta^{a}-m}{\sigma}\right)}\,\frac{k^{\text{ж}}}{k^{\text{ж}}+k^{a}}\exp\left(-\frac{k^{a}}{168}\left(\alpha\tau^{\text{ж}}-\tau^{a}\right)\right)\Big)$$

here $\delta^{\text{ж}}, k^{\text{ж}}, \tau^{\text{ж}}, \delta^{a}, k^{a}, \tau^{a}$

are ticket prices, train (flight) numbers a week and time of movement by train and aircraft.

q, m, σ, α are some constants, $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{x^2}{2}}dx$.

The total number of passengers who want to use air transport in the period of time $(t, t+\tau)$ is given as $\mathcal{N}^{a}(t,t+\tau)=\mathcal{N}(t,t+\tau)\,P^{+}\,P^{a}$ when $\mathcal{N}(t,t+\tau)$ is the total number of potential passengers between A and B towns. The $\mathcal{N}(t,t+\tau)$ value as constants which are in the formula (1) - (3) are obtained by means of statistical prehistory processing.

As a result the input prediction of their change (concerning the transportation between A-B town) as

$$Y(t)=G(t)\,\rho(t)+C(t)$$ where

$G(t), C(t)$ are determined functions which depend on inputs, and $\rho(t)$ is stationary random process of Puasson type.

The input prediction desighns are rather complicated. It is necessary to point out that 45000 primary data are introduced into computer for studying of the object traffic prehistory. The result of calculation is about 1000 data.

That means that the prediction must be performed on a comparatively distant prospect. At present the prediction is performed a year before the plan is put into operation.

The design block of long-term planning.

As a whole the process of input change, given by (5), is nonstationary. However there are comparatively long periods of time. $G(t)=const$, $C(t)=const$ there are stationary demand sections. Naturally the moments of the stationary demand disturbance are the moments, when the control effect is required.

The perspective plan calculation consists of the following stages:

a/The control period division into the stationary demand sections.

b/The required airline calculation for the section of the greatest stationary demand.

c/The required flight number calculation for every section of stationary demand.In that case the input and output connection is taken into account.

We shall consider the given stages in detail.The division into stationary demand sections is performed on the basis of object traffic prehistory investigation,in that case the previous year regularities are transfered to the following year completely.

Usually the section of the greatest stationary demand comes on August.The airline calculation is performed for this month. Considering the system tendentionality and the delay time assumption of the control effects with respect to the plan no more than for a year (the question is about the plan of short-term prospect),the airline list is put into the calculation basis.These airlines functioned previous year.The search of new airlines,according to the town pairs,which had no direct air communication,is performed on the basis of special algorithms.The airline list thus compound is taken as a basis of further calculation.

Then the neccesary number design of the flight having the same name is made.The transportation demand of all the passengers who want to perform the flight is put into the calculation basis.There is a reverse connection with inputs and airline lists with the given airline calculation.The connection with the inputs is based on the use of successive interation calculation process.These interations consider that the demand function is convex as to the flight number.The linear programming model of the type:

a/ limitation $\sum_{i\in\omega_j} z_i \geqslant b_j \quad ; \qquad j = 1,2,\dots m \; ;$

$$z_i \geqslant 0 \quad ; \qquad i = 1,2,\dots n \; ;$$

b/ the minimum functional of the transportation cost

$$\sum_{i=1}^{n} c_i z_i$$

is used at every interation step.

Here z_i are the auxiliary unknowns,the so-called airline loading variants.

ω_j -the multitude of all the loading variants,which supply the j - town pair.

The linear problem solution makes it possible to indicate the necessary flight number.Then the return to the prediction formula (2),(3),(4) is made again etc.The interation process coinsides quickly and two interations are practically enough.

The received solution is analysed in computer to define accurately the airline list.For this purpose the number of the transported vacant chair is calculated for every airline,the composition of intermediate landing is defined exactly and new airlines are set.After all it is necessary that the chair usage percent should be about 80 % for every airline section.

The calculation process of the flight number with interations is performed again to define the airliner composition accurately.

The next stage of plan calculation is connected considering the available aircraft park.The required aircraft numbers of all types are defined according to the calculated plan. These values are compared with the disposed ones.Then the plan is defined accurately by the equalization ratio method. In this case the flight number decreases in the case of necessity and flights are redistributed between aircraft types.In this equilization ratios we consider that it is necessary to provide the destant passengers and the passengers who cannot use railway delivery when the flight number decreases.The total calculation circuit is given in Fig.3.

Output cell block of the current control and pulse calculation.

As it was noted above the current control necessity appears with unpredicted input change,note that the current control necessity substantially depends on the input behaviour prediction and the perspective plan development quality.

When some input behaviour differs a little from predicted one there is a certain plan reserve which permits not to use the current control.The planned chair occupation ratio (80%) is a reserve.It is obvious that the increase of demand for transportations against the planned one,even when it is 10%, doesn't require the flight number increase.

It is necessary to use the current control with great input behaviour differences (15% and more).Here the question arrises about the necessary control volume.The total aviation connection graph is rather great.When calculating current control it is necessary to mark out the connections subgraph in which the required flight number recalculation and airline composition change are performed.Denote the passenger airstream between A_{s_1} and A_{s_2} towns by Y_{s_1,s_2} .Denote the passenger stream transported along the line with k-number between A_{s_1} and A_{s_2} towns by $Y^{(k)}_{s_1,s_2}$ .The airlines having $k_1, k_2, \dots k_\ell$ number are named connected along $(s_1 s_2)$ stream if

$$\sum_{i=1}^{\ell} Y^{(k_i)}_{s_1,s_2} = Y_{s_1,s_2} ; \quad Y^{(k_i)}_{s_1,s_2} > 0 ; \quad i = 1,2,\dots \ell$$

The airlines for which $Y^{(k_i)}_{s_1,s_2}$ is zero are isolated on the $(s_1 s_2)$ section.

Let the Y_{s_1,s_2} stream changed substantially according to the predicted one.Then the current control cell consisting of airlines for which $Y^{(k_i)}_{s_1,s_2} > 0$ is extracted and those airlines are connected with them along airline streams.

As the calculations show the current control cell volume for the perspective plan performed correctly is not large and usually makes no more than 5 airlines.

The whole calculation given in the description of the perspective plan calculation block is made within the limits of the current control cell.

The calculation cell volume is small and requires little time (no more than an hour).The control volume is not comparatively large.It doesn't exceed 3-5 pulses.

Conclusion.

The given control system is used in practice.The experience shows that many parts of the system require further development.This concerns both the prediction block and the perspective plan calculation block.However the control principles proved their value completely.These principles are suitable for the other tendentional systems connected with demand supply.

References.

1.I.B.Gertcbach,H.B.Kordonsky,A.B.Nikitina

Some comments on statistical value estimation of unperformed demand for passenger air-transportations.
RIGA Reports,issue 73.
"Application of mathematical methods in control planning".Riga,1966.

2.V.K.Linis.

Notes on optimal planning of transport activity.
RIGA Reports,issue 73.
Riga,1966.

3.V.K.Linis.

Organization of aviation passenger transportations.
Works to ALl-Union conference on application of economic-mathematical methods in branch planning and control.
Section No.1.,M.,1966.

4.Zh.Mot.

Statistical previsions and solutions at enterprises.
"Progress",M.,1966.

5.Y.M.Paramonov,L.V.Molchanova.

Passenger number distribution analysis before holidays.
RIGA Reports,73,Riga,1966.

6.Y.M.Paramonov,O.R.Frolov.

Passenger transportation analysis and extrapolation works to All-Union conference on application of economic-mathematical methods in branch planning and control.
Section No.1.,M.,1966.

7.E.Eltete.

Character and property investigation of income distribution.
Collected articles "Standard of Living".Statistics,M.,1966.

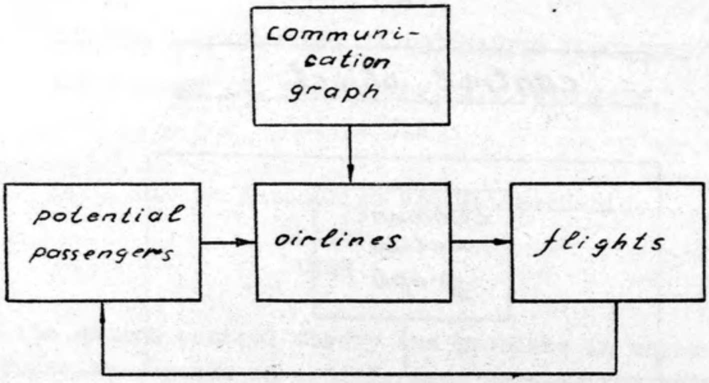

communi-
cation
graph
potential
passengers
airlines
flights

fig. 1

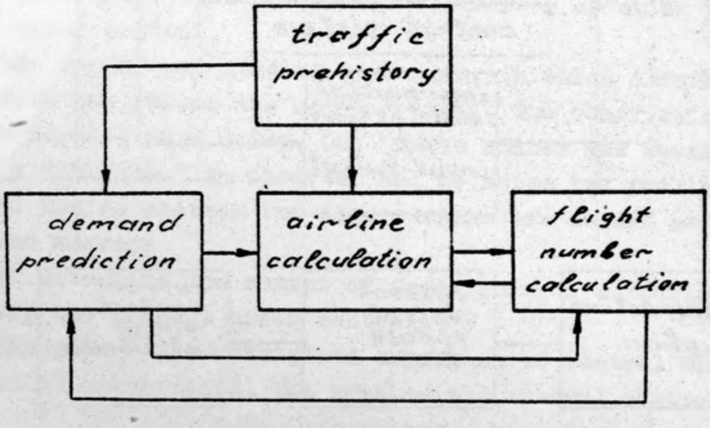

traffic
prehistory
demand
prediction
airline
calculation
flight
number
calculation

fig. 3

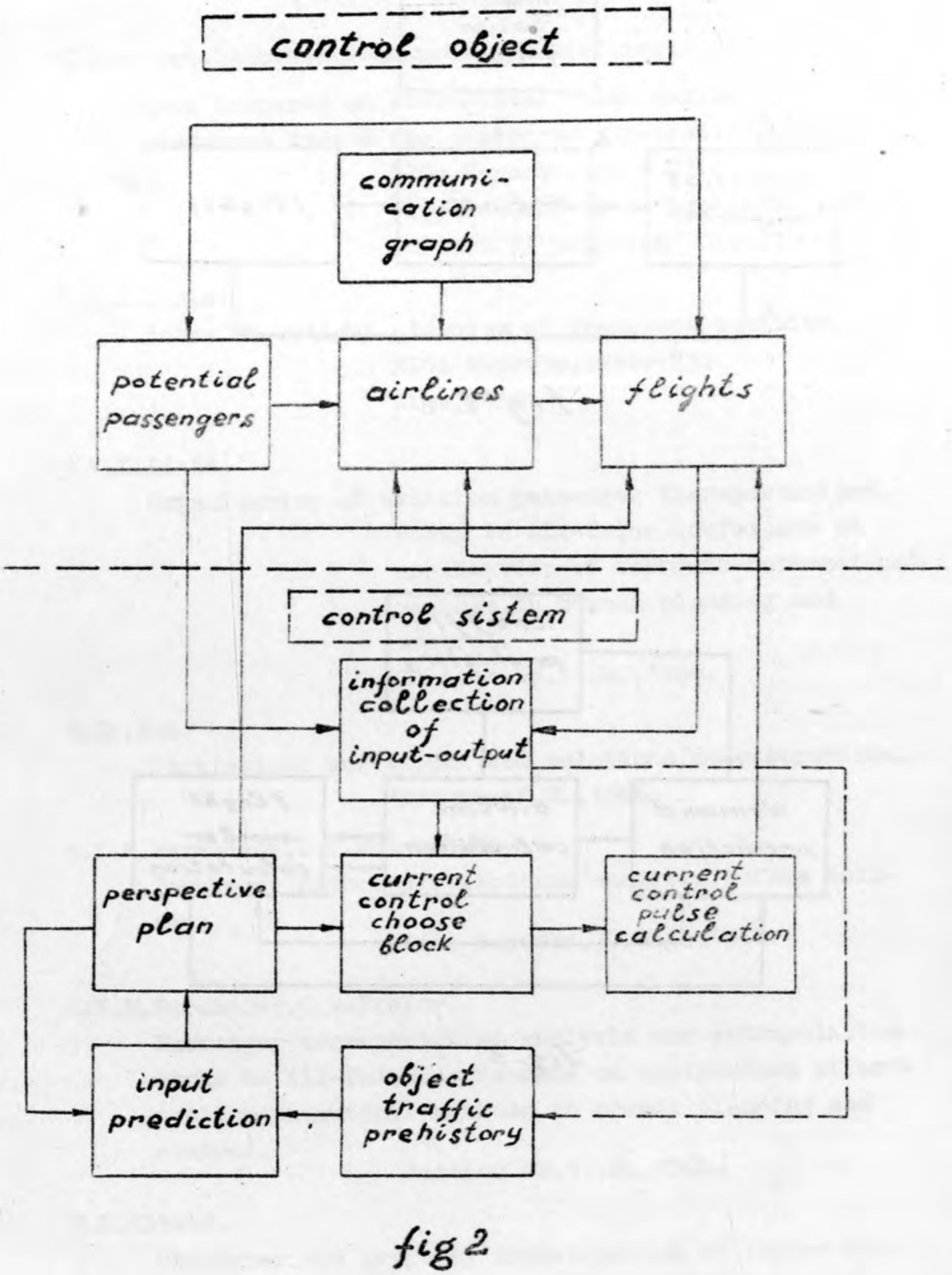

control object
communication graph
potential passengers
airlines
flights
control sistem
information collection of input-output
perspective plan
current control choose block
current control pulse calculation
input prediction
object traffic prehistory

fig 2

56.4

CONTROL AND CHECKING ON STOCHASTIC PROCESSES

A.A.KLEMENT'EV, E.P.MASLOV, A.M.PETROVSKY,
A.I.YASHIN

Institute on Automatics and Telemechanics
Moscow
USSR

In the modern control theory the problems in which a loss function depends on a difference between the actual and reference signals, on a control action and on a cost of checking are of great interest.

Such problems are frequently met in the practice, for example, in the mass production industry.

The necessity of synthesis of optimal algorithms of control and checking was remarked in several papers.[1-4] But only a few of them dealt with the problems, where a checking process was under control.

In the paper[4] was studied a problem, in which a markov process with two states was under control. The observations of markov process were noisy. One of the states was considered as a "failure." An observer had to solve two problems:

1) to decide whether the observations are needed at the current moment;

2) to determine the moment of appearance of a "failure" and to stop the process under control.

In the paper[5] was considered a problem of control with informational constraints. The synthesized optimal controller had to divide the region of observations, according to their costs.

In this paper an optimal algorithm of control and checking on a discrete random process is synthesized. The costs of control, checking and difference between the actual and reference signals are taken into account.

A general statement of the problem is as follows.

1. The system under consideration is discrete-continuous. All variables are considered only at discrete moments of time, n=0,1,2... . The value of a quantity at a n-th discrete time instant carried the subscript n. The N-stage control process is considered, where the finite member N is fixed.

2. It is a Bayesian problem that is considered (it is assumed given the a- priori density of each random variable). The errors of control and checking are negligible.

3. The process $\{\eta_n\}$ under control is characterized by a random vector of parameters $\bar{\Lambda}$. The a priori distribution of vector $\bar{\Lambda}$ is assumed to be given.

4. The result of checking coincide with the coordinate of process under control at the moment of checking.

It is assumed also, that a reference process $\{\theta_m\}$ is deterministic.

The physical essence of the procedure of control and checking on the process $\{\eta_n\}$ amounts to the following. At the end of some arbitrary (n-1)'st stage (moment of time) the controller possesses some information on the past behaviour of the process $\{\eta_k\}$, $k \leq n-1$. On the basis of this information the controller decide between two following alternatives:

1. Don't check the process $\{\eta_k\}$ at the moment t=n-1 and compute the optimal control action as a function of the past information only.

2. To check the process $\{\eta_k\}$ at the moment t=n-1. In this case the controller observe the value η_{n-1} and compute an optimal control action as a function of the result of checking and past information.

Note, that a point t=n-1 is simultaneously the end of the (n-1)'st stage and the beginning of the n'st stage, and introduce a random variable.

$$\chi_n = \begin{cases} 1, \text{ if it was decided to check the process } \{\eta_k\} \text{ at the moment } t=n-1; \\ 0, \text{ otherwise.} \end{cases}$$

Let us introduce the following designation:

η_n - a coordinate of the process under control at the moment t=n,

u_n - a control action at the n'th stage.

It will be clear from the following that the problem in hand is solved by means of dynamic programming method. In accordance with this method[2,4], minimization of a functional is accomplished by a "reverse" movement from the last stage to the first one.

Thus, a decision about checking χ_n and a value of control action u_n at the n'th stage are some functions of a past information and so depend, as a matter of fact, on the nonselected vectors of decisions $\vec{\chi}_{n-1}=(\chi_1,..., \chi_{n-1})$, control actions $\vec{u}_{n-1}=(u_1,..., u_{n-1})$, and also on a vector of observations, which structure and dimension yet are not selected. The structure of the vector of observations depends on a structure of decision vector $\vec{\chi}_{n-1}$.

The best, what it can be done in this case, is to synthesize a dependence of χ_n and u_n on the vectors of observations, control actions and decisions in general.

To avoid the uncertainty, related to an unknown structure of observation vector, and to create an algorithm of checking and control in a closed form, we suggest a formalistic scheme of synthesis of sequence of observations.

The essence of this scheme amounts to the following. Let us decide to check the process under control, i.e. $\chi_n=1$

In this case the result of observation coincides with a coordinate η_{n-1}. The decision $\chi_n=0$ excludes the checking on a coordinate η_{n-1}. But the fact of nonentering of any information about the process $\{\eta_k\}$ is equivalent, from the point of view of accumulation of information about $\{\eta_k\}$, to entering of some information about a hypothetical process, independent in any way on the process $\{\eta_n\}$. In particular, in this work we suppose that this hypothetical process is a random sequence $\{\varepsilon_n\}$, independent from $\{\eta_n\}$.

Thus, in both cases, when it was decided to check the process $\{\eta_n\}$ and when it was decided to exclude the checking, we can formally consider, that the controller accomplished at the moment t=n-1 an observation. Let us name this observation as a "generalized" observation and designate it by symbol y_n. So we can formally write an expression

$$y_n = \begin{cases} \eta_{n-1}, & if \ x_n = 1; \\ \varepsilon_{n-1}, & if \ x_n = 0, \end{cases} \tag{1}$$

or

$$y_n = x_n \eta_{n-1} + (1 - x_n) \varepsilon_{n-1} \tag{2}$$

At the end of some arbitrary (n-1)'st stage the controller possesses a decision vector $\vec{X}_{n-1} = (x_1, \dots, x_{n-1})$, a vector of control actions $\vec{u}_{n-1} = (u_1, \dots, u_{n-1})$and a vector of generalized observations $\vec{y}_{n-1} = (y_1, \dots, y_{n-1})$. The controller also knows a sequence of values of a reference deterministic process $\{\vec{\theta}_m\}$ for an arbitrary moment m.

On the basis of this information the controller come to a decision

$$x_n = x_n \left(\vec{x}_{n-1}, \vec{y}_{n-1}, \vec{u}_{n-1}, \vec{\theta}_m \right), \ n = 1, \dots, N \tag{3}$$

about the checking on a coordinate η_{n-1} . Let the controller comes to a decision $x_n = 0$. Then a coordinate η_{n-1} is not observed, $y_n = \varepsilon_{n-1}$ and an optimal control action u_n^* is computed as a function of the past information only:

$$u_n^* = u_n^* \left(\vec{x}_{n-1}, x_n = 0, \vec{y}_{n-1}, \vec{u}_{n-1}, \vec{\theta}_m \right) \tag{4a}$$

If the controller comes to a decision $x_n=1$; then the coordinate η_{n-1} is checked, $y_n=\eta_{n-1}$, and an optimal control action u_n^* is computed as a function of the result of checking and past information:

$$u_n^* = u_n^*(\vec{x}_{n-1}, x_n=1, \vec{y}_{n-1}, \eta_{n-1}, \vec{u}_{n-1}, \vec{\theta}_m) \tag{4b}$$

At the next stage everything is repeated.

As the criterion of optimality for the present paper we chose the criterion of minimal total risk. It is formed as follows. At each stage we determine threе types of losses.

1. A loss related to a difference between the controlled and reference processes. This type of losses is determined by a function, depending on the values θ_n, η_n and, in general case, also on time:

$$C_{1n} = C_1(n, \theta_n, \eta_n), \quad n=1,2,\dots,N. \tag{5}$$

2. A loss related to a control action. This type of losses is determined by a function

$$C_{2n} = C_2(n, u_n), \quad n=1,2,\dots,N. \tag{6}$$

3. A loss related to checking on a coordinate η_{n-1}. This type of losses is determined by a function

$$C_{3n} = x_n C_3(n), \quad n=1,2,\dots,N. \tag{7}$$

Let us call

$$C_n = C_1(n, \theta_n, \eta_n) + C_2(n, u_n) + x_n C_3(n), \quad n=1,\dots,N \tag{8}$$

as a specific loss function. Then the general loss function is equal to

$$C_\Sigma = \sum_{n=1}^{N} \left[C_1(n, \theta_n, \eta_n) + C_2(n, u_n) + x_n C_3(n) \right] \tag{9}$$

We shall consider as optimum that system for which the total risk (the expectation of the variable C_Σ)

$$R_\Sigma = M\{C_\Sigma\} = M\{C_1(1,\theta_1,\eta_1) + C_2(1,u_1) + x_1 C_3(1)\} + M\{C_1(2,\theta_2,\eta_2) + C_2(2,u_2) + x_2 C_3(2) + \ldots + M\{C_1(N,\theta_N,\eta_N) + C_2(N,u_N) + x_N C_3(N)\} \tag{10}$$

is minimal.

It is required to find a sequence of decision rules

$$\Gamma_n = \Gamma_n^x(x_n \mid \vec{x}_{n-1}, \vec{y}_{n-1}, \vec{u}_{n-1}, \vec{\theta}_m)\cdot\Gamma_n^u(u_n \mid \vec{x}_n, \vec{y}_n, \vec{u}_{n-1}, \vec{\theta}_m) \triangleq \Gamma_n^x \cdot \Gamma_n^u, \quad n = 1, 2, \ldots, N \tag{11}$$

and, accordingly, a sequence of pairs (x_n, u_n), for which the total risk R_Σ is minimal.

Let us write down an expression for a specific risk Rn:

$$R_n = \int \{C_1(n,\theta_n,\eta_n) + C_2(n,u_n) + x_n C_3(n)\}\cdot P(\bar{\lambda})\cdot \Omega(\vec{x}_n, \vec{u}_n, \vec{y}_n, \eta_n, \bar{\lambda}) \cdot P(\eta_n \mid \bar{\lambda}, \vec{x}_n, \vec{u}_n, \vec{y}_n)\cdot \prod_{i=1}^{n} P(y_i \mid \bar{\lambda}, \vec{x}_i, \vec{y}_{i-1}, \vec{u}_{i-1})\cdot \prod_{i=1}^{n}(\Gamma_i^x\cdot\Gamma_i^u)\,d\Omega, \quad n = 1, 2, \ldots, N \tag{12}$$

Here and in what follows $\Omega(\cdot)$ denotes a region of joint variation of the variables within the parentheses, with $d\Omega$ being an infinitesimal element of this region. We shall also agree, that the functions $P(\cdot)$, having different arguments, will, in general, represent different functions. The function $P(\cdot)$ will denote the probability density of a random variable within the parentheses.

The total risk is equal

$$R_{\Sigma} = \sum_{n=1}^{N} R_n \tag{13}$$

The problem of choice of 2N-dimensional vector $(\vec{x}_N, \vec{u}_N)$ which minimized the total risk R_{Σ}, is solved by dynamic programming method.^{2}

At first the last pair (x_N, u_N) is determined. In the expression (13) from (x_N, u_N) only the last term depends. So, an optimal pair (x_N^*, u_N^*) is found from the condition that the risk R_N be minimal. Let the 2(N-1) dimensional vector $(\vec{x}_{N-1}, \vec{u}_{N-1})$ is given. Then

$$\begin{aligned} R_N = \int_{\Omega(\vec{x}_{N-1}, \vec{y}_{N-1}, \vec{u}_{N-1})} \prod_{i=1}^{N-1} (\Gamma_i^x \cdot \Gamma_i^u) \Big\{ \int_{\Omega(\eta_N, \bar{\lambda}, x_N, y_N, u_N)} & [C_1(N, \theta_N, \eta_N) + C_2(N, u_N) + \\ + x_N C_3(N)] \cdot P(\bar{\lambda}) \cdot P(\eta_N \,|\, \bar{\lambda}, \vec{x}_N, \vec{y}_N, \vec{u}_N) \cdot & \\ \cdot \prod_{i=1}^{N} P(y_i \,|\, \bar{\lambda}, \vec{x}_i, \vec{y}_{i-1}, \vec{u}_{i-1}) \cdot \Gamma_N^x \cdot \Gamma_N^u \, d\Omega \Big\} d\Omega & \end{aligned} \tag{14}$$

Where

$$\begin{aligned} \Gamma_i^x &= \Gamma_i^x (x_i \,|\, \vec{x}_{i-1}, \vec{y}_{i-1}, \vec{u}_{i-1}, \vec{\theta}_m); \\ \Gamma_i^u &= \Gamma_i^u (u_i \,|\, \vec{x}_i, \vec{y}_i, \vec{u}_{i-1}, \vec{\theta}_m), \quad i = 1, 2, \dots, N \end{aligned} \tag{15}$$

At each stage the controller decides between two alternatives:
$x_n = 0$ or $x_n = 1$. Therefore, at n=N the decision rule

$$\Gamma_N^x(x_N \mid \vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) = \gamma(x_N = 1 \mid \vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) \cdot \delta(x_N - 1) + \gamma(x_N = 0 \mid \vec{x}_{N-1}, u_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) \cdot \delta(x_N - 0), \tag{16}$$

Where $\gamma(x_N = j \mid \vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m)$, $j = 0, 1$, - the probabilities to accept the corresponding decisions in the presence of information $\vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m$, and $\delta(x_N - j)$ - delta-functions.

Substituting (16) into (15) and integrating the last expression on x_N, we find

$$\begin{aligned}
R_N = &\int_{\Omega(\vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1})} \prod_{i=1}^{N-1} (\Gamma_i^x \cdot \Gamma_i^u) \Big\{ \gamma(x_N = 1 \mid \vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) \cdot \\
&\cdot \int_{\Omega(\eta_N, \bar{\lambda}, y_N, u_N)} [C_1(N, \theta_N, \eta_N) + C_2(N, u_N) + C_3(N)] \cdot P(\bar{\lambda}) \cdot P(\eta_N \mid \bar{\lambda}, \\
&\vec{y}_N, \vec{x}_{N-1}, x_N = 1, \vec{u}_N) \cdot \prod_{i=1}^{N} P(y_i \mid \bar{\lambda}, \vec{x}_i, \vec{y}_{i-1}, \vec{u}_{i-1}) \cdot \Gamma_N^u(u_N \mid \vec{x}_{N-1}, \\
&x_N = 1, \vec{y}_N, \vec{u}_{N-1}, \vec{\theta}_m) d\Omega + \gamma(x_N = 0 \mid \vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) \cdot \\
&\cdot \int_{\Omega(\eta_N, \bar{\lambda}, y_N, u_N)} [C_1(N, \theta_N, \eta_N) + C_2(N, u_N)] \cdot P(\bar{\lambda}) \cdot P(\eta_N \mid \bar{\lambda}, \vec{x}_{N-1}, x_N = 0, \\
&\vec{y}_N, \vec{u}_N) \cdot \prod_{i=1}^{N} P(y_i \mid \bar{\lambda}, \vec{x}_i, \vec{y}_{i-1}, \vec{u}_{i-1}) \cdot \Gamma_N^u(u_N \mid \vec{x}_{N-1}, x_N = 0, \\
&\vec{y}_N, \vec{u}_{N-1}, \vec{\theta}_m) d\Omega \Big\} d\Omega
\end{aligned} \tag{17}$$

Let us examine in detail the last formula. It follows from (2), that if $x_N = 1$: $y_N = \eta_{N-1}$. Therefore

$$P(\eta_N \mid \bar{\lambda}, \vec{y}_N, \vec{x}_{N-1}, x_N = 1, \vec{u}_N) = \\ = P(\eta_N \mid \bar{\lambda}, \vec{y}_{N-1}, \eta_{N-1}, \vec{x}_{N-1}, x_N = 1, \vec{u}_N) \tag{18}$$

$$P(y_N \mid \bar{\lambda}, \vec{x}_{N-1}, x_N = 1, \vec{y}_{N-1}, \vec{u}_{N-1}) = \\ = P(\eta_{N-1} \mid \bar{\lambda}, \vec{x}_{N-1}, x_N = 1, \vec{y}_{N-1}, \vec{u}_{N-1}) \tag{19}$$

Further, if $x_N = 0$: $y_N = \varepsilon_{N-1}$. The processes $\{\eta_n\}$ and $\{\varepsilon_n\}$ are independent. Therefore

$$P(\eta_N \mid \bar{\lambda}, \vec{y}_N, \vec{x}_{N-1}, x_N = 0, \vec{u}_N) = P(\eta_N \mid \bar{\lambda}, \vec{y}_{N-1}, \varepsilon_{N-1}, \\ \vec{x}_{N-1}, x_N = 0, \vec{u}_N) = P(\eta_N \mid \bar{\lambda}, \vec{y}_{N-1}, \vec{x}_{N-1}, x_N = 0, \vec{u}_N) \tag{20}$$

$$P(y_N \mid \bar{\lambda}, \vec{x}_{N-1}, x_N = 0, \vec{y}_{N-1}, \vec{u}_{N-1}) = P(\varepsilon_{N-1}), \tag{21}$$

where $P(\varepsilon_{N-1})$ a priori density of the random variable ε_{N-1}.

It was remarked before, that in the case when it was decided to exclude the checking of a coordinate η_{n-1}, an optimal control action u_n is determined as a function of past information only. It means, that the decision rule

$$\Gamma_N^u(u_N \mid \vec{x}_{N-1}, x_N = 0, \vec{y}_N, \vec{u}_{N-1}, \vec{\theta}_m) = \\ = \Gamma_N^u(u_N \mid \vec{x}_{N-1}, x_N = 0, \vec{y}_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m) \tag{22}$$

Thus in the second summand of (17) only the density $P(\varepsilon_{N-1})$depends on $y_N = \varepsilon_{N-1}$. Let us substitute the expressions (18)+(22) into (19), integrate (19) on ε_{N-1} and introduce two functions

$$\alpha_N = \int\limits_{\Omega(\eta_N, \bar{\lambda})} [C_1(N, \theta_N, \eta_N) + C_2(N, u_N) + C_3(N)] \cdot P(\bar{\lambda}) \cdot P(\eta_N | \bar{\lambda}, \vec{y}_{N-1}, \eta_{N-1}, \vec{x}_{N-1}, x_N = 1, \vec{u}_N) \cdot P(\eta_{N-1} | \bar{\lambda}, \vec{x}_{N-1}, x_N = 1, \vec{y}_{N-1}, \vec{u}_{N-1}) \cdot \prod_{i=1}^{N-1} P(y_i | \bar{\lambda}, \vec{x}_i, \vec{y}_{i-1}, \vec{u}_{i-1})] d\Omega \qquad (23)$$

$$\beta_N = \int\limits_{\Omega(\eta_N, \bar{\lambda})} [C_1(N, \theta_N, \eta_N) + C_2(N, u_N)] \cdot P(\bar{\lambda}) \cdot P(\eta_N | \bar{\lambda}, \vec{y}_{N-1}, \vec{x}_{N-1}, x_N = D, \vec{u}_N) \cdot \prod_{i=1}^{N-1} P(y_i | \bar{\lambda}, \vec{x}_i, \vec{y}_{i-1}, \vec{u}_{i-1}) d\Omega \qquad (24)$$

Taking into account these functions, we found the following expression for the risk R_N

$$R_N = \int\limits_{\Omega(\vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1})} \prod_{i=1}^{N-1} (\Gamma_i^x \cdot \Gamma_i^u) \cdot \Phi_N(\vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) d\Omega \qquad (25)$$

where

$$\Phi_N = \Phi_N(\vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) = \gamma(x_N = 1 | \vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) \cdot \int\limits_{\Omega(\eta_{N-1}, u_N)} \alpha_N \Gamma_N^u (u_N | \vec{u}_{N-1}, \vec{x}_{N-1}, x_N = 1, \vec{y}_{N-1}, \eta_{N-1}, \vec{\theta}_m) +$$

$$+ \gamma(x_N = 0 | \vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) \cdot \int\limits_{\Omega(u_N)} \beta_N \Gamma_N^u (u_N | \vec{u}_{N-1}, \vec{x}_{N-1}, x_N = 0, \vec{y}_{N-1}, \vec{\theta}_m) d\Omega \qquad (26)$$

Optimization of R_N concerning (x_N, u_N) is reduced to optimization of Φ_N . Let us choose the optimal control action u_N at first. We impose the following limitation. We shall seek our controller in the class of systems which possess regular strategies concerning to optimal control actions u_n, $n=1,2,\dots,N$.

Let u_n^{0*} - an optimal control action, corresponding to the decision $x_n=0$, u_n^{1*}- an optimal control action, corresponding to the decisión $x_n=1$. Therefore

$$\Gamma_N^u(u_N \mid \vec{x}_{N-1}, x_N=0, \vec{y}_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m) = \delta(u_N - u_N^{0*});$$

$$\Gamma_N^u(u_N \mid \vec{x}_{N-1}, x_N=1, \vec{y}_{N-1}, \eta_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m) = \delta(u_N - u_N^{1*}). \qquad (27)$$

Substituting (27) into (26) and integrating the last formula on u_N we find that in the expression for Φ_N only α_N depends on the optimal action u_N^{1*} , and only β_N depends on the optimal action u_N^{0*} .

Therefore, the optimal control action u_N^{1*} is equal to a value u_N, which minimizes the function α_N . Analogically, the optimal control action u_N^{0*} coincides with a value u_N, which minimizes the function β_N .

Let

$$\alpha_N^* = \min_{u_N \in \Omega(u_N)} \alpha_N \quad ; \quad \beta_N^* = \min_{u_N \in \Omega(u_N)} \beta_N \qquad (28)$$

Where Ω (U_N) a region of permissible control actions at the n'th stage.
Substituting (28) into (26), we find

$$\Phi_N^* = \min_{u_N} \Phi_N =$$

$$= \gamma(x_N = 1 | \vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) \cdot \int_{\Omega(\eta_{N-1})} \alpha_N^* d\Omega + \tag{29}$$
$$+ \gamma(x_N = 0 | \vec{x}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1}, \vec{\theta}_m) \cdot \beta_N^*$$

Let us suppose that our controller possesses regular strategies concerning the decisions x_n also: from
$$\gamma(x_n = 1 | \vec{x}_{n-1}, \vec{y}_{n-1}, \vec{u}_{n-1}, \vec{\theta}_m) = 1$$
follows (a.s.) $\gamma(x_n = 0 | \vec{x}_{n-1}, \vec{y}_{n-1}, \vec{u}_{n-1}, \vec{\theta}_m) = 0$ and on the contrary.

Thus optimization of Φ_N^* concerning x_N is reduced to comparison of two functions $\int_{\Omega(\eta_{N-1})} \alpha_N^* d\Omega$ and β_N^* and to choice the lesser of them. When

$$\int_{\Omega(\eta_{N-1})} \alpha_N^* d\Omega > \beta_N^* \tag{30}$$

the decision $x_N = 0$ is reached; when

$$\int_{\Omega(\eta_{N-1})} \alpha_N^* d\Omega < \beta_N^* \tag{31}$$

the decision $\chi_N = 1$ is reached. If these functions are equal, the choice is arbitrary. The result of such double minimisation Φ_N is a function

$$\Phi_N^{**} = \min_{\chi_N, u_N} \Phi_N = \min\Big[\beta_N^*, \int\limits_{\Omega(\eta_{N-1})} \alpha_N^* d\Omega\Big] \qquad (32)$$

Let us now determine a pair (χ_{N-1}, u_{N-1}). We remark at first that by analogy with (23), (24), (26) the functions $\alpha_n, \beta_n, \Phi_n$ can be constructed for each stage $n = 1, 2, \dots, N$.

Let us suppose that a 2(N-2)-dimensional vector ($\vec{\chi}_{N-2}, \vec{u}_{N-2}$) is known, and a pair (χ_N^*, u_N^*) is chosen optimally. In the expression (13) from (χ_{N-1}, u_{N-1}) depends the sum $S_{N-1} = R_{N-1} + R_N^*$, where $R_N^* = \min\limits_{\chi_N, u_N} R_N$. This sum is equal to

$$S_{N-1} = \int\limits_{\Omega(\vec{\chi}_{N-2}, \vec{u}_{N-2}, \vec{y}_{N-2})} \prod_{i=1}^{N-2} (\Gamma_i^{\chi} \cdot \Gamma_i^{u}) \cdot \Phi_{N-1}(\vec{\chi}_{N-2}, \vec{y}_{N-2}, \vec{u}_{N-2}, \vec{\theta}_m) d\Omega +$$

$$+ \int\limits_{\Omega(\vec{\chi}_{N-1}, \vec{u}_{N-1}, \vec{y}_{N-1})} \prod_{i=1}^{N-1} (\Gamma_i^{\chi} \Gamma_i^{u}) \cdot \Phi_N^{**}(\vec{\chi}_{N-1}, \vec{y}_{N-1}, \vec{u}_{N-1}, \vec{\theta}_m) d\Omega = \qquad (33)$$

$$= \int\limits_{\Omega(\vec{\chi}_{N-2}, \vec{y}_{N-2}, \vec{u}_{N-2})} \prod_{c=1}^{N-2} (\Gamma_i^{\chi} \cdot \Gamma_i^{u}) \cdot F_{N-1}(\vec{\chi}_{N-2}, \vec{u}_{N-2}, \vec{y}_{N-2}, \vec{\theta}_m) d\Omega,$$

where

$$\begin{aligned} F_{N-1} &= F_{N-1}(\vec{x}_{N-2}, \vec{y}_{N-2}, \vec{u}_{N-2}, \vec{\theta}_m) = \gamma(x_{N-1}=1 \mid \vec{x}_{N-2}, \vec{y}_{N-2}, \vec{u}_{N-2}, \vec{\theta}_m)\cdot \\ &\cdot \Big\{ \int\limits_{\Omega(\eta_{N-2}, u_{N-1})} [\alpha_{N-1} + \Phi_N^{**}(\vec{x}_{N-2}, x_{N-1}=1, \vec{y}_{N-2}, \eta_{N-2}, \vec{u}_{N-1}, \vec{\theta}_m]\cdot \\ &\cdot \Gamma_{N-1}^{u}(u_{N-1} \mid \vec{x}_{N-2}, x_{N-1}=1, \vec{y}_{N-2}, \eta_{N-2}, \vec{u}_{N-2}, \vec{\theta}_m)\, d\Omega\Big\} + \qquad (34) \\ &+ \gamma(x_{N-1}=0 \mid \vec{x}_{N-2}, \vec{y}_{N-2}, \vec{u}_{N-2}, \vec{\theta}_m)\cdot \Big\{ \int\limits_{\Omega(u_{N-1})} [\beta_{N-1} + \int\limits_{\Omega(\varepsilon_{N-2})} \Phi_N^{**}(\vec{x}_{N-2}, \\ &\quad x_{N-1}=0, \vec{y}_{N-2}, \varepsilon_{N-2}, \vec{u}_{N-1}, \vec{\theta}_m)\, d\Omega]\, \Gamma_{N-1}^{u}(u_{N-1} \mid \vec{x}_{N-2}, x_{N-1}=0, \\ &\quad \vec{y}_{N-2}, \vec{u}_{N-2}, \vec{\theta}_m)\, d\Omega\Big\} \end{aligned}$$

In the expression (34) from (x_{N-1}, u_{N-1}) depends the function F_{N-1} only. Therefore optimization of S_{N-1} concerning (x_{N-1}, u_{N-1}) is reduced to optimization of F_{N-1} . Let us introduce two functions

$$\mathcal{H}_{N-1} = \alpha_{N-1} + \Phi_N^{**}(\vec{x}_{N-2}, x_{N-1}=1, \vec{y}_{N-2}, \eta_{N-2}, \vec{u}_{N-1}, \vec{\theta}_m) \quad (35)$$

$$\varphi_{N-1} = \beta_{N-1} + \int\limits_{\Omega(\varepsilon_{N-2})} \Phi_N^{**}(\vec{x}_{N-2}, x_{N-1}=0, \vec{y}_{N-2}, \varepsilon_{N-2}, \vec{u}_{N-1}, \vec{\theta}_m)\, d\Omega \quad (36)$$

It follows from the formulae (34)+(36) that the optimal control action u_{N-1}^{1*} coincides with a value u_{N-1}, which minimizes the function $\mathcal{H}_{N-1}$, and the optimal control action u_{N-1}^{0*} coincides with a value u_{N-1}, which minimizes the function φ_{N-1}.

Let us denote

$$\mathcal{H}_{N-1}^{*} = \min_{u_{N-1} \in \Omega(u_{N-1})} \mathcal{H}_{N-1} \quad , \quad \varphi_{N-1}^{*} = \min_{u_{N-1} \in \Omega(u_{N-1})} \varphi_{N-1} \qquad (37)$$

Therefore

$$F^{*}_{N-1} = \min_{u_{N-1}} F_{N-1} = \gamma(\chi_{N-1}=1 \,|\, \vec{x}_{N-2}, \vec{y}_{N-2}, \vec{u}_{N-2}, \vec{\theta}_m) \cdot \int_{\Omega(\eta_{N-2})} \mathcal{H}^{*}_{N-1}\, d\Omega + \gamma(\chi_{N-1}=0 \,|\, \vec{x}_{N-2}, \vec{y}_{N-2}, \vec{u}_{N-2}, \vec{\theta}_m) \cdot \varphi^{*}_{N-1} \qquad (38)$$

Optimization of F^{*}_{N-1} concerning χ_{N-1} is reduced to comparison of $\int_{\Omega(\eta_{N-2})} \mathcal{H}^{*}_{N-1} d\Omega$ and φ^{*}_{N-1} and to choice of lesser of them. If

$$\int_{\Omega(\eta_{N-2})} \mathcal{H}^{*}_{N-1}\, d\Omega > \varphi^{*}_{N-1}$$

the decision $\chi_{N-1} = 0$ is reached, and when

$$\int_{\Omega(\eta_{N-2})} \mathcal{H}^{*}_{N-1}\, d\Omega < \varphi^{*}_{N-1}$$

the decision $\chi_{N-1} = 1$ is reached. If these functions are equal, the choice is arbitrary.

In a similar manner other pairs (χ_{N-k}, u_{N-k}), $k=1,\dots,N-1$ are determined:

1. Two functions

$$\mathcal{H}_{N-k} = \alpha_{N-k} + F^{**}_{N-k+1}(\vec{x}_{N-k-1}, x_{N-k}=1, \vec{y}_{N-k-1}, \eta_{N-k-1}, \vec{u}_{N-k}, \vec{\theta}_m)$$

$$\varphi_{N-k} = \beta_{N-k} + \int_{\Omega(\varepsilon_{N-k-1})} F^{**}_{N-k+1}(\vec{x}_{N-k-1}, x_{N-k}=0, \vec{y}_{N-k-1}, \varepsilon_{N-k-1}, \vec{u}_{N-k}, \vec{\theta}_m)\, d\Omega$$

2. Minimizing $\mathcal{H}_{N-k}$ and φ_{N-k} concerning u_{N-k}, we find optimal control actions u^{1*}_{N-k} and u^{0*}_{N-k}, respectively.

3. Substituting u^{1*}_{N-k} and u^{0*}_{N-k} in $\mathcal{H}_{N-k}$ and φ_{N-k}, we find

$$\mathcal{H}^{*}_{N-k} = \min_{u_{N-k} \in \Omega(u_{N-k})} \mathcal{H}_{N-k} \quad ; \quad \varphi^{*}_{N-k} = \min_{u_{N-k} \in \Omega(u_{N-k})} \varphi_{N-k}$$

4. Comparing the functions φ^{*}_{N-k} and $\int\limits_{\Omega(\eta_{N-k-1})} \mathcal{H}^{*}_{N-k}\, d\Omega$ and choosing the lesser of them, we find an optimal decision x_{N-k} .

<u>An example</u>. Let $\{\eta_n\}$ be a random process which is the absense of the control meets the expression $\eta_n = \lambda n + h_n$, where $\{h_n\}$ is a sequence of independent normal values with statistics $(0, \sigma^2)$ and λ is an unknown random parameter, distributed normally with statistics (λ_0, σ_1^2). Let a random initial condition of the process $\{\eta_n\}$ is equal to η_0 . The distribution of η_0 is assumed to be normal with statistics $(0, \sigma_2^2)$. Let, for the aid of simplicity, we consider a two-stage process, N=2; at the first stage we put $x_1 = 1$. Then $y_1 = \eta_0$. In the presense of control there is formed a sequence

$$\begin{aligned} &\eta_1 = u_1 + \lambda + \eta_0 + h_1, \\ &\eta_2 = u_2 + u_1 + 2\lambda + \eta_0 + h_2, \end{aligned} \qquad (39)$$

To each term of the sequence (39) a generalized observation is put in accordance

$$y_n = x_n \eta_{n-1} + (1 - x_n)\varepsilon_{n-1}, \quad n = 1, 2,$$

Let us assume that a hypothetical sequence $\{\varepsilon_n\}$ is distributed normally with the statistics $(0, \sigma_\varepsilon^2)$. We introduce the following loss functions

$$c_{1n} = \eta_n^2 \;; \qquad c_{2n} = u_n^2 \;; \qquad c_{3n} = 1 \qquad (40)$$

and assume that on $x_n = 0$: $u_n = 0$

It can be shown that the following two strategies meet the problem:

1. $$x_1^* = 1, \quad u_1^{1*} = -0{,}6\eta_0 - 0{,}8\lambda_0;$$
$$x_2^* = 1, \quad u_2^{1*} = -0{,}5(u_1^{1*} + \eta_0) - \frac{\sigma^2\lambda_0 + \sigma_1^2(\eta_1 - \eta_0 - u_1^{1*})}{\sigma^2 + \sigma_1^2} \qquad (41)$$

The minimal total risk corresponding to this strategy is equal

$$R_{\Sigma 1}^* = 2 + 2\sigma^2 + 3\sigma_1^2 + \frac{2\sigma_1^2\sigma^2}{\sigma_1^2 + \sigma^2} + \frac{7}{5}\lambda_0^2 + \frac{3}{5}\sigma_2^2.$$

2. $$x_1^* = 1, \ u_1^{1*} = -\frac{2}{3}\eta_0 - \lambda_0; \qquad x_2 = 0, \ u_2^{0*} = 0 \qquad (42)$$

In this case the minimal total risk is equal

$$R_{\Sigma 2}^* = 1 + 2\sigma^2 + 5\sigma_1^2 + 2\lambda_0^2 + \frac{2}{3}\sigma_2^2$$

If $0 < \sigma_1^2 \leq 1/2$, the second strategy is optimal. Otherwise the first strategy is optimal.

References.

1. Пугачев В.С. Теория случайных функций и её применение к задачам автоматического управления. Изд.3-е, Физматгиз, 1962.

2. Фельдбаум А.А. Основы теории оптимальных автоматических систем. Изд.2-е, Изд-во "Наука", 1966.

3.Цыпкин Я.З.Адаптация,обучение и самообучение в автоматических системах. Автоматика и телемеханика,1966, №1.

4.Стратонович Р.Л.Условные марковские процессы и их применение к теории оптимального управления.Изд-во МГУ,1966.

5.Стратонович Р.Л.Экстремальные задачи теории информации и динамическое програмирование.Техническая кибернетика,1967,№5.

57.1

TWO-ROTOR GYROORBIT THEORY

V.A.Besekersky, Doctor of Science
V.G.Gordeev, Doctor of Science
J.G.Ostromuhov, Candidate of Science
Institute of Fine Mechanics and Optics
Leningrad
USSR

I. Introduction

A theory of a one-rotor gyroorbit, used in artificial earth satellites for making up the plane of their orbit, has been developed and published lately.[I-4] The gyroorbit together with the local vertical measuring element of infrared or any other type permits to erect on board the satellite a current orbital coordinate system, which can be used as a basis or reference coordinate system when the satellite is performing its various tasks required. These tasks may include, for instance, geodetic survey, meteorological reconnaissance, providing communication and navigation of terrestrial objects etc.

When a memory mode is needed, i.e. when working with the vertical measuring element switched off, an extra vertical gyro is erected on board the satellite, which can simultaneously serve as a smoothing device increasing the accuracy of the gyroorbit operation. One of composite gyro devices of this type is described in the literature.[2,3] This device has a certain disadvantage associated with the necessity to provide the vertical gyro precession motion in the normal orientation mode and in the gyroscopic memory mode. It results in an essential reduction of the accuracy of the gyro device operation due to the non-linearity and instability of the precession-realizing devices.

Essentially free from the above mentioned disadvantage is the two-rotor gyroorbit, reported by Tokar.[5] This gyroorbit is characterized by the use of two gyroscopes (Fig.I) with angular momenta H_1 and H_2, positioned in the orbit plane at an angle of 90° to each other.

Fig.I represents a current orbital coordinate system $x_oy_oz_o$, the axes of which are directed along the transversal, the current vertical and the binormal. Three successive rotations through the Eulerian angles of the satellite revolution - a heading (yaw) angle ψ , a pitch angle ϑ and a roll (rotation) angle γ - give a satellite-fixed coordinate system $x_c y_c z_c$. A rotation through an angle ε_ρ yields an instrumented orbit plane. A rotation through an orbital motion angle φ^* in the instumented orbit plane gives a reference coordinate system $\dot{x}_2 y_2 z_2$. The vectors of the angular momenta H_1 and H_2 are misaligned from this coordinate system by the angles α_1, β_1 and α_2, β_2.

The two-rotor gyroorbit is also a composite gyro device providing the orbit making up in the normal orientation mode and the orbit and current vertical making up in the gyroscopic memory mode. Here, however, the vectors of the angular momenta appear positioned in the inertial space (with the accuracy up to the orbit precession), which greatly reduces the gyro device errors due to the elimination of the need to produce the precession motion after a reference coordinate system has been established on board the satellite.

Further discussed is the establishing of the orbit plane. The process of the current vertical erection does not differ from that of the conventional vertical gyro except for the gyro precession motion elimination. The orbital angle φ^*(Fig. I) in this case is made up by a precise drive, the speed of which changes in accordance with the satellite circular or elliptic motion.

2. Initial equations

Taking into account only the precession theory and considering the satellite control system independent and ideal in relation to its axes, one can write the equations of motion of the first and second gyros as follows:

$$\left.\begin{aligned} H_1 p_1 &= M_{z1} + M_{Bz1} , \\ -H_1 r_1 &= M_{x1} + M_{Bx1} , \end{aligned}\right\} \qquad (I)$$

$$\left.\begin{aligned} -H_2 q_2 &= M_{z2} + M_{Bz2} , \\ H_2 q_2 &= M_{y2} + M_{By2} \end{aligned}\right\} \qquad (2)$$

where H_1 and H_2 are the angular momenta, M_{z1}, M_{x1}, M_{z2} and M_{y2} - the control torques; $M_{Bz1}, M_{Bx1}, M_{Bz2}$ and M_{By2} - the disturbing torques projected upon the correspoding axes.

The projections of the absolute angular velocity on the Resal axes are given by

$$\left.\begin{aligned} p_1 &= \| \omega_e \| \times \| e_{1j} \| + \omega_{x1} \,, \\ r_1 &= \| \omega_e \| \times \| e_{3j} \| + \omega_{z1} \,, \\ q_2 &- \| \omega_e \| \times \| e'_{2j} \| + \omega_{y2} \,, \\ r_2 &= \| \omega_e \| \times \| e'_{3j} \| + \omega_{z2} \,, \end{aligned}\right\} \quad (3)$$

$$j = 1, 2, 3,$$

where $\| e_{1j} \|$, $\| e_{3j} \|$, $\| e'_{2j} \|$ and $\| e'_{3j} \|$ are matrices determining the direction cosines; the angular velocities are determined as follows

$$\left.\begin{aligned} \omega_{x1} &= \dot{\alpha}_1 - \dot{\varepsilon}_p \sin(\varphi^* + \beta_1) \\ \omega_{z1} &= \dot{\beta}_1 \cos\alpha + \dot{\varepsilon}_p \sin\alpha_1 \cos(\varphi^* + \beta_1) \\ \omega_{y2} &= \dot{\alpha}_2 + \dot{\varepsilon}_p \cos(\varphi^* + \beta_2) \\ \omega_{z2} &- \dot{\beta}_2 \cos\alpha_2 + \dot{\varepsilon}_p \sin(\varphi^* + \beta_2) \sin \alpha_2 . \end{aligned}\right\} \quad (4)$$

The control torques for the lateral motion can be formulated according to the expressions

$$\left.\begin{aligned} M_{z1} &= -k'_1 \Delta\gamma \sin(\varphi^* + \varphi_0) - k'_2 \int_0^t \Delta\gamma \sin(\varphi^* + \varphi_0)\, dt \,, \\ M_{z2} &= -k'_3 \Delta\gamma \cos(\varphi^* + \varphi_0) - k'_4 \int_0^t \Delta\gamma \cos(\varphi^* + \varphi_0)\, dt , \\ \Delta\gamma &= \alpha_1 \cos(\varphi^* + \beta_1) - \alpha_2 \sin(\varphi^* + \beta_2) - (\gamma_B \cos\varepsilon_p + \vartheta_B \sin\varepsilon_p) \end{aligned}\right\} \quad (5)$$

where k'_1 and k'_3 are the transfer factors of the correction channel, k'_2 and k'_4 - the transfer factors of the integral correction, φ_0 - a constant angle introduced for transient improving, γ_B and ϑ_B - the signals from the local vertical indicator.

The correction signals in the orbital plane are expressed as follows

$$\left.\begin{aligned} M_{x1} &- k'_5 \frac{\beta_1 - \beta_2}{2} \\ M_{y2} &= -k'_5 \frac{\beta_1 - \beta_2}{2} \end{aligned}\right\} \quad (6)$$

The equations for the gimbal providing gyro tracking angles ε_p and φ^* are

$$\begin{aligned} \dot{\varepsilon}_p &= k_6\left[\alpha_2 \cos(\varphi^* + \beta_1) + \alpha_1 \sin(\varphi^* + \beta_2)\right], \\ \dot{\varphi}^* &= k_7 \frac{\beta_1 + \beta_2}{2}, \end{aligned} \qquad (7)$$

where k_6 and k_7 are the Q-factors of the gimbal servo-systems.

The Ox_c , Oy_c and Oz_c - axes satellite stabilization system is fed by input signals proportional to the angles

$$\begin{aligned} \Delta\psi &= \varepsilon_p, \\ \Delta\vartheta &= (\varphi_{np} - \varphi^*)\cos\varepsilon_p - [\alpha_1 \cos(\varphi^* + \beta_1) - \alpha_2 \sin(\varphi^* + \beta_2)]\sin\varepsilon_p, \\ \Delta\gamma &= [\alpha_1 \cos(\varphi^* + \beta_1) - \alpha_2 \sin(\varphi^* + \beta_2)]\cos\varepsilon_p + (\varphi_{np} - \varphi^*)\sin\varepsilon_p. \end{aligned} \qquad (8)$$

The programmed value of the φ_{np} angle for the circular or near-circular orbits can be produced according to the expression

$$\dot{\varphi}_{np} = \Omega_n + c_1 \Delta\vartheta + c_2 \int_0^t \Delta\vartheta \, dt, \qquad (9)$$

where Ω_n is the programmed value of the orbital velocity c_1 and c_2 - the transfer factors of the drive circuit producing φ_{np} ; $\Delta\vartheta = (\vartheta_B \cos\varepsilon_p - \gamma_B \sin\varepsilon_p) - \vartheta^*$ and $\vartheta^* = \varphi_{np} - \varphi^*$.

The simultaneous equations from (I) to (9) can be but computer analysed because of their high order and essential nonlinearities.

Consider only the solution of the linearized equations. Suppose that $H_1 = H_2 = H$ and designate $k_1 = k_1' H^{-1}$, $k_2 = k_2' H^{-1}$ assuming that $k_1 = k_2 = k$.

Designate the angles between the gyroscopic plane and the orbital plane measured in the horizontal and vertical planes as ε and α . These angles are essentially the unknown variables determining the accuracy of the orbit making up.

$$\left.\begin{aligned} \varepsilon &= \psi^* - \varepsilon_p - \alpha_1 \cos\varphi^* - \alpha_2 \sin\varphi^*, \\ \alpha &= \gamma^* - \alpha_2 \cos\varphi^* + \alpha_1 \sin\varphi^*. \end{aligned}\right\} \qquad (I0)$$

Consider that the angles γ^* , ϑ^* and ψ^* are small.

The equations describing the orbital plane making up and the errors of the satellite orientation angles γ^* , ψ^* can be considered independently of the equations describing the gyroscope motion in the plane, formed by the vectors of the angular momenta and the motion of the φ_{np} generating drive. Furthermore, consider the case witnout using an integral correction, that is, assume $k_3=k_4=0$.

Then the linearized simultaneous equations for the lateral motion of the gyroorbit reduce to

$$
\begin{aligned}
&(\Omega_1\varepsilon+\dot\alpha)\cos\varphi^{*}\quad \Omega_2\alpha-\dot\varepsilon)\sin\varphi^{*}=\\
&\qquad\qquad = k\gamma_B\sin(\varphi^{*}+\varphi_0)+a\\
&(\Omega_1\varepsilon+\dot\alpha)\sin\varphi^{*}+(\Omega_2\alpha-\dot\varepsilon)\cos\varphi^{*}=\\
&\qquad\qquad = k\gamma_B\sin(\varphi^{*}+\varphi_0)+b\\
&\Omega_1=\Omega_0-\dot\vartheta-\omega_{np}\cos i\\
&\Omega_2=\Omega_0-\omega_{np}\cos i\\
&a=\omega_{np}\sin i\cos\Delta\varphi+\omega_1\\
&b=\omega_{np}\sin i\sin\Delta\varphi+\omega_2\\
&\gamma_B-\alpha+\Delta\gamma_B.
\end{aligned}
\qquad (II)
$$

Here ω_{np} is the orbit precession rate, i - the orbit inclination angle, ω_1 and ω_2 - the gyro drift rates, $\Delta\gamma_B$ - the error of the local vertical measuring element.

For the case of low eccentricity orbits $(e \leqslant 0{,}3)$ one can assume $\varphi^{*}=\Omega_0 t$. Then from the simultaneous equations (II) we can obtain the equations of the Carson transforms taking into account the non-zero initial conditions

$$
\left.
\begin{aligned}
&\Omega_1\varepsilon(p)+(p+k\cos\varphi_0)\alpha(p)=A(p),\\
&-p\,\varepsilon(p)+(\Omega_2+k\sin\varphi_0)\alpha(p)=B(p),\\
&A(p)=\frac{(ap+b\Omega)p}{p^2+\Omega^2}+p\,\alpha(0)+\Delta\gamma_B k\cos\varphi_0,\\
&B(p)=\frac{(bp-a\Omega)p}{p^2+\Omega^2}+p\,\varepsilon(0)+\Delta\gamma_B k\sin\varphi_0.
\end{aligned}
\right\}
\qquad (I2)
$$

The solution of these simultaneous equations can be written as follows:

$$\varepsilon(p)=\frac{A(p)(\Omega_2+k\sin\varphi_0)-B(p)(p+k\cos\varphi_0)}{p^2+k\cos\varphi_0 p+\Omega_1(\Omega_2+k\sin\varphi_0)}\,,$$

$$\alpha(p)=\frac{B(p)\Omega_1+pA(p)}{p^2+k\cos\varphi_0 p+\Omega_1(\Omega_2+k\sin\varphi_0)}\,. \tag{I3}$$

From (I3) follows in particular the two-rotor gyroorbit stability condition $-\frac{\pi}{2}<\varphi_0<\frac{\pi}{2}$.

The characteristic equation (I3) coincides with the characteristic equation of a one-rotor gyroorbit[I-4] , if it is assumed that the correction coefficients are $k_1=k\cos\varphi_0$ and $k_2=k\sin\varphi_0$.

3. Main components of error in making up orbital plane

In accordance with equations (I2) and (I3) the ε heading and α roll errors in making up the orbital plane are expressed by

$$\left.\begin{aligned}\varepsilon&=\varepsilon_\delta+\varepsilon_{np}+\varepsilon_\gamma+\varepsilon_0\,,\\ \alpha&=\alpha_\delta+\alpha_{np}+\alpha_\gamma+\alpha_0\,,\end{aligned}\right\} \tag{I4}$$

where ε_δ and α_δ are the errors due to gyro drift; ε_{np} and α_{np} are the errors caused by orbital precession; ε_γ and α_γ are the errors due to the noise in the vertical measuring element signal; ε_0 and α_0 are the errors due to the non--zero initial conditions.

The orbital precession errors are not further discussed as they are assumed to be negligible.

Non-zero initial condition errors. For simplification suppose $\Omega_1=\Omega_2=\Omega$ that is allowed for the angle ϑ^* steady transient. Hence

$$\varepsilon_0(p)=\frac{p\alpha(0)(\Omega+k\sin\varphi_0)-p\varepsilon(0)(p+k\cos\varphi_0)}{p^2+pk\cos\varphi_0+\Omega^2+\Omega k\sin\varphi_0} \tag{I5}$$

$$\alpha_0(p)=\frac{p\varepsilon(0)\Omega+p^2\alpha(0)}{p^2+pk\cos\varphi_0+\Omega^2+\Omega k\sin\varphi_0}\,.$$

From equations (I5) it is seen that at $t \to \infty$ the angle $\varepsilon_0 \to 0$ and $\alpha_0 \to 0$.

The nature of the transient is determined by the roots of the characteristic equation

$$p_{1,2} = \frac{1}{2}\left(-k\cos\varphi_0 \pm \sqrt{k^2\cos^2\varphi_0 - 4\Omega k\sin\varphi_0 - 4\Omega^2}\right). \tag{I6}$$

In particular for the minimum time of the transient with multiple roots we have

$$\varphi_0 = \arcsin\left(1 - \frac{2\Omega}{k}\right). \tag{I7}$$

<u>Gyro drift errors</u>. From equations (I2) and (I3) we obtain

$$\begin{gathered}
\varepsilon_g(p) = \frac{\omega_2 p(p^2 + A_1 p + B_1)}{(p^2+\Omega^2)\left(p + \frac{k\cos\varphi_0}{2}\right)^2}, \\
\alpha_g(p) = \frac{\omega_1 p(p^2 + A_2 p + B_2)}{(p^2+\Omega^2)\left(p + \frac{k\cos\varphi_0}{2}\right)},
\end{gathered} \tag{I8}$$

where

$$\begin{aligned}
&A_1 = k\cos\varphi_0 - \frac{\omega_1}{\omega_2}(2\Omega + k\sin\varphi_0), \\
&B_1 = -\Omega\left(\Omega + k\sin\varphi_0 + \frac{\omega_1}{\omega_2}\cos\varphi_0\right), \\
&A_2 = 2\Omega\frac{\omega_2}{\omega_1}, \\
&B_2 = -\Omega^2.
\end{aligned}$$

From equation (I8) one can find the function originals

$$\begin{gathered}
\varepsilon_g(t) = \varepsilon_m \sin(\Omega t + \lambda) + (\varepsilon^* t + \varepsilon^{**})e^{-\frac{t}{\tau}}, \\
\alpha_g(t) = \alpha_m \sin(\Omega t + \lambda_1) + (\alpha^* t + \alpha^{**})e^{-\frac{t}{\tau}}.
\end{gathered} \tag{I9}$$

The values of the parameters for $\varepsilon_g(t)$ are

$$\varepsilon_m = \frac{\sqrt{\omega_1^2 + \omega_2^2}}{\Omega}\sqrt{1+\cos^2\varphi_0},$$

$$\lambda = \operatorname{arctg}\frac{\omega_1\sin\varphi_0 - \omega_2\cos\varphi_0}{\omega_2\sin\varphi_0 + \omega_1\cos\varphi_0} - 2\operatorname{arctg}\frac{\cos\varphi_0}{1+\sin\varphi_0},$$

$$\varepsilon^* = \omega_1 \operatorname{tg}\varphi_0 - \omega_2,$$

$$\varepsilon^{**} = \frac{\omega_1(\sin\varphi_0 - \cos\varphi_0)(1+\cos\varphi_0) - 2\omega_2\cos^2\varphi_0\sin\varphi_0}{\Omega}$$

Accordingly for $\alpha_\delta(t)$

$$\alpha_m = \frac{\sqrt{\omega_1^2+\omega_2^2}}{\Omega}(1-\sin\varphi_0)$$

$$\lambda_1 = -arctg\,\frac{\omega_2}{\omega_1} - 2\,arctg\,\frac{\cos\varphi_0}{1+\sin\varphi_0}$$

$$\alpha^* = \omega_1\sin\varphi_0 - 2\,\omega_2\cos\varphi_0$$

$$\alpha^{**} = -\frac{\omega_2\sin\varphi_0 + \omega_1\cos\varphi_0}{\Omega}(1-\sin\varphi_0)$$

It results from the above expressions that the two-rotor gyroorbit errors caused by the constant gyro drifts change according to the harmonic law with the satellite orbital period. The amplitude of the ε_m error cannot be made smaller than

$$\varepsilon_m^0 = \frac{\sqrt{\omega_1^2+\omega_2^2}}{\Omega}\,. \tag{20}$$

With φ_0 properly chosen the amplitude of the $\alpha_m$error can be reduced to any value.

Errors due to vertical measuring element noises. From equations (I2) and (I3) it follows that

$$\varepsilon_\gamma(p) = -\Delta\gamma_B\,\frac{k(\Omega\cos\varphi_0 - p\sin\varphi_0}{p^2+pk\cos\varphi_0+\Omega(\Omega+k\sin\varphi_0)}$$

$$\alpha_\gamma(p) = \Delta\gamma_B\,\frac{k(\Omega\sin\varphi_0 + p\cos\varphi_0)}{p^2+pk\cos\varphi_0+\Omega(\Omega+k\sin\varphi_0)} \tag{2I}$$

For constant noise $\Delta\gamma_B = \Delta\gamma_{B-} = const$ the stable values of the errors are

$$\varepsilon_{ycm} = -\frac{\Delta\gamma_B k\cos\varphi_0}{\Omega+k\sin\varphi_0} \approx -\Delta\gamma_{B-}\,ctg\,\varphi_0\,,$$

$$\alpha_{ycm} = \frac{\Delta\gamma_B k\sin\varphi_0}{\Omega+k\sin\varphi_0} \approx \Delta\gamma_{B-}\,. \tag{22}$$

For the case of harmonic noise in the vertical measuring element signal $\Delta\gamma_B = \Delta\gamma_m\sin\omega_B t$ or when the satellite has a rolling hunting in the insensitivity region of the vertical measuring element the gyroorbit errors will also change according to the harmonic law.

$$\varepsilon_{m\gamma}=\Delta\gamma_m k\sqrt{\frac{\Omega^2\cos^2\varphi_0+\omega_B^2\sin^2\varphi_0}{[\Omega(\Omega+k\sin\varphi_0)-\omega_B^2]^2+k^2\omega_B^2\cos^2\varphi_0}}\;,$$

$$\alpha_{m\gamma}=\Delta\gamma_m k\sqrt{\frac{\Omega^2\sin^2\varphi_0+\omega_B^2\cos^2\varphi_0}{[\Omega(\Omega+k\sin\varphi_0)-\omega_B^2]^2+k^2\omega_B^2\cos^2\varphi_0}}\;. \qquad (23)$$

It is interesting to note a particular case when the vertical error is caused by the Earth ellipticity. Then

$$\Delta\gamma_B=\operatorname{arctg}\frac{a}{b}\frac{x}{\sqrt{a^2+x^2}}-\operatorname{arctg}\frac{b}{a}\frac{x}{\sqrt{a^2+x^2}}\;, \qquad (24)$$

where a - the Earth equatorial radius; b - the Earth polar radius; x - the coordinate of a point on the Earth surface, measured in the equatorial plane.

The coordinate that corresponds to the maximum error is

$$x_m=\frac{a}{\sqrt{2}}\;. \qquad (25)$$

Hence

$$\Delta\gamma_{max}=\operatorname{arctg}\frac{a}{b}-\operatorname{arctg}\frac{b}{a}\;. \qquad (26)$$

For Krasovsky's ellipsoid

$$\Delta\gamma_B=\Delta\gamma_{max}\sin\varphi \qquad (27)$$

where $\Delta\gamma_{max}$ = II ang.min; φ - an orbital angle determined from the nodal line. From equations (23) one can find $\varepsilon_{m\gamma}$ = $\alpha_{m\gamma}$ = $\Delta\gamma_{max}$ = II ang.min for this case.

Effect of random noise.Assume that the spectral density of the vertical measuring element noise can be written as

$$S_B(\omega)=\frac{2\Delta\gamma_{ck}^2 T_k}{1+\omega^2T_k^2}$$

Then, expressing the spectral densities of the errors from equations (21) and integrating them for all the frequencies, we obtain

$$\varepsilon_{ck}^2=\Delta\gamma_{ck}^2\,\frac{T_k(k_1^3+k_2^2\Omega+k_2^3T_k\Omega)}{k_1k_2(1+k_2T_k+k_1T_k^2\Omega)}\approx$$

$$\approx\Delta\gamma_{ck}^2\,T_k\,k\,\operatorname{tg}\varphi_0\sin\varphi_0\;, \qquad (28)$$

$$\alpha_{ck}^2 = \Delta\gamma_{ck}^2 \frac{T_k(k_2^2 + k_1\Omega + k_1 k_2 T_k \Omega)}{k_2(1 + k_2 T_k + k_1 T_k \Omega)} \approx$$

$$\approx \Delta\gamma_{ck}^2 T_k k \cos\varphi_0 , \qquad (29)$$

where $k_1 = k\cos\varphi_0$ and $k_2 = k\sin\varphi_0$.

The above expressions can be used to determine the correction factor k which provides the desired noise smoothing of the current vertical measuring element.

4. Conclusions

I. The principal dynamic properties of a two-rotor gyroorbit are similar to those of a composite gyro device, containing a one-rotor gyroorbit and a precessing vertical gyro. In contrast to the latter, however, the two-rotor gyroorbit operates under more favourable conditions and does not contain the errors associated with the necessity to provide the vertical gyro precession motion. It brings about an essential increase of the accuracy in the gyroscopic memory mode.

2. Constant gyro drifts result in orbit plane making up errors which change according to a harmonic law. It gives an evidence of a modulating property of the two-rotor gyroorbit. The different spectra of the one-rotor ond two-rotor gyroorbits permit to achieve their effective combining.

3. The two-rotor gyroorbit is capable to provide the required smoothing of the noise contained in the local vertical measuring element signal, which increases the accuracy of making up the reference coordinate system on board the satellite.

References

I. Роберсон Р.Э. Измерение угла рыскания спутника с помощью гироскопа. Труды ИФАК.АН СССР. I960.
(Measuring Satellite Yaw Angle with the Help of a Gyroscope. Trans. IFAC. Academy of Sciences. USSR. 1960).

2. Алексеев К.Б., Бебенин Г.Г. Управление космическими летательными аппаратами, Изд."Машинстроение", I964.
(Attitude Control of Space Vehicles. Published by "Mashino - stroenije").

3. Chatkoff M.L., Lynch L.G. Attitude Control of a Space Vehicle by a Gyroscopic Reference Unit. Aero Space Engineering, No 5, 1960.

4. Раушенбах Б.В., Токарь Е.Н. Некоторые вопросы управления в межпланетном пространстве, сб. "Искусственные спутники Земли", АН СССР, вып.5, I960.
(Some Control Problems in Space. A book of collected articles "Iskusstvennie Sputniki Zemli". Academy of Sciences. USSR. Issue 5, 1960).

5. Токарь Е.Н. Возможные принципы ориентации космического аппарата относительно вращающейся системы координат, "Космические исследования", том IУ, вып. 3, I966.
(Possible Principles of Space Vehicle Orientation Relative to a Rotating Coordinate System. Book "Kosmicheskije Issledovanija", v. IV, iss.3, I966).

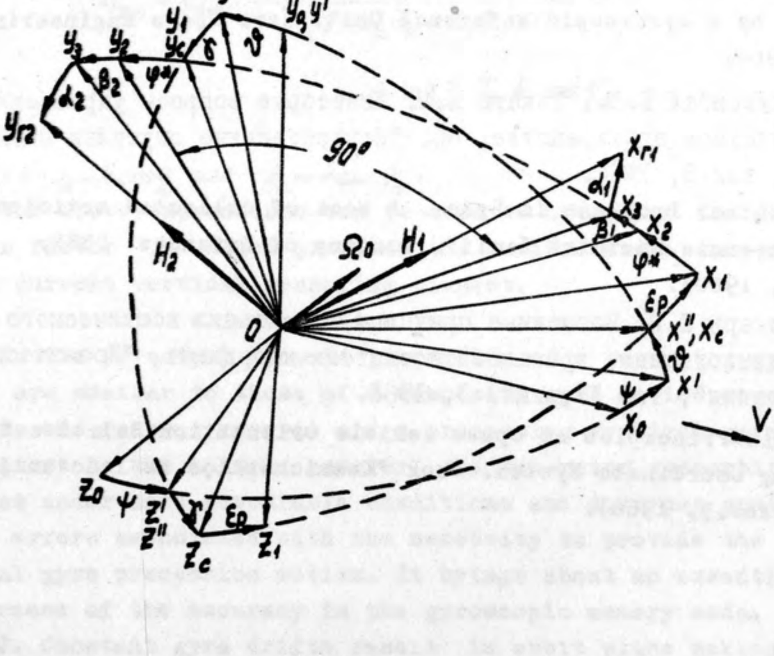

y''
y1
y0, y'
y3
y2
yc
γ
ϑ
β2
ψ**
α2
yr2
90°
xr1
α1
x3
β1
x2
ψ*
x1
H2
Ω0
H1
O
εp
x'', xc
ϑ
x'
ψ
x0
V
z0
ψ
z'
γ
εp
z''
z1
zc

Fig.1

57.2

INVESTIGATION OF MULTIPLEX AUTO-OSCILLATIONS OF SPACECRAFT

E.V.Gaushus.

Moscow, USSR.

A specific feature of motion in space is nearly complete absence of damping forces, therefore the spacecraft free motion about the centre of mass is mainly conservative. Apart from controlling moments of the active control system the spacecraft is affected by comparatively small disturbing moments which, however, significantly influence the spacecraft motion.

Under the influence of disturbing moments the spacecraft usually undergoes multiplex asymmetric auto-oscillations.

These facts along with string requirements of minimizing the control system propellant consumption lead to a necessity of precise and strict investigation of the spacecraft stationary motion, which often excludes the possible use of approximate methods.

The most effective method of the spacecraft dynamics study is a method of point transforms /1/,/2/. It should be noted that the necessity of taking into account the main non-idealities of the control system even in the simpliest case of plane oscillations leads to the use of self-contained dynamic systems with multivalent phase plane.

Multiplex periodic motions which are represented by corresponding closed trajectories with multiple intersections in a phase plane may exist in such systems.

The theoretical problems of multiplex periodic motion study are considered in /3/.

The spacecraft plane oscillations about the centre of mass are considered below, i.e.,the equation describing the motion about one of its axes of inertia is supposed to be

$$J\frac{d^2\varphi}{dt^2} = \mathcal{H}(u)M_{\text{упр}} - M_{\text{в}} , \qquad (1)$$

where I - moment of inertia,

$M_{\text{упр}}, M_{\text{в}}$ - controlling and disturbing moments, respectively,

φ - deviation angle from given direction,

$\varkappa(u)$ - controlling function.

The control system forms a control signal $u = m\frac{d\dot{\varphi}}{dt} + n\varphi$. The control function is of a relay responce type with a dead zone 2a and a hysteresis loop b. The disturbing moment is supposed to be constant in magnitude and direction. Changing variables system (1) may be reduced to

$$\frac{dz}{d\tau} = \varkappa(\sigma)F - \frac{1}{4}p\,, \quad \frac{dy}{d\tau} = z \tag{2}$$

$$\tau = 2sz + y\,,$$

where

$$\tau = \frac{2nb}{m(a+b)}t\ ; \quad y = \frac{n}{a+b}\varphi\,; \quad \sigma = \frac{1}{a+b}\delta\,,$$

$$p = \frac{\tilde{M}m^2(a+b)}{Jnb^2}\ ; \quad s = \frac{b}{a+b}\ ; \quad F\left[\frac{4nb^2J}{Mm^2(a+b)}\right]^{-1}.$$

Construction of Point Transform.

Let us choose an edge of the first sheet of a phase plane as a segment without contact and construct the transform of this line (L) into itself. As a coordinate of this straight line we shall use

$$x = z - \frac{1}{2}ps\,.$$

There are several types of phase trajectories which perform point transform T of the straight line L into itself (Fig.1) and, hence, the correspondence function f(x) of this transform has the first-order discontinuities.

It is easily seen that to find periodic motions it is sufficient to consider only two branches of this function defined by phase trajectories A and B (Fig.1).

Correspondence functions are

$$\theta(X) = -\sqrt{X^2 + Q}$$
$$u(X) = -\Omega + \sqrt{(X+\Omega)^2 + 2\Omega - Q} \tag{3}$$

for trajectory A and

$$\alpha(X) = \sqrt{X^2 - E}$$
$$v(X) = \Omega - \sqrt{(X-\Omega)^2 + 2\Omega + Q}$$
$$\beta(X) = -\sqrt{X^2 + E}$$
$$u(X) = -\Omega + \sqrt{(X+\Omega)^2 + 2\Omega - Q}\ , \tag{4}$$

for trajectory B where $Q = ps$; $E = p(1-s)$; $\Omega = 2Fs$.

Further investigations will use the following parameters

$$Q, E \quad , \quad H = \frac{1}{\Omega}$$

Let us denote transforms defined by the trajectories A and B as T_ε and T_η and the correspondence functions of these transforms as $\varepsilon(X)$ and $\eta(X)$, respectively

$$\varepsilon(X) = u\,[\Theta(X)]$$

$$\eta(X) = u[\,\beta\{v[\alpha(X)]\}] \tag{5}$$

Thus, the correspondence function f(x) of the point transform is

$$f(X) = \begin{matrix} \varepsilon(X) & \text{at} & X < \sqrt{E} \\ \eta(X) & \text{at} & X > \sqrt{E} \end{matrix} \tag{6}$$

Let us analyse first the dynamics of system (2) at very large controlling moments taking H=0. When H=0 the correspondence functions(u) and (v) are substantially simplified and become

$$u\,(X) = 1 + X$$

$$V\,(X) = 1 - X$$

and the correspondence functions $\varepsilon(X)$ and $\eta(X)$ may be written as

$$\varepsilon\,(X) = 1 - \sqrt{X^2 + Q}$$

$$\eta\,(X) = 1 - \sqrt{(1 - \sqrt{X^2 - E}\;)^2 + E} \tag{7}$$

The multitude òf possible periodic motions of dynamic system (2) is defined by the aggregate of fixed points (simple and multiple) of the point transform T.

Multiple fixed points of the transform T are determined as simple fixed points of the transforms which represent all possible sequences of the transforms T_ε and T_η:

$$T_K = T_\varepsilon^{i_1}\, T_\eta^{i_3}\, T_\eta^{i_4} \;\ldots \tag{8}$$

Let us consider first the transform:

$$\Pi_n = T_\varepsilon\, T_\eta^{\,n-1} \tag{9}$$

Definition of Periodic Motions.

The point transform T_h has no fixed points, for at $E>0$ $h(x)<x$

The transform T_ε has one simple fixed point

$$x_* = \frac{1}{2}(1-a). \tag{10}$$

This point is stable at any values of parameters, for its characteristic root

$$\lambda = \frac{d\varepsilon}{dx}\Big|_{x=x_*} = -\frac{1+a}{1-a} \tag{11}$$

does not exceed unity in module.

The bifurcation of the fixed point x_* may occur only at the transition through the T-transform discontinuity boundary, i.e., at

$$\frac{1}{2}(1-a) = \sqrt{E}$$

Thus, at $a>1-2\sqrt{E}$ the point transform T has the only stable simple fixed point, and at $a<1-2\sqrt{E}$ the transform T has no simple fixed points.

The limit cycle corresponding to fixed point (10) is given in Figure 3.

To find the fixed points of the transform $\Pi_n = T_\varepsilon T_h^{-1}$ we shall analyse the correspondence functions of these transforms

$$f_n(x) = h_{n-1}[\varepsilon(x)] \tag{12}$$

where $h_{n-1}(x)$ is the (n-1)th iteration of the function $h(x)$.

As the transform T_h has no fixed points it is enough to consider the transform Π_n in the interval $(-\sqrt{E}, +\sqrt{E})$. In spite of the T-transform discontinuity it will allow to use some theorems of /3/ for continuous transforms.

Let us consider the main properties of the transforms Π_n and the correspondence functions $f_n(x)$.

The function $f_n(x)$ is precise due to the $\varepsilon(x)$ function precision. The correspondence function $f_n(x)$ has the only maximum at the point x=0.

The derivative function $f_n(x)$ is

$$\frac{df_n(x)}{dx}=\frac{d\varepsilon}{dx}\prod_{i=1}^{n-1}\frac{d\eta}{dx}\left[f_i(x)\right] \tag{13}$$

and is the monotonically decreasing function $\left(\frac{d^2 f_n(x)}{dx}<0\right)$.

The fixed points of the transform Π_n are found from the equation

$$f_n(x)-x=0. \tag{14}$$

It follows from the said that this equation has no more than two roots, i.e., the transform Π_n has no more than two fixed points.

The problem of determining the fixed points of the transform Π_n is closely connected with the investigation of the dependence of these transforms on the parameters Q and E.

We shall carry out this investigation changing Q at a fixed E. The correspondence function $f_n(x,E,Q)$ is a monotonically decreasing function in the parameter $Q\left(\frac{\partial f_n}{\partial Q}<0\right)$.

To analyse the point transforms $\Pi_n=T_\varepsilon T_\eta^{n-1}$ let us introduce the transforms T_ε^{-1} and T_η^{-1} inverse to the transforms T_ε and T_η and the functions which are inverse to the functions $\varepsilon(x)$ and $\eta(x)$. We shall denote them as $\bar\varepsilon(x)$ and $\bar\eta(x)$. It should be noted, generally speaking, that these functions are not one-valued, but later on their positive branches will be used.

$$\bar\varepsilon(x)=\sqrt{(1-x)^2-Q}$$
$$\bar\eta(x)=\sqrt{(1-\sqrt{(1-x)^2-E})^2+E} \tag{15}$$

It is evident that the n-th iteration of the $\bar\eta(x)$ function is inverse to the n-th iteration of the function $\eta(x)$.

The inverse function $\bar\eta(x)$ and its iteration may be represented by the straight lines as follows

$$\bar\eta_n(x)=1-\eta_n(1-x) \tag{16}$$

The correspondence function of the transform $\Pi_n^{-1}=T_\eta^{-(n-1)}T_\varepsilon^{-1}$ is

$$\bar f_n(x)=\bar\varepsilon\left[\bar\eta_{n-1}(x)\right]. \tag{17}$$

Each transform Π_n is determined only at sufficiently small values of the parameter Q.

At $Q>(1-\sqrt{E})^2$ for any $x\in(-\sqrt{E},+\sqrt{E})$, $\varepsilon(x)\in(-\sqrt{E},+\sqrt{E})$ and, therefore, the transforms Π_n and the function $f_n(x)$ are not defined except for the transform Π_1 with the correspondence function $\varepsilon(X)$.

At $Q=(1-\sqrt{E})^2$ the transform with the correspondence function $f_2(X)=h[\varepsilon(X)]$ is born at the point x=0.

The functions $\varepsilon(X)$ and $f_2(X)$ increase with the further decrease of Q, the determination zone of the latter expanding. When $f_2(0)=\sqrt{E}$ the transform Π_3 with the correspondence function $f_3(X)=h\{h[\varepsilon(X)]\}$ is born at the point x=0, etc.

It follows that the parameters for any transform Π_n being changed, the following bifurcation moments of its topology changes may take place:

moment A_n corresponding to the Π_n transform birth;

moment B_n corresponding to the fixed points of the Π_n transform birth;

moment C_n corresponding to the death of a smaller fixed point at the transition through the Π_n transform discontinuity boundary;

moment D_n corresponding to the expansion of the determination zone of the transform Π_n up to the interval $(-\sqrt{E},+\sqrt{E})$. In this case the larger fixed point of the transform Π_{n-1} will die.

moment F_n corresponding to obtaining the value of $\sqrt{E}$ by the correspondence functions $f_n(x)$, which is accompanied by the Π_{n+1} transform birth;

moment H_n corresponding to the death of the larger fixed point of the transform Π_n at the transition through the discontinuity boundary $X=+\sqrt{E}$.

These moments are illustrated in Figure 4.

The equations of the bifurcation curves corresponding to the moments mentioned above may be written in an explicit form except for the curve B_n which can be obtained from the equations

$$f_n(X)-X=0$$

$$\frac{df_n}{dX}-1=0. \qquad (18)$$

The bifurcation curves of the rest moments are

$$\begin{aligned}
&A_n \quad \eta_{n-2}[\varepsilon(0)] = \sqrt{E}\\
&C_n \quad \eta(\sqrt{E}) = -\bar{\varepsilon}[\bar{\eta}_{n-2}(\sqrt{E})]\\
&\text{Д}_n \quad \eta_{n-2}[\varepsilon(\sqrt{E})] = \sqrt{E}\\
&F_n \quad \eta_{n-1}[\varepsilon(0)] = \sqrt{E}\\
&H_n \quad \eta_{n-1}[\varepsilon(\sqrt{E})] = \sqrt{E}
\end{aligned} \tag{19}$$

These equations may be solved in terms of the parameter Q as follows

$$\begin{aligned}
&Q_{A_n} = [1-\bar{\eta}_{n-2}(\sqrt{E})]^2\\
&Q_{C_n} = [1-\bar{\eta}_{n-2}(\sqrt{E})]^2 - (1-\sqrt{1+E})^2\\
&Q_{\text{Д}_n} = [1-\bar{\eta}_{n-2}(\sqrt{E})]^2 - E\\
&Q_{F_n} = [1-\bar{\eta}_{n-1}(\sqrt{E})]^2\\
&Q_{H_n} = [1-\bar{\eta}_{n-1}(\sqrt{E})]^2 - E .
\end{aligned} \tag{20}$$

The analysis of these curves shows that the bifurcation moments A_n.....H_n follow as it was listed above. The important fact of the coincidence of the moments A_n and F_{n-1} as well as D_n and H_{n-1} should be noted.

It follows that no more than two transforms Π_n may exist simultaneously, i.e., only the points of either the (n+1)th or (n-1)th multiplicity may exist simultaneously with the fixed points of the n-multiplicity.

The bifurcation curve form also shows that for any n the curves A_n.....H_n exist in the $Q>0$ zone only at sufficiently small values of E and when $E \to 0$ $n \to \infty$.

Thus, any multiplex periodic motions may occur at different values of the parameters of the system (fixed points of any multiplicity from 1 to ∞).

Investigation of Stability of Multiplex Periodic Motions

Bifurcations of the point transforms Π_n connected with the existence of multiplex periodic motions were considered earlier. Now let us analyse bifurcations of stability of these motions. The stability of the fixed point C of the transform Π_n is determined by the value of the correspondence function $f_n(x)$ at this point /3/. Let us call this value a characteristic root and denote it as λ_n. The derivative of the function $f_n(x)$ is

$$\frac{df_n}{dx} = \frac{d\varepsilon}{dx}(x)\prod_{i=1}^{n-1}\frac{dh}{dx}\left[f_i(x)\right] \tag{21}$$

It should be noted that the fixed points equation (14) cannot be solved in terms of x.

Therefore to study the stability let us exclude the parameter Q from the characteristic root and use the coordinate of the fixed point C instead of it.

After transforms the characteristic root may be written as follows

$$\lambda = -\frac{C}{1-C}\left[\prod_{i=1}^{n-1}\left(\frac{1}{\bar{h}_i(C)}-1\right)\left(\frac{1}{\sqrt{[1-\bar{h}_{i-1}(C)]^2-E}}-1\right)\right]^{-1} \tag{22}$$

This expression is convenient for constructing the bifurcation boundaries B_n.

The characteristic root λ_n is a monotonic function of the parameter C and, thus, it takes extreme values of x on the boundaries of the zone of changing C. At the moment B_n a semi-stable fixed point is born (with $\lambda_n = +1$) which then is split into two points, the smaller one being unstable (its characteristic root $\lambda_n > 1$).

The characteristic root of the larger fixed point decreases with Q. The range of changing Q is limited by the value of Q=0 and, thus, the coordinate C is limited by the value of $C=C_0$ corresponding to the value of Q=0.

Let us define λ_n at the point $C=C_0$. At Q=0 the fixed point equation (14) may be solved in terms of C

for $\quad n = 2K+1 \quad , C_0 = h_K(1/2) \qquad K = 1, 2, 3 \ldots$

for $\quad n = 2m \qquad C_0 = h_{m-1}\left(1-\sqrt{1/4-E}\right) \quad m = 1, 2, 3 \ldots$ (23)

Substituting (23) into (22) after these transforms gives

$$\lambda(C_0) = -1 \tag{24}$$

This means that with varying the parameters Q and E the fixed points of the point transforms $\Pi_n = T_\varepsilon T_h^{\,n-1}$ are born and die without changing their stability, the smaller point being always unstable and the larger one being always stable.

On Existence of Other Types of Periodic Motions.

The multiplex periodic motions corresponding to the fixed points of the transforms $\Pi_n = T_\varepsilon T_\eta^{n-1}$ were considered above. The results received allow to turn to the question of existence of other types of periodic motions.

Let us divide all possible sequences (8) of the point transforms T_ε and T_η into the following classes:

I - transforms $\Pi_n = T_\varepsilon T_\eta^{n-1}$;

II - transforms containing T_ε^κ as a factor where $\kappa > 1$;

III- transforms containing a group of factors $T_\varepsilon T_\eta^n T_\varepsilon T_\eta^m$; $(n \neq m$

IV - transforms $(T_\varepsilon T_\eta^{n-1})^\kappa$ which are multiplicities of the first class transforms.

The first class transforms were considered above. The second class transforms do not exist as the derivative of the correspondence function $\varepsilon(X)$ does not exceed unity in module.

The third class transforms for $|n-m| > 1$ do not exist due to coincidence of the bifurcations A_n and F_{n-1} as well as D_n and H_{n-1}. Let us consider the transform $T_\varepsilon T_\eta^{n-1} T_\varepsilon T_\eta^n$. If the parameters Q and E are such that the transforms $T_\varepsilon T_\eta^{n-1}$ and $T_\varepsilon T_\eta^n$ exist simultaneously then the first of them is defined in the interval $(-\nu, +\nu)$. , and the second one is defined in the interval $(-\sqrt{E}, \nu)$ and $(+\nu, +\sqrt{E})$ where $\nu = \bar{\varepsilon}[\bar{\eta}_{n-1}(\sqrt{E})]$. The existence of the transforms $T_\varepsilon T_\eta^{n-1} T_\varepsilon T_\eta^n$ is conditioned by the existence of such points in the interval $(+\nu, +\sqrt{E})$ which are transferred into the interval $(-\nu, +\nu)$ by means of the transforms $T_\varepsilon T_\eta^{n-1}$. It should be noted that for any Q and E in the zones considered the inequality $f_n[f_n(\sqrt{E})] \geqslant \sqrt{E}$ is valid, and the equality exists only on the bifurcation boundary H_n.

It follows that due to the monotonic function $f_n(x)$ in the interval $(\nu, \sqrt{E})$ the transforms $T_\varepsilon T_\eta^{n-1} T_\varepsilon T_\eta^n$ do not exist for any $X \in (\nu, \sqrt{E})$ $f_n(X) \in (\nu, E)$.

Let us consider the existence of two-fold fixed points of the transforms Π_n . The second iteration of the correspondence function $f_n(x)$ is an even function having no more than three extrema (x=0 and roots of the function $f_n(x)$).

It was already shown that the characteristic root λ_n does not exceed unity in module. It follows that the inequality

$f_n[f_n(\sqrt{E})] > \sqrt{E}$ being taken into account, the transform Π_n either has no two-fold points at all or has an even number of two-fold cycles. On the other hand, the function $f_n[f_n(X)]$ has no more than one bending point and the existence of two-fold cycles requires at least three points.

Hence, the transform Π_n has no two-fold cycles and basing on the theorem 4 /3/ we may conclude that it has no cycles of greater multiplicity either.

Thus, the multitude of possible periodic motions of the dynamic system /2/ is exhausted by the determined fixed points of the point transforms $\Pi_n = T_\varepsilon T_\eta^{n-1}$.

The dynamic system may have either one or two stable limit cycles (which are neighbouring in a multiplicity value) depending on E and Q values. The multiplicity of the fixed points may take any values from 1 to ∞ depending on the parameters, i.e., any multiplex periodic motions may take place in equation (2). Zones of multiplex periodic motion existence and stability are presented in the bifurcation diagram of Figure 6.

It should be noted that the value Q=0 is bifurcate. Really, at Q=0 the function $\varepsilon(X)=1-X$ and the correspondence function of the transform Π_n is $f_n(X) = \eta_{n-1}(1-X)$.

Using (16) the second iteration of this function

$$f_n[f_n(X)] = \eta_{n-1}[1-\eta_{n-1}(1-X)]$$

may be written as

$$f_n[f_n(X)] = \eta_{n-1}[\bar{\eta}_{n-1}(X)] = X$$

Thus, the point transform $\Pi_n^2 = [T_\varepsilon T_\eta^{n-1}]^2$ is identical at Q=0.

This means that a continual multitude of the fixed points of 2n multiplicity is born out of a stable n-fold fixed point of the transform T. This case is analysed in /4/ in detail.

H - Dependence Analysis.

Dynamic system (2) was considered above at H=0. When H is other than zero the problem becomes more complicated. Therefore, we shall give only main qualitative results without carrying out complete and strict analysis. It is easy to see that the correspondence function $\eta(X)$ increases with H, and

H being large enough, we shall have the inequality $\hbar(\sqrt{E})>0$. It follows that in this case the diagram of bifurcations $A_n \ldots\ldots H_n$ ceases to exist, for at the moment of birth of the transform Π_n at the point x=0 the value of the correspondence function $f_n(X)=\hbar(\sqrt{E})>0$. Consequently, the transform Π_n has no more than one fixed point, this one being unstable on the boundary of the Π_n transform definition zone at the moment of its birth, for its characteristic root at the moment of its birth equals to $-\infty$.

Further change of parameters causes increasing the characteristic root up to the value of $\lambda_n=-1$. At the moment of Π_n transform fixed point stability bifurcation an unstable two-fold cycle of the transform Π_n is born at this point, i.e., the fixed points of the transform $\Pi_n^2=(T_\varepsilon T_\hbar^{n-1})^2$. Then this unstable two-fold cycle increases and dies at the transition through the discontinuity boundary $X=\sqrt{E}$. With further change of parameters the stable fixed point of the transform Π_n also increases and dies at the transition through the discontinuity boundary of the transform T $X=\sqrt{E}$. Such is the process of birth and death of the Π_n transform fixed points at sufficiently large values of H.

It should be noted that in contrast to the case H=0 the transform $T_\hbar$ has fixed points at sufficiently large values of H. At some values of the parameters the function $\hbar(x)$ touches the coordinate angle bisectrix. This is accompanied by rather unusual bifurcation, i.e., a simple periodic motion (of the fixed points of the transform $T_\hbar$) is born out of an infinitely multiplex one (infinitely multiple fixed points of the transform T). This bifurcation is shown in Figure 7. Thus, as parameter H increases the fixed points of the transform Π_n of higher multiplicities are born according to the scheme given above. When H approaches some value of H_* the multiplicity of the fixed points of the transform T tends tc infinity and at $H=H_*$ the infinitely multiple cycle dies and a simple semi-stable fixed point of the transform $T_\hbar$ is born which is then split into two points: stable and unstable. With further increase of H the unstable point dies at the transition through the discontinuity boundary $X=\sqrt{E}$.

At values of H large enough but not exceeding $\frac{2}{Q}$ the transform T has only simple fixed points which correspond either to the transform T_{ε} or the transform $T_{\flat}$ (Fig.8). At $H > \frac{2}{Q}$ the system becomes absolutely unstable, for in this case the disturbing effect exceeds the controlling one.

REFERENCES.

I. Андронов А.А., Витт А.А., Хайкин С.Э. Теория колебаний. Гостехиздат I954 г.

2. Неймарк Ю.И. Метод точечных преобразований в теории нелинейных колебаний. Известия ВУЗов. Радиофизика т.I, №№ I,2,5-6, I958г.

3. Гаушус Э.В. К теории точечного преобразования. Автоматика и телемеханика, I966, № I2.

4. Гаушус Э.В. Исследование одного типа автоколебаний космического аппарата. Искусственные спутники Земли,I963, вып. I6.

FIGURES

Figure 1. The phase plane of the dynamic system.

Figure 2. The correspondence function of the transform T.

Figure 3. The limit cycle corresponding to the fixed point of the transform T_{ε} .

Figure 4. The bifurcation of the point transforms $\Pi_{\flat}$.

Figure 5. The limit cycle corresponding to the fixed point of the transform Π_{ε} .

Figure 6. The existence and stability zones of multiplex periodic motions.

Figure 7. The birth of a simple periodic motion out of an infinitely multiplex one.

Figure 8. The limit cycle corresponding to the transform $T_{\flat}$.

57.3

DYNAMIC STUDIES OF PRELIMINARY STABILIZATION SYSTEM OF A GRAVITY-STABILIZED SATELLITE WITH TAKING INTO ACCOUNT TRANSDUCERS CONSTRAINTS AND BENDING OSCILLATIONS OF STABILIZER.

V.I. Popov
Moscow Bauman Higher Engineering School,

V.Yu. Rutkovskii
Institute of Automation and Telemechanics (Engineering Cybernetics)

Moscow

USSR

A gravity stabilization system (GSS) is suggested for orientation of an artificial earth satellite (AES)[1,2]. In order to obtain a gravity-stabilized satellite a stabilizer is attached to it in the form of one or two bars with weights at the ends.

Initial Perturbations of the satellite at the moment of its separation from carrier rocket are such that GSS is not able to prevent the spin of the satellite. For some projects it is required that the satellite be quickly stabilized, after its separation, with the aid of a passive or active preliminary stabilization system (PSS).

The passive PSS's normally use force fields for creation of control moments[2]. The active PSS's can be constructed with the use of rotatable wheels or gas-reaction nozzles[3,4]. Other principles of constructing passive or active PSS's are possible.

Before the moment of separation of the satellite from carrier rocket the gravity stabilizer is in the folded position and rigidly connected with the satellite body. After separation of the satellite the stabilizer must be open which is possible to do either right after the satellite separation or after the end of operation of the PSS.

This paper analyzes the dynamics of a relay gas-reaction PSS in the phase plane with regard to constraints of transducers. The question of use of transducers constraints in formation of nonlinear control rules is considered. In order to decrease the amplitude of auto-oscillations in the PSS having a relay characteristic with hysteresis loop it is suggested that the internal feedback compensating the lag should be included. Spin motions of the satellite with opened bars are studied. The equations are derived of flat bending oscillations in the satellite-stabilizer system and the effect of natural and artificial damping in bars and bending oscillations of the system during the operation of the PSS are studied.

1. Equations of Motion of the PSS

For damping of initial disturbances occuring at the moment of satellite separation from carrier rocket we shall use an active relay PSS whose control moments are created by gas--reaction mozzles. As an initial position relative to which the satellite must stabilize we choose the orbital coordinate system $0X_oY_o\ Z_o$. Introduce for consideration the bound coordinate system 0_{xyz} whose axes are directed along the main central axes of satellite's inertia. The axes of the system $0xyz$ in the set position of equilibrium of the AES coincide with the axes of orbital coordinate system. Both systems have their origin in the center of mass 0 of the satellite. It is known that the position of the axes $0xyz$ relative to $0x_oy_oz_o$ is determined by setting three independent angles: ϑ - pitch angle, Ψ - yaw angle, and γ - angle of roll.

In considering the satellite's motion round the center of mass, when the PSS operates, we shall neglect all the perturbing moments due to their smallness compared with control moments. The equations of motion of the satellite round the center of mass in the bound coordinate system have conventional form of the Eulerian dynamic equations.

In a general case the control moment affecting the satellite depends on the angles γ, ψ, ϑ , on projections of absolute angular spin velocity onto the bound axes $\omega_x, \omega_y, \omega_z$ and on time t that is

$$M = \Phi(\gamma, \psi, \vartheta, \omega_x, \omega_y, \omega_z, t)$$

Thus, three Eulerian dynamic equations combine six independent functions $\gamma, \psi, \vartheta, \omega_x, \omega_y, \omega_z$. In order to make the problem definite it is necessary to obtain three more equations with the help of Eulerian kinematic equations which would establish the connection between $\omega_x, \omega_y, \omega_z$, the angles γ, ψ, ϑ and the orbital velocity ω_0 .

Since a short-term operating PSS is considered with sufficiently great control moments at low angular velocities $\omega_x, \omega_y, \omega_z$ and angular deviations of the satellite the Eulerian dynamic and kinematic equations are simplified. Linearization of these equations leads as is known to division of the Eulerian dynamic equations into three independent equations of the second order

$$\mathcal{J}\ddot{x} = -M, \tag{1}$$

where $\mathcal{J}$ - moment of inertia with respect to the desired axis, x - satellite angular deviations, M - control moment.

This paper discusses the results of dynamics study of the PSS on the basis of equation (1).

Consider the PSS motion equations. Combine relay characteristics of the output cascade of an inertia-free amplifier and electropneumatic valve into one nonlinear function $\Phi_\tau(\sigma)$ containing the total controller lag τ . The equivalent relay characteristic $\Phi(\sigma)$ may have the dead zone or hysteresis loop. The argument of the function $\Phi_\tau(\sigma)$ is the output voltage from the linear cascade of the amplifier. The amplifier is assumed to be perfect.

We shall take into account the constraints in transducers characteristics.

If the amplifier receives the feedback signal $K_3\,\mathrm{sign}\,\Phi(\sigma)$ ~~Sign~~ compensating the controller lag then the equation for input coordinate of relay characteristic may be written in the following form

$$\sigma = K_1 \rho_1(x) + K_2 \rho_2(\dot{x}) + K_3 \, \mathrm{sign}\, \Phi(\sigma),$$

where K_1, K_2, K_3 - constant coefficients,

$\rho_1(x)$ - signal from the angle transducer,

$\rho_2(\dot{x})$ - signal from the angular velocity transducer.

The equations for $\rho_1(x)$ and $\rho_2(\dot{x})$ are of the form

$$\rho_1(x) = \begin{cases} x_{lim} & for \quad x > x_{lim}, \\ x & for \quad |x| \leq x_{lim}, \\ -x_{lim} & for \quad x < -x_{lim}, \end{cases}$$

$$\rho_2(\dot{x}) = \begin{cases} \dot{x}_{lim} & for \quad \dot{x} > \dot{x}_{lim}, \\ \dot{x} & for \quad |\dot{x}| \leq \dot{x}_{lim}, \\ -\dot{x}_{lim} & for \quad \dot{x} < -\dot{x}_{lim}, \end{cases}$$

where $x_{lim}, \dot{x}_{lim}$ - constraint values relative to coordinate and its velocity, respectively.

The control moment of the system is the control action. If we assume that, in switching on and switching off the electropneumatic valve, the nozzle thrust inereases and ~~ax~~ decreases instantaneously and/do not to take into account the transport lag then the equation of the control mement may be written in the form

$$M = M_{max} \Phi(\sigma)$$

where M_{max} - maximum moment created by the nozzle.

The motion of one channel of the PSS can be described in a simplified form, the lag being taken into account, by the set of equations

$$\ddot{x} = -\delta,$$
$$\delta = \Phi[\sigma(t-\tau)],$$
$$\sigma = K_1 \varphi_1(x) + K_2 \varphi_2(\dot{x}) + K_3 \operatorname{Sign} \Phi(\sigma), \qquad (2)$$
$$\Phi(\sigma) = \begin{cases} m & \text{for } \sigma > \varepsilon, \\ 0 & \text{for } |\sigma| \leq \varepsilon, \\ -m & \text{for } \sigma < -\varepsilon \end{cases}$$

Here the following notations are introduced:

δ - value of controller action ;

σ - control function; m - constant in quantity values taken by the relay characteristic

$(m = \dfrac{M_{max}}{J})$, ε - value of controller dead zone;

τ - lag. The other notations are given earlier. The structural scheme of the PSS, corresponding to the equations of system's motion (2), is given in Fig. 1.

2. The Use of Transducers Constraints for Formation of Nonlinear Control Rules

If in the PSS there are used transducers of angle and angular velocity which have constraints then they can be used for formation of a nonlinear control rule for construction of a system close to optimal one from the point of view of minimum of flow of working body or minimum of thrust impulse. Note that for system (2) the minimal thrust impulse required for damping of initial angular velocity $\dot{x}_0$

$$(Pt)_{min} = \frac{J}{\ell}\dot{x}_0 , \qquad (3)$$

where $P = \text{Const}$ - nozzles thrust, ℓ - arm ($P\ell = M_{max}$)

For solution of the given problem let us study the phase plane for system (2) when $\tau = K_3 = 0$.

The type of the phase plane considerably depends on the relationship of parameters of control rule

$$\sigma = K_1 \rho_1(x) + K_2 \rho_2(\dot{x}) \qquad (4)$$

and on the value of dead zone of the relay function $\Phi(\sigma)$.

It is necessary to consider the following combinations of parameters

a) $K_1 x_{lim} - K_2 \dot{x}_{lim} < -\varepsilon,$

b) $K_1 x_{lim} - K_2 \dot{x}_{lim} = -\varepsilon,$

c) $\varepsilon > K_1 x_{lim} - K_2 \dot{x}_{lim} > -\varepsilon,$

d) $K_1 x_{lim} - K_2 \dot{x}_{lim} = \varepsilon,$

e) $K_1 x_{lim} - K_2 \dot{x}_{lim} > \varepsilon.$

The ype of the phase plane which is three-sheeted plane (on sheet I - $\Phi(\sigma) = m$, on sheet II $\Phi(\sigma) = 0$ and on sheet III - $\Phi(\sigma) = -m$) is presented in Fig. 2a-e, correspondingly. Note, that the relationships of the parameters "b" and "d" are in effect bifurcational due to the fact that at infinitely small variation of any parameter the kind of switching lines varies qualitatively. Apart from this, in these cases we confront a new interesting fact when switching "lines" at some regions are, in fact, no longer lines and occupy certain parts of the phase plane (shaded areas in Fig. 2 b, d).

According to (2) the switching lines wholly lie on sheet II and, consequently, the shaded parts refer also to sheet II.

From consideration of the phase plane it follows that the PSS will be more close to optimal one in case of relationship of the parameters "a", "b" and "c". In fact the thrust impulse in this case

$$Pt = \frac{J}{l}\left(\dot{x}_0 + 2\,\frac{K_1 x_{lim} - \varepsilon}{K_2}\right) \qquad (5)$$

and for sufficiently great K_2 it practically coincides with its minimal value.

Thus, using only natural constraints of parameters and applying no logical elements one can construct the system close to optimal with respect to flow of working body.

It should be noted that in this case we have long time of regulation but as a rule, during the regulation time the strict constraints for PSS are not assigned.

It is interesting to compare the suggested system with respect to thrust impulse with the optimal quick response system. It is easy to calculate that for the latter, when $\mathcal{E} = 0$.

$$Pt = \frac{\mathcal{J}}{\ell}(1+\sqrt{2})\dot{x}_0 \qquad (6)$$

i.e. the thrust impulse is two times greater than minimal one.

Consider the phase plane for the case "b" when $\tau \neq 0$. Its kind is shlown in Fig. 3. There are sections of superposition of sheets due to the lag. The equations of switching lines are the following

$$\begin{aligned}
&(L_1) \quad \dot{x} = -\frac{\kappa_1 x_{lim} - \mathcal{E}}{\kappa_2} - m\tau, \\
&(L_2) \quad (\kappa_2 - \kappa_1\tau)\dot{x} + \kappa_1 x = \mathcal{E} - \kappa_2 m\tau + \kappa_1 m\frac{\tau^2}{2}, \qquad (7) \\
&(L_3) \quad (\kappa_2 - \kappa_1\tau)\dot{x} + \kappa_1 x = \mathcal{E}, \\
&(L_4) \quad x = -x_{lim} + \dot{x}\tau
\end{aligned}$$

The lines (L'_i) are symmetrical with (L_i) relative to the origin $i = 1, 2, 3, 4$. It is known that in the presense of lag the system (2) will have oscillation at any relations of parameters. Without taking into account the auto-oscillations the thrust impulse will be

$$Pt = \frac{\mathcal{J}}{\ell}\left(\dot{x}_0 + 2\,\frac{\kappa_1 x_{lim} - \mathcal{E}}{\kappa_2}\right) + 2P\tau \qquad (8)$$

In conclusion note that since ε usually is small then in practice the relationship of the parameters "a" must be chosen since in case "c" strict constraints must be imposed on stability of values of $K_1, K_2, x_{lim}, \dot{x}_{lim}$.

With lag taken into account it is evident that condition "a" should be supplemented with the condition

$$2\varepsilon > m\tau \qquad (9)$$

because, if (9) is not fulfilled, the describing point "slips" the dead zone and the thrust impulse increases sharply.

In the steady-state motion the system is in auto-oscillating condition. Here x and $\dot{x}$ do not reach their constraints ($|x| < x_{lim}$ and $|\dot{x}| < \dot{x}_{lim}$). Auto-oscillations in system (2) in absense of transducers constraints are studied in detail. In investigation of auto-oscillations one may make use of results obtained in reference 5.

In conclusion let us discuss the operation of the PSS with relay characteristic having hysteresis loop. It has been proved[6] that in this case the amplitude of auto-oscillations can be decreased due to inclusion of feedback ($K_3 \neq 0$) which compensates the lag. On the phase plane the decrease of amplitude of auto-oscillations is explained by approaching of switching lines.

At almost complete lag compensation the controller approaches the ideal one but here all disqdvantages of such a controller are revealed. The control organ in this case switches at a very high frequency. This frequency may become inadmissible for the given controller and controlled plant. At high frequences the breakdown of auto-oscillations may occur which is equivalent to that the system will operate without controller. Therefore, .if theoretically, the lag can be compensated completely it is not expedient to do this in practice. Realization of compensation should be done within admissible limits.

3. Equations of Elastic Oscillations of the Satellite-Stabilizer System

In deriving the approximate equations of the satellite spin motion due to stabilizer bending oscillations we neglect destributed mass of bars and weights movements due to longitudinal and torsional oscillations taking into account weights movements associated only with transverse bending oscillations. We also neglect all perturbing moments because of their smallness compared with moments from oscillating weights and control moments from PSS. We neglect also the perturbing Coriolis forces occuring in the result of interaction of transfer and relative movements of weights[7].

For obtaining the above mentioned differential equations consider kinematic scheme of motion of only the i-th bar with the weight (Fig. 4). It is assumed that the center of mass of the system is located at the point "O" and the bar and weight are rigidly mounted at points A_i and B_i, respectively.

During the stabilizer opening due to the action of Cariolis forces or opening mechanism the weights receive initial deflections. In this case the satellite will oscillate round the center of mass under the action of moments received from the weights oscillating under the action of elastic forces of the bars.

If the weight on the i-th bar deviates by the angle ψ_i then the satellite body will rotate by the angle $-\varphi$.

Hence, if the satellite-stabilizer system is the satellite and "n" bars with weights at the ends then it may be considered as the system with "n+1" degrees of freedom. The generalized coordinates will be $\varphi, \psi_1, \psi_2, \dots \psi_n$ respectively.

In the case under discussion the Lagrangian equations of the second order are of the form

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\varphi}} - \frac{\partial \mathcal{L}}{\partial \varphi} = -M_\varphi , \qquad (10)$$

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\psi}_i} - \frac{\partial \mathcal{L}}{\partial \psi_i} = -M_{\psi_i}\ , \quad i=1,2,\dots n.$$

where $\mathcal{L}$ - Lagrangian function ($\mathcal{L} = T - V$);

T and V - kinematic and potential energy of the system, respectively, M_{φ} - control moment of the PSS, M_{ψ_i} - moment from internal friction forces in material of the i-th bar.

Introduce the following notations (Fig. 4): $OA_i = a_i$, $A_iB_i = \ell$, $OB_i = \rho_i$, v_i - absolute velocity of motion of the i-th weight; $m_1 = m$ - mass of the i-th weight

$$v_i^2 = (\dot{\varphi}\rho_i)^2 + (\dot{\psi}_i \ell)^2 - 2\dot{\psi}_i \ell \dot{\varphi} \rho \cos(\alpha_i + \psi_i - \varphi) \qquad (11)$$

The kinematic energy of the satellite-stabilizer system is:

$$T = \frac{1}{2} I_c \dot{\varphi}^2 + \frac{1}{2}\sum_{i=1}^{m} m v_i^2, \qquad (12)$$

where I_c - moment of inertia of the satellite.

The expression for potential energy of the i-th weight is of the form[8]

$$V_i = \frac{3}{2}\frac{E_i I_i}{\ell}\psi_i^2, \qquad (13)$$

where $E_1 = E$ - Young's modulus, $I_1 = I$ - moment of inertia of the bar cross-section.

Substituting expressions (12), (13) into Lagrangian function and taking into account (11) we obtain, according to (10), the differential equations of the motion of the system under discussion[7].

$$[I_c + mna^2 + mn\ell^2 + 2a\ell\sum_{i=1}^{n}\cos\psi_i]\ddot{\varphi} - \qquad (14)$$

$$-2ma\ell\dot{\varphi}\sum_{i=1}^{n}\sin\psi_i\,\psi_i - m\sum_{i=1}^{n}(\ell^2+a\ell\cos\psi_i)\ddot{\psi}_i +$$

$$+ma\ell\sum_{i=1}^{n}\sin\psi_i\cdot\dot{\psi}_i^2 = -M_\varphi,$$

$$m\ell^2\ddot{\psi}_i - m(\ell^2+a\ell\cos\psi_i)\ddot{\varphi} + ma\ell\dot{\varphi}^2\sin\psi_i +$$

$$+\frac{3EI}{\ell}\psi_i = -M_{\psi_i}\,,\quad i=1,2,\ldots n$$

Note that equations (14) describe the motion of the satellite-stabilizer system conditioned only by oscillations of weights. At small deviations the total angular motion of the system can be considered as the sum of the above mentioned motion and the solid body motion.

To the system (14) the equations describing dynamics of the PSS must be added.

4. Studies of Motion of the Satellite-Stabilizer System Caused by Weights Oscillations

Simplify equations (14) neglecting the moments from forces of internal friction in bars since it has been proved that they are small [9]. Linearizing the equations, assuming the amplitudes of auto-oscillations to be small, we obtain

$$[I_c+nm(a+\ell)^2]\ddot{\varphi} - m\ell(a+\ell)\sum_{i=1}^{n}\ddot{\psi}_i = -M_\varphi,$$

(15)

$$m\ell^2\ddot{\psi}_i - m\ell(a+\ell)\ddot{\varphi} + \frac{3EI}{\ell}\psi_i = 0,$$

$$i=1,2,\ldots n.$$

Since the PSS is a relay system the control moment may be assumed to be constant at each region of addition, and at these regions it is possible to find the solution of the set of equations (15).

For a stabilizer with one bar the solution of the set of equations (15) at the initial conditions $t=0,\ \varphi=\varphi_0,$ $\dot\varphi=\dot\varphi_0$, $\psi=\psi_0$ and $\dot\psi=\dot\psi_0$ has the form

$$\begin{aligned}\varphi=\varphi_0-A_1\psi_0+(\dot\varphi_0-A_1\dot\psi_0)t+A_1\psi_0\cos\omega_1 t &+\\ +\frac{A_1}{\omega_1}\dot\psi_0\sin\omega_1 t-B_1M_\varphi t^2&,\end{aligned}\qquad(16)$$

$$\psi=\psi_0\cos\omega_1 t+A_2\dot\psi_0\sin\omega_1 t-C_1M_\varphi$$

For a stabilizer with two bars with weights the solution of the set of equations (15) for the most disadvantageous conditions $t=0,\ \varphi=\varphi_0,\ \dot\varphi=\dot\varphi_0,\ \psi_1=\psi_2=\psi_0$ and $\dot\psi_1=\dot\psi_2=0$ has the form

$$\begin{aligned}&\varphi=\varphi_0-A\psi_0+\dot\varphi_0 t+A\psi_0\cos\omega_2 t-BM_\varphi t^2,\\ &\psi_1=\psi_0\cos\omega_2 t-C\,M_\varphi,\\ &\psi_2=\psi_1\end{aligned}\qquad(17)$$

The constant coefficients $A, A_1, A_2, B, B_1, C, C_1$ and frequences ω_1 and ω_2 depend on parameters of the satellite-stabilizer system and are defined by the following formulae

$$A=\frac{2m\ell(a+\ell)}{I_c+2m(a+\ell)^2},\quad A_1=\frac{m\ell(a+\ell)}{I_c+m(a+\ell)^2},\quad A_2=\frac{1}{\omega_1},$$

$$B=\frac{1}{2[I_c+2m(a+\ell)^2]},\quad B_1=\frac{1}{2[I_c+m(a+\ell)]},$$

$$C=\frac{m\ell^2(a+\ell)}{3EI[I_c+2m(a+\ell)^2]},\ C_1=\frac{m\ell^2(a+\ell)}{3EI[I_c+m(a+\ell)^2]},$$

$$\omega_1^2=\frac{3EI}{m\ell^3}\left[1+\frac{m(a+\ell)^2}{I_c}\right],\ \omega_2^2=\frac{3EI}{m\ell^3}\left[1+\frac{2m(a+\ell)^2}{I_c}\right]$$

It is seen from (16) and (17) that when the PSS is switched off ($M_\varphi=0$) after opening the stabilizer, the satellite at $\varphi_0\neq 0$ will slowly spin and simultaneously oscillate at frequency depending on bars rigidity. The angular deviation increases until the perturbation is balanced by gravity moment. Due to the fact that the forces of internal friction in bars are negligibly small the gravity-stalibized system satellite-stabilizer will make slow undamped oscillations under the effect of gravity moment and quick oscillations from the shaking weights on bars. The undesirable high-frequency oscillations in the system can be damped if the internal friction forces are introduced in bars.

Since the roots of characteristic equation of the system of the first approximation (15) are pure imaginary then the judgement of motion stability from these equations is not lawful. For the strict analysis it is necessary to study the equations of motion (14).

As an example the set of equations (14) has been numerically integrated on a digital computer for a number of particular values of parameters of the satellite and stabilizer[2]. Comparison of results of numerical integration with analytical solution of simplified equations has shown that frequences and amplitudes of the satellite and stabilizer in both cases practically coincide.

The effect of moment from the internal friction forces in material of bars and damping devices has been also investigated on the digital computer. The damping moment was taken into account according to the formulae $M_\psi=K\dot{\psi}$.

The energy dissipation in the bar (K = 0.001; 0.005; 0.01) do not practically influence on system's oscillations. If the bar is equipped with damping devices (K = 1; 5; 10; 100) then the oscillations in the system damp very quickly though the satellite continues to deviate from the set position up to the moment when the equilibrium state is reached due to the gravity moment. After this the gravity-stabilized system satellite-stabilizer will slowly oscillate under the effect of gravity moment. However, the angle of deviation will be smaller due to introduction of artificial damping in bars. Thus, the satellite cannot be damped within small time intervals at the expense of bending oscillations.

On gravity-stabilized satellites launched in USA the stabilizer opens after final stabilization of the satellite with the aid of "yaw-yaw" system and besides there is magnetic damping [2]. At such a choice of the moment of opening the stabilizer the satellite's angular velocity, not damped by the PSS, decreases due to the increase of the moment of inertia of the satellite-stabilizer system.

In equations (16) and (17) the value of control moment M_φ has very small coefficient. If the time of operation of the PSS is short and control moment is small then one may expect that it will not render essential effect during its operation on satellite's dynamics. Therefore a conclusion can be made on the expediency of opening the stabilizer after switching off the PSS. The conclusion obtained requires additional studies. If the stabilizer does not possess sufficient rigidity then it is reasonable to open it after preliminary stabilization of the satellite. The influence of the PSS on dynamics, when the gravity-stabilizer is open; should be judged of by the results of studying motion equations (14).

Complete equations of motion of bending oscillations of the satellite-stabilizer system have been investigated on a digital computer with taking into account the PSS operation for the chosen parameters of satellite and stabilizer [2]. The studies have shown that if the PSS has relay characteristic with a dead zone then the bending oscillations of the satellite-stabilizer system can be damped within acceptable time interval at small control moments.

References

I.Д.Е.Охоцимский, В.А.Сарычев. Система гравитационной стабилизации искусственных спутников",Сб."Искусственные спутники Земли, вып.I6,Из-во АН СССР, I963.

2.Проблемы ориентации искуственных спутников Земли, под ред.С.Ф. Сингера. Из-во "Наука", I966.

3.Б.В.Раушенбах, Е.Н.Токарь. Некоторые вопросы управления в межпланетном пространстве", Сб."Искусственные спутники Земли,вып.5.Из-во АН СССР, I960.

4.В.П.Легостаев, Б.В.Раушенбах. Система одноосной ориентации по солнцу кораблей-спутников "Восток".Космические исследования, т.IУ,вып.3,Из-во АН СССР,I966.

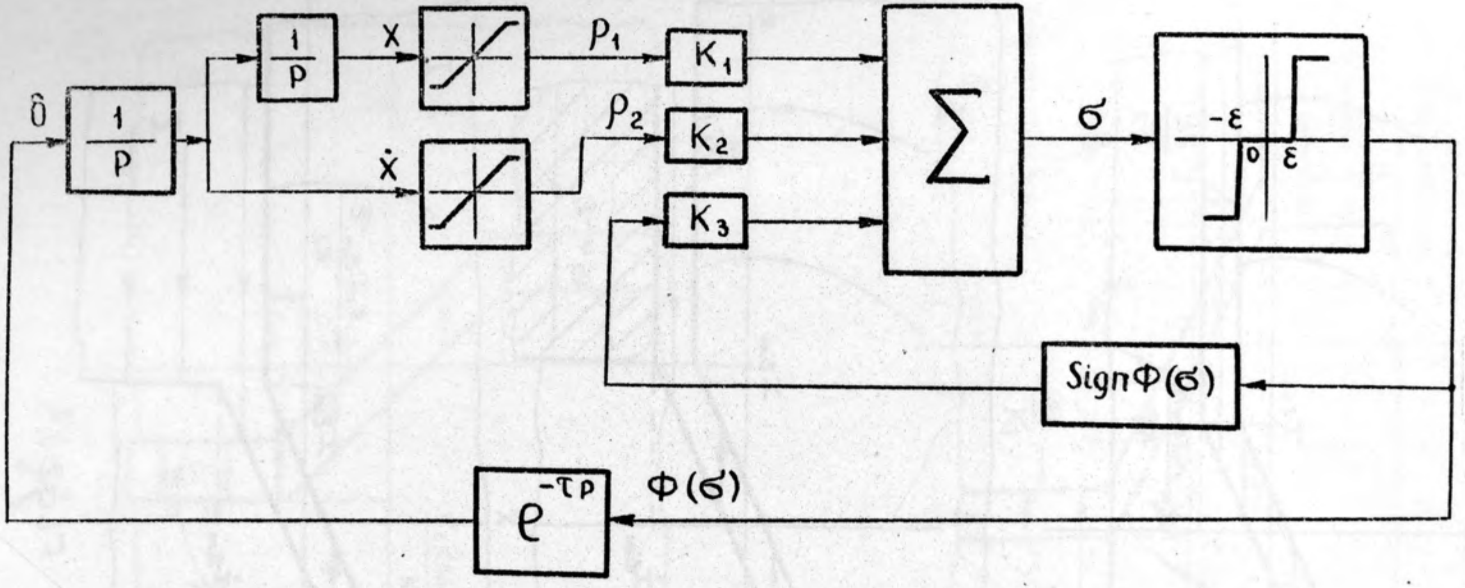
5.В.В.Петров, В.Ю.Рутковский. Теория простейших релейных сервомеханизмов с запаздыванием, Изв. АН СССР, ОТН, № 4, I956.

6.В.И.Попов. Исследование одной релейной системы с изменяющимся во времени регулирующим воздействием и переключением закона регулирования. Сб."Теория и применение автоматических систем". Из-во "Наука", I964.

7.В.И.Попов, В.Ю.Рутковский. Исследование плоских изгибных колебаний гравитационно-устойчивой системы спутник стабилизатор. Космические исследования, т.3, вып.5, Из-во АН СССР, I965.

8.Ю.А.Шиманский. Динамический расчет судовых конструкций. Судпромгиз, I963.

9.Я.Г.Пановко. Внутренее трение при колебаниях упругих систем. Физматгиз, I960.



$\hat{0}$
$\frac{1}{p}$
$\frac{1}{p}$
x
$\dot{x}$
ρ_1
ρ_2
K_1
K_2
K_3
$\sum$
σ
$-\varepsilon$
0
ε
$Sign\Phi(\sigma)$
$\Phi(\sigma)$
$e^{-\tau p}$

Fig 1.

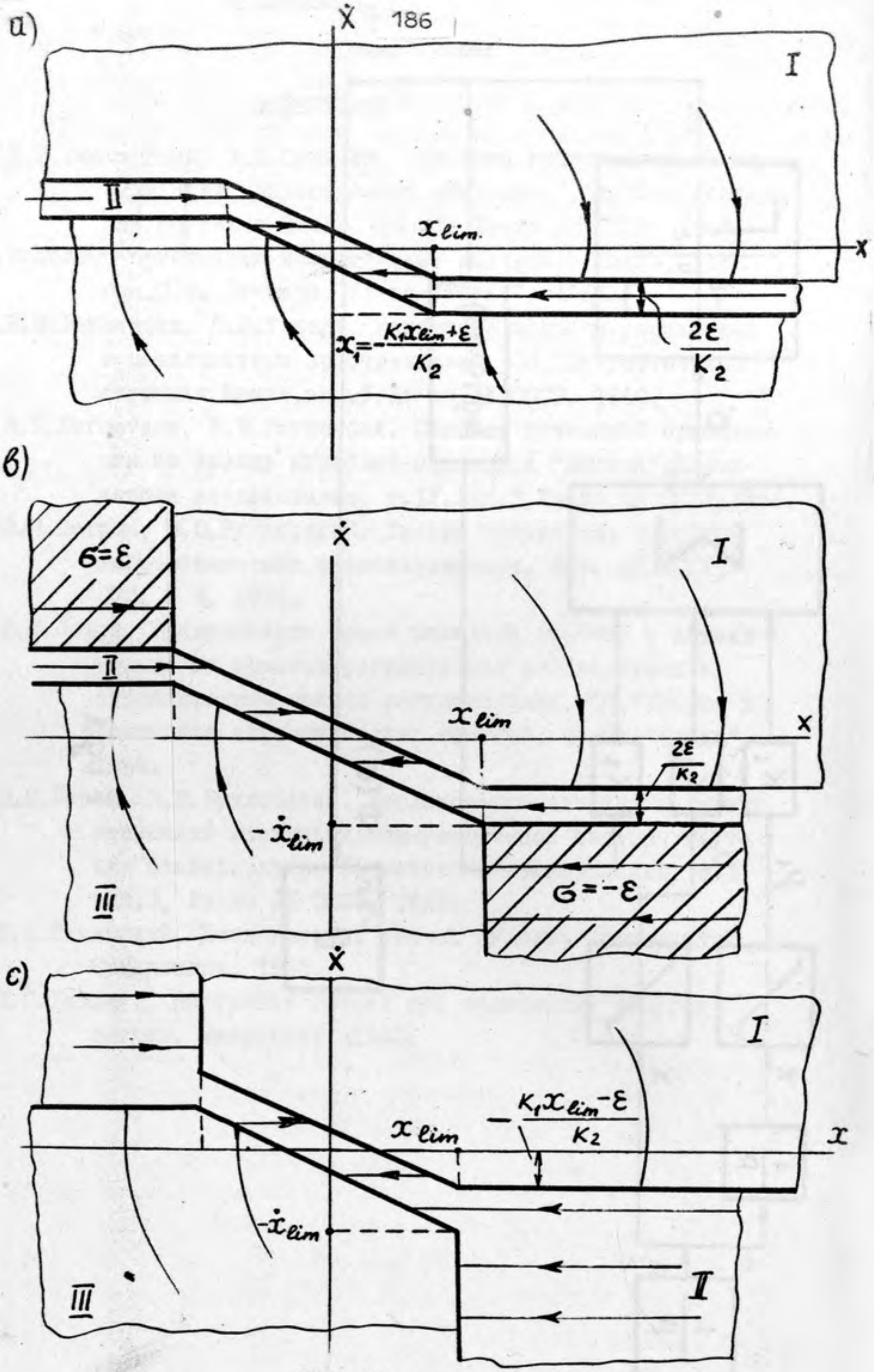

a)
$\dot{X}$
I
II
x
$x_{lim.}$
$x_1 = -\frac{K_1 x_{lim} + \varepsilon}{K_2}$
$\frac{2\varepsilon}{K_2}$
в)
$\sigma = \varepsilon$
$\dot{x}$
I
II
x
x_{lim}
$\frac{2\varepsilon}{K_2}$
$-\dot{x}_{lim}$
III
$\sigma = -\varepsilon$
c)
$\dot{x}$
I
$-\frac{K_1 x_{lim} - \varepsilon}{K_2}$
x_{lim}
x
$-\dot{x}_{lim}$
II
III

Fig2.

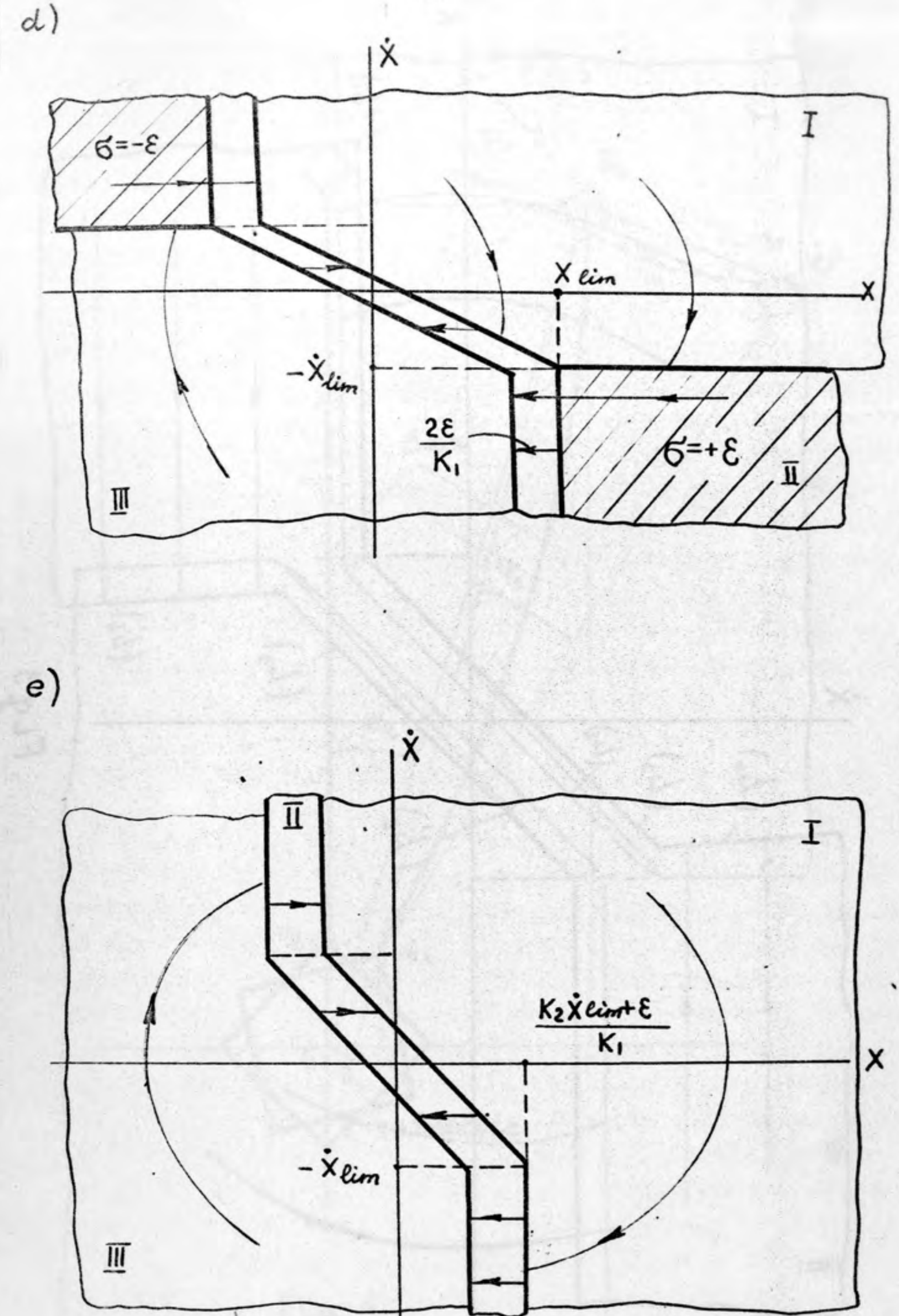


Fig2

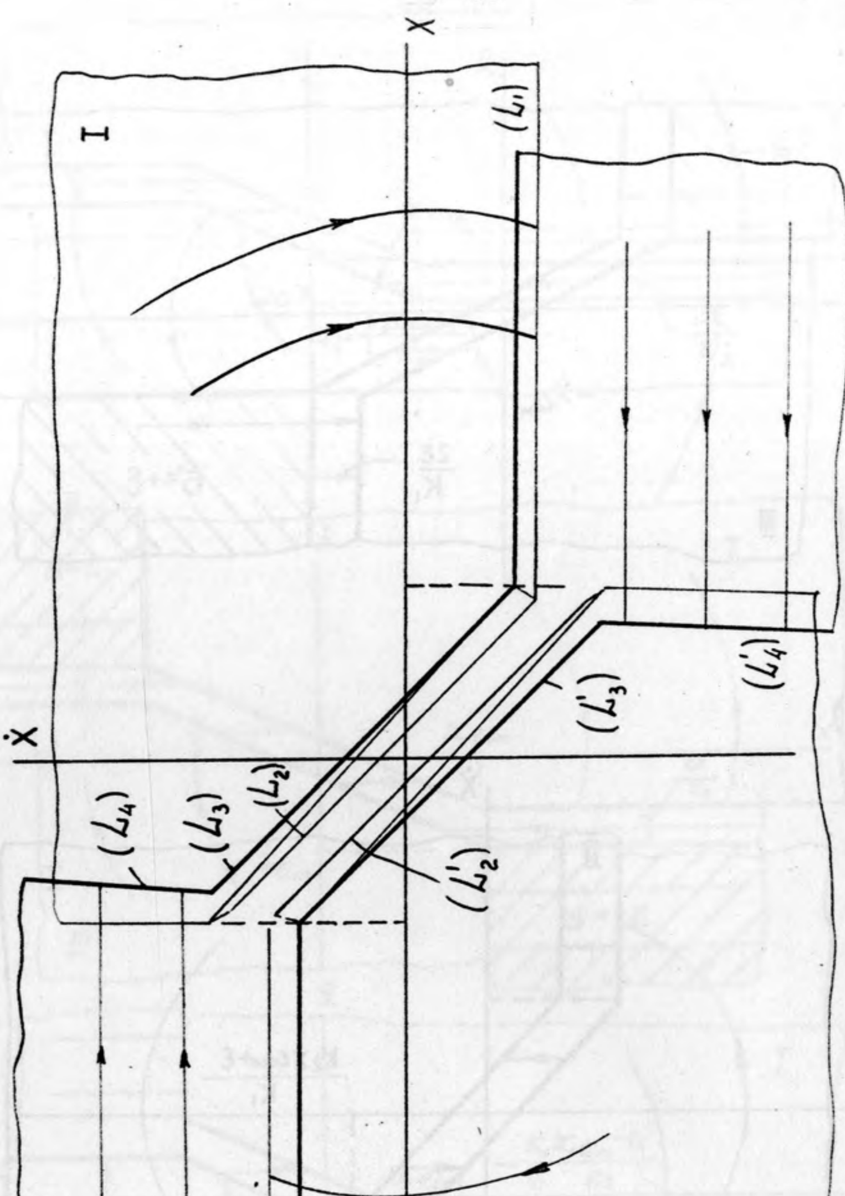


Fig 3

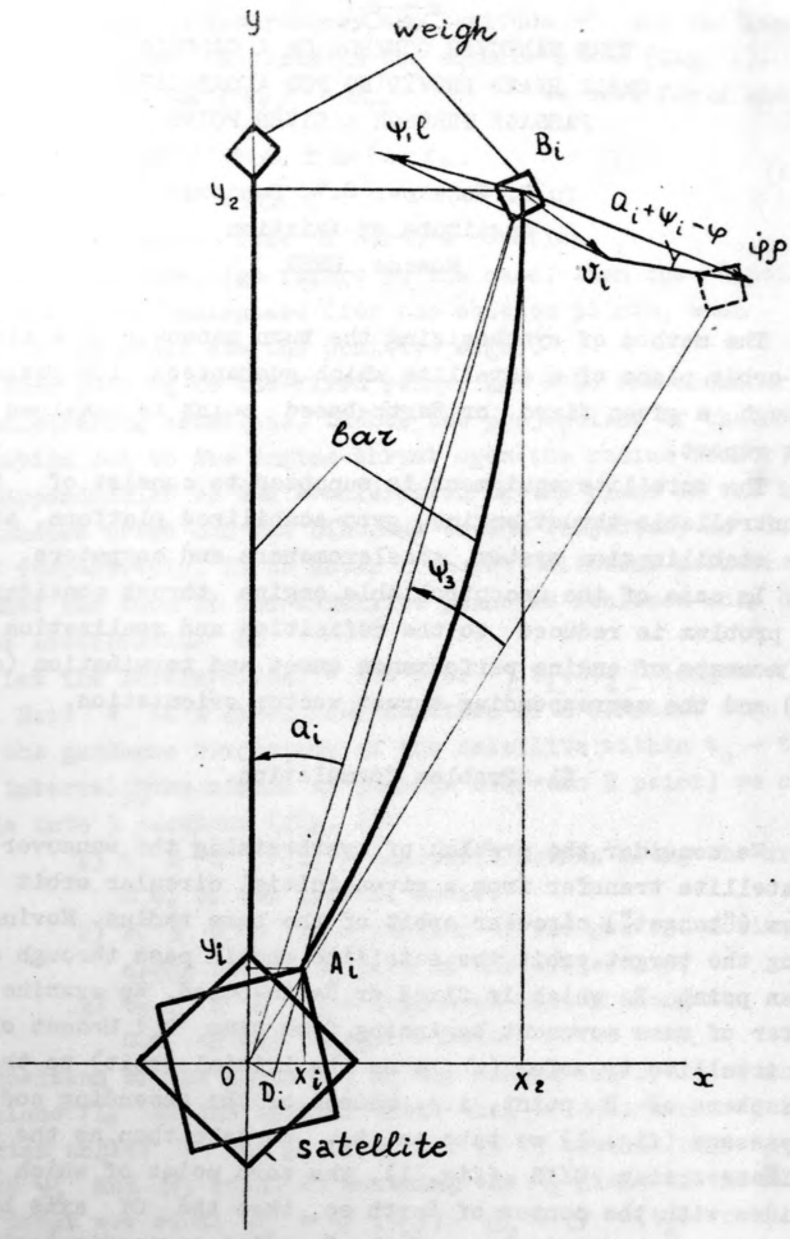

y
weigh
$\psi_1 \ell$
B_i
y_2
$a_i+\psi_i-\varphi$
$\dot{\varphi}\rho$
v_i
bar
ψ_3
a_i
y_i
A_i
0
D_i
x_i
x_2
x
satellite

Fig. 4.

57.5

TURN MANEUVER CONTROL OF A CIRCULAR ORBIT PLANE PROVIDING FOR A SATELLITE PASSAGE THROUGH A GIVEN POINT

Yu.P. Gouskov, S.V. Bunjakin
Institute of Aviation
Moscow USSR

The method of synthesizing the turn maneuver of a circular orbit plane of a satellite which guarantees its passage through a given fixed or Earth-based point is examined in this report.

The satellite equipment is supposed to consist of the uncontrollable-thrust engine, gyro-stabilized platform, attitude stabilization system, accelerometers and computers.

In case of the uncontrollable engine thrust considered, the problem is reduced to the definition and realization of the moments of engine performance onset and termination (cut off) and the corresponding thrust vector orientation.

§1. Problem formulation.

We consider the problem of synthesizing the maneuver of a satellite transfer from a given initial circular orbit to a new ("target") circular orbit of the same radius. Moving along the target orbit the satellite should pass through a given point R which is fixed or Earth-based. We examine the center of mass movement beginning from time t_0. Moment of the satellite transfer (while on its initial orbit) to the hemisphere of R point, i.e. moment of the ascending node C_0 passage (fig. 1) we take as t_0. We take then as the coordinate system OXYZ (fig. 1), the zero point of which coincides with the center of Earth so, that the OY axis be directed to the hemisphere of the R point perpendicularly to the equator plane Π_3, OX axis - to the ascending node C_0 of the initial orbit Π_0 and OZ axis - perpendicularly to the OXY plane.

Position of the radius vector r_R of the R point on the

Earth's surface is governed by the latitude φ and the longitude α about OX axis in the equator plane (fig. 1). Provided that $\alpha(t_0) = \alpha_0$ we have for α and φ

$$\begin{aligned} &\alpha(t) = \alpha_0 \pm \omega(t - t_0) \\ &\varphi = const \end{aligned} \qquad (1.1)$$

where ω is angular rate of Earth's rotation.

Here negative sign refers to the case, when the R point is in the south hemisphere (for the equator points, when $\varphi =$ 0, we shall use the positive sign).

When guiding to the fixed point, $\omega = 0$. Now consider the maneuvering satellite. Denote the projections of the acceleration due to the engine thrust upon the radius vector, the perpendicular to the acceleration in the plane of the instantaneous orbit and the binormal to the trajectory as S, T and W respectively. It is known from the celestial mechanics [1] that the turn of the satellite plane is realized only due to the acceleration W.

Let the acceleration W act over $[t_1, t_2]$ length of time. Here W is a given time function of a constant sign. Then the guidance trajectory of the satellite within $t_0 - t_3$ time interval (t_3 the moment of passage over the R point) we can divide into 3 portions (fig. 2):

1) $t_0 \leq t < t_1$ - unpowered motion along the arc C_0C_1 of the initial orbit;
2) $t_1 \leq t \leq t_2$ - portion of the powered motion along the C_1C_2 part of the trajectory;
3) $t_2 < t \leq t_3$ - unpowered motion along the arc C_2R of the "target" orbit.

The position of the plane Π_1, of the osculating orbit against the plane Π_0 of the initial orbit will be determined by the Eulerian angles ψ and θ (fig. 1). It is obvions that the angles ψ_2 and θ_2 (fig. 2) defining the Π_2 plane of the "target" orbit are equal $\psi_2 = \psi(t_2)$, $\theta_2 = \theta(t_2)$ respectively.

The position of the radius vector r_c of the satellite center of mass (fig. 1) on the osculating and "target" orbits is defined by the argument of the latitude

The change of angles Ψ, θ and u when $t > t_1$ is characterized by the differential equations of the osculating elements [1]

$$\frac{d\Psi}{dt} = W\frac{r}{\sqrt{p\mu}}\sin u \csc\theta$$

$$\frac{d\theta}{dt} = W\frac{r}{\sqrt{p\mu}}\cos u \tag{1.2}$$

$$\frac{du}{dt} = \frac{\sqrt{p\mu}}{r^2} - W\frac{r}{\sqrt{p\mu}}\sin u \, ctg\,\theta$$

where μ - gravitational constant,

r - current radius ,

p - parameter of the osculating orbit.

The change of the parameters p and r is governed by the equations [2]

$$\frac{dy}{d\vartheta} = z$$

$$\frac{dz}{d\vartheta} = -y + \frac{1}{p}\left(1 + n_S + \frac{z}{y}n_T\right) \tag{1.3}$$

$$\frac{dp}{d\vartheta} = \frac{2}{y}n_T$$

where $y = r^{-1}$, $n_S = S/\mu y^2$, $n_T = T/\mu y^2$ are the components of the longitudinal acceleration at the satellite height, and ϑ (angular distance of the satellite from the axis fixed to the osculating orbit) is governed by the equation

$$\frac{d\vartheta}{dt} = r^{-2}\sqrt{p\mu} \tag{1.4}$$

If the shape of the initial circular orbit is maintained unchanged during the maneuver, then $p = r$ = const., and the system (1.2) can be rewritten as

$$\frac{d\Psi}{dt} = \Omega n_w \sin u \csc \theta$$
$$\frac{d\theta}{dt} = \Omega n_w \cos u \qquad (1.5)$$
$$\frac{du}{dt} = \Omega (1 - n_w \sin u \operatorname{ctg} \theta)$$

where $\Omega = r^{-\frac{3}{2}} \sqrt{\mu}$ is orbital angular rate; $n_W = \frac{Wr^2}{\mu}$ is side acceleration at the satellite height.

§ 2. Maneuver programming.

The first step in the maneuver synthesizing is the formulation of the control acceleration program providing for a coincidence of the radius vectors of the R and C points, moving according to the (1.1), (1.5) rules at a time t_3.

a) Kinematics of the powered motion. According to the problem in view the control acceleration dependence on time for $t > t_1$ is preset. Therefore, as a result of the equation (1.5) solution with the initial requirements $\Psi(t_1) = \Psi_1$, $\theta(t_1) = u(t_1) = 0$, the following relationships can be found:

$$\theta = \theta(\tau), \quad \Delta \Psi = \Psi - \Psi_1 = \Delta \Psi(\tau), \quad u = u(\tau) \qquad (2.1)$$

where $\tau = t - t_1$.

If n_w = const. these functions according to [3] take form:

$$\theta = 2 \arcsin \frac{n_w \sin(0{,}5 \Omega \tau \sqrt{n_w^2 + 1})}{\sqrt{n_w^2 + 1}}$$

$$\Delta \Psi = u = \operatorname{arctg} \frac{\operatorname{tg}(0{,}5 \Omega \tau \sqrt{n_w^2 + 1})}{\sqrt{n_w^2 + 1}}$$

We are not able to ascertain the form of a decision (2.1) for the general case of n_w assignement. It is possible only to point out the following features of this solution resulting from the pattern of the system (1.5) second equation, when the sign of n_w is supposed to be constant:

1) the angle of orbit plane rotation cannot exceed a certain value θ_{max}, dependent or n_w;

2) the maximum value of the angle of orbit plane rotation is attained when $u = \pi/2$;

3) the θ variation with time is proceeding monotonously.

Thanks to the mentioned above features, when $0 \leq u \leq \pi/2$ it is possible to pass in the solution (2.1) from the independent variable t to the independent variable θ

We have as a result:

$$\tau = \tau(\theta), \quad \Delta\Psi = \Delta\Psi(\theta), \quad u = u(\theta) \qquad (2.2)$$

For n_w = const. the relations (2.2) are written as:

$$\tau = \frac{2}{\Omega}\left(\cos\frac{|\theta max|}{2}\right) arc\sin\frac{\sin\frac{|\theta|}{2}}{\sin\frac{|\theta max|}{2}}$$

$$u = \Delta\Psi = arctg\left(\cos\frac{|\theta max|}{2} tg\, arc\sin\frac{\sin\frac{|\theta|}{2}}{\sin\frac{|\theta max|}{2}}\right)$$

where $\theta_{max} = 2$ arctg n_w.

In the case of a pulse maneuver of the orbit plane rotation ($n_w = \infty$)

$$\tau(\theta) = \Delta\Psi(\theta) = u(\theta) = 0 .$$

b) Necessary relationships.

Considering the motion in the central gravitational field we shall set up the requirements sufficient for a trajectory of the satellite guidance to the line of the radius vector of the R point.

Suppose that the radius vectors r_c and r_R (fig.2) coincide as a result of the maneuver at a time t_3. We shall in that case suppose for certainty, that at $0 \leq \alpha_3 < \pi$ a satellite passes over the R point (with respect to the time t_0) during the second revolution, but when $\pi \leq \alpha_3 < 2\pi$ it passes over the R point during the second semi revolution (we mean the 1st revolution here).

Then the balance of time and the requirements of the corresponding equality of the two spherical coordinates of the R and C points against the Π_0 plane and the OX axis in it should be met. As can be seen from fig. 2, these requirements give three relations:

$$\sin\Theta_2 = \frac{\sin\varphi\cos i - (-1)^k\cos\varphi\sin i\sin\alpha_3}{\sin u_3}$$

$$S_3 = K\pi + \operatorname{arctg}\frac{\cos\varphi\cos\alpha_3}{\sin\varphi\sin i + (-1)^k\cos\varphi\cos i\sin\alpha_3} + \qquad (2.3)$$

$$+ u_3 - \operatorname{arcctg}\frac{\operatorname{ctg} u_3}{\cos\Theta_2}$$

$$\alpha_3 = \alpha_0 \pm \frac{\omega}{\Omega}\left(S_3 - \Delta\Psi(\Theta_2) - u(\Theta_2) + \Omega\tau(\Theta_2)\right)$$

where $S_3 = \Psi_2 + u_3$, $K = \begin{cases} 2 & 0 \leq \alpha_3 < \pi \\ 1 & \pi \leq \alpha_3 < 2\pi \end{cases}$

and functions $\Delta\Psi(\Theta), u(\Theta), \tau(\Theta)$ are defined by the relations (2.2).

c) Trajectory optimization.

Three relations obtained above tie together the four unknown quantities - $\alpha_3, S_3, \Theta_2, u_3$. One is free thesefore, to choose arbitrarily one of these parameters. Let us exclude this i uncertainty with the requirement of energy optimum (minimum energy losses) for the maneuver.

Under preset engine thrust an energetically optimal maneuver is to be one with a minimum length of its active part. As the variation of a latitude argument u within the active part of energetically optimal maneuver with a single engine performance onset is always not to exceed $\pi/2$, so in view of monotonousness in the Θ variation, the minimization of engine performance time is equivalent to the minimization of the absolute value of the angle Θ_2 required for the maneuver.

The determination of an optimal trajectory of guidance is therefore in our case reduced to the problem of minimization of the function $\mathcal{J} = |\Theta_2|$ under conditions of (2.3).

This problem is most simply solved carrying out a series of numerical solutions of equations (2.3) for different values of $\mathcal{U}_3$ (in the neighbourhood of optimum supposed) with the subsequent finding the relation $|\Theta_2| = f(\mathcal{U}_3)$ minimum The process of iterating equations (2.3) converges rather fast if one takes $\alpha_3^{(1)} = \alpha_0$ as the first approximation.

With the found optimum values $\mathcal{U}_3^*$, Θ_2^*, S_2^* of parameters $\mathcal{U}_3$, Θ_2, S_3 the program of an engine performance is easily defined. An engine, the thrust action line of which is directed along the binormal to the trajectory is to be engaged at a moment t_1, when the angular range Ψ_1 of a satellite from the ascending node reaches the value of $\Psi_1^* = S_3^* - \mathcal{U}_3^* - \Delta\Psi(\Theta_2^*)$. The period of engine performance Δt is defined as the time of orbit plane rotation to the angle Θ_2 i.e. $\Delta t = \tau(\Theta_2^*)$.

§ 3. Maneuver control.

The final stage in the process of a maneuver synthesis is the realization of the programmed guidance trajectory. To realize it a control system is employed. The feedback system examined here is based on the measurement of accelerations n_s, n_τ, n_w. The feedback introduction is dictated by the necessity to counterbalance accidental deviations of registered accelerations from their programmed values. These deviations arise in account for engine thrust fluctuations and the error of a satellite angle stabilization.

Assuming that accelerations n_s, n_τ, n_w are measured by means of mounted on the gyroplatform accelerometers we shall form the program correcting control as a function of these measurements. The correction δ of the side motion is in that case formed discretly, to the account of the moment of engine cutoff with regard to the programmed value; the longitudinal perturbed motion control is realized with a relay by means of a proper thrust vector deviation with respect to the binormal. To the account of that, the thrust vector components, which act in accordance with the conservation of the initial circular orbit form requirement shall be projectted upon the radius vector and the normal to it in the orbit plane.

a) Quality criterions.

As the result of perturbations the satellite in a designed moment of encounter t_3 shall be at an angle distance ρ from the target point R. The satellite radius vector attitude at that moment can be characterized (fig.3) by the Eulerian angles

$$\begin{aligned} \psi_2 &= \psi_2^* + \delta\psi_2 \\ \theta_2 &= \theta_2^* + \delta\theta_2 \\ u_3 &= u_3^* + \delta u_3 \end{aligned} \qquad (3.1)$$

or angles

$$\begin{aligned} E_3 &= \mathcal{E}_3^* + \delta E_3 \\ \lambda_3 &= \lambda_3^* + \delta\lambda_3 \end{aligned} \qquad (3.2)$$

Here and further on the programmed motion parameters are marked with an asterisk.

Considering deviations δE_3 and $\delta\lambda_3$ to be small we obtain from the spherical triangle C_oCR (fig. 3) for the angular deflection (a miss)

$$\rho^2 = \delta E_3^2 + \sin^2\mathcal{E}_3^*\,\delta\lambda_3^2 \qquad (3.3)$$

From the spherical triangle KC_oC (fig. 3) we define in a linear approximation

$$\begin{aligned} \delta E_3 &= c_1\delta\psi_2 + c_2\delta\theta_2 + c_3\delta u_3 \\ \delta\lambda_3 &= e_1\delta\psi_2 + e_2\delta\theta_2 + e_3\delta u_3 \end{aligned} \qquad (3.4)$$

where

$$c_1 = -\cos u_3^* \sin\psi_2^* \csc\mathcal{E}_3^* (1 - \operatorname{tg} u_3^* \operatorname{ctg}\psi_2^* \cos\theta_2^*)$$

$$c_2 = \csc\mathcal{E}_3^* \sin u_3^* \sin\psi_2^* \sin\theta_2^*$$

$$c_3 = \sin u_3^* \cos\psi_2^* \csc\mathcal{E}_3^* (1 - \operatorname{ctg} u_3^* \operatorname{tg}\psi_2^* \cos\theta_2^*)$$

$$e_1 = -\frac{\sin^3\lambda_3^* \cos\varepsilon_3^* \sin\varepsilon_3^* \sec\lambda_3^*}{\sin^2\theta_2^* \; \sin^2 u_3^*} \, c_1$$

$$e_2 = ctg\,\theta_2^* \, tg\,\lambda_3^* + \frac{c_2}{c_1} e_1$$

$$e_3 = ctg\, u_3^* \, tg\,\lambda_3^* + \frac{c_3}{c_1} e_1$$

Let us introduce for symmetry, designations

$$\begin{aligned} x_1 &= \psi(t) - \psi^*(t) \\ x_2 &= \theta(t) - \theta^*(t) , \\ x_3 &= u(t) - u^*(t) \end{aligned} \qquad (3.5)$$

We have then

$$\delta\psi_2 = x_1(t_3), \quad \delta\theta_2 = x_2(t_3), \quad \delta u_3 = x_3(t_3) \qquad (3.6)$$

Inserting expressions (3.4) into (3. 3) we obtain accounting for designations (3.6)

$$\rho^2 = (c \cdot x(t_3))^2 + \sin^2\varepsilon_3^* \, (e \cdot x(t_3))^2 \qquad (3.7)$$

where points designate scalar product of vector $x = \{x_1, x_2, x_3\}$ with vectors $c = \{c_1, c_2, c_3\}$ and $e = \{e_1, e_2, e_3\}$.

Considering (3.7) as a criterion of side control quality submit the side correction synthesis to the aim of the functional (3.1) minimization.

The synthesis of a longitudinal motion stabilization is dependent on the requirement of rapidity of action criterion. This requirement being accomplished the accumulated deviations will be set aside at a minimum of time possible.

b) Side correction.

The synthesis of a control which defines a moment of engine cutoff is performed in assumption that the circular orbit form distartions may be neglected.

Let us linearize equations (1.5) in the vicinity of the programmed motion at the length $[t_1, t_3]$. Accounting for (3.5) we obtain

$$\dot{x} = A(t)x + b(t)(\xi_w + q_w) \qquad (3.8)$$

where ξ_w is a chance deviation of acceleration n_w, and q_w is the controlling force,

$$A = \begin{Vmatrix} 0 & a_{12} & a_{13} \\ 0 & 0 & a_{23} \\ 0 & a_{32} & a_{33} \end{Vmatrix}, \qquad b = \begin{Vmatrix} b_1 \\ b_2 \\ b_3 \end{Vmatrix}$$

Coefficients a_{ij}, b_K are

$$a_{12} = -\Omega n_w^* \sin u^* \operatorname{ctg}\theta^* \csc\theta^* \qquad a_{33} = -\Omega n_w^* \cos v^* \operatorname{ctg}\theta^*$$

$$a_{13} = \Omega n_w^* \cos u^* \csc\theta^* \qquad b_1 = \Omega \sin u^* \csc\theta^*$$

$$a_{23} = -\Omega n_w^* \sin u^* \qquad b_2 = \Omega \cos u^*$$

$$a_{32} = \Omega n_w^* \sin v^* \csc^2\theta^* \qquad b_3 = -\Omega \sin u^* \operatorname{ctg}\theta^*$$

The solution of vector equation (3.8) with zero initial conditions at a moment t_2 takes form

$$x(t) = \int_{t_1}^{t} N(t,\tau) b(\tau) \Delta n_w(\tau) d\tau$$

where t_1 is the engine cutin moment, $\Delta n_w = \xi_w + q_w$, $N(t,\tau)$ is a matrix weight function of equation (3.8). Hence we obtain for the moment t_3

$$x(t_3) = \int_{t_1}^{t_2} N(t_3,\tau) b(\tau) \xi_w(\tau) d\tau + \int_{t_1}^{t_3} N(t_3,\tau) b(\tau) q_w(\tau) d\tau \qquad (3.9)$$

where t_2 is the engine cutoff moment.

As the controlling force q_w is created at the expence of the moment t_2 shift with the respect to its programmed value t_2, for the second integral at the right side of (3.9) we have in approximation:

$$\int_{t_1}^{t_3} N(t_3,\tau) b(\tau) q_w(\tau) d\tau \approx N(t_3,t_2^*) b(t_2^*) n_w^*(t_2^*) \delta t_2 \qquad (3.10)$$

where $\delta t_2 = t_2 - t_2^*$.

The problem of correcting control synthesis is thus reduced to the search of such a shift δt_2 that provides a minimum for the functional (3.7).

Insert (3.9) into (3.7) we find, taking into account (3.10)

$$\rho^2 = (c\cdot P + c\cdot Q\,\delta t_2)^2 + \sin^2\varepsilon_3^* (e\cdot P + e\cdot Q\,\delta t_2)^2 \qquad (3.11)$$

where

$$P = \int_{t_1}^{t_2} N(t_3,\tau)\, b(\tau)\, \Delta n_w(\tau)\, d\tau$$

$$Q = N(t_3, t_2^*)\, b(t_2^*)\, n_w^*(t_2^*) .$$

The shift value δt_2 which provides minimum for the criterion (3.11) is equal to

$$\delta t^\circ = -[(c\cdot Q)^2 + \sin^2\varepsilon_3^* (e\cdot Q)^2]^{-1} [(c\cdot Q)(c\cdot P) + \sin^2\varepsilon_3^* (e\cdot Q)(e\cdot P)]$$

We obtain hence, that the engine cutoff must proceed at the moment t_2 when the condition

$$t_2 - t_2^* = \int_{t_1}^{t_2} \Phi(\tau)\, \Delta n_w(\tau)\, d\tau \qquad \text{is realized;}$$

here $\Phi(\tau) = \dfrac{[(c\cdot Q)c + \sin^2\varepsilon_3^* (e\cdot Q)e]\cdot N(t_3,\tau)\, b(\tau)}{(c\cdot Q)^2 + \sin^2\varepsilon_3^* (e\cdot Q)^2}$.

c) Longitudinal motion control.

In view of the relay control synthesis we linearize the system (1.3) in the vicinity of a circular orbit with the radius r^*. Introducing designations $\eta_1 = \delta y$, $\eta_2 = \delta z$, $\eta_4 = \delta\rho$ and neglecting the term $r^{*-2}\,\delta_\rho$ we obtain

$$\frac{d\eta_1}{d\vartheta} = \eta_2$$

$$\frac{d\eta_2}{d\vartheta} = -\eta_1 + \frac{n_s}{r^*}$$

$$\frac{d\eta_3}{d\vartheta} = \eta_4 \qquad (3.12)$$

$$\frac{d\eta_4}{d\vartheta} = 2r^* n_T$$

$$n_s = \xi_s + q_s$$

$$n_T = \xi_T + q_T$$

where ξ_s and ξ_T are the accidental components of accelerations, q_s and q_T are controlling forces.

As follows from the equation of motion (3.12) the height and the velocity of flight stabilization are realized independently - the first with the help of the control q_S, the second with the help of q_T.

In synthesizing controls q_S and q   we shall consider only a free motion, not taking into consideration
and ξ_T. The last ones are considered here only as the sources of deviations η_1, η_2, η_3, η_4 rise; the q_S and q_T role is to serve to the most rapid elimination of these deviations.

Making use of the results, obtained in [4] , and introducing the zero zones $\mathcal{E}$ and δ we define the laws of controls q_S and q_T formation as follows:

$$q_S = \begin{cases} -q_S^{max} & \eta_2 \geqslant -(\mathrm{sign}\,\eta_1)\sqrt{a^2-(\eta-a\,\mathrm{sign}\,\eta_1)^2} \\ 0 & \eta_1^2+\eta_2^2 \leq \mathcal{E}^2 \\ q_S^{max} & \eta_2 < -(\mathrm{sign}\,\eta_1)\sqrt{a^2-(\eta_1-a\,\mathrm{sign}\,\eta_1)^2} \end{cases} \qquad (3.13)$$

$$q_T = \begin{cases} -q_T^{max} & \eta_4 \geq -(\mathrm{sign}\,\eta_3)\sqrt{2b\,\eta_3\,\mathrm{sign}\,\eta_3} \\ 0 & \eta_3^2+\eta_4^2 \leq \delta^2 \\ q_T^{max} & \eta_4 < -(\mathrm{sign}\,\eta_3)\sqrt{2b\,\eta_3\,\mathrm{sign}\,\eta_3} \end{cases} \qquad (3.14)$$

where q_S^{max} and q_T^{max} are modules of the maximum values of controls q_S and q_T,

$$a = \frac{q_S^{max}}{\tau^*} \quad , \qquad b = 2\tau^*\, q_T^{max}$$

Note, that the law (3.13) is written in assumption that $|\eta_1| \leq 2a$ or $|\delta\tau/\tau| \leq q_S^{max}$. The controls (3.13), (3.14) counter deviations with accuracy to $\mathcal{E}$ and δ. One can obtain an expression of η_1, η_2, η_3, η_4 through accelerations n_S and n_T - by analogy with the side motion - by means of equations (3.12) solution. The transform from ϑ to t one can realize by means of approximate ratio $d\vartheta \approx \Omega dt$ resulting from (1.4) when $p \approx \gamma^*$.

References

I. Дубошин Г.И. Небесная механика. Основные задачи и методы. Физматгиз, I963.

2. Копнин Ю.М. К задаче о повороте плоскости орбиты спутника. Космические исследования, I965, вып. 4.

3. Гуськов Ю.П. Метод управления поворотом плоскости круговой орбиты спутника ПММ, I963, вып. 3.

4. Понтрягин Л.С. и другие Математическая теория оптимальных процессов. Физматгиз, I96I.

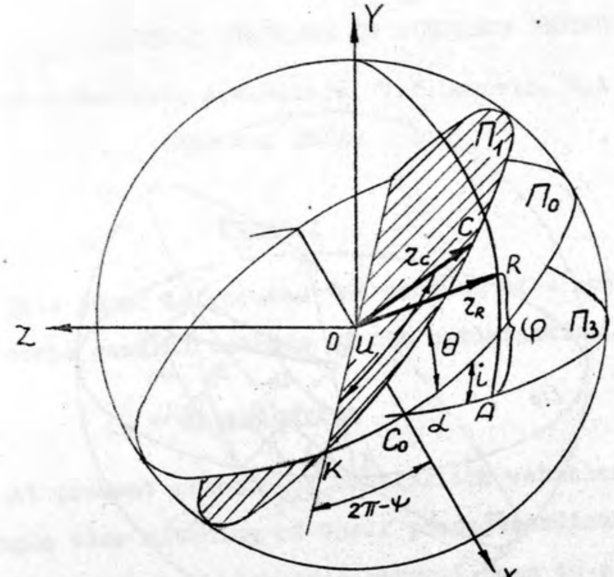


Figure 1

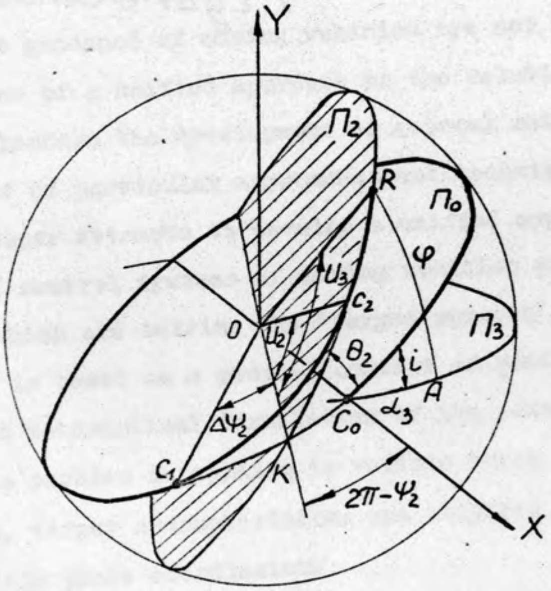


Figure 2

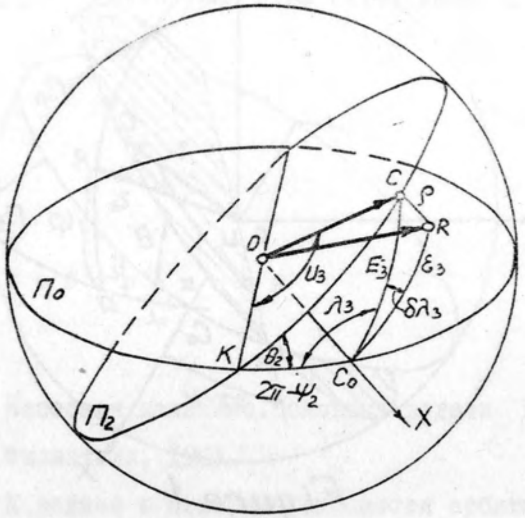

C
ρ
R
O
U_3
E_3
ε_3
$\delta\lambda_3$
λ_3
Π_0
K
θ_2
$2\pi-\psi_2$
C_0
X
Π_2

Figure 3.

57.8

GENERAL PROBLEMS OF GUIDANCE THEORY

E.A.Fedosov, A.M.Batkov, V.F.Levitin, V.A.Skripkin
(Moscow, USSR)

Summary

This paper introduces the concept of a guidance system and presents general methods of its optimization.

Introduction

At present methods of controlling vehicles to induce a single time matching of their phase coordinates with those of another vehicle which in a general case is a moving one are being developed independently for vehicles of various types. So for instance theories of remote control, homing and self-sufficient guidance of moving vehicles are not interconnected. The absence of a unified approach to the solution of these problems hinders the development of general methods and mutual enrichment of particular approaches and techniques.

This paper attempts to develop a unified approach to the design of control systems of moving vehicles of different classes which are solving the "target pursuit" task. The approach is based on a general concept of guidance system and a unified mathematical formulation of the control system synthesis problem as applied to various types of moving vehicles, target characteristics and utilized data concerning the vehicle phase coordinates.

Modern variational methods permit to get the solution for the synthesis problem with account of possible constraints reflecting specific features of various controlled vehicles.

I. Guidance System Concept

Assume there are two controlled vehicles whose position in phase coordinate space varies in time. Let one of the vehicles be a guided one and the other be a target.

The relative position of the two vehicles is described with some functional I of their phase coordinates. The simplest case of the functional I is the distance between the guided vehicle and the target in the three dimensional space. Sometimes their relative position is also determined with coordinate derivatives and angular positions of the vehicles.

Let the control of the guided vehicle aimed at changing the minimum value of I at least at one moment of the finite or infinite time interval T be defined as the guidance of the vehicle to the target and the moment when the lower bound is reached, $I(t)$, be defined as the rendezvous time t_{zv}. The rendezvous time is determined with the condition

$$I(t_{zv}) = \inf_{t \in T} I(t), \tag{I}$$

The control system ensuring the guidance of the vehicle to the target in accordance with its capabilities and limited information about phase coordinates of both vehicles is called a guidance system.

An essential distinction of the guidance system is that the aim of the control is achieved at least at a moment that is unknown during the control process and depends on the relative position of the vehicles.

The main feature of the guidance system is the accuracy determined with the minimum value of the functional $I(t)$ in the guidance process or, in other words, with its value at the rendezvous time $I(t_{zv})$.

To control the guided vehicle it is important to determine the current measure $I(t_{zv})$ as function $x(t)$ of the phase coordinate vector of both the guided vehicle and the target at an arbitrary moment $t < t_{zv}$ of the guidance interval T on the basis of a known law of phase coordinate variation of both vehicles in the interval (t, t_{zv}). Rational current measures should satisfy the following conditions:

I) $x(t_{zv}) = I(t_{zv})$.

2) For every moment t of the guidance interval the current measure $x(t)$ is a monotone nondecreasing odd function of $I(t_{zv})$.

3) Under the conditions of available information about the relative coordinates and $x(t) = 0$ the condition $I(t_{zv}) = 0$ is achieved at the minimum (or zero) values of the control in the interval t, t_{zv}.

If $I(t_{zv})$ determines the distance between the vehicles ther for $x(t)$ it is convenient to use the miss vector $\bar{h}(t)$ defining the distance between the target and the relative trajectory that is constructed on the assumption that the moving vehicle and the target are in a relative rectilinear motion. The miss is correlated with the distance vector $\bar{R}$ and the relative velocity $\bar{V}$ by the formula

$$\bar{h} = \bar{R} - \bar{v}_o (\bar{R} \cdot \bar{V}_o).$$

where the index "o" denotes unit vectors.

By differentiating the miss it is possible to find the correlation for the miss module change rate

$$\dot{h} = j_h \cdot t_y$$

and the angular rate of the vector h in a plane normal to the relative velocity vector

$$\Omega = \frac{j_\Omega}{h} t_y ,$$

where j_h is projection of the total relative acceleration vector on the miss direction; j_Ω is projection of the total relative acceleration vector on the direction normal to the miss and relative velocity, and $t_y = -\frac{R(\bar{R}_o \cdot \bar{v}_o)}{v}$ is time till rendezvous with the current values of distance and relative velocity.

Relative acceleration vector projection on the relative velocity direction equal to $\dot{v}$ affects the miss only through the quantity t_y .

Thus it follows that the moving vehicle and the target can achieve maximum miss variation velocity if their acceleration vectors are collinear to the miss.

The specific features of guidance systems are determined with information and energy constraints that are described with irreversible nonlinear transforms having no inverses. Information constraints are connected with specific characteristics of information transducers and their position in phase coordinate space as well as with the impossibility of certain vehic phase coordinate measurement and measurement errors.

Energy constraints are determined by vehicle types, guidance process environment, control energy reserve and control force variation range. Suppose further that differential equations describing the relative motion of vehicles in the guidance process are known:

$$\dot{x} = F(x, u, j_t, \xi, t) \tag{2}$$

In general case the vector x includes coordinates of the guided vehicle and the target as well as the current guidance accuracy measure, hence equation (2) is usually called coupling equation. Controls u and j_t determine vehicle motions and in general case they depend on measured values Y of phase coordinates x.

$\xi(t)$ is a random function including control realization errors and reflecting information incompleteness about the vehicles.

Energy constraints of the guidance system determine essential nonlinearities of function F as well as additional conditions under which equation (2) is solved.

Information constraints define dependance of vector Y on x:

$$Y = G\left[Hx + v\right]. \tag{3}$$

Matrix H allows to single out the measured components of vector x and function G defines transducer nonlinearities and information interrupts.

The block diagram of the guidance system is shown in Fig.I.

2. Guidance System Optimization Methods

Guidance system optimization problem consists in defining control vector U depending on vector $Y(t)$ from the condition

$$\min M\left\{\Phi\left[I(t_{zv})\right]\right\}, \qquad (4)$$

where M is mathematical expectation operation; Φ is function of guidance system error, and the rendezvous moment t_{zv} is determined by the condition (I). Energy constraint presence in minimizing (4) necessitates taking into account conditions of the form

$$u(t)\in V, \qquad (5)$$

where U is a closed region of isoperimetrical restrictions

$$\int_{t_0}^{t_{zv}} f(u,x,t)\,dt \le C \qquad (6)$$

or of the mathematical expectations of conditions (6) / t_0 is a starting moment of guidance/.

It should be noted that control must be physically realizable that is,it should depend only on the past values of the observed process realization Y defined by expression (3).

An essential feature of the considered variational problem is determined by condition (I). In a determinate case and with Φ = I it reduces to a variational problem with a parameter.

Assuming the rehdezvous moment t_{zv} is defined by the condition

$$g\left[x(t_{zv}),u,j_t\right]=0, \qquad (7)$$

necessary conditions for control optimality by applying the maximum principle can be written in the form /see ref. [I]/

$$\max_{u\in U} M\left[H(\Psi, x, u, j_t, t)/Y\right], \tag{8}$$

where

$$H(\Psi, x, u, j_t, t) = \Psi^T F(x, u, t), \tag{9}$$

$$\frac{d\Psi_i}{dt} = -\frac{\partial H}{\partial x_i}, \tag{IO}$$

$$\Psi_0(t_t) = -1, \quad \Psi_i(t_{zv}) = -\frac{\partial \Phi[x(t_{zv})]}{\partial x_i(t_{zv})} \quad (i=0,1,\dots,n) \tag{II}$$

Conditions (6) are taken into account by phase coordinate vector expansion x and the moment t_{zv} is defined by the condition (7).

If the target also uses an optimal control maximizing guidance errors condition (8) can be rewritten in a form generalizing the results [2]:

$$\max_{j_t\in V} \min_{u\in V} M\left[H(\Psi, x, u, j_t, t)/Y\right] \tag{I2}$$

or

$$\max_{u\in U} \min_{j_t\in V} M\left[H(\Psi, x, u, j_t, t)/Y\right] \tag{I3}$$

depending on the amount of information that the target and the guided vehicle possess concerning the strategies of each other. It is assumed that for control purposes in the guidance process the target also uses vector Y or some of its components.

It follows from optimality conditions that control $u(t)$ at the moment t is determined by the function of mathematical expectation of an error at the rendezvous moment t_{zv} with a known realization of vector Y in the interval (t_o, t).

In a particular case

$$\Phi\left[I(t_{zv})\right] = x^T(t_{zv}) Q x(t_{zv}) \qquad (I4)$$

of the constraint

$$M\left[\int_{t_0}^{t_1} u^T P u\, dt\right] = C \qquad (I5)$$

for a linear coupling equation

$$\begin{aligned} \overset{\circ}{x} &= Ax + Gu + \varkappa(t) j_t(t), \\ x(t_0) &= C_0. \end{aligned} \qquad (I6)$$

the necessary conditions for optimality reduce to a nonlinear integral equation of special kind

$$u(t) = S_g P^{-1} G^T k^T(t_{zv}, t) Q M\left[x(t_{zv}) / Y(\tau), \tau \leq t\right], \qquad (I7)$$

where Q is a positive and P is a definitely positive matrices, respectively; P^{-1} is inverse P matrix; $K(t_{zv}, t)$ is weight function matrix of equation (I6)

$$S_g z = \begin{cases} z & z \in U, \\ 1 & z \notin U. \end{cases} \qquad (I8)$$

$x(t_{zv})$ is defined by the solution of equation (I6) at the moment t_{zv}.

At present most fully developed are design methods for optimal linear systems based on the minimum of miss root mean square at a fixed time t_1. Available results are briefly summarized below.

Suppose phase coordinate vector of the guided vehicle satisfies set of equations (I6). The result of the vehicle coordinate measurements by the information transducers is described by p-dimensional vector

$$\underset{(p\times1)}{Y(t)} = \underset{(p\times n)}{H(t)}\,\underset{(n\times1)}{x(t)} + \underset{(p\times1)}{v(t)} \tag{I9}$$

For simplicity assume that vectors c_0, $u(t)$, $v(t)$ are independent at $t \geqslant t_0$ and

$$M\left[u(t)\,u^{T}(\tau)\right] = Q(t)\,\delta(t-\tau),$$
$$M\left[v(t)\,v^{T}(\tau)\right] = R(t)\,\delta(t-\tau). \tag{20}$$

Define the moving vehicle control system. After the linear processing of the measurement results this system determines the value of the prescribed linear form $\underset{(1\times n)}{\varkappa^{*}}\,\underset{(n\times1)}{x(t_1)}$ at time $t = t_1$ $(t_1 \leq t_{zv})$ with minimum root mean square error. Such systems can be represented by weight function matrix $K(t,\tau)$.

System error $\varepsilon(t_1)$ is expressed as

$$\varepsilon(t_1) = \int_{t_0}^{t_1} \varkappa^{*} K(t_1,\tau)\,H(\tau)\,d\tau + \int_{t_0}^{t_1} \varkappa^{*} K(t_1,\tau)\,v(\tau)\,d\tau - \varkappa^{*} x(t_1). \tag{2I}$$

Block diagram of this system is presented in Fig.2.

Using the property of conjugate systems [2] we transform expression (2I) for the error to

$$\varepsilon(t_1) = C_0^{*} z(t_0) + \int_{t_0}^{t_1} z^{*}(\tau)\,\mathcal{L}(\tau)\,j_t(\tau)\,d\tau + \int_{t_0}^{t_1} z^{*}(\tau)\,G(\tau)\,u(\tau)\,d\tau + \int_{t_0}^{t_1} \varkappa^{*} K(t_1,\tau)\,v(\tau)\,d\tau, \tag{22}$$

where $z(t)$ satisfies the set of equations

$$\begin{cases} \mathring{z} = -A^{T}(t)\,z(t) - H^{T}(t)\,K(t_1,t)\,\varkappa, \\ z(t_1) = -\varkappa \end{cases} \tag{23}$$

Averaging yields

$$\bar{\mathcal{E}}^2(t_1) = \int_{t_0}^{t_1} z^T(\tau) G(\tau) Q(\tau) G^T(\tau) z(\tau)\, d\tau + \\ + \left\{ \int_{t_0}^{t_1} z^T(\tau) \varkappa(\tau) j_t(\tau)\, d\tau \right]^2 + z^T(t_0) \overline{C_0 C_0^T} z(t_0) + \\ + \int_{t_0}^{t_1} \varkappa^T k(t_1,\tau) R(\tau) k^T(t_1,\tau) \varkappa\, d\tau . \tag{24}$$

After introducing new variables satisfying the set of equations

$$\begin{aligned} \overset{o}{z}_{n+1} &= z^T(t) \varkappa(t) j_t(t), & z_{n+1}(t_0) &= 0; \\ \overset{o}{z}_{n+2} &= z^T(t) G(t) Q(t) G^T(t) z(t), & z_{n+2}(t_0) &= 0; \\ \overset{o}{z}_{n+3} &= \varkappa^T k(t_1,t) R(t) k^T(t_1,t) \varkappa, & z_{n+3}(t_0) &= 0, \end{aligned} \tag{25}$$

the synthesis problem can ultimately be formulated as follows:

Given the set of differential equations (23) and (25), it is necessary to choose matrix $k(t_1, t)$ which after satisfying necessary boundary conditions would minimize error mean square

$$\mathcal{E}^2(t_1) = z^2_{n+1}(t_1) + z_{n+2}(t_1) + z_{n+3}(t_1) + z(t_0) \overline{C_0 C_0^T} z(t_0). \tag{26}$$

It should be noted that if vector $\varkappa$ can take arbitrary values we have a filtration problem.

By using Pontryagin maximum principle for solving the problem we find that the synthesis problem solution reduces to the solution of the following set of equations:

$$\overset{o}{z} = -A^T z - H^T(t) k^T(t_1,t) \varkappa, \qquad z(t_1) = -\varkappa;$$

$$\overset{o}{z}_{n+1} = z^T(t) \mathcal{L}(t) j_t(t),. \qquad z_{n+1}(t_0) = 0;$$

(27

$$\overset{\circ}{z}_{n+2} = z^T(t)GQG^Tz(t), \qquad z_{n+2}(t_0)=0;$$

$$\overset{\circ}{z}_{n+3} = \varkappa^T k(t_1,t)R(t)k^T(t_1,t)\varkappa, \qquad z_{n+3}(t_0)=0;$$

$$\overset{\circ}{\psi} = A(t)\psi - 2z_{n+1}(t_1)\mathcal{L}(t)j_t(t)+2GQG^T, \quad \psi(t_0)=2\overline{C_0C_0^T}z(t_0)$$

$$k(t_1,t)\varkappa^T = -\frac{1}{2}R^{-1}(t)H(t)\psi(t),$$

$$\overline{\varepsilon^2}(t_1) = -\frac{1}{2}\varkappa^T\psi(t_1),$$

Equations (27) allow to define optimal weight function of the system not only with the prescribed $j_t(t)$ but also with the worst behaviour of the opponent. In this case the guidance system can be determined from the conditions [2]:

$$\text{I} \quad \max_{j_t\in U}\ \min_{k}\ \overline{\varepsilon^2}\left[k(t_1,t), j_t(t)\right]; \tag{28}$$

$$\text{II} \quad \min_{k}\ \max_{j_t\in U}\ \overline{\varepsilon^2}\left[k(t_1,t), j_t(t)\right]. \tag{29}$$

To solve problem I it is necessary to maximize over $j_t \in V$ for the set of equations (27); the solution of problem II changes equation (9) only for z_{n+1} and ψ which are written as follows:

$$\begin{aligned} z_{n+1} &= \max_{j_t\in U} z^T(t)\mathcal{L}(t)j_t(t), \\ \overset{\circ}{\psi} &= A(t)\psi - \psi_{n+1}\frac{\partial}{\partial z^T}\left[\max_{j_t} z^T\mathcal{L}(t)j_t(t)\right]. \end{aligned} \tag{30}$$

Suppose all the inputs to the control system and measurement errors are random with zero mathematical expectation and vector is arbitrary then substitution in (27)

$$\begin{aligned} \psi(t) &= 2P(t)z(t), \\ P(t_0) &= \overline{C_0C_0^T} \end{aligned} \tag{31}$$

yields a known equation that should be satisfied by variance matrix of the optimal system and by optimal estimation of phase coordinates

$$\frac{dP}{dt} = PA^T + AP - PH^T R^{-1} HP + GQG^T,$$
$$P(t_0) = \overline{C_0 C_0^T}, \tag{32}$$
$$\frac{d\hat{x}(t)}{dt} = \left[A(t) - k(t_1 t)H(t)\right]\hat{x}(t) + k(t_1 t)\left[Hx(t) + v\right].$$

In the absence of random forces $u(t)$ applied to the vehicle equations defining optimal weight function sections can be written as

$$\overset{o}{z} = -A^T z - H^T k^T(t_1, t)\varkappa,$$
$$z(t_1) = -\varkappa, \tag{33}$$
$$\dot{\psi} = A(t)\psi, \qquad \psi(t_0) = 2\overline{C_0 C_0^T}\, z(t_0),$$

$$k^T(t_1, t)\varkappa = -\frac{1}{2}R^{-1}(t)H(t)\psi(t),$$

$$\bar{\varepsilon}^2(t_1) = -\frac{1}{2}\varkappa^T\psi(t_1).$$

The solutions of this set of equations are

$$\varkappa^T k(t_1, t) = \varkappa^T\Phi(t_1, t_0)\left[E + \overline{C_0 C_0^T}\mathcal{D}\right]^{-1}\overline{C_0 C_0^T}\,\Phi^T(t, t_0)H^T R^{-1}(t) \tag{34}$$

$$\bar{\varepsilon}^2(t_1) = \varkappa^T\Phi(t_1, t_0)\left[E + \overline{C_0 C_0^T}\mathcal{D}\right]^{-1}\overline{C_0 C_0^T}\,\Phi^T(t_1, t_0)\varkappa, \tag{35}$$

where $\Phi(t, \tau)$ is transition matrix of the vehicle

$$\mathcal{D} = \int_{t_0}^{t_1}\Phi^T(\mu, t_0)H^T(\mu)R^{-1}(\mu)H(\mu)\Phi(\mu, t_0)\,d\mu. \tag{36}$$

At present problems of optimal control system realization in the form of simple correcting devices have not yet been adequately investigated. Data processing algorithms produced with account of additional control constraints are practically realizable only by utilizing computers, and, besides, sometimes optimal systems are critically sensitive to inputs different from design ones.

However guidance systems that are optimal in accuracy at a fixed time can be realized by using stationary correcting devices.

Consider as an example block diagram of a guidance system shown in Fig.3 where $W_v(t,\tau)$ is a known weight function of the common part; $M_i(t,\tau)$; $P_i(t,\tau)$; $R_i(t,\tau)$ are weight functions of the prescribed links and $\Phi_i(t-\tau)$ are weight functions of the correcting links.

Suppose that as a result of solving an accuracy optimal fixed time problem we find weight function sections $Q_i(t_1,\tau)$ $(i=1,\dots,n)$ connecting inputs at points a_i $(i=1,\dots,n)$ with the output at point c and defined in interval $(-\infty \leq \tau \leq t_1)$. It can be shown that integral equations connecting closed system weight functions and separate links become

$$Q_i(t_1,\tau) = W \circ \Phi_i \circ P_i \circ \mathcal{M}_i - \sum_{q=1}^{n} Q_q \circ \mathcal{M}_q^{-1} \circ R_q \circ W_c \circ \Phi_i \circ P_i \circ \mathcal{M}_i . \qquad (37)$$

The Laplace transform over the variable $t_1-\tau$ in both sides of (37) in case of stationary correcting filters $\Phi_i(t-\tau)$ yields the following relationship defining

$$\Phi_i(s) = \frac{\mathcal{L}[Q_i \circ \mathcal{M}_i^{-1} \circ P_i^{-1}]}{W(s,t_1) - \sum_{q=1}^{n} \mathcal{L}[Q_q \circ \mathcal{M}_q^{-1} \circ R_q \circ W]} \qquad (38)$$

Synthesis methods presented here are common for such seemingly different guidance systems as homing guidance systems, remote control and self-sufficient systems.

Examples of block diagrams of homing guidance systems, remote control and self-sufficient systems are shown in Fig.4.

It is an inherent feature of all the guidance systems to measure the moving vehicle coordinates (MVC) directly on board the moving vehicle.

A homing guidance system distinction is that relative coordinates (RC) are measured on board the moving vehicle.

A specific feature of remote control is that absolute or relative coordinates of the target and the moving vehicle are being measured in a third point (moving or fixed). Control commands are transmitted on board over data link (DL).

A characteristic feature of a self-sufficient control system is that predetermined coordinates of the target are stored on board the moving vehicle and corresponding control commands are produced on the basis of calculated relative coordinates.

Block Diagram Notations

TC - target coordinates;

KR - kinematic relationships relating the coordinates of the target and the moving vehicle with directly measured quantities;

ME - measurement errors;

CF - correcting filters;

RC - relative coordinates;

D - disturbances acting on the moving vehicle;

MV - moving vehicle;

MVC - moving vehicle coordinates;

DL - data link.

3. An Example of Accuracy Optimal Guidance System Synthesis (Homing Guidance System)

Consider an example of synthesis of an optimal homing guidance system in a plane.

Let kinematic coupling link (assuming constant closing velocity $\dot{D}$ of the vehicles) be described by the differential equation

$$\dot{\omega}_{\ell_0} = \frac{2}{t_{\tau v}-t}\omega_{\ell} + \frac{j_t - j_v}{|\dot{\mathfrak{D}}|(t_{\tau v}-t)} \quad , \tag{39}$$

where $t_{\tau v}$ is homing guidance time; ω_{ℓ_0} is line if sight angular rate; j_t , j_v are acceleration components normal to line of sight of the target and the guided vehicle, respectively

We shall take into account the following inputs:

I. Initial guidance error. Instantaneous values of line of sight and miss angular rates are related by the expression

$$h(t) = \frac{\mathfrak{D}^2(t)}{|\dot{\mathfrak{D}}|}\omega_{\ell_0}(t) = |\dot{\mathfrak{D}}|(t_{\tau v}-t)^2\omega_{\ell_0} \tag{40}$$

Assume the initial error is a random variable with given variance $\overline{\omega^2}_{\ell_0}$, $M[\omega_{\ell_0}] = 0$.

2. Target maneuver. We shall consider it as a random quantity characterized with its variance $\overline{j}^2_t$, $M[j_t] = 0$; this random quantity is constant during all the time for which the optimal system is designed.

3. Constant measurement error characterized by variance $\overline{\ell}^2_0$, $M[\ell_0] = 0$ and representing an additional line of sight pseudo angular rate.

4. Interference appearing in measuring line of sight angular rate. It is assumed that the interference is an additive signal in line of sight angle with the correlation function

$$R_{vv}(t,\tau) = \sigma_1\delta(t-\tau) + \frac{\sigma_2\delta(t-\tau)}{|\dot{\mathfrak{D}}|^2(t_{\tau v}-t)^2} \tag{4I}$$

5. Disturbances acting on the moving vehicle will not be

taken into account. Accordingly the transducer measuring own coordinates of the moving vehicle will be supposed nonexisting. Such a simplification is reasonable if in design process the target maneuver is taken into account.

In accordance with the above mentioned assumptions concerning the homing guidance system synthesis problem the equations of the vehicle become

$$\begin{cases} \overset{\circ}{x}_1 = \dfrac{2}{t_{rv}-t}x_1 + \dfrac{x_4}{|\dot{D}|(t_{rv}-t)}, \\ \overset{\circ}{x}_2 = x_1, \\ \overset{\circ}{x}_3 = 0, \\ \overset{\circ}{x}_4 = 0 \end{cases} \tag{42}$$

The measured quantity is

$$Y = x_2 + t x_3 + V(t), \tag{43}$$

$$H(t) = \|0, 1, t, 0\|. \tag{44}$$

The transition matrix of the set of equations (42) is easily defined and can be written as

$$\Phi(t,\tau) = \left\| \begin{matrix} \dfrac{(t_r-\tau)^2}{(t_{rv}-t)^2} & 0 & 0 & \dfrac{1}{2|\dot{D}|}\left[\dfrac{(t_{rv}-\tau)^2}{(t_{rv}-t)^2}-1\right] \\ \dfrac{(t_{rv}-\tau)(t-\tau)}{t_{rv}-t} & 1 & 0 & \dfrac{(t-\tau)^2}{2|\dot{D}|(t_{rv}-t)} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix} \right\| \tag{45}$$

We shall find the section of the weight function connecting line of sight angle $Y(t)$ with the miss from the miss mean square minimum at the moment t_1 separated from the rendezvou by the period $\Delta = t_{rv} - t_1$.

Since the miss is related with the coordinate $x_1 - \omega_\ell$ by formula (40) vector $\mathfrak{X}$ should be expressed as

$$\mathcal{X} = \begin{Vmatrix} |\dot{\mathcal{D}}| \ \Delta^2 \\ 0 \\ 0 \\ 0 \end{Vmatrix} \tag{46}$$

and

$$\mathcal{X}^T K(t,\tau) = |\dot{\mathcal{D}}| \Delta^2 K_{11}(t_1,\tau) \tag{47}$$

Assuming the initial target maneuver error and measurement error ω_ℓ are not correlated the variance matrix of initial conditions can be written as

$$\overline{C_0 C_0^T} = \begin{Vmatrix} W_{\ell_0}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \overline{b_0^2} & 0 \\ 0 & 0 & 0 & \overline{j^2} \end{Vmatrix} \tag{48}$$

Substituting expressions from (44) to (48) into (33) and making some transforms to determine weight function $K_{11}(t_1,\tau)$ relating $Y(t)$ with line of sight angular rate yields the equation

$$|\dot{\mathcal{D}}| \Delta^2 K_{11}(t,\tau) = \frac{1}{G_1 + G_2/|\dot{\mathcal{D}}|^2 (t_{tv}-\tau)^2} \Big\{ \frac{|\dot{\mathcal{D}}| t_{tv}^3}{t_{tv}-\tau} \overline{W_{\ell_0}^2} + \frac{\tau^2 (t_{tv}^2 - \Delta^2)}{4|\dot{\mathcal{D}}| (t_{tv}-\tau)} \overline{j_t^2} -$$
$$- \int_0^{t_1} |\dot{\mathcal{D}}| \Delta^2 K_{11}(t_1 \mu) \frac{t_{tv}^2 \mu}{t_{tv}-\mu} d\mu \frac{\bar{\omega}_{\ell_0}^2 \tau^2}{t_{tv}-\tau} - \int_0^{t_1} |\dot{\mathcal{D}}| \Delta^2 K_{11}(t_1 \mu) \frac{\mu^2}{4|\dot{\mathcal{D}}|(t_{tv}-\mu)} d\mu \frac{\tau^2}{(t_{tv}-\tau)} \overline{j_t^2} - \tag{49}$$
$$- \int_0^{t_1} |\dot{\mathcal{D}}| \Delta^2 K_{11}(t_1 \mu) \bar{b}_0^2 \mu \, d\mu \, \tau \Big\} \quad (0 \le \tau \le t_1)$$

It is assumed here that the homing guidance system is switched on at $t_0 = 0$.

To simplify optimality equation (49) divide both sides of it by $\frac{|\dot{D}|\Delta}{t_{zv}-\tau}$; denote $\overline{a}_0^2 = t_{zv}^2\,\overline{\omega}_{\ell o}^2$; $a_1^2 = \overline{\dot{j}^2}/|\dot{D}|^2$ and substitute

$$W_z(t_1,\tau) = \frac{\Delta}{t_{zv}-\tau} K_{11}(t_1,\tau) \tag{50}$$

The weight function $W_z(t_1,\tau)$ relates the quantity $\ell/|\dot{D}|$, linear mismatch, divided by closing velocity module with line of sight angular velocity multiplied by the time $\Delta = t_{zv} - t_1$ left till the moving vehicle meets the target.

In accordance with the accepted notations equation (I3) is rewritten as

$$W_z(t_1,\tau) - \frac{1}{G_1(t_{zv}-\tau)^2+\frac{G_2}{|\dot{D}|^2}}\Big\{\overline{Q_0^2}\,\tau\int_0^{t_1} W_z(t_1,\mu)\,d\mu + \overline{a_1^2}\,\frac{\tau^2}{4}\int_0^{t_1} W_z(t_1,\mu)\mu^2 d\mu + \overline{b_0^2}\,\tau(t_{zv}-\tau)\int_0^{t_1} W_z(t_1,\mu)(t_{zv}-\mu)\mu\,d\mu\Big\} = \tag{5I}$$

$$= \frac{1}{G_1(t_{zv}-\tau)^2+G_2/|\dot{D}|^2}\Big\{\overline{a_0^2}\,\frac{t_{zv}\tau}{t_{zv}-t_1} + \overline{a_1^2}\,\frac{\tau^2}{4}\cdot\frac{2t_{zv}-t_1}{t_{zv}-t_1}\Big\}.$$

The solution of (5I) or of the equivalent (49) is given by the formulas from (34) to (36).

Optimal system variance can be written as

$$\overline{\varepsilon^2}(t_1) = \overline{\omega_{\ell o}^2}\,t_{zv}^2\,|\dot{D}|^2\Big[t_{zv}^2 - \Delta^2\int_0^{t_1} W_z(t_1,\mu)\mu\,d\mu\Big] + \overline{\dot{j}_t^2}\,\frac{t_{zv}-\Delta^2}{4}\Big[(t_{zv}^2-\Delta^2) - \int_0^{t_1} W_z(t_1,\mu)\mu^2 d\mu\Big]. \tag{52}$$

4. Some Problems of Guidance Theory

At present guidance systems are designed in two phases. At the first phase of kinematic investigation a guidance law is chosen that defines the relationship between the phase coordinates of the guided vehicle and those of the target under idealized conditions without taking into account information element errors, certain energy constraints and with the prescribed law of target motion. At the second dynamic phase parameters of smoothing filters are only refined.

The method presented in this paper allows to solve the problem and simultaneously it takes into account kinematic and dynamic aspects. Taking into account phase coordinate measurement errors and energy constraints defines not only the parameters of the guidance law but its structure too. Without taking into account random errors there exist infinite number of guidance laws and with taking into account these random errors the optimization problem has a unique solution. To our mind guidance laws existing at present are only successful particular solutions of the general problem. The investigation of this problem may become a subject of further development of guidance theory.

The investigation of the problems considered in this paper shows that the hypothesis concerning the target motion considerably affects the guidance system structure and xxx characteristics of its errors. At present assumptions regarding the nature of the target motion are based on typical laws of its phase coordinate variation. From the point of view of the considered methods it is more expedient to define the hypothesis concerning the target motion depending on the degree of its knowledge of phase coordinates of the guided vrhicle and the control system of the latter. The solution of this problem can be based on methods of theory of games and their development as applied to optimal guidance problems in the presence of random disturbances.

On the one hand the accuracy of the optimal guidance systems that are designed without taking into account information and energy constraints is sufficiently high and it considerably reduces when the constraints are taken into account. Onthe other hand taking into account constraints considerably complicates the guidance systems and makes them more sensitive to the assumptions regarding the system functioning conditions. In this case guidance systems are described by nonlinear operators even in the presence of Gauss inputs. At the same time engineering methods of guidance problem solving with due account of constraints and realization simplicity have not yet been elaborated as it would be necessary. There are no approximate general synthesis methods which could have been used to get correcting devices in nonlinear automatic control systems with random inputs.

The attempts to unify various existing guidance system design methods entail great difficulties due to the absence of automatic control system optimization methods in case of restrictions in regard to their structure. These problems include optimization of nonstationary systems by using stationary correcting devices or of nonlinear systems by using linear links. Automatic control system simplicity and reliability requirements often reduce to the solution of the above mentioned problems.

The solution of these problems is one of the fundamental tasks of the guidance theory.

REFERENCES

I. Modern Automatic Control System Design Methods. Analysis and Synthesis. Moscow, "Mashinostroyenie" Publishing House, I967.

2.ALEXANDROV V.M., Minimax Approach to the Solution of the Information Processing Problem. USSR Academy of Sciences Proceedings, Engineering Cybernetics, No.5, 1966.

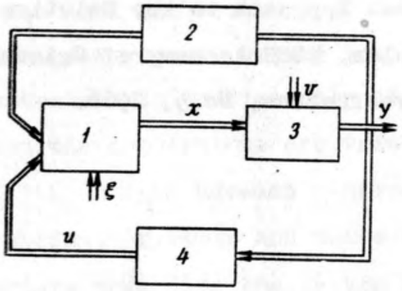


Fig.1.

I - coupling equation ; 2 - target control operator ;
3 - information elements ; 4 - guided vehicle control operator.

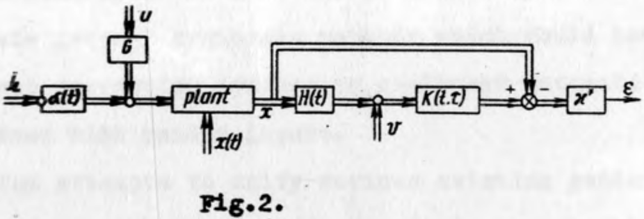


Fig.2.

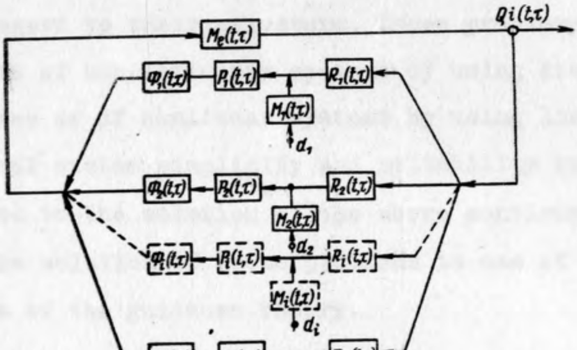


Fig.3.

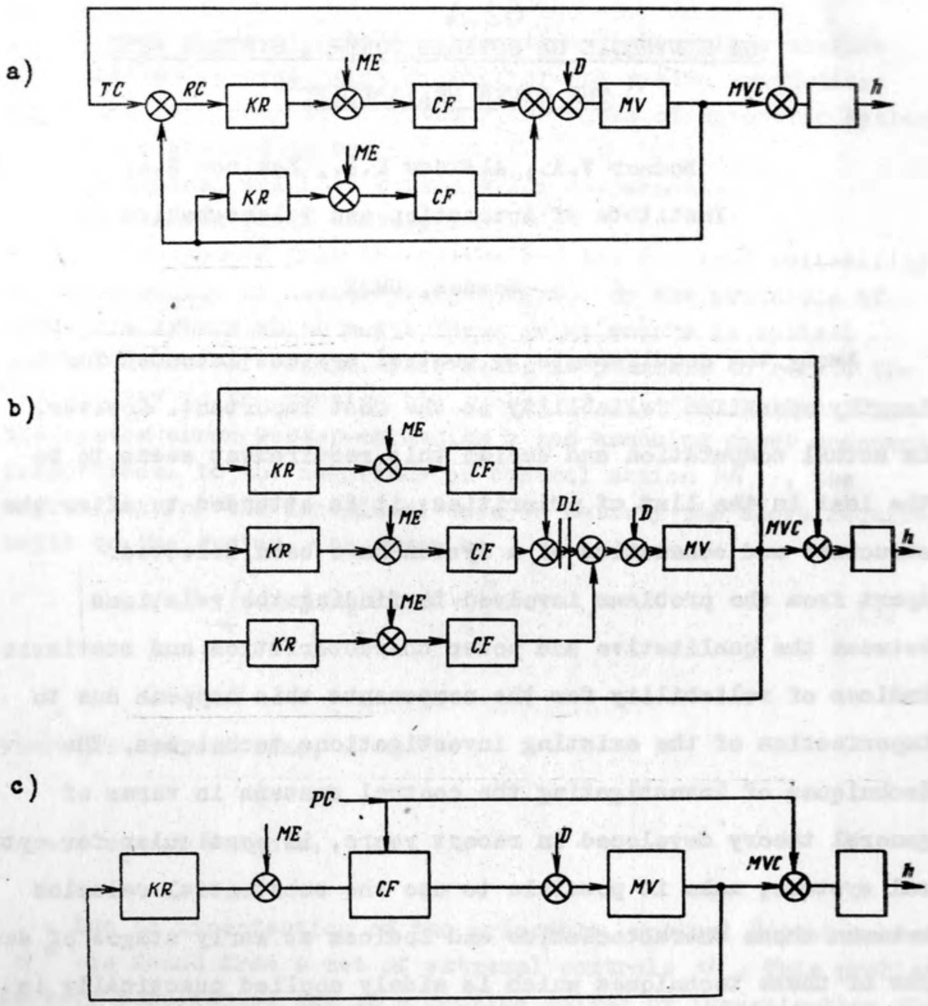

a)
TC
RC
KR
ME
CF
D
MV
MVC
h
ME
KR
CF
b)
ME
KR
CF
ME
DL
D
KR
CF
MV
MVC
h
ME
KR
CF
c)
PC
ME
D
KR
CF
MV
MVC
h

Fig.4.

62.1

ON SYNTHESIS OF OPTIMAL CONTROL SYSTEMS WITH THE GIVEN RELIABELITY[x)]

Bodner V.A., Alexeev K.B., Zakirov R.A.
Institute of Automation and Telemechanics

Moscow, USSR

Among the requirements to control systems intended for lengthy operation reliability is the most important. However, in actual computation and design this requirement seems to be the last in the list of priorities; it is attended to after the structure and schematics of a system have been selected. Apart from the problems involved in finding the relations between the qualitative and power characteristics and statistic indices of reliability for the components this happens due to imperfection of the existing investigations techniques. The techniques of investigating the control systems in terms of general theory developed in recent years, in particular for optimal systems, make it possible to use the substantial relation between those characteristics and indices at early stages of design. One of these techniques which is widely applied practically is Pontriagin's principle of maximum.

This paper is a first attempt to synthetize a certain class of systems by using a generalized optimality criterion which would include requirements to accuracy, power consumption and reliability.

1. Method of study

The essence of Pontriagin's principle is in finding

[x)] Reliability is understood in a narrowed sense as probability of faultless operation.

an extremal value of a certain functional which is a measure of quantitative evaluation of whether the system synthetized meets the requirements. For the given class of automatic systems these are assumed to be:

1) minimal weighted system error dispersion;

2) minimal power consumption.

It is assumed that the system has the required reliability P_3 (probability of faultless operation). By the principle of maximum a system which meets these requirements is optimal and structured of relays. This makes it possible to relate the reliability to the permissible number of switchings. Denoting x the system error vector-column as x and assuming power consumption proportional to the magnitude of control action $|u|$, the expression for the functional used to satisfy the above requirements to the system is given by

$$\mathcal{J} = \int_0^{t_y} (c_1 x^2 + c_2 \dot{x}^2)\,dt + \int_0^{t_y} |u|\,dt \qquad (1)$$

where$t_y$is control time,

c_1 and c_2 are weights.

Due to imperfection of the principle, optimal 8ntrol control u^* is found from a set of extremal controls u . This problem is very complicated and as a routine solved by investigating the control process in phase space. For closed-loop automatic systems the inverse time technique is used in such an investigation. The optimal paths are found as solution to a set of differential equations in inverse time originating in a certain point in the phase space. By the computing procedure of the maximum principle, regions of optimal controls are found where functional (1) acquires a relative minimum. Phase paths of the system state in this region differ in instants when control u is switched. Each component can have one of three possible values +1,0 -1. When the initial differential equation is integrated in inverse time these instants of switching can be expressed by timeτwhich sets a limit e.g. to the action of the control $u = +1$. Then the absolute minimum of functional (1) is found from the cōndition

$$\frac{\partial T}{\partial \tau_1} = 0 \qquad (2)$$

at $\frac{\partial^2 J}{\partial \tau_1^2} > 0$,

where $J(x_0, u) \equiv J(\tau_1)$.

x_0 is the initial value of the error (t=0) or the value of the error at which the relay responds (control switched).

Selection of τ_1 from condition (2) determines the optimal control u^* unambiguously.

Reliability R can also be determined as a function of τ_1. To know control time t_y and total time T of normal system functioning .Then is sufficient the optimal control is found when the problem of the minimum of functional (1) at the additional constraint $\mathcal{P}(\tau_1)=\mathcal{P}_3$ is solved

The Lagrnage technique can be used for this purpose. Condition (2) is then written as

$$\frac{\partial}{\partial \tau_1}\left(J(\tau_1)+\lambda \mathcal{P}(\tau_1)\right)=0, \tag{3}$$

where λ is the Lagrange factor.

It is characteristic that τ_1 is in this case a function of the control error initial value and the required probability P_3 of the system operating without faults. When the control time t_y is also to be found, the optimal control is found from these conditions

$$\frac{\partial[J(t_y,\tau_1)+\lambda\mathcal{P}(t_y,\tau_1)]}{\partial t_y}=0\,; \qquad \frac{\partial[J(t_y,\tau_1)+\lambda\mathcal{P}(t_y,\tau_1)]}{\partial \tau_1}=0\,. \tag{4}$$

Conditions (3) and (4) facilitate selection of the system structure that would meet the requirements, the required reliability included.

2. Application of the method

Let us take up orientation control of a space vehicle. The attitude of the vehicle is given by the orbital base earth centered system OXYZ. Denote the angular rotation velocity $\varkappa$ around the axis OZ as ν (Fig.1). The problem of control to be discussed below is to make the axis OY of the coordinates system OXYZ related to the vehicle rigidly coincide with the geocentric vertical line in a plane which crosses the axis OY of the base system and the velocity vector $\vec{V}$ of the vehicle center of masses.

The orientation system incorporates:

- a one - coordinate vertical to measure the angular error (control error between the axes OY and oy)
- two pairs of jet nozzles fixed on the vehicle body so as to apply the control moment M_y;

amplifiers-converters and solenoid type electromagnetic ~~type~~ valves to control the operation of the nozzles.

Assuming that the vehicle is an absolutely solid body and oy is the main axis of inertia, the dynamic propertied of the plant controlled are described by the equation

$$I\,\frac{d^2\varphi(t)}{dt^2}=M_y(t)+M_B(t), \tag{5}$$

where $\varphi(t)$ is the angular deviation of the vehicle;

I - is the moment of inertia

M_ℓ is the disturbance moment.

Due to the orbital motion, the signal at the input to the orientation system will be

$$\varphi_{\ell x} = \nu t + \varphi_0 , \qquad (6)$$

where φ_0 is the initial value of $\varphi_{\ell x}(t)$ at $t = 0$.

Consequently the orientation error is found by the equation

$$\theta = \varphi_{\ell x}(t) - \varphi(t) = \nu t + \varphi_0 - \varphi. \qquad (7)$$

Fig.2. shows the schematic of a closed-loop orientation system. The requirements that this system should meet have been discribed above.

The orientation system operates in the following manner. Then the vehicle deviates bz the angle $\theta \geqslant \theta_g$ where θ_g - is the permissible orientation error, jet nozzles are switched on. However, the orbital motion and the effect of the disturbance moment lead again to accumulation of another error and action of the mozzles.

Denote the time during which the error $\theta = \theta_g$ is corrected as t_y and the time of the error accumulation as t_n. Then the number n of the cycles of correction and accumulation during time T of the vehicle flight will be

$$n = \frac{T}{t_y + t_n} . \qquad (8)$$

In a relay orientation system there will two switchings (of the relay and the valve); during the flight there will be 2n switchings. If this is associated with the probability of faultless operation $P < P_3$, the requirement to reliability is not met and the orientation accuracy at the given structure of the system does not meet the requirements to the vehicle.

Assuming the angular velocity of the vehicle constant (ν = const.) we find from equations (7)

$$\frac{d\theta}{dt} = \nu - \frac{d\varphi}{dt} , \quad \frac{d^2\theta}{dt^2} = - \frac{d^2\varphi}{dt^2} .$$

Then we can write equation (5) as

$$\frac{d^2\theta}{dt^2} = -\frac{M_y}{I} - \frac{M_B}{I}, \qquad (9)$$

or

$$\frac{d^2\theta}{dt^2} = -a u(t) + m_B(t),$$

where

$$a = \frac{M_g(t)}{I}, \quad u(t) = \frac{M_y(t)}{I}, \quad m_B(t) = \frac{M_B(t)}{I},$$

M_g-is the utmost permissible value of the control moment.

Since $M_y \gg M_B$, the effect of the disturbing moment on (t_y) angular movements of the vehicle during the entire time of the nozzles operation can be neglected During error accumulation the disturbing moment changes the angular velocity of the vehicle by the amount $M_B t_n$ at $M_B = \text{const}$.

Because the time interval t_n is small we will consider $M_B(t)$ as pulsed disturbance at the instant when the nozzles start to act $(t = 0)$, i.e.

$$M_B(t) = M_{B0}\,\delta(t), \qquad (10)$$

where M_{B0} is the pulse area;

$\delta(t)$ is the Dirak function.

Denote

$$x_1(t) = \theta(t) = \nu t + \varphi_0 - \varphi(t),$$

$$x_2(t) = \frac{d\theta}{dt} = \nu - \dot{\varphi}(t),$$

Represent equation (9) as a set of two first-order equations

$$\frac{dx_1(t)}{dt} = x_2(t);$$

$$\frac{dx_2(t)}{dt} = -a u(t), \qquad (11)$$

at the following initial conditions

$$x_1(0)=x_{10}=\varphi_o\,, \quad x_2(0)=x_{20}=\nu-\dot{\varphi}(0) \tag{12}$$

Considering expression (10) for the disturbing moment the latter condition can be given by

$$x_{20}=\nu-m_{bo}\,,$$

where

$$m_{bo}=\frac{M_{bo}}{I}\,.$$

It is characteristic that the action of each input signal and disturbing moment can be seen in initial values of the system state vector

$$\bar{x}(0)=(x_{10},\ x_{20})\,.$$

The values of that vector at time $t=t_y$ are assumed zero, i.e.

$$\bar{x}(t_y)=(0,0) \tag{13}$$

while t_y has not been found yet.

Thus the boundary conditions of eq. (11) or one of eq. (9) written in new notations as

$$\frac{d^2x_1(t)}{dt^2}=-au(t) \tag{14}$$

are specified completely.

By the principle of maximum we will find the expression for the Hamiltonian

$$H=\frac{1}{2}C_1x_1^2(t)+\frac{1}{2}C_2x_2^2(t)+|u(t)|-x_2(t)P_1(t)\,. \tag{15}$$

The equations for conjugated variables ($P_1(t)$ and $P_2(t)$) are given by

$$\frac{dP_1(t)}{dt}=-\frac{\partial H}{\partial x_1}=-C_1x_1(t)\,;$$

$$\frac{dP_2(t)}{dt}=-\frac{\partial H}{\partial x_2}=-C_2x_2(t)-P_1(t)\,. \tag{16}$$

Because the initial and final conditions of the state vector $\bar{x}(t)$ are given, the boundary conditions for $p_1(t)$ and $p_2(t)$ from eq. (16) can be taken arbitrarily. Set of equations (16) can be expressed by one second order equation

$$\frac{d^2 p_2(t)}{dt^2} = c_1 x_1(t) + c_2 a u(t) \tag{17}$$

From eqs (11) and (1y) follows that before we can determine the optimal control u(t) we have to know the conjugated variable $P_2(t)$ which in its turn depends on the current value of the state vector.

Let us first express u(t) as a function of $P_2(t)$. The minimal Hamiltonian will be obtained if

$$\left[|u^*| - a u^* p_2\right] = min\left[|u| - a u p_2\right], \tag{18}$$

at

$$-1 < u < 1,$$

where $u^*(t)$ is the optimal control.

From condition (18) we find

at

$$u^*(t) = \begin{cases} +1 . & p_2(t) \geqslant \frac{1}{a}; \\ 0 & |p_2(t)| < \frac{1}{a}; \\ -1 & p_2(t) \leqslant -\frac{1}{a} . \end{cases} \tag{19}$$

at

at

Introduce inverse time $\tau = t_y - t$, originating at the end of the control process. Because $dt = -d\tau$, equs (14) and (17) will be given by

$$\frac{d^2 x_1(t)}{d\tau^2} = -a u(p_2),$$

Besides (20)

$$\frac{d^2 p_2(\tau)}{d\tau^2} = c_1 x_1(\tau) + a c_2 u(p_2);$$

$$x_1(\tau)\Big|_{\tau=0} = \frac{dx_1(\tau)}{d\tau}\Big|_{\tau=0} = 0; \quad \frac{dx_1(\tau)}{d\tau} = -x_2(\tau).$$

The phase plane origin of coordinates can be reached at $u^* = \pm 1$. which corresponds to $\rho_2(\tau) \geq \frac{1}{2}$. Eqs (20) will then be given by

$$\frac{d^2x_1(\tau)}{d\tau^2} = -a, ; \qquad \frac{d^2\rho_2(\tau)}{d\tau^2} = c_1 x_1(\tau) + ac_2 . \tag{21}$$

Integration of eqs (21) in the interval of inverse time $0 \leq \tau \leq \tau_1$,

(22)

yields

$$\begin{aligned} \frac{dx_1(\tau)}{d\tau} &= \frac{dx_1(0)}{d\tau} - a\tau = -a\tau; \\ x_1(\tau) &= x_1(0) - \frac{a\tau^2}{2} = -\frac{a\tau^2}{2}; \\ \frac{d\rho_2(t)}{d\tau} &= -\frac{c_1\tau^3}{6} + ac_2\tau + \rho_{20}; \\ \rho_2(\tau) &= \rho_{10} + \rho_{20}\tau - \frac{ac}{24}\tau^4 + \frac{ac_2}{2}\tau^2; \end{aligned} \tag{23}$$

where

$$\rho_{20} = \frac{d\rho_2(\tau)}{d\tau}\Big|_{\tau=0} ; \qquad \rho_{10} = P_2(\tau)\Big|_{\tau=0}$$

The control time τ_1 corresponds to switching from control $u = \pm 1$ to $u = 0$.

The describing point until time τ_1 moves along the parabola described by the following equation (Fig. 3)

$$x_1(\tau) = -\frac{1}{2a} x_2^2(\tau) . \tag{24}$$

Since selection of P_{10} and P_{20} is arbitrary we can assume that $P_2(\tau) > \frac{1}{a}$ and then the parabole of eq. (21) will be the optimal path. At time $\tau = \tau_1$ the nature of the describing point motion changes; this moves parallel to the abscissa axis. Then $P_2(\tau) = \frac{1}{a}$ which makes it possible to write

$$\begin{aligned} P_2(\tau) &= P_{10} + P_{20}\tau_1 - \frac{ac_1}{24}\tau_1^4 + \frac{ac_2}{2}\tau_1^2 = \frac{1}{a}, \\ \frac{dP_2(\tau)}{d\tau}\Big|_{\tau=\tau_1} &= P_{20} + ac_2\tau_1 - \frac{ac_2}{6}\tau_1^3 < 0 \end{aligned} \tag{25}$$

The latter inequality follows from the fact that at $u = +1$ and $\tau < \tau_1$

$$P_2(\tau) > \frac{1}{a} ,$$

and at $u = 0$ and $\tau = \tau_1$

$$P_2(\tau) = \frac{1}{a}$$

The optimal path equation for $\tau > \tau_1$ is found from integrated eq (2 at u = 0.

From the first equation

$$\frac{d^2 x_1(\tau)}{d\tau^2} = a$$

and with eq. (21) in mind we find

$$\frac{d^2 x_1(\tau)}{d\tau^2} = -a\tau_1$$

or

$$x_2(\tau) = a\tau_1 , \qquad x_1(\tau) = -a\tau_1 (\tau - \tfrac{1}{2}\tau_1) \tag{26}$$

From the second equation

$$\frac{d^2 p_2(\tau)}{d\tau^2} = -ac_1\tau_1 (\tau_1 - \tfrac{1}{2}\tau_1)$$

we obtain

$$\frac{dp_2(\tau)}{d\tau} = p_{20} - \frac{ac_1\tau_1}{2}\tau(\tau-\tau_1) - \frac{1}{6}ac_1\tau_1^3 + ac_2\tau_1 ,$$

$$p_2(\tau) = \frac{1}{a} + p_{20}(\tau-\tau_1) - \frac{ac_1\tau_1}{12}(\tau-\tau_1)(2\tau^2 - \tau_1\tau + \tau_1^2) + ac_2\tau_1(\tau-\tau_1). \tag{27}$$

The switching of control from u = 0 and u =-1 occurs at $\tau = \tau_2 > \tau_1$. Switching is easily shown to be unfeasible for
At time $\tau = \tau_2$

$$p_2(\tau_2) = -\frac{1}{a} , \tag{28}$$

by the above considerations.

Assuming in equalities (27) that $\tau = \tau_2$ and bearing in mind eq. (28) we have

$$-\frac{1}{a} = \frac{1}{a} + p_{20}(\tau_2-\tau_1) - \frac{1}{6}ac_1\tau_1^3(\tau_2-\tau_1)(2\tau_2^2 - \tau_1\tau_2 + \tau_1^2) + ac_2\tau_1(\tau_2-\tau_1). \tag{29}$$

Hence

$$p_{20} - \frac{ac_1\tau_1}{2}\tau_2(\tau_2-\tau_1) - \frac{1}{6}ac_1\tau_1^3 + ac_2\tau_1 < 0 , \qquad p_{20} = -\frac{2}{a(\tau_2-\tau_1)} + \frac{ac_1\tau_1}{12}(2\tau_2^2 - \tau_1\tau_2 + \tau_1^2) - ac_2\tau_1 , \tag{30}$$

and also

$$p_{20} < \frac{1}{6}ac_1\tau_1(3\tau_2^2 - 3\tau_1\tau_2 + \tau_1^2) - ac_2\tau_1$$

The latter inequality is already satisfied by eq. (25) because $\tau_2^2 > \tau_1\tau_2$. Therefore by using eqs (25) and (30) we find

$$\tau_1(\tau_2-\tau_1)^2(2\tau_2+\tau_1) < \frac{24}{a^2c_1} \cdot \qquad (31)$$

The points on the phase plane for which the inequality is true lie between the parabola $x_1(\tau) = -\frac{1}{2a}x_2^2(\tau)$ and the line for which

$$\tau_1(\tau_2-\tau_1)^2(2\tau_2+\tau_1) = \frac{24}{a^2c_1} \cdot$$

These calculated expressions make it possible to find a solution to the problem formulated. Indeed, at known τ_2 and τ_1 of eqs (25) and (30) P_{10} and P_{20} are found such that the optimal control finds their value unambiguously. The values of τ_2 and τ_1 are found from condition (4) for the quality functional minimum (4) with the required reliability of the system.

Synthesis computation

For illustration of the synthesis technique described we will briefly describe the computing procedure. For physical illustrativeness (Fig. 4) assume that

$$x_{10} < 0 \ , \quad x_{20} = 0 \, .$$

Note that the latter condition is equivalent to the equation $\nu = m_{60}$. Write the expression for optimal control as a function of switching instants in inverse time

$$u^*(\tau) = \begin{cases} +1 & 0 \leq \tau \leq \tau_1 \\ 0 & \tau_1 < \tau < \tau_2 \\ -1 & \tau > \tau_2 \, . \end{cases} \qquad (32)$$

By integrating eq. (11) we obtain

$$x_2(\tau) = \begin{cases} a\tau & 0 \leq \tau < \tau_1 \quad \text{at} \\ a\tau_1 & \tau_1 \leq \tau < \tau_2 \quad \text{at} \\ a\tau_1(\tau_1+\tau_2) - a\tau & \tau > \tau_2 \quad \text{at} \end{cases}$$

$$x_1(\tau) = \begin{cases} -\frac{1}{2}a\tau^2 & 0 \leq \tau \leq \tau_1 \quad \text{at} \\ -a\tau_1(\tau - \frac{1}{2}\tau_1) & \tau_1 < \tau < \tau_2 \quad \text{at} \\ -a(\tau_1+\tau_2)\tau + \frac{1}{2}a(\tau_1^2+\tau_2^2+\tau_3^2) & \tau > \tau_2 \quad \text{at} \end{cases}$$

It is easily seen that

$$\begin{aligned} t_y &= \tau_1 + \tau_2 , \\ x_{20} &= x_2(\tau_1 + \tau_2) = 0 , \\ x_{10} &= x_1(\tau_1 + \tau_2) = -a\tau_1\tau_2 . \end{aligned} \tag{33}$$

from eq. (33) we find

$$\tau_2 = -\frac{x_{10}}{a\tau_1} . \tag{34}$$

Represent functional (1) as

$$J = \frac{1}{2} c_1 \int_0^{\tau_1+\tau_2} x_1^2(\tau)d\tau + \frac{1}{2} c_2 \int_0^{\tau_1+\tau_2} x_2^2(\tau)d\tau + 2\tau_1 .$$

It is interesting to note that this functional can be computed directly by integration for a fixed value of x_{10}. The optimization problem has thus been reduced x to finding such a value of τ_1 which minimizes the quality functional.

Then we find the n control cycles which correspond to the required faultless operation of a valve, P_3. With time t_n known we find t_y by eq. (13) and assume the value of x_{10}.

The switching of control at τ_1 and $\tau_2 = \frac{|x_{10}|}{a\tau_1}$ ensures the minimal error dispersion and fuel consumption at the required reliability of the system.

This synthesis technique can be extended to control systems with a human operator. With higher requirements to the control quality in such system the load on the operator also increases, which fatigues him quicker and reduces the reliability. Therefore such systems must be synthetized when the quality criterion is minimized and the required reliability of the operator maintained.

Reference

I, Понтрягин Л.С. и др. Математическая теория оптимальных процессов. Физматгиз, I960.

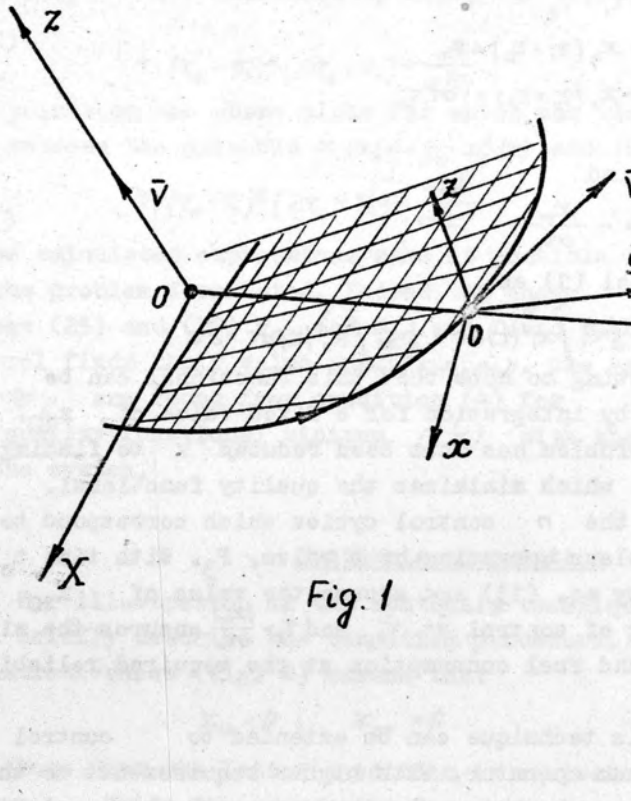

Z
$\bar{V}$
O
z
$\bar{V}$
y
O
Y
x
X

Fig. 1

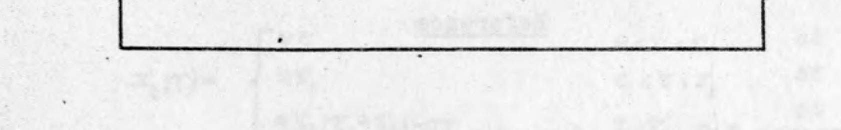
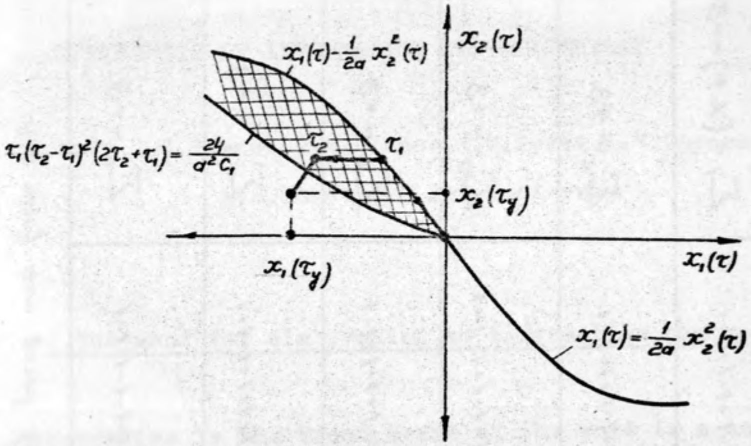
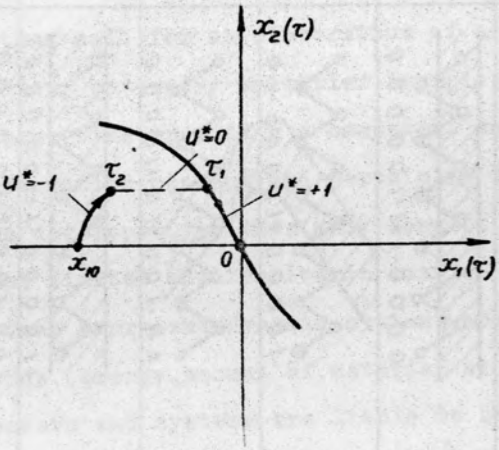

M_B
$\varphi_{Bx}(t)$
$\vartheta(t)$
Controller
M_y
$\frac{1}{J p^2}$
$\varphi(t)$

Fig. 2



$x_2(\tau)$
$x_1(\tau)-\frac{1}{2a}x_2^2(\tau)$
$\tau_1(\tau_2-\tau_1)^2(2\tau_2+\tau_1)=\frac{24}{a^2c_1}$
τ_2
τ_1
$x_2(\tau_y)$
$x_1(\tau_y)$
$x_1(\tau)$
$x_1(\tau)=\frac{1}{2a}x_2^2(\tau)$

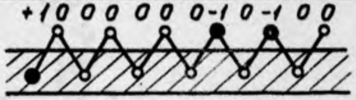
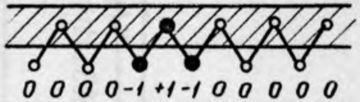
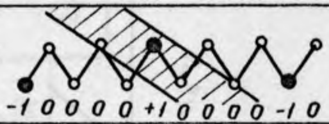
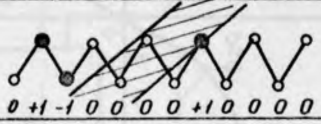
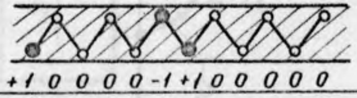
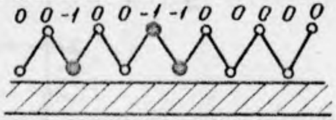
Fig.3



$x_2(\tau)$
$U^*=0$
$U^*=-1$
τ_2
τ_1
$U^*=+1$
x_{10}
0
$x_1(\tau)$

Fig.4

Table 1. Six basic situations and their codes

Situation	Example of signal x	Prototype of situation r	Scalar product
„Shift of part characteristic downwards"		+1 -1 +1 -1 +1 -1 +1 -1 +1 -1 +1 -1	$\sum_1 -(x\cdot r) - -3$
„Shift of part characteristic upwards"		-1 +1 -1 +1 -1 +1 -1 +1 -1 +1 -1 +1	$\sum_2 = +3$
„Turn of part characteristic clockwise"		-1 +1 -1 +1 -1 +1 +1 +1 -1 +1 -1 +1	$\sum_3 = +3$
„Turn of part characteristic counter clockwise"		+1 -1 +1 -1 +1 +1 +1 +1 +1 -1 +1 -1	$\sum_4 = +3$
„Good enough, steady"		+1 +1 +1 +1 +1 +1 +1 +1 +1 +1 +1 +1	$\sum_5 = +3$
Increase of part characteristic peak amplitude		-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	$\sum_6 = +3$

Note. Hatched circles denote peak where representing point was later than in others.

63.1

ELEMENTS OF INFORMATION CONTROL THEORY

B.N.Petrov,V.V.Petrov,G.M.Ulanov,V.M.Ageev,A.V.Zaporozhets,
A.S.Uskov,I.D.Kotchubievsky

I.The need for elaborating an information control theory

Cybernetics in the broad sense of the word is a science dealing with information and control within purposively functioning dynamic systems.

The contemporary theory of control describes fairly well only relatively simple cases.Thus,a systemic approach to complex automation of productive processes as well as development of multimeasurable systems call for an elaboration of a uniform information control theory governing multiplex dynamic systems.

And at the same time,providing a competent general theory based on an information concept and pertaining to a linkage system has been developed,yet this very concept has so far enjoyed a rather limited application in control systems.

Control systems represent dynamic systems operating with limited resources (energy,amount of material etc.).These properties of elements and systems are liable to be reflected in a transmissive function distinctive from one.The motion of such systems is described by variables,restricted according to module,examined at the terminal band of frequencies within a limited time interval.Hence,in such systems processes are represented as a sequence of interdependent states.

Purposively dynamic systems reveal both determined and casual stationary and nonstationary signals and different combinations of the latter.Under these conditions arises a general problem of discerning states and of observing dynamic precision in the reproduction of required processes and in the identification of elements and systems.For the purpose of solving the indicated problems it is natural to refer to the facilities of the information theory.

II.An information approach to the theory of dynamic systems

I.Discernability of state in control object

Any object appears as a complex aggregation in a general case of heterogeneous elements.Therefore,the state of such object may be distinguished merely under the condition,when the dynamic variables,describing this entire object differ in a certain value $\varepsilon \neq 0$ designated as discernability threshold (2,3).The given method of describing the behaviour of a dynamic system incorporates the properties of a continuous and discrete notion.The introduction of discernability permits to produce an adequate description of objects at different hierarchic levels of organization and to determine the maximal amount of information required for the functioning of the control system.

The choice of the discernability threshold introduced into the mathematical model of the dynamic system may also be based on the required accuracy of the investigation contrary to the object's natural physical properties.This permits to operate with a minimally required amount of information for the solution of a fixed problem.

The introduction of discernability thresholds and a careful account of limitations related to phase variables of the dynamic system determines the threshold properties of all the parameters of its mathematical model.

2.The theorem of readings at a fixed dynamic accuracy

In real systems,under conditions of limited resources,thrre always occurs a dynamic inaccuracy,which leads to a certain loss of information,and,consequently,to a reduction of demands towards exactness of reproduction.Thus,it appears advisable to reduce the demands of minimal numeral readings applied to dynamic systems as compared to the established theory of readings adopted for dymamic systems (3,4).It is felt adequate to reflect the values of the process at time intervals where

$$\Delta t = \frac{1}{2(W-k)}$$

and K denotes the reduction of the frequency band,which is determined from the conditions of the fixed dynamic accuracy:

$$\int_{2\pi(W-2k)}^{2\pi W} S_x(\omega) d\omega \leq \pi \delta^2 \qquad (3,I)$$

where S_x (ω)-is the signal's spectral density

δ — is the maximally permissible meansquare error

From the condition $\delta = 0$ it follows that K=0 and,hence,we arrive at conditions of Kotelnikov's theorem.

3.The information control theory is based on an enthropic description of complex and multimeasurable dynamic systems (5). A similar enthropic approach enables to appreciate different control processes.

4.The operation of the automatic system proceeds owing to a specially arranged compensation in the control of casual disturbances.The quality of the control process depends on the rate of this compensation.

The basic conditions of control in its information language represents a balance of enthropies exhibiting a compensative control performance.In its general appearance the result may be expressed as (I):

$$H_t(X)=H_t(V)-H_t(V/X)-H_t(Z)+H(Z/X,V) \quad (3,2)$$

In this case the H_t index designates dynamic enthropy,which characterizes indeterminancy of some value throughout a minute time interval and corresponds the time discernability threshold.

In case of a complete compensation a total invariance of the control system is attained.Under this situation the balance equation of the enthropies' control and disturbance effects may be represented as

$$H_t(V)-H_t(V/X)=H_t(Z)-H_t(Z/X,V). \quad (3,3)$$

At insignificant deviations away from a complete balance (3,3) the invariance condition is observed up to $\mathcal{E}$ (2,6,7,8).

The indicated information conditions are indispensable for a purposeful functioning of dynamic systems.

III.Problems of information control theory and control

Let us consider the results of some developments in this range,which reflect the specificity of control systems mentioned above.

I.Introduction of the diversity number measure in the dynamic system.

The necessity of introducing this measure is involved with restrictions in applying enthropy for the estimation of control processes.

In this paper it is intended to elucidate only the basic notion of this measure having demonstrated it in a continuous model.

Suppose x(t) is an arbitrary (in its general case nonstationary) control process fixed as a product of spaces $X \otimes T$.This process may be represented as:

$$X(t) = f(t) + \overset{o}{X}(t) \qquad (4,I)$$

where f(t) is the function of mathematical expectation of the process determining the distribution of $f \ni F$ values throughout the range of determining T and x (t) is a centred casual process determined by a probabilistic distribution of values $\overset{o}{X} \ni \overset{o}{\mathcal{X}}$

for each $t \ni T$.

Suppose ~~Increase~~ we estimate the distribution of values of x(t)process onto $X \otimes T$ - the measure of diversity of multiple values of the process,which is uniform for determined and casual functions.

It is sensible to expect that this measure of diversity revealed an additivity capacity and failed to depend on concrete values of the process and scales and took into account merely the character of distribution of the process values i.e. displayed a feature similar to those of enthropy's casual values (processes) in the information theory.

Supposing the density of probabilities distribution is $P_t(\mathring{X})$, providing values are $\mathring{x} \ni \mathring{X}$ and the continuous process is $\mathring{x}(t)$ and each single $t \ni T$ in this case the dynamic enthropy of the process /i.e. the enthropy at t/ (I) moment may be put down as:

$$H_t(\mathring{X}) = -\int_{\mathring{X}} P_t(\mathring{x}) \log[\varepsilon_x P_t(\mathring{x})] dx \qquad (4,2)$$

where ε_x - is the discernability threshold of values determined in set X.

Having assumed that the available constant shifting fails to change the appearance of distributions it turns out that

$$P_t(\mathring{x}) = P_t(x)$$

$$H_t(\mathring{X}) = -\int_{X} P_t(x) \log[\varepsilon_x P_t(x)] dx = H_t(t) \qquad (4,3)$$

Hence,dynamic enthropy represents a measure of diversity valu $x \ni X$ of process $x(t)$ merely for X for each $t \ni T$ and does not take into account the function of mathematical expectation f (t) i.e. an important dynamic characteristic of the process x (t).

Hence, a complete measure of diversity in the distribution of the process values x (t) onto X⊗T calls for an introduction of the characteristics of distribution of the process values T determined f (t).

For the purpose of taking into account merely the character of distribution of the process values T regardless of its scale and physical character we shall examine a normalized process

$$\frac{X(t)}{X_{max}} = \frac{f(t)}{X_{max}} + \frac{\overset{o}{X}(t)}{X_{max}} \qquad (4,4)$$

Let us introduce the density of its values distribution T as a characteristics for the distribution of values of a continuous normalized function of mathematical expectation

$$\overset{*}{f}(t) = \frac{|f_t'(t)|}{|X_{max}|}, \qquad (4,5)$$

$\overset{*}{f}(t)$ characterizes the intensity of changing the process values in time.

In this case the density of diversity of process values x (t) may be represented as

$$\mu(x,t) = \frac{\overset{*}{f}(t)}{|X_{max}|} \log[\frac{|X|}{\varepsilon_x}] - K(t) P_t(x) \log[\varepsilon_x P_t(x)],$$

where K (t) is the function of readings.

Having introduced as a complete characteristics of the process values x (t) at a time moment t T the rate velocity of building up diversity V_t (x)

$$V_t(X) = -\int_X \mu(x,t)\,dx = V_t(F) + \overset{*}{K}(t) H_t(X) \qquad (4,7)$$

~~where~~ providing

$$V_t(F) = \overset{*}{f}(t) \log[\frac{[X]}{\varepsilon_x}] \qquad (4.8)$$

is the rate of evolving diversity of a determinating component (function of mathematical expectation) of the process x (t) at a time moment t T.

$H_t(x)$ is the dynamic enthropy of determination in accordance with (4.3)

$K^*(t)\,H_t(X)$ is the dynamic diversity of the process x (t) per t ∋ T at moment t.

Having introduced functionar R (X,T) in the capacity of a complete measure of diversity of x (t) process values per X T in the form of

$$R(X,t) = -\int_T\int_X \mu(x,t)\,dx\,dt = R(F,T) + R(\overset{\circ}{X}), \qquad (4.9)$$

providing that

$$R(F,T) = -\int_T f^*(t)\,\log\left[\frac{|X|}{\varepsilon_x}\right]dt \,. \qquad (4.10)$$

The diversity of the determinative component (function of mathematical expectation) of the process x (t) per T

$$R(\overset{\circ}{X}) = \int K^*(t)\,H_t(\overset{\circ}{X})\,dt \,. \qquad (4.11)$$

The diversity of the centred component x (t).

Thus, the suggested measure of diversity in an arbitrary control process represents a sum of measures ef diversity of determinative and centred components of a process.

The suggested evaluation may serve as a basis of information analysis of control systems and control.

A detailed review of diversity features may be regarded as ~~xx~~ a topic of a separate communication.

2.Potential characteristics of elements and control systems

One of the basic problems of the information theory is to ascertain "what may and may not be attained through the services of our automatic systems and what are the latter's potential capacities"~~of~~.

The solution of this problem ought to be based on a fundamental notion of theoretical nature,namely,potential characteristics. It is suggested that The value characterizing the limit~~ed~~ing dynamic features should be designated as potential characteristics of element (system). In our particular case

$$C\ (x)\ =\ MAX\ R\ (\ X\ ) \tag{4.I2}$$

where the maximum is surveyed according to all the possible effect-ual values y (t) causing process x (t).

The very notion of a potential characteristics adopted for automatic systems is greatly relative.

First of all it depends on the routine of the system's operation,criteria of its efficiency and on its "input" and"output".

A dynamic potential characteristics reflecting the system's peculiar features at moment t or,to be more precise,at an interval equalling the time discernability threshold ε_t appears to be the most specific trait of automatic systems.

The introduction of the notion of "potential characteristics" permits to formulate and substantiate the basic theorem applicable to dynamic systems,which in the theory of information[4] corresponds to Shennon's theorem.

The general formulation of such theorem applicable for dynamic systems may be represented as follows:

Supposing at the object's input,providing the discernability is E the input effect y (t) exhibits a dynamic diversity R_t (y),in this case there is a possibility of obtaining at of the very same the object's output a similar diversity of states $X(t), [R_t(X) = R_t(Y)]$ induced by a very same effect y (t) providing

$$V_t(Y) \leq C_t(X) \tag{4.I3}$$

and it appears to be impossible if

$$V_t(Y) > C_t(X) \tag{4.I4}$$

In the suggested model this theorem is distributed onto arbitrary control processes including those of the determined and nonstationary type.And,indeed,having represented the expression of dynamic diversity (4.7) according to its components,we are to obtain:

$$V_t(F_y) + K_y^*(t) H_t(Y) \leq C_t(X) \tag{4.I5}$$

instead of (4.I3).And,hence,it follows that the possibilities of both components are restricted by the very same potential characteristics and,therefore,the increase of diversity of one of the components is achievable only at the cost of reducing the diversity of the other one.Thus,within the range both the determined and casual effect may be conveyed within the limits of a very same potential characteristics.

In case some faults and distortions are observed within the system,their diversity in dynamics may be taken into account by respective additive members without altering the very essence of the abovecited formulation of the basic theorem.

~~The basic theorem may be formulated in different ways with due account of different~~ The basic theorem may vary in its formulation in relation to particular automatic systems.The formulation of such theorem applied to stabilization systems ~~xxx~~ is given (I).

It ought to be noted that E-values connected with discernability threshold either generally with some $\mathcal{E}$ value characterizing dynamic precision are used in the formulation of the basic theorem.Diversity are diminishing functions of $\mathcal{E}$.

Thus,a dynamic potential characteristics is a limiting characteristics restricting choice between intensity of the effect and the accuracy of its reproduction.

Hence,any control is restricted in its possibilities by a potentiã characteristics.'It is this very value that ~~is responsible for~~ responses to the basic question as to what may be and what may not be achieved by control systems.

3.The measure of enthropic stability in control processes

Conditions of enthropic and information stability may be regarded as a criteria of definiteness in the course of control processes (IO,II,I2).

The specific feature of an enthropic notion comprising a vast aggregation of events or processes is the singling out of a highly-probable group from the latter,which permits to realize real operational regimes in analysis and calculations.

This very specificity of group enthropy and information ought to permit to single out characteristic features of dynamic systems in the course of their information description.

Under these conditions it seems appropriate to introduce ~~the~~ a measure of definiteness in the progress of the process.In case of an uncontrolled system the unfavourable case is that of a regular distribution and dynamic enthropy,which is determined as

$$H_t(X)=\log\frac{|X|}{\varepsilon_x} \qquad \text{for } t \ni T$$

This situation naturally arises a question as to how much an enthropic evaluation is distinguished providing the object is controlled and the distribution differs from a regular one.

The solution of this question,along with an enthropy notion as a mathematical expectation of enthropic density $/-\log\varepsilon_x P(x)/$ it is expedient to expose its dispersion in the following way

$$D_{Ht}=M[-\log\varepsilon_x P_{(x)}-H_t(X)]^2$$

In this case

$$D_{Ht}=-\int_{-\infty}^{\infty}P_t(x)\log^2[P_t(x)]dx-H_0^2(X)$$

and $H_o(X)$ is the differential enthropy.It follows from this expression that the D_H value does not depend on the step of quantization according to level.It is determined only by the function of distribution of a casual value and is a restricted value,whereas an absolute enthropy of a continuous casual value tends to infinity under conditions of unlimited reduction of the quantization step according to level.

It may be shown that at ~~regular~~ a uniform distribution of $D_H=0$ the D_H value may characterize the rate of definiteness in the course of control processes.

The increase of D_H reflects the role of highly-probable states of the control process and may be used as a measure of information stability.

4. Problems of signal filtration within a restricted frequency band

In some problems of automatic control system (ACS) and control it appears expedient to reproduce the signal in a precise way merely in a W_1 part of the frequency band,

($W_1 = W-K$; W- siganl frequency band)

For this purpose, according to a generalized theorem of readings (4), it is suffice to ~~perform~~ make measurements in a lapse of interval

$$\Delta t = \frac{1}{2(W-K)}$$ sec. (W,K - in hertz)

and the function of readings appears as

$$U_{(t)} = \frac{\sin[2\pi(W-2K)t]}{2\pi(W-K)t}$$

In real machines interferences are imposed on a useful signal.

In the presence of a high-frequency interference the optimal rectangular filter may be obtained from the minimum condition of the root-meansquare (RMS) error.

For signals, limited in frequency, the expression of RMS error is thus:

$$\overline{\mathcal{E}}^2 = \frac{1}{\pi}\Big[\int_{2\pi(W-K^*)}^{2\pi W} S_m(\omega)\,d\omega + \int_0^{2\pi(W-K^*)} S_n(\omega)\,d\omega\Big]. \qquad (4,I6)$$

Having minimized the expression (4,I6) according to parameter K it follows that

$$S_m(W-K^*) = S_n(W-K^*) \qquad (4.I7)$$

The relation (4.I7) may be regarded ~~similarly~~ as well as in the generalization of V.A.Kotelnikov's theorem in the case, when an interference is imposed upon a useful signal$^{x)}$.

Note: For signals with an unlimited spectrum expression (4.I7) appears as $S_m(W^*) = S_n(W^*)$ where W^* is a frequency band of a rectangular filter.

An appraisable filter (fig.I) is determined from codition

$$\pi\delta^2_{max\,max} \geqslant \int\limits_{W-K}^{W} S_m(\omega)\,d\omega + \int\limits_{0}^{W-K} S_n(\omega)\,d\omega$$

The method of integral quadratic approximation of rectangular characteristics may be used for the realization of aprraisable filter with a W-K band and expressed through

$$\Phi(i\omega) = K\,\frac{\beta_m(i\omega)^m + \beta_{m-1}(i\omega)^{m-1} + \ldots + \beta_1 i\omega + 1}{\alpha_n(i\omega)^n + \alpha_{n-1}(i\omega)^{n-1} + \ldots + \alpha_1 i\omega + 1}$$

The described method is not connected with the regularization problem and permits to synthesize filters at a fixed dynamic accuracy with a minimal band transmission.

In restricting the band transmission by an appraising value W-K and precisely calculated spectral densities of the useful signal and interferences,the synthesis problem may be reduced to the solution of the Kolmogorov-Wiener problem:

$$\varphi(i\omega) = \frac{1}{2\pi\psi(i\omega)}\int\limits_{0}^{\infty} e^{-i\omega t}\int\limits_{-\infty}^{\infty}\frac{\varphi_0(i\omega)S_m(\omega)}{\psi^*(i\omega)}\,e^{i\omega t}\,d\omega, \tag{4.I8}$$

where $\psi(i\omega)$, $\psi^*(i\omega)$ are complex-conjugated multipliers with no zeroes and no poles in the lower and upper semiplanes of the complex ω plane,respectively.

$$\psi(i\omega)\,\psi^*(i\omega) = S_m(\omega) + S_n(\omega)$$

$\varphi_0(i\omega)$- ideal operator and

$$|\varphi_0(i\omega)| = \begin{cases} e^{-\alpha[\omega - W + K]}, & \omega \geqslant (W-K)\cdot 2\pi \\ 1, & \omega \leq (W-K)\cdot 2\pi \\ e^{\alpha[\omega + W - K]}, & \omega \leq -(W-K)\cdot 2\pi \end{cases} \tag{4.I9}$$

Having assumed thay in expression (4.I9)$\alpha > \alpha^*$ where α^* is an adequately large number,we obtain a family of operators akin to the appraised one (fig.2).

Providing α =0 formula (4.I8) represents a characteristics of an optimal filter.Having supposed that in expression (4.I9) α =0 and having replaced in the first integral (4.I8) 0 by ∞ we obtain a physically unrealizible optimal transfer function

$$\Phi(i\omega) = \frac{S_m(\omega)}{S_m(\omega) + S_n(\omega)} \qquad (4.20)$$

A respective value of RMS error is estimated by formula

$$\varepsilon^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_m(\omega)\, S_n(\omega)}{S_m(\omega) + S_n(\omega)} \qquad (4.2I)$$

Expression (4.2I) may be used as an appraising formula in the projection of optimal filters.Formulas (4.2O), (4.2I) are of principal significance.Having applied expression (4.20) and (4.2I) it ~~may be~~ is easy to shown ~~that~~ the connection of Shennon's results with those obtained through the method of statistical optimization of filters.

Having examined from this angle the error in the class of information signals i.e. having restricted its approximation by the frequency band of spectrum W and time extension T we obtain enthropy of error in band Δ f.

$$H_{\Delta f} = T \Delta f \log 2\pi e S_\varepsilon(f) \Delta f \qquad (4.22)$$

Integrating (4.22) in regard to the band and having used expression (4.2I) we obtain enthropy related to one degree of freedo of the signal's error

$$H_n(\varepsilon)=H_n(m)-\frac{1}{W}\int_{W}\log\frac{S_m(f)+S_n(f)}{S_n(f)}\,df \qquad (4.23)$$

The integral of the right part coincides with Shennon's expression of the maximal rate of transmitting information along the channel with W band.

The same result may be obtained through using Shennon's formula for the loss of enthropy in a linear filter.

Thus,we succeeded in obtaining a correspondence between Shennon's formula and the optimal filter of Kolmogorov-Wiener.As long as in seeking the latter the condition of physical implementation has not been taken into account,the maximal rate of transmitting information according to Shennon proves to be unrealizible.In other words, Shennon's formula provides for an excessive estimation of maximal rate in transmission of information.

5. Amount of information and transitional processes

Let us evaluate the change of amount of information in case of serious errors in the system observed in transitional patterns.

Suppose there is a fixed ensemble of signals x(t) receiving values within a limited set with metrics $\rho(X_1,X_2)=|X_1-X_2|$ which in a time moment t=0 is transmitted onto an input of a linear dynamic system with an impulse transitional function u (t). It is assumed the initial conditions are zero.At the system's output is: an ensemble $y(t)$ with a metrics $\rho(y_1,y_2)=|y_1-y_2|$.Let the signals x (t) and y (t) be represented thus:

$$X_{i+1}(t)=X_i(t)+\Delta X\,\xi(t)$$

Where

$$y_{i+1}(t)=y_i(t)+\Delta y(t)$$

$$\Delta y(t)=\Delta X\int^{t}\xi(t-\tau)u(\tau)\,d\tau$$

At any time moment

$$\lim_{\Delta y \to 0}\left(\frac{\Delta X}{\Delta y}\right)_t = \left(\frac{\partial X}{\partial y}\right)_t = \mathcal{J}_t\left(\frac{X}{y}\right) = \frac{1}{\int_{-\infty}^{t}\xi(t-\tau)\,u(\tau)\,d\tau} = K(t)$$

providing $\mathcal{J}_t\left(\frac{X}{y}\right)$ is the Jacobian transformation of co-ordinates at t time.

In the information theory data related to some casual X value produced as a result of observing casual y value change its indefiniteness. The latter is characterized by a substitution of an unconditional x value enthropy for a mean conditional enthropy of x value relative to y value. With due account of linearity transformation, the amount of information at any time moment is described by a following relation:

$$\mathcal{J}_t(Y,X) \approx H_t(X) - \log K(t)$$

It is necessary to note that in the measurement of casual values k (t) function is easily determined:

$$K(t) = \frac{1}{\int_0^t u(\tau)\,d\tau} = \frac{1}{A(t)}$$

Having estimated the transitory function of system A (t) by using ~~means of~~ any of the known methods, one can calculate the numeral change of information throughout the transitory process.

~~Having~~ When dealing with stationary casual signals with a limited frequency band F exhibiting information capacity at any time moment H(x) it turns out that

$$\mathcal{J}_y(x,t) = H(x) + \frac{1}{2F}\int_0^F \log|W(f)|^2 df - \log K(t) + \log\frac{\Delta X}{\Delta y}$$

~~The~~ In its general case ~~the~~ it is rather difficult to obtain ~~the~~ initial correlations for the obtained expression. However, in some particular cases (Gauss processes, the filter being described by a differential equation of the first order) the problem is solved relatively easy.

IV. Perspectives and tasks of information theory

The application of ~~the~~ postulates of the classical information theory to problems involved with control encounters considerable difficulties.The latter may be removed through ~~xx~~ a further indispensable elaboration of the basic ideas of information theory and the establishment of quantitative correlations between losses of information and the precision of its reproduction in dynamic systems.

The basic problem to be solved to built up an informtion control theory and control is to be in the capacity to solve the problem of transmitting diversity in systems with different types of ~~xxx xxxx~~ feedbacks,particularly in multi-linked and multi-measurable systems.

The second,no less important,problem is the study of systems ~~withxxx~~ of a branched hierarchic structure and a priority of commands.

The third urgent problem which arose presently in connection with the demand for an implementation of the results of the information control theory consists in the development of computing methods of information analysis and synthesis of complex multi-measurable systems of multi-storeyed structure.

An information approach permits to view ~~the~~ a complex of control measurement systems and control from a uniform theoretical angle regardless of its designation and methods of realization.Thus, scientifically-grounded solutions may be adopted in earlt stages of projection in cases where up to late there has been a domain of experience and intuition.

References

I.B.N.Petrov,I.D.Kotchubievsky,G.M.Ulanov.Information aspects of control over technological processes.USSR Academy of Science Publishers"Technical Cybernetics",~~N9~~N4,1967.

2.B.N.Petrov,I.D.Kotchubievsky,G.M.Ulanov,E.B.Dudin.Discernability, invariance and information in systems of rigid and variable structure.Paper "Multi-link and invariant systems."Nauka" Publishers, 1968.

3.V.V.Petrov,A.V.Zaporozhets.Information appraisal of dynamic preeision in information systems and control in optimal systems.Statistical ~~systems~~ methods."Nauka" Publishers,1967.

4.V.V.Petrov.Appraisal of dynamic precision in information control systems.In the book "Current methods in projecting automatic control systems (ACS)",Moscow,Machine-building Publishers,1967.

5.A.A.Krasovsky.Change of *enthropy* in continuous dynamic systems. USSR Academy of Sciences Publishers"Technical Cybernetics" ,N 5,1964.

6.B.N.Petrov."The invariance principle and conditions of its application in calculating linear and nonlinear systems".Trudy of of Ist International Congress,Moscow,196I.

7.V.V.Petrov,V.MAgeev,A.VZaporozhets.Some links problems - invariance and dynamic precision in automatic control systems.Trudy of 3rd All-Union Conference on the theory of invariance and its application in automatic machines.Kiev,1966.

8.I.D.Kotchubievsky,G.M.Ulanov.Information methods in invariance theory.Trudy of 3rd All-Union Conference on the theory of invariance and its application in automatic machines.Kiev,1966.

9.K.Shennon.Mathematical link theory.In the book "Papers on the theory of information and cybernetics",Foreign literature Publishers 1963.

IO.R.A.Dobrushin.General formulation of Shennon's basic theorem in the information theory.

II.I.S.Pinsker.Information and information stability of casual values and processes.USSR Academy of Sciences Publishers,Moscow, I960.

I2.A.A.Krasovsky.Enthropic stability of linear continuous systems of automatic control.Izvestia USSR Academy fm of Sciences. "Technical cybernetics" N 5,I963.

63.4

Correctness, Regularization and Minimal Complexity Principle in statistical Dynamics of Automatic Systems.

V.V. Solodovnikov

V.L. Lensky

Moscow High Technical School of Baumann.

I. The design of optimal automatic control systems in most cases consists of two steps.

On the first step a mathematical synthesis of the system is made, which usually requires the solving of some variational problem.

On the second step a physical realization of the determined solution is made.

Both these steps are connected with several difficulties.

Let a functional $J(x)$ which is the index of performance of the system is given on a class of operators X. By the use of the conditions for the extremum existence for functionals the optimality condition is often received as a functional equation

$$Ax = y \qquad (1)$$

for the unknown operator x.

For instance, the solving of optimization problem in the class of polynominal filters:

$$F[y(\tau)]=\int_0^T k_1(t,\tau)y(t-\tau)d\tau+\ldots+\int_0^T\!\!\int_0^T k_n(t,\tau_1,\ldots\tau_n)y(t-\tau_1)\ldots y(t-\tau_n)d\tau_1\ldots d\tau_n, \quad (2)$$

if the functional $F[y(\tau), t-T\leqslant\tau<t]$, which gives the minimum for the mean square value of the difference between transformed $F[y]$ and desired $x(t)$ signals, yields the system of integral equations

$$J\{F\}=M\{x(t)-F[y(\tau),t-T<\tau\leqslant t]\}^2=\min, \qquad (3)$$

$$\sum_{i=1}^{n}\int_0^T\ldots\int_0^T k_i(t,\tau_1,\ldots\tau_i)\Gamma_y(t,\tau_1,\ldots\tau_i;\theta_1,\ldots\theta_j)d\tau_1\ldots d\tau_j=\Gamma_{xy}(t,\theta_1,\ldots\theta_j),$$

$$j=1,2,\ldots n \qquad (4)$$

where $$\Gamma_y = M\{y(t-\tau_1)\dots y(t-\tau_i)y(t-\theta_1)\dots y(t-\theta_j)\},$$
$$\Gamma_{xy} = M\{x(t)y(t-\theta_1)\dots y(t-\theta_j)\} \tag{5}$$

and k_i ($t, \tau_1, \dots \tau_i$) are multidimensional impulse response functions.

This system of equations can be written in the abbreviated form

$$\Gamma_y\, k = \Gamma_{xy} \tag{6}$$

where k and T_{xy} are the elements of the Hilbert space.

In many cases the exact solution of the equation (1) is impossible and so some numerical technique for the finding of the approximate solution must be used.

Besides the initial data for the solution of the equation (1) usually are known with some error.

All this points out that it is necessary to investigate the stability of the solution of the equation (1) relative to the accepted numerical technique and errors of the initial data.

The degree of stability of the equation (1) relative to the variations of its right-hand side is defined by the modules of a continuity for the inverse represantation

$$\omega(\delta, X) = \sup \rho(x, x_1) \ \text{при}\ x, x_1 \in X,\ \rho(Ax, Ax_1) \leqslant \delta \tag{7}$$

where the function ρ (x, x_1) defines a metrics in class X.

It is evident that the possible error in evaluation of the optimal operator growa together with the expansion of a class in which the optimal operator is to be determined, that is to say the following inequality holds

$$\omega(\delta, X_1) \leqslant \omega(\delta, X_2)\ , \quad X_1 \subset X_2\ . \tag{8}$$

If equations (1) is incorrect the solution becomes unstable relative to errors in the initial data and so a synthesis problem of the optimal system becomes principally impossile. But even if a problem is correct the numerical procedure often is very complicated and time and memory required for the determining of the solution with a given accuracy by the aid of a computer are very great.

For the possibility of the physical realization of the solution it is usually necessary to approximate it in some way.

A tendency to diminish a loss in the index of performance makes it necessary to increase the accuracy of its approximation and the class of operator in which the extremum of functional J (x) is determined. This leads to a loss in technological properties of the control systems.

Let for every operator $x \in X$ the minimum necessary cost of the practical realization of a control system is defined by the function $C_\varepsilon(x)$, an operator of which approximates x with the error ε. It is easy to see, that if

$$X1 \subset X_2$$

then

$$\max_{x \in X_1} C_\varepsilon(x) \leqslant \max_{x \in X_2} V_\varepsilon(x) \qquad (9)$$

The same inequality holds for the every function $V_\varepsilon(x)$ giving the measure of computing work
necessary to determine the operator x with the accuracy ε

$$\max_{x \in X_1} V_\varepsilon(x) \leqslant \max_{x \in X_2} V_\varepsilon(x) \qquad (10)$$

Let for every $x \in X$, the function τ_ε (x) is determined.

This function defines the maximum possible probability faultless work of the control system.

During some fixed interval of time, an operator of which approximates operator x with the errer ε.

It is also evident, that if $X_1 \subset X_2$, then

$$\min_{x \in X_1} \tau_\varepsilon (x) \geqslant \min_{x \in X_2} \tau_\varepsilon (x)$$

Let us call the operator θ, corresponding to the absence of control system as a null-operator.

For null-operator it is naturally to propose $C_\varepsilon(\theta) = 0$, $V_\varepsilon(\theta) = \theta$ and τ_ε (x) = 1.

Let a family of classes $\mathfrak{M}$ satisfies to the conditions

$$\bigcap_{X \in \mathfrak{M}} X = \theta$$

Paying attention to the ineqalities it is evident that it is desirable to determine the operator of a control system belonging to the most contracted class of the family. But the contraction of the class leads to a poorer index of performance.

This contradiction may be solved if the control system synthesis problem will be defined as follows:

Among all operators having the admissible index of performance, which is supposed to be known, to find the operator belonging to the narrowest class of a family of operators considered.

This synthesis problem may be formulated in a more compact form as follows.

Consider two operators x_1, x_2:

$$x_1 \in X_1 \; ; \; x_2 \in X_2$$

and

$$X_1 \subset X_2$$

It is evident that it is a more difficult problem to determine and to realize with a given accuracy the oprrator x_2 than the operator x_1.

Accordingly let us admit that the operator x_2 as a more complicated operator than the operator x_1 if there is no information exept that

$$x_1 \in X_1 \; ; \quad x_2 \in X_2 \quad \text{and} \quad X_1 \subset X_2$$

In this case a family of classes M may be considered as a complexity scale for a set being the combination of all classes of a family.

Now the synthesis problem may be defined as a minimal complexity principle as follows :

Among all operators having the admissible index of performance to find the operator of minimal complexity relative to the given scale.

As well the principle of limited complexity may be used:

For the definite class belonging to the scale of complexity considered to find an operator to which correspond the extremum value of the index of performance.

2. To apply the minimal complexity principle it is necessary to have the methods for the designing of a complexity scale.

Consider a continuous functional G(x) which has the absolute minimum on the null-operator.

Then a one-parametric family of classes $X_t = \{x \mid G(x) \leqslant t\}$ has the properies of the family of classes M defined above.

In this case a compression of the class corresponds to the minimization of the functional G(x).

The application of the minimum complexity principle leads to the conditional extremum problem:

To minimize the functional G(x) subject to the constraint $J(x) = q$.

The solution of this problem require the minimization of the functional

$$\lambda G(x) + J(x)$$

where λ is Lagrange multiplier.

The other method to define the complexity scale is as follows.

Consider the increasing system finite dimensional classes

$$\theta \subset X_1 \subset X_2 \subset \ldots \subset X_n \subset \ldots \subset X$$

where index means the dimensionality of the class.

This system of classes may be considered as the compexity scale and the application of complexity principles in this case redûces to extremal problems for the functions of many variables. The design of such complexity scale is possible, for instance, for the class X with basis $x_1, x_2, \ldots . x_n, \ldots$In this case finite dimensional class X_n is a set of all possible linear combinations of basis elements.

The other design methods of complexity scalec are also possible. The systems of classes included in the complexity scale generally is not unique. Therefore every specific synthesis problem require a construction of the complexity scale which take into account a peculiarity of the problem considered and possibilities of the physical realization of a system.

3. The application of complexity principles is useful not only for that the synthesized systems have better technological properties, but as well for that the necessary conditions of extremum existence in the form of equation (1) are correct in Tichonov sense, if the complexity scale is correspondingly chosen. For the correctness of conditions for the extremum existence in the form (1) it is sufficient that the classes $X_t=\{x \mid G(x)\leqslant t\}$ are compact and the Euler operator for the functional G(x) is completely continuous.[1] In this case the functional G(x) is the functional of regularization for the eq.(1). The incorrectness of eq.(1) is not only theoretical possibility which practically may not be taken into account. Many problems of statistical dynamics of control systems reduce to linear integral equations of the first kind which are incorrect. Actually the synthesis of the system which is optimal in rms sense require the minimization of thequadratic functional

$$J(x)=(Ax,x)-2(x,y) \tag{12}$$

where A is a positive self-adjoint linear operator. It is known that an element x minimize the functional J)(x) provided it satisfies the lineat eq. (1).

The solution of eq.(1) may be written as Stiltijes integral expansion of the self-adjoint operator A:

$$x=\int_0^\infty \frac{1}{\lambda}\, dE_\lambda y, \tag{13}$$

is used, where E_λ is the expansion of the unity corresponding to the self-adjoint operator.

The solution exists in that and only in that case, if

$$\int_0^\infty \frac{1}{\lambda^2}\, d\,(E_\lambda y,y) < \infty \tag{14}$$

If the zero is a limit point of the spectre of the operator A, then the integral (14) may not exist.

This impuls that the eq.(1) may not have the solution with a finite norm.

Such situation arises for instance in all cases of linear filters synthesis, when optimal impulses response includes - function and its derivatives not integrable in quadrature sense. These filters are physically nonrealizable. Let the element y with some error h is given , then, taking into account theadditivity of the inverse operator, the square of the error norm pf the solution may be represented as follows:

$$\|\delta x\|^2 = \int_0^\infty \frac{1}{\lambda^2} d\,(E_\lambda h, h) \tag{15}$$

The eq.(15) implies that the square of the error norm may have any value, depending of the spectral measure $(E_\lambda h, h)$ distribution.

A mathematical problem is correct provided the solution of this problem exist, is unique and is stable relative to variations of the initial data. In this sense the problems of statistical dynamics in the form of eq. (1) are incorrect.

For the application of the complexity principles the complexity functional may be given, for instance, in form

$$G(x) = (Bx, Bx) \tag{16}$$

where B is a positive and continuous operator.

In special case, when B is the unity-operator we have

$$G(x) = (x, x) = \|x\|^2$$

and the application of the minimal complexity principle leads in this case to the following problem of the calculus of variations.

To find the minimal value of the functional G(x) = (x,x), provided

$$I(x) = (Ax, x) - 2(x, y) = q$$

An element representing the solution satisfies to the equation of the second kind

$$x + Ax " y \qquad (17)$$

In this case the application of minimal complexity prineiple is equivalent to the weak regularization in Tichonov sense. If B is a differential operator then it is equivalent to a strong regularization.

Using spectral expansion of the operator , the complexity scale may be designen as follews

$$X_t = \left\{ x \middle| x = \int_t^\infty \frac{1}{\lambda} dE_\lambda u \right\}$$

where $\|u\| < \infty$

Then $X_{t_1} \subset X_{t_2}$, provided $t_1 > t_2$.

As it may be shown, the solution of minimization problem for the functional (7) in class X_t is given by the formula

$$R_\delta[y] = \bar{x} = \int_t^\infty \frac{1}{\lambda} dE_\lambda y$$

In the discrete spectre case the algorithm of regularization is equivalent to the determination of the solution in the form of a linear combination of elements with eigenvalues exeeding t.

Let us now prove that the regularization in Tichonov sense is equivalent to the application of complexity principles.

The method of regularization for eq.(1) requires the solution of the equation.

$$\lambda Bx + Ax = y$$

where B is Euler-operator of a regularization functional, λ-parameter of regularization chosen depending of the error norm initial data.

It is evident that the equation (18) results from the minimization of the functional

$$\lambda G(x) + J(x). \tag{19}$$

The minimization of functional (19) may be considered as the minimization of functional G(x) for the given value of functional J (x).

Consider the set $X_t = \{x \mid G(x) \leqslant t\}$

Minimization of the functional G(x) is equivalent to the contraction of the set A_t in the sense that for t t_2 we have $A_{t_1} \subseteq A_{t_2}$.

Hence, the regularization implies the minimization of complexity.

4. Let us consider some examples application.

Consider a linear control system the input of which consists of: deterministic signal g(t), random signal m(t) with zero mean-value and correlation function R_m (t,T), and noise n(t) with zero mean-value and correlation function R_n(t,T). As an index of performance for a control system let us accept the functional

$$J = \mathcal{E}_g^2 + \mu^2 \mathcal{E}^2_{rms} \tag{20}$$

where $\mathcal{E}_g$ - dynamical error, $\mathcal{E}_{rms}$ - rms error and λ weighting multiplier.

It is easy to show that

$$J = g^2(t) + \mu^2(t) R_m(t,t) - 2\int_0^t [g(t)g(\tau) + \mu^2(t) R_m(t,\tau)] k(t,\tau) d\tau +$$
$$+ \int_0^t \int_0^t \{ g(\theta) g(\tau) + \mu^2(t) [R_m(t,\theta) + R_n(t,\theta)] \} k(t,\theta) k(t,\tau) d\tau \, d\theta \tag{21}$$

where k(t,T) is the impulse response of the system.

The minimization of J gives the necessary and sufficient condition for minimum of J in the form of the integral equation for k(t,T):

$$\int_0^t \left\{ g(t)g(\tau) + \lambda^2(t)\left[R_m(\tau,\theta) + R_n(\tau,\theta)\right]\right\} k(t,\theta)\,d\theta = g(t)g(\tau) + \mu^2(t)R_m(t,\tau), \quad t > \tau \qquad (22)$$

This is parametric Fredholm integral equation of the first kind. The solution of it may include -functions and be connected therefore with essential difficulties for the practical realization.

Let us suppose that

$$J = \mathcal{E} g^2 + \mu^2 \mathcal{E}^2_{rms} = q(t) \qquad (23)$$

where q(t) is the given admissible value of J.

As the complexity functional we choose

$$G = \int_0^t k^2(t,\tau)\,d\tau \qquad (24)$$

In this case the necessary and sufficient condition for the minimum of J reduces to the parametric Fredholm integral equation of the second kind:

$$\lambda k(t,\tau) + \int_0^t \left\{ g(0)g(\tau) + \lambda^2(t)\left[R_m(\tau,\theta) + R_n(\tau,\theta)\right]\right\} k(t,\theta)\,d\theta = g(t)g(\tau) + \mu^2(t)R_m(t,\tau), \quad t > \tau \qquad (25)$$

The right-hand side of this equation is the restricted function Hence its solution is contained in a class of restricted functions as well, that is to say, it does not contain -functions,

If we choose as the complexity functional the integral of the square of a differential operator:

$$G = \int_0^t \left[\sum_{i=0}^{n} a_i k^{(i)}(t,\tau)\right]^2 d\tau \qquad (26)$$

then the minimum condition fof has the form of an integro-differential equation, the solution of which has at least $k(t,\tau)$ restricted derivatives.

Hence, by choosing of the complexity functional the differential properties of impulse response may be controlled.

This is very important for the determination of the system structure. To illustrate this, let us suppose that an impulse response k(t,T) may be approximated with any accuracy required by the sums of the form

$$P_m(t,\tau) = \sum_{i=1}^{m} C_i\, \varphi_i(t)\, \psi_i(\tau) \tag{27}$$

In this case the dynamical system may be formed from the dynamical of the first order.

The differential equation of each element may be represented as

$$D_i(p,t)\, r(t) = M_i(p,t)\, n(t), \quad i = 1, \dots, n \tag{28}$$

The differential operator D_i(p,t) and $M_{i\ (p,t)}$ are defined by the equations

$$\left.\begin{aligned} D_i(p,t)\, r(t) &= \frac{dr(t)}{dt} - \frac{1}{\varphi_i(t)} \cdot \frac{d\varphi_i(t)}{dt}\, r(t) \\ C_i\, \psi_i(\tau) &= M_i^{*}(p,\tau)\, \frac{1}{\varphi_i(\tau)} \end{aligned}\right\} , \tag{29}$$

where operator Mi (p,T) is adjoint to operator M_i (p,t).

To get the simpliest structure it is naturally to choose among the set of sums of degree m such sum which give the best approximation:

$$E_m(k) = \inf \| k(t,\tau) - P_m(t,\tau) \| \tag{30}$$

It is known that the rapidity of decrease of E_m(k) completely depends of differential properties of k(t,T).

In the case of stationary signals the determination of linear optimal stationary system with complexity functionals which have the form of integrals from the square of impulse response and its derivatives leads to systems with minimum frequency band with.

But the less is the band width the simpler is the physical realization of the system.[5]

5. Let us consider another example of application of complexity principles which is based on the second method of com-

This is a synthesis problem of nonlinear discrete filters a finite memory, which transform the given stationary random signal in the desired stationary signal in the best possible way. A general solution of this problem is unknown and probably can not be received.

Consider the complexity scale $\mathfrak{M} = \{F_m\}$ where F_m is the class of polynomials

$$\sum_{i=0}^{n-1} h_i x_{t-i} + \sum_{0 \leqslant i_1 \leqslant i_2}^{n-1} h_{i_1 i_2} x_{t-i_1} x_{t-i_2} + \ldots + \sum_{0 \leqslant i_1 \leqslant i_m}^{n-1} h_{i_1 \ldots i_m} x_{t-i_1} x_{t-i_2} \ldots x_{t-i_m} . \quad (31)$$

It is easy to see that this system may represent the complexity scale. This is evident from the facts that every continuous function may be approximated by the polinomial (31) of sufficiently high degree with any desired accuracy and the union VF_m contains the class of continuous functions. Filters (31) are called Kolmogorov-Gabor filters.[6]

According to complexity principles a synthesis of a nonlinear filters requires the solution of a variation problem in the class F_m. This problem is equivalent to the determination of a projection of a random quantity y_t on a subspace formed by all possible linear combinations of random quntities $x_{t-i}, x_{t-i} \; x_{t-i_2}, \ldots$

The solution of this problem reduces to the solution of the normal system equations for the weighting coefficients $h_i, h_{i_1 i_2}, \ldots$ by the method of the least squares.

However this method of a nonlinear lifter synthesis still is connected with some difficulties.

The first difficulty is a very rapid increase of the number of weighting coefficients to be calculated with the increase of the polinomial lifter degree M, and memory n.

It may be proved, that this number N is defined by

$$N = \sum_{k=1}^{m} \frac{n(n+1)\dots(n+k-1)}{k!} \qquad (32)$$

A great number of weighting coefficients implies a severe requirements to memory volume and speed of discrete filter. Besides, with an increase of a number of the weighting coefficients very rapidly increases the volume of computations.

The second difficulty consists of that many methods of the determination of the weighting coefficients are unstable relative to errors of approximate computational procedures.

For instance, the determination of the weighting coefficients with the aid of the normal equations method leads to a poorly defined system of linear algebraic equations and this effect increases with an increase of the system order. So a limited accuracy of approximate computations leads to a practically incorrect problem. Hence the natural measure of complexity for Kolmogorov-gabor filters is the number of weighting coefficients to be determined.

According to complexity principles, it is necessary in space formed by all possible linear combinations of random quantities

to find the subspace of a minimal dimensionality for which the admissible value of the error holds. Or for given the dimensionality to find the subspace for which the error is minimum. The determination of such subspaces may be effected by the trial-and-error process. But it is connected with a very great volume of a computations.

To minimize the complexity of filters (31) the numeric-theoretical procedures of approximate analysis may be used.

Actually, in many cases the discrete random signal Y_t considered and discrete random signal Y_t desired may be put in correspondence with some continuous signals x(t) and Y(t). This holds, for instance, in the case when x_t and Y_t are times which we get from random continuous signals by a sampling process. If this is not so, then continuous signals x(t) and y(t) having in discrete moments of time the same values us discrete signals considered may always be formed.

Without any restriction on generality it may be supposed that $(n - 1)\Delta = 1$, where Δ is the sampling interval of a random continuous signal forming a discrete signal.

Consider one of the product of the sum (31) which has a degree s :

$$x_{t-i_1}\ x_{t-i_2} \cdots x_{t-i_s}$$

The aforementioned statement imples that this member may represent a value of a stochastic field determined on a unity hypercube of s dimensions in the point M $(\Delta i_1, \Delta i_2 \dots \Delta i_s)$ The field itself is given by

$$f(t,\tau_1,\dots,\tau_s) = x\,(t-\tau_1)\dots x(t-\tau_1). \qquad (33)$$

If a continuous signal x(t) has a continuous in the mean derivative of the order , then it can be proved that the stochastic field (33) may be approximated by

$$x(t-\tau_{.})\dots x\left(t-\tau_s\right) = \sum_{k=0}^{n-1} x_{(t-a_1 k(\mathrm{mod}\,n))}\dots x_{(t-a_s k(\mathrm{mod}\,n))}\,\theta_k(\tau_1,\dots,\tau) + R,\dots \qquad (34)$$

where $\theta_k(\tau_1,\dots\tau_s)$ are some basis-functions, $a_1,\dots,a_s$ - optimal coefficients determined according to ^{7}, p mod n - the remainder which is get by division of p by n, and R is estimated by the inequality

$$\overline{R}^2 < C\,\frac{\ell n^{\gamma}\, n}{n^{\alpha-1/2}} \qquad (35)$$

The field value may be represented in any point by the equation (34). Hence, it may be also determined in thepoints with coordinates:

$$\tau_1 = \frac{1}{n-1}\, i_1,\ \tau_2 = \frac{1}{n-1}\, i_2,\ \ldots\ \tau_s = \frac{1}{n-1}\, i_s.$$

Therefore, any product of degree s may be approximately represented as a linear combination of n products of degree s having the form

$$x_{t-a_1k(\mathrm{mod}n)}\, x_{t-a_2k(\mathrm{mod}n)} \cdots x_{t-a_sk(\mathrm{mod}n)}, \qquad (36)$$

that is to say

$$x_{t-i_1} \cdots x_{t-i_s} = \sum_{k=0}^{n-1} \theta_k(i_1, \ldots, i_s)\, x_{t-a_1k(\mathrm{mod}n)} \cdots x_{t-a_sk(\mathrm{mod}n)} + R \ . \qquad (37)$$

The equation (37) implies that the products not coinciding with (36), are not needed, for they it may be approximately determined from (37).

Hence the optimal subspace is the subspace formed by all possible linear combinations of the system of random quantities:

That is to say, the synthesis of nonlinear discrete filter is reduced to the solution of the variational problem :

in the class of filters given by $\sum_{k=0}^{n-1} h_k^1 x_{t-k} + \sum_{k=0}^{n-1} h_k^2 x_{t-a_1k(\mathrm{mod}n)} x_{t-a_2k(\mathrm{mod}n)} +$

$$+ \ldots + \sum_{k=0}^{n-1} h_k^m x_{t-a_1k(\mathrm{mod}n)} \cdots x_{t-a_mk\,(\mathrm{mod}n)}. \qquad (39)$$

To illustrate the effectivness of the complexity minimizathon the table is given here which contains the numbers N and N ^{1} of weighting coefficients for the complete Kolmogorov-Gabor filter and for the minimum complextity filter, when the memory n of the filters equals t$^\circ$ IO (n = 10)

m	1	2	3	4	5	6
N	10	65	285	1000	3oo2	8007
N_1	10	20	30	40	50	60

The results of computations show that the essential complexity minimization leads to a loss in a quality of a reproduction of a desired signal which equals only to parts of percent. Moreover, the increase of the complexity of the filter given by (39) does not result in the essential imprevement of a quality of reproduction. On the contrary in many cases it implies the loss in the quality which comes from the poorer definiteness of the normal equations system.

Conclusion.

The optimal control theory developed up to the present time did not take into account the difficulties connected with a necessity of realization, with a complexity of algorithms for optimal control systems determination.

But essentially these two sides of the synthesis procedure may not be considered independently.

Hence, in the present time the investigations having in new to take into account in the formulation of optimal synthesis problems the complexity of optimization algorithms as well as the complexity of a physical realization of these algorithms have a great theoretical and practical significance.

64.1

ON ONE SYSTEM OF AUTOMATIC CONTROL OF MICROCIRCUITS MANUFACTURING PROCESSES

V.M.Glushkov, V.P.Derkach, G.T.Makarov

For the last 10-15 years the number of computers, which are the most powerful means for improving the efficiency of the man versatile activity, was increased hundreds of times. At the same time they, as a rule, are assembled manually, which rises their cost, reduces reliability and restricts their application. The time required for development and construction of computers is sometimes commensurable with, or in some cases even exceeds, the time of their depreciation, the latter being about 3 or 4 years nowadays.

In spite of numerous attempts on the part of investigators to automate production of circuits employing discrete components, significant results have not yet been obtained due to rapid complication of equipment and great variety of utilized parts. The search of essentially new physical and technological methods of constructing this equipment was needed that resulted in replacement of discrete parts and components used as constructing elements by microcircuits,well known now, produced in the form of thin films or hard semiconductor monoblocks.

Together with a rapid reduction in size and increase in service life of the devices the most important advantage of such circuits is the simplification of a task to automate stages of their production.

Although physical phenomena which are sufficiently studied, and such known methods as the vacuum depositing of matter on a backing, the diffusion, recrystallization, oxidation, epitaxial growth of monocrystal films,etching, photolithography, heat treatment, etc., are laid down to the microelectronics, the practical use of those methods are complicated due need of space localization of processes taking place on numerous microareas of material. Therefore, the development of ways to bring about a strictly controllable selective occuring of physical-technological processes is the most important problem of this new science.

One of such a method which is currently used in practice more often is the influence upon matter through a mask. Its use led to significant results which strengthened the microelectronics as a progressive science.

However, difficulties arose in the utilizing of this method. It is sometimes necessary to place the masks of different configurations in the chamber and mechanisms for their movements, which should oftentimes be heated to degas them and attain a deep vacuum. Noticeable errors appear in repeated matching of masks and, consequently, precision of geometric dimensions of components being made is reduced and reproducibility of their characteristics becomes worse. In many cases the mask-making of a required configuration is impossible altogether.

Precision of dimensions is increased by employing precision contact masks made on backings processed by photolithography. But application of photolithographic processes in handling semiconductor materials leads to increase in the variety of types of process operations used which causes additional errors. Because of random distribution of faulty components over the plate the wiring pattern in each copy of the microcircuit made in this way is individual. Consequently, a new mask is to be fabricated for each plate. To avoid sharp rise in production cost and great time loss they now restrict a number of elements made on one plate. Still the release of sound products does not exceeds several dozens of percent, at best.

A great variety of operations, the considerable portion of which are performed manually at that, materially hampers the solving of problems of complete automation of manufacturing microcircuits. It should also be taken into account that new processing methods rapidly coming into being cause essential alterations of costly automatic lines. That is why in many countries an intensive search is undertaken for such processing methods and tools, the parameters and characteristics of which would more completely comply with the problems of making devices with elements of micron dimensions.

Encouraging prospects for a successful solving those problems are opened by results of studying interactions between electron and ion beams and a hard body. The electron and ion (elion) processing of materials is justifiably said to be the most advanced of all current methods of creating components of microcircuits. P-n junctions, transistors, microwelding, polymerization, decomposition of chemical compositions aimed at restoration of chemical elements on local areas of a backing, recrystallization, exposition of photoresistive layers, deposition of films, scribing, hermetization, measurement and check of parameters of processing operations and articles, determination of chemical composition of materials, as well as a number of main operations necessary and performed in practice for fabricating microcircuits are made at present with the aid of electron and ion beams.

The advantage of elion methods consists in the possibility of accomplishing all processing stages in a vacuum chamber without its dehermetization, i.e. under conditions providing a high reproducibility of article parameters. A power density of electron and ion beams is readily adjusted and amounts to millions of kilowatts per cm^2. They permit to localize physical and technological processes in a very small space. A minimum diameter obtained in practice makes fractions of the micron.

Since the processes are localized not by mechanical movements of elements, say, masks, but by electric and magnetic fields actuating charged particles, one can choose for processing any point of a fabricated microcircuit almost in a moment and to a great accuracy, obtain any configurations of processed areas having rejected masks or, make masks proper, if required, to a best accuracy now attained.

The elion technology is one of examples of such a field of industry, the rapid development of which is conditioned by cybernetic means of control. Manually, through use of optical devices, it is possible to make only single laboratory samples of elements. No one can make in this way, during reasonable time, large blocks-multi-component circuits with an acceptable reproducibility of their parameters.

From the standpoint of variety of processing character and a number of elements which are to be acted upon by the control circuit, the elion installations are advantageous controlled plants . changing in time but little, therefore the principles of automatic control and technical means utilized for their embodiment, provided they are selected on the base of the newest scientific and engineering achievements, should have sufficiently vital capacity.

Developed at the first stage at the Institute of Cybernetics of Academy of Sciences, Ukr. SSR, is an open system of automatic control of processes manufacturing components of elion microcircuits ("Киів-67"), which is now operated at one of works and has already shown high fidelity and efficiency. The digital control method is laid down to the system as less time-consuming, and possessing better potentialities than the method using a continuous scanning.

In the electron-beam fabrication of microcircuit components, for instance, it becomes advantageous more often to employ a pulse mode of influencing the material by the beam. It is required to specify lengths of pulses and breaks, as well as a number of process pulses at each point. In the case of a continuous scanning of the entire raster, as it is done, say, in using a photocopy, the lengths of breaks are always related to pulse lengths by the equation:

$$t_n = \frac{LHt_u}{\mathcal{D}d_n}\ ,\qquad \text{where}$$

L and H = width and height of raster;
t_u= pulse length of processing;
D = permissible distance at which the beam can move during the time of action of the processing pulse, obtained from the condition of ensuring a sufficient resolving power;
d_n= beam diameter.

The formula shows that the relation between t_n and t_u cannot be chosen at random. Moreover, irradiated and shielded areas are scanned at the same velocity. This necessitates to waste time for useless motion of "blocked" beam, which is accumulated with an increase in a number of pulses and variety of time conditions of processing, and is determined by the equation:

$$t_{в.в.} = \sum_{k=1}^{m} \sum_{i=1}^{n} a_{ik} t_{nk},\qquad \text{where:}$$

a = relative number of unprocessed points on the backing;
m= number of varions lengths of processing pulses required for making the circuit;
n = maximum number of pulses of the same length, needed for processing at one point of the raster.

The continuouş scanning method is implied to be applicable only then when a change in time modes of processing within the raster is not needed. In overwhelming majority of other cases provision should be made for the possibility of quick automatic directing of the beam toward any area of the backing, and irradiating of a chosen point by immovable beam with any ratio t_n/t_u.

To perform this, the most convenient is the contour scanning obtained by converting digital codes, corresponding to coordinates of points, to deflecting currents or voltages. The use of the pitch contour scanning leads to the necessity of applying the digital system for controlling the electron-beam installation.

Existing universal digital computers are not adapted to such a purpose since they cannot ensure the minimal speed necessary for realization of advantages of the electron-beam technology and require excessive complication of programming processing.

Our system can be applied in laboratory investigations and job-lot production. It can also be employed for electron-beam welding, scribing and in many other cases when it is needed to act upon an object of control at the great speed through several channels simultaneously.

One of the most important features of the system is the simplicity of programming production tasks. One instruction can be used for processing any geometrical figure of five encountered more often of arbitrary dimensions within the raster (Fig.1). In this case the power and time modes of irradiation are the same for all points. Changing parameters a and b of "Point raster" used in manufacturing circuits with regularly positioned components (e.g. diode matrices) the distances between processed points are established. For $a=1$ or $b=1$ a series of parallel lines is obtained necessary, for instance, to create the so called current-carrying snake. If $a=1$ and $b=1$, the "Point raster" is converted to a rectangular area.

To make a large number of regularly positioned circuits on one silicon plate by a simplified program, the possibility of processing a series of regularly positioned areas by one instruction is provided. For this purpose serves a frame "Series of rectangles". If this frame is specified within one row, on the backing a broken line will be made encountered in fabricating resistors made in the form of a thin-film snake.

Lines and areas of arbitrary shape required, for instance, in welding bodies, mask-making, gating photo-resistive layers,etc., are processed by means of frames "Inclined line", "Circumference" ("Arc") and "Area". In the latter, two edges of the figure,or one of them, can be limited by an inclined line or a portion of the circumference, forming a triangle, equilateral or rectangular trapezium, circle, segment, etc.

Combining these geometric figures and links among them on the backing, and specifying their processing in a requisite time sequence, it is possible to make complex circuits, construct various functional devices.

Fof the general view of the "КиІв -67 " see Fig.2, and the block-diagram and functions of units БУО and БО - Figs.3 and 4.

Each instruction (frame code) is represented by ten twelxe-digit binary words stored in a magnetic memory in a permanent sequence. These words contain data on beam power modes, time parameters of processing, and directions on the law of moving the beam over the area surface, and all initial quantities necessary to obtain figures of required sizes. Registers СчНХ , СчКХ , РНУ and РКУ serve for specifying the commencement and end of beam movements. With a 5-μ pace and an additional reverse digit the maximum area of the surface being processed equals 10 X 10 mm.

The memory capacity is 4096 twelve-digit words which makes more than 400 codes of frames. This number obviously is sufficient for many applications, however the possibility is provided of replacement or replanishment of programs contained in the memory in the course of accomplishing process operations by means of a punch card input device or other computer used,for instance, for the automatic compiling of programs to fabricate microcircuits or to correct them according to feedback signals.

The nature and number of the principle geometrical figures had a decisive effect on constructing the computing unit (БУО) the base of which represents the two-integrator linearly-circular interpolator reconstructed for processing various frames (according to a code in the frame indication register). Each of them in turn consists of a counter and a counter-type adder.

In processing the point raster, for instance, distance a expressed in a number of pitches, between points on axis X is given by a code in the СчХ$_0$, and distance b on axis Y - by a code in the СчУ$_0$. These codes are stored in adders ΣХ and ΣУ during the calculation of all frame points.

Upon irradiation of the first point, the processing time of which is determined by time parameter shaping unit (БФВП, the transfer to the next point is done for which purpose a code in deviation counter СчХ (deflection unit БО) is changed by unit and, unit is subtracted from the code in the СчХ$_0$ simultaneously, provided the ТССХ is at "0", i.e. the row processing is not completed. Calculating the unprocessed points is accomplished up to zero state of the СчХ$_0$, which is seen by changing over the coincidence flip-flop ТСХ$_0$ to "1". This is followed by restoration of the code, characterizing the distance between places of processing, by its transferring from the ΣХ to the СчХ$_0$ and from the ΣУ to the СчУ$_0$ and a signal to switch on the БФВП is delivered. On processing the line (ТССХ at "1") the beam moves along axis Y as a result of addition (or subtraction) of units to (from) counter СчУ and simultaneous subtraction of units from the code in the СчУ$_0$. After reaching the specified distance between lines (ТСУ$_0$ is at "1") the beam is set to the initial point of the next line owing to code transfer from the СчНХ to the СчХ with simultaneous restoration of codes in the СчХ$_0$ and СчУ$_0$. Control is transferred to unit БФВП again. In such a sequence the control operations of processing and calculation of coordinates are altered until the code in the СчХ becomes equal the СчКХ code, and СчУ code equals the code in register РКУ, whereupon a signal "Frame end" is produced serving as an indication for delivering the next instruction from the memory (card input). For algorithms of processing various frames refer to Fig.5.

The time parameter shaping unit adjusts the lengths of pulses and breaks in between with any ratios of them in the range from 2 sec to 10.2 sec. A number of pulses irradiating each point of material can be taken from 1 to 2047. A continuous mode of operation is also provided.

To convert the codes to a deflecting current a principle is utilized of summation on a load of established, by the binary law, currents generated by stabilizers Ст.1 through Ст.10 (Fig.6) according to the equation

$$J_H = a_0 2^0 J_0 + a_1 2^0 J_0 + \ldots + a_n 2^n J_0 = J_0 \sum_{m=0}^{n} a_m 2^m$$

where: a_m = 0 and 1 depending on state of m-th digit of the deflection counter.

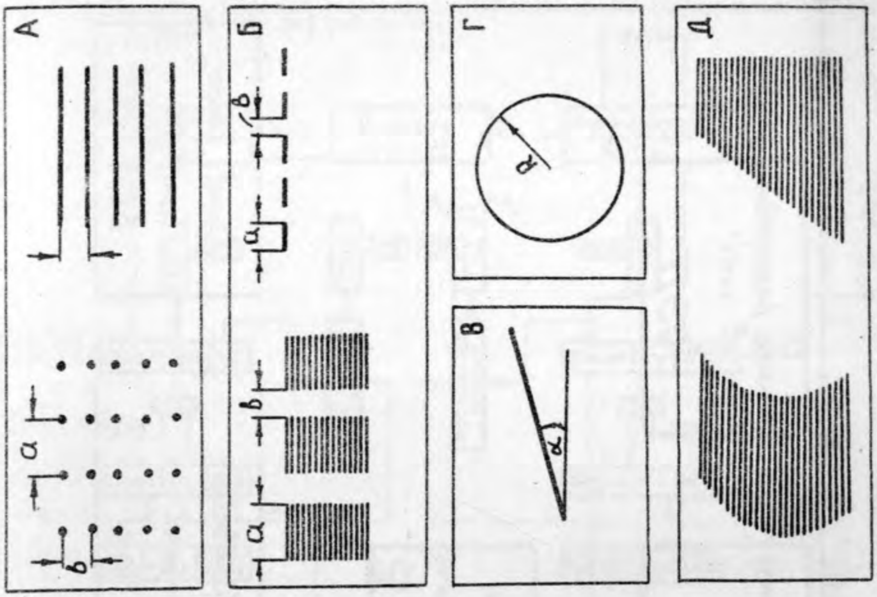
Aiming at maintenance of a constant dimension of processed region with variation of accelerating voltage the output current of the converter is corrected by a change in the reference voltage of the stabilizers. The deflecting coils are reversed by magnetically-controlled lug relays having a high speed and long service life.

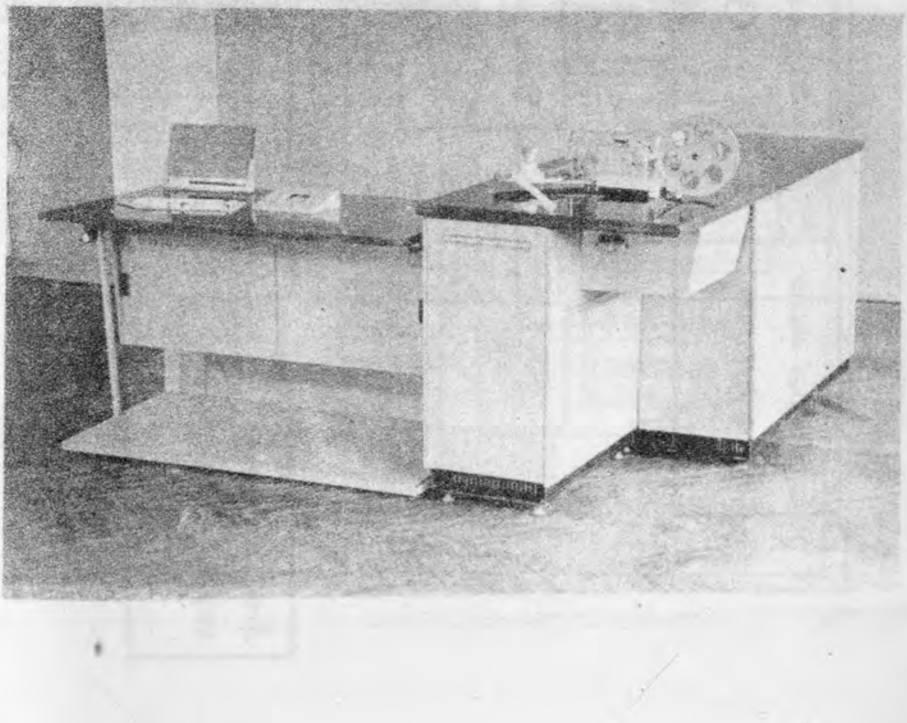
The magnitude of the beam current (a hundred of possible levels, is given in percent of a rated value) and of the accelerating voltage (16 stages, is given in кV) is adjusted by means of the power parameter control unit (БУЭП).

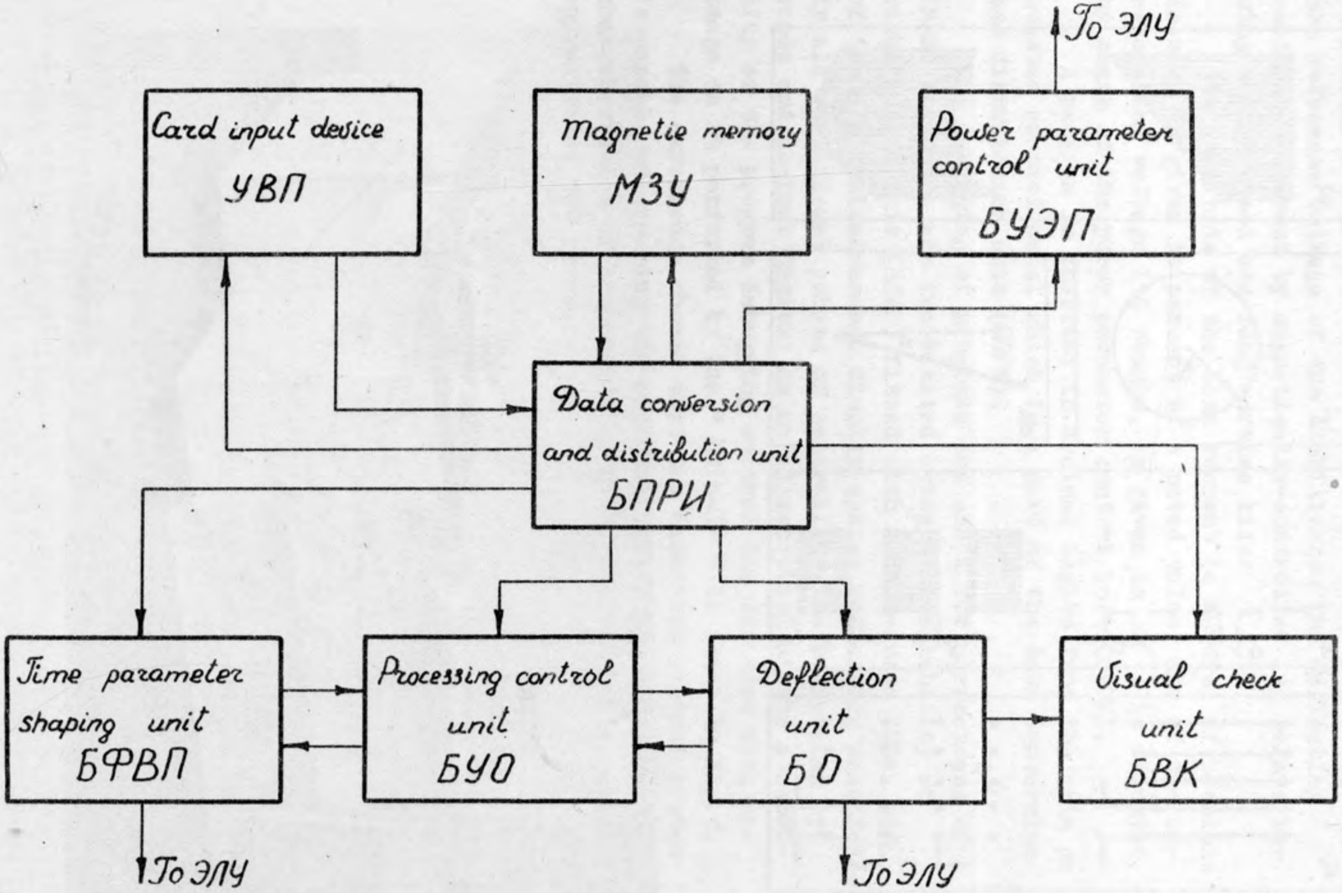
A program is inserted in decimal digits from the main or mounted control panel which is a part of the data conversion and distribution unit (БПРИ).

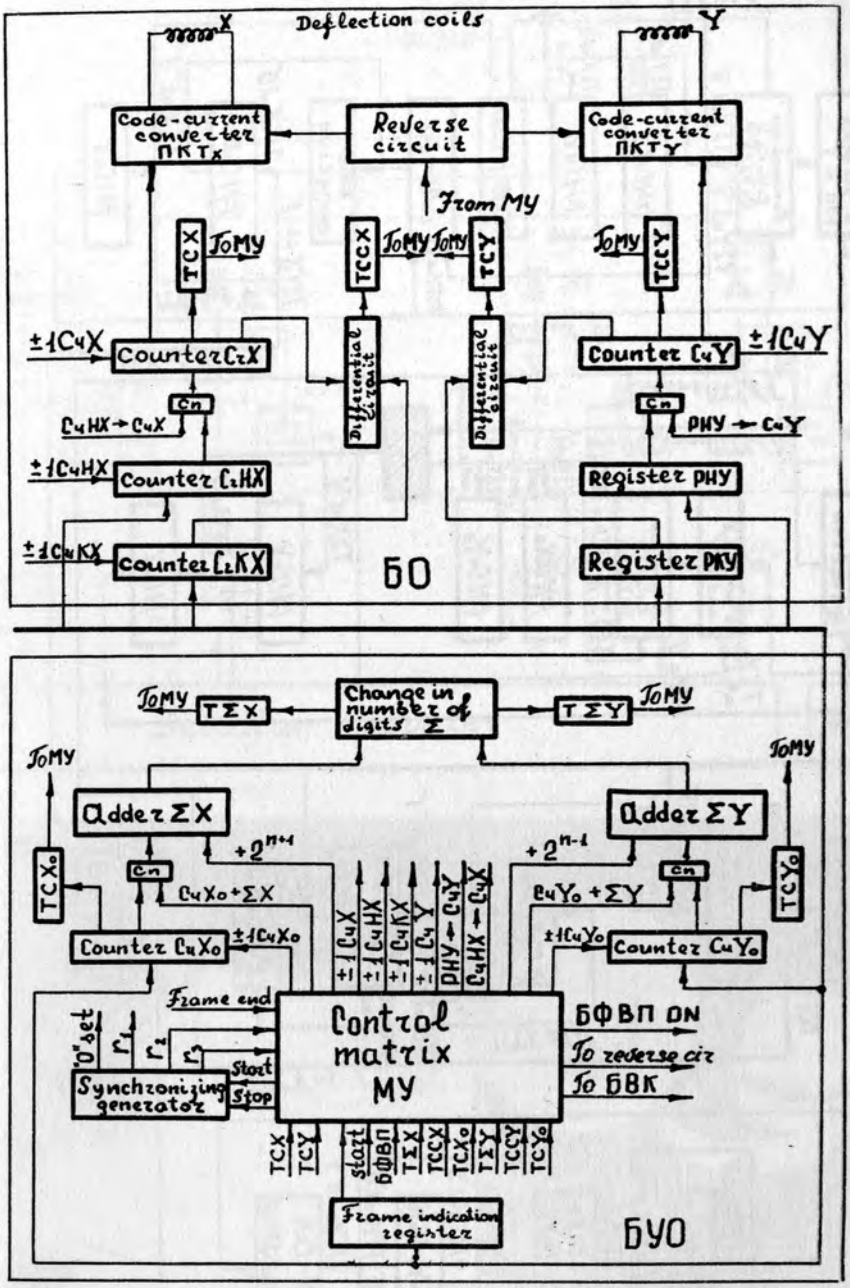
The debugging of programs and check for correctness of their insertion are facilitated owing to availability of visual check unit (БВК) fitted with a dark-trace tube, and of the electroluminescent display system reflecting practically all significant points of controlling the processing by words and decimal digits. As an illustration of the possibility of the program debugging control the skiatron with the image on it performed by the " КиЇв-67 " is shown in Fig.7.

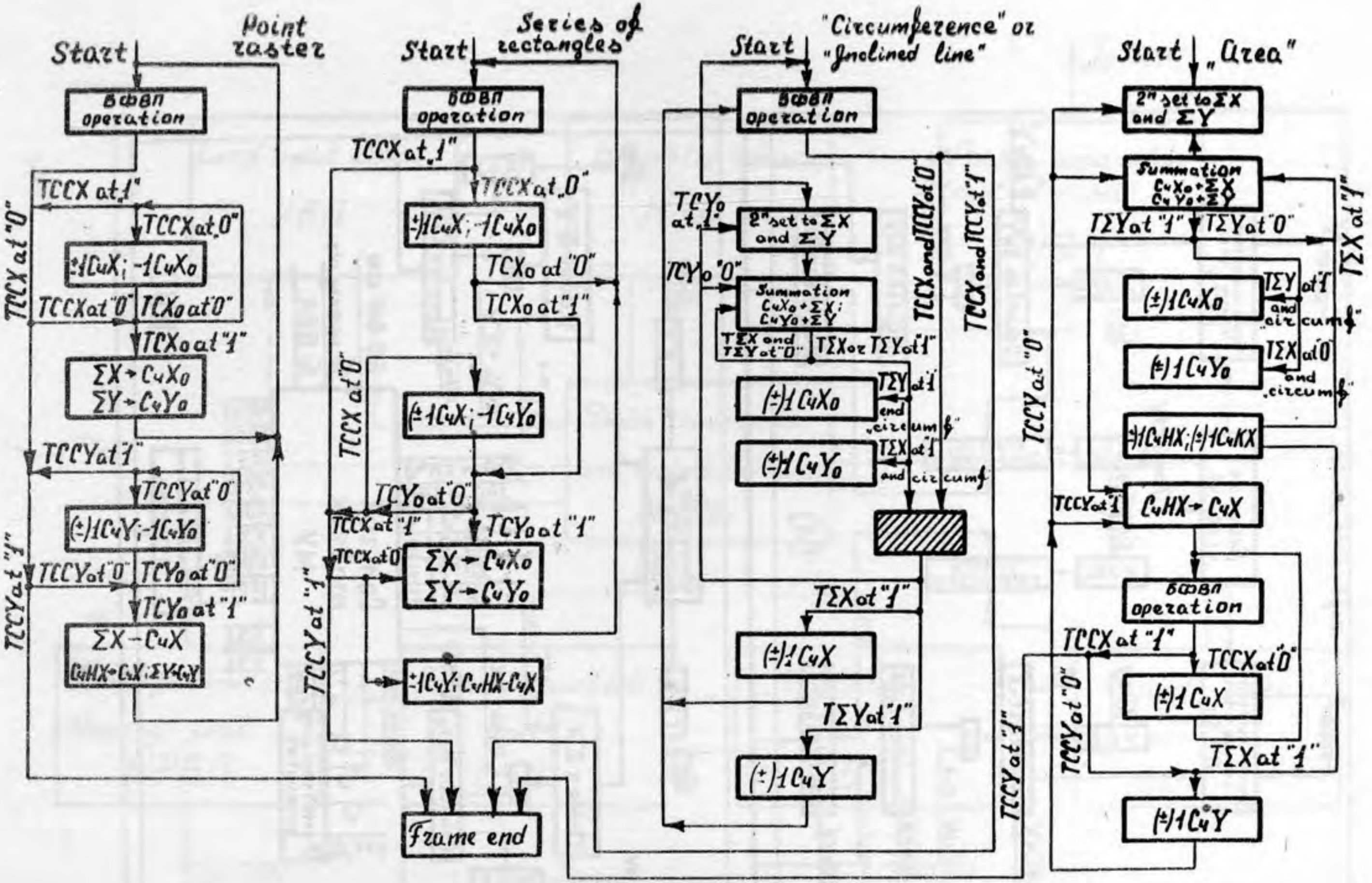
The experience showed that the described control system is capable of ensuring the reproducibility of electrical characteristics of components of elion microcircuits, which approximates 100 percent.


А
a
b
Б
a
b
a
b
В
α
Г
a
Д

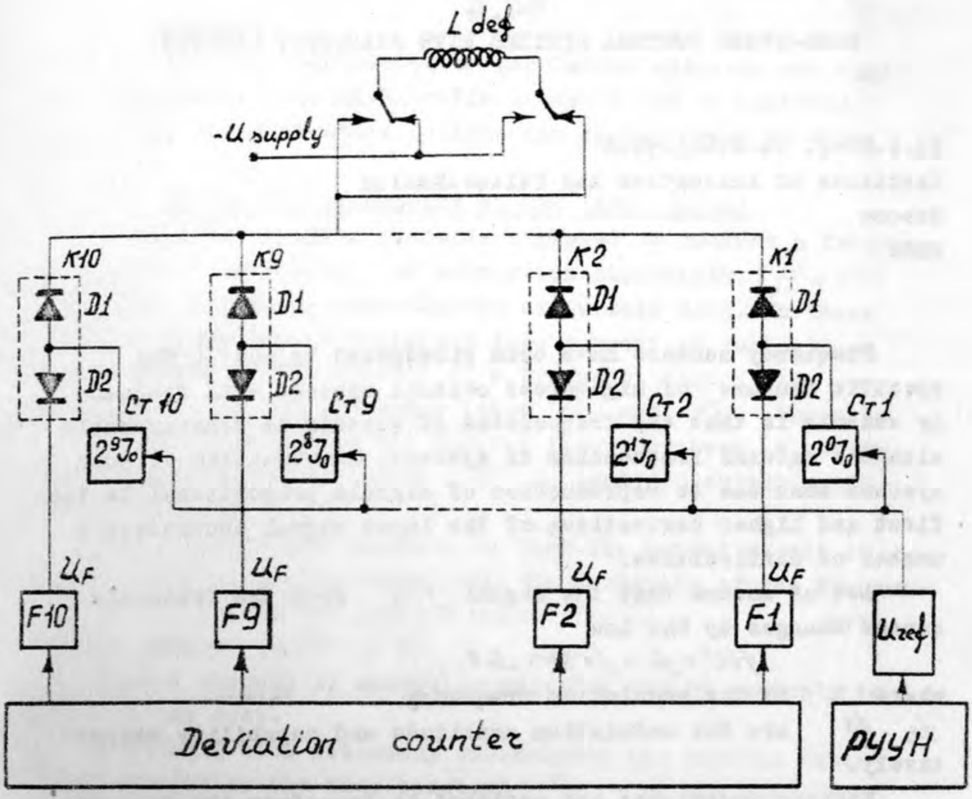



To ЭЛУ
Card input device
УВП
Magnetie memory
МЗУ
Power parameter control unit
БУЭП
Data conversion and distribution unit
БПРИ
Time parameter shaping unit
БФВП
Processing control unit
БУО
Deflection unit
БО
Visual check unit
БВК
To ЭЛУ
To ЭЛУ


Deflection coils
X
Y
Code-current converter ПКТx
Reverse circuit
Code-current converter ПКТy
From MY
TCX
ToMY
TCCX
ToMY
ToMY
TCY
ToMY
TCCY
±1C4X
CounterCzX
Differential circuit
Differential circuit
Counter C4Y
±1C4Y
Cn
C4HX → C4X
Cn
PHY → C4Y
±1C4HX
Counter CzHX
Register PHY
±1C4KX
Counter CzKX
БО
Register PKY
ToMY
TΣX
Change in number of digits Σ
TΣY
ToMY
ToMY
ToMY
Adder ΣX
Adder ΣY
+2^n+1
+2^n-1
TCX0
Cn
C4X0+ΣX
±1C4
±1C4HX
±1C4KX
±1C4Y
PHY → C4Y
C4HX → C4X
C4Y0 +ΣY
Cn
TCY0
Counter C4X0
±1C4X0
±1C4Y0
Counter C4Y0
Frame end
"0" set
r1
r2
r3
Start
Stop
Synchronizing generator
Control matrix MY
БФВП ON
To reverse cir
To БВК
TCX
TCY
start
БФВП
TΣX
TCCX
TCX0
TΣY
TCCY
TCY0
Frame indication register
БУО



Start
Point raster
БФВП operation
TCCX at "1"
TCCX at "0"
(±)1CuX; -1CuX0
TCCX at "0"
TCX0 at "0"
TCX0 at "1"
ΣX → CuX0
ΣY → CuY0
TCCY at "1"
TCCY at "0"
(±)1CuY; -1CuY0
TCCY at "0"
TCY0 at "0"
TCY0 at "1"
TCCY at "1"
ΣX → CuX
CuHX → CuX; ΣY → CuY
Start
Series of rectangles
БФВП operation
TCCX at "1"
TCCX at "0"
(±)1CuX; -1CuX0
TCX0 at "0"
TCX0 at "1"
TCCX at "0"
(±)1CuX; -1CuY0
TCY0 at "0"
TCCX at "1"
TCY0 at "1"
TCCX at "0"
ΣX → CuX0
ΣY → CuY0
TCCY at "1"
±1CuY; CuHX → CuX
Frame end
Start
"Circumference" or "Inclined line"
БФВП operation
TCY0 at "1"
2n set to ΣX and ΣY
TCY0 "0"
Summation CuX0 + ΣX; CuY0 + ΣY
TΣX and TΣY at "0"
TΣX or TΣY at "1"
(±)1CuX0
TΣY at "1" and "circumf"
(±)1CuY0
TΣX at "1" and circumf
TCCX and TCCY at "0"
TCCX and TCCY at "1"
TΣX at "1"
(±)1CuX
TΣY at "1"
(±)1CuY
TCCY at "1"
Start
"Area"
2n set to ΣX and ΣY
Summation CuX0 + ΣX; CuY0 + ΣY
TΣY at "1"
TΣY at "0"
TΣX at "1"
TCCY at "0"
(±)1CuX0
TΣY at "1" and "circumf"
(±)1CuY0
TΣX at "0" and "circumf"
(±)1CuHX; (±)1CuKX
TCCY at "1"
CuHX → CuX
БФВП operation
TCCX at "1"
TCCX at "0"
TCCY at "0"
(±)1CuX
TΣX at "1"
(±)1CuY


L def
-U supply
K10
K9
K2
K1
D1
D2
Ст 10
Ст 9
Ст 2
Ст 1
2⁹J₀
2⁸J₀
2¹J₀
2⁰J₀
U_F
F10
F9
F2
F1
U_ref
Deviation counter
РУУН

64.4

HIGH-SPEED CONTROL SYSTEMS WITH FREQUENCY SENSORS

Ye.K.Krug, Ye.A.Legovich
Institute of Automation and Telemechanics
Moscow
USSR

Frequency sensors have been widespread by now [1]. The specific feature of high-speed control systems with frequency sensors is that the frequencies of sensors is conmeasurable with the natural frequencies of systems. Construction of such systems when due to reproduction of signals proportional to the first and higher derivatives of the input signal encounters a number of difficulties.

Let us assume that the signal $f(t)$ from the frequency sensor changes by the law

$$f(t) = f_0 + f_A \sin \Omega t$$

where Ω is the modulation frequency
f_A, $\frac{f_A}{f_0}$ are the modulation amplitude and capability respectively.

Various techniques are employed to transform the frequency signal and construct units that would ensure corrective actions proportional to frequency deviation from the desired value, the integral of the deviation, the first, second, etc. derivatives of the deviation.

Those techniques are largely dependant on how the frequency signal is measured. There are two basic approaches:

1. Continuous (analog) measurements whereby frequency is measured by direct detection of the frequency signal;

2. Discrete (digital) measurements whereby frequency is measured

   a) by a count of pulses of the frequency to be measured for a given time interval θ (This approach will be termed integral), or

   b) by a count of the reference frequency pulses for the period $T_t = \frac{1}{f(t)}$ of the frequency to be measured. In this approach the input frequency period is measured.

The systems constructed by the latter approach use digital instrumentation; high static accuracy can be achieved. Let us see how corrective actions can be performed in each approach.

Continuous processing of the input signal

Among the existing circuits employed to measure a frequency signal at relatively low modulation capability. $\left(\frac{f_{\Omega}}{f_0} \le 0{,}2\right)$ those with frequency demodulators are widely used. In these circuits frequency is converted into current or voltage, an analog quantity. The circuit of a frequency demodulator contains (Fig. 1a) two resonance loops F_1 and F_2 tuned to boundary frequencies of the desired frequency band, a rectifier with a smoothing filter SF and d/c output amplifiers with gain K.

The choppers are adjusted so that the output signal of the demodulator is proportional to the deviation of the frequency $f(t)$ from its desired value $f_d(t)$ presumably equals $f_d(t) = f_0(t)$.

Such a circuit if carefully adjusted can be accurate to 0.5 to 1 per cent.

A circuit with frequency demodulator has inertia determined chiefly by the time constant T_{SF} of the smoothing filter elements.

The second column of Table 1 presents frequency responses of various correcting units where a frequency demodulator is used. To obtain a derivative signal at the output of the frequency demodulator a differentiating element is connected which consists of a loop $R_1C_1 = T_1$ and an amplifier with gain K_1 $(T_{d1} = KT_1)$ To obtain the second derivative signal two such elements are connected in series (Fig. 1b). Frequency conversion with a frequency demodulator practically cannot be used in systems where the frequency $f(t)$ is conmeasurable with the modulation frequency $\Omega(t)$ due to great phase errors introduced by the filtering element.

For instance if at the output the frequency $f(t)$ variation amplitude should not exceed 1 per cent $(2\pi f_0 T_{SF} = 100)$ and relatio $\frac{2\pi f_0}{\Omega}$ equals 0.01, then the phase error will exceed 40° . Assuming that the phase error is 10° we will have the relation $\frac{2\pi f_0}{\Omega} < 570$.

Investigations have revealed[9]that if the frequencies $f(t)$ and $\Omega(t)$ are conmeasurable and the modulation capability $\frac{f_\Omega}{f_0} \leq 0{,}1$, it is reasonable to use frequency multipliers with overall multiplication factor $N \geq 100$ (fig. 1b). In that case the inertia of filters can be reduced N fold. If $N \geq 100$ the time constant of the transducer can be neglected; the dynamic properties of the smoothing filter depend practically on delay introduced by the multiplier equal to $\frac{1}{f_0}$. The frequency responses of corrective units where a frequency multiplier is used are presented in the thirs column of Table 1.

Such a transducer can operate in the range $0 \leq \frac{\Omega}{f_0} \leq \frac{2\pi}{60}$; its phase error will not exceed $5^\circ - 6^\circ$ (e.g. at f_0 = 50 c.p.s. $\frac{\Omega_{max}}{2\pi} \leq 1{,}0$ c.p.s).

Design of corrective units that yield the first and the second derivatives as well as the integral of deviation encounters the well-known difficulties inherent in analog conversion techniques. The accuracy of analog integrators is below 0.5 to 1.0 per cent; this is also the case of analog amplifiers.

In practice amplifiers and differentiating loops enable to obtain signals of a narrow range and proportional to the first and second derivatives. The maximal values of differentiating constants depend on permissible phase distortions and maximal gain K which can be obtained at (ΩT) max$\leq$0.2 or $(\Omega T_d)_{max} \leq 0{,}2K$. If $K \geq 20$, $T_{d\,max} \leq \frac{4}{\Omega_{max}}$.

Thus we can say that the continuous (analog) conversion technique can be employed provided that the frequencies $f(t)$ and $\Omega(t)$ are conmeasurable only when a frequency multiplier with the multiplication factor $N \geqslant 100$ is used.

The static accuracy of the corrective units under consideration can be improved to 0.5 to 1 per cent.

Phase error due to conmeasurability of frequencies depends on the quantity $\frac{\Omega}{f_0}$. The corrective units which yield signals proportional to the first and second derivatives have additional phase errors inherent in analog differentiation techniques.

The integral processing of frequency signal

High static accuracy and noise-immunity in frequency measurements can be obtained if frequency pulses are counted for a certain time interval θ at discrete time instances nT. The expression for frequency in this technique is

$$f_{nT} = \int_{nT}^{nT+\theta} f(t)\,dt$$

The corrective unit which realizes the desired algorithm incorporates a pulse counter C a logical element LE and a pulse generator PG (Fig. 2a). The latter ensures that at time instances nT the logical element is fed a voltage for an interval θ during which the counter receives pulses of frequency $f(t)$.

Frequency responses of various corrective units acting by this technique are given in the second column of Table 1.

We can see that all systems with integral conversion of a frequency signal have considerable phase distortions caused by the measurement interval θ. Furthermore, the units which yield signal describing the deviations of frequency from its desired value and the units which compute the derivatives have phase errors dependent on quantization in time[3].

The output signal amplitudes of separate corrective units are also dependent on the magnitude of T. Since phase errors are large, the integral conversion technique seems to be advisable when frequencies $f(t)$ and $\Omega(t)$ are not conmeasurable.

Indeed, to achieve high accuracy of measurements, it is normally assumed that , and

should meet the condition $\frac{f_A}{\Omega} \geqslant 10000$.

It the modulation capability is low $\left(\frac{f_A}{f_0} \leq 0{,}1\right)$ frequency multipliers Fm (fig. 2a) are advisable if the interval θ to be reduced. In this case (see Table 2, third column) the interval can be made N times shorter and reach $\theta' = \frac{1}{\delta f_A \cdot N}$, where δ is the frequency measurements error.

In proportional units the cycle T can be reduced to θ' However, in systems where the corrective units ensure signals proportional to derivatives the time T cannot be changed arbitrarily. A salient feature of such units is that the time intervals have to be chosen according to possible range within which the modulation frequency Ω vary. Our analysis yielded specific relations between the input signal amplitude $A_p = f_A \cdot \theta$, the amplitude and phase errors and the quantity ΩT. Fig. 3 shows the relations to a log. seale. Solid lines denote the variations of the quantity ΩT against the ampliutude A_p at the input of a corrective element which reproduces a signal proportional to the first derivative. With this signal the amplitude at the output of that element equals A_{d_1} = 20 units, A_{d_1} = 50 units and A_{d_1} = 100 units. The dotted line denotes the relation of ΩT against A_p at the same values of amplitudes at the output of the corrective element that represents a second derivative signal.

Fig. 3 and Table 2 convince us that quantization in level which is inherent in all digital systems is the lower bound of the frequency range $\Delta(\Omega T)$ while the permissible phase error due to quantization in time is the upper bound of this corrective element. Fufthermore, the higher the order of correction the narrower the frequency range and the higher the requirements to the accuracy of measurements and adjustment of the parameter T.

Thus when the deviation is measured to the accuracy of 0.1 per cent and $f_A\theta$ = 1000 units, the first derivative can be found to the accuracy of 5 per cent (A_{d_1} = 20 units) at phase error 11° ($\Omega T \leq 0.2$) in the range $\Delta(\Omega T)_1$, ($0.02 \leq \Omega T \leq 0.2$) which le the derivative can be found with the accuracy of 5 per cent at phase error 18° ($\frac{3}{2}\Omega T = 0.20$) only at $f_A\theta \geq 10000$ within the range $\Delta(\Omega T)_2$ ($0.14 \leq \Omega T \leq 0.2$).

Therefore we can say that in the digital measurements technique the first - order derivative is obtained at a limited accuracy, while derivatives of high (upwards of the second order are practically impossible to obtain.

We have to note, however, that corrections proportional to the deviation of the frequency from its desired value and to the integral of deviation are obtained without difficulty. The integrated digital technique is very accurate in calculation of these corrections.

Frequency signal processing when its period is measured

When the period of the frequency to be measured is filled with reference frequency f_e, the measurement can be very accurate. A block-diagram of device using this approach is shown in Fig. 2b. The device includes a pulse counter C, a logical element LE and a reference frequency generator PG. The logical elements control loop feeds the frequency f_e pulses to the counter during an interval $T_t = \frac{1}{f(t)}$. As a result the counter records the number $N^* = [f_e \cdot \frac{1}{f(t)}]$.

Frequency responses of separate corrective units recorded in this measurements technique are presented in Table 3.

The left - side columns of the Table refer to the case when each period is measured successively and the differences between neighboring periods are measured (deviation $\Delta x_n = [f_e \cdot T_n]^* - [f_e \cdot T_{n-1}]^*$

The right - side columns present data for units which operate discontinuously with constant cycle time T.

It should be noted that systems based on measurements of period are non-linear. Therefore the data cited are true for low modulation depth $(\frac{f_A}{f_0} \leq 0.1)$ or when the systems have been linearized.

We can see that successive measurement of each period is the speediest technique which can be employed to obtain the deviation signal. However, derivatives cannot be thus measured. The accuracy of their measurements is very low, since the difference proper between neighboring frequency periods is very small, while the amplitudes of derivatives are proportional to $\frac{\Omega}{f_0}$.

If the measurement is taken with the accuracy 0.1 per cent which corresponds to $\frac{f_e}{f_0}$ = 1000, while the output signal modulation capability is $\frac{f_A}{f_0} \leq$ 0.1 and $\frac{\Omega}{f}$ 0.1 (e.g. f_0 = 50 c.p.s., f_A = 5 c.p.s., $\frac{\Omega}{2\pi}$ = 0.5 c.p.s.), the deviation amplitude corresponds to 100 units, while the amplitude of the first difference will not exceed one unit $A_{d1} = A_D \frac{\Omega}{f_0}$ = 1 units. This means that in such a system even the sign of the derivative cannot be accurately found.

To calculate the derivatives it is practical to introduce discrete time. Then the cycle time T has to be chosen in accordance to the input signal frequency range of the graphs in Fig.3

In practice we can design digital corrective units calculating the first derivative when the period is measured and the appropriate magnitude of T has been chosen. Both in this case and in the integrated technique the frequency band of the input signal at which the required accuracy of the derivative is achieved has the quantization in level (by the quantity $\frac{f_A}{f_0}$) as its lower bound and the phase error introduced by quantization in time as its upper bound.

Linearization is not essential in design of such correcting units since the measurement for signals proportional to the derivatives is not great.

It should be noted that if there is no linearization in systems with integrated action, static errors can appear if $\frac{1}{f(t)}$ is measured at discrete time instants nT rather than at each period at frequency $f(t)$ [4]. If each period is measured successively the integrated corrective unit can be regarded as a continuous integrator whose equation can be written as

$$y = \int [f(t) - f_e(t)]\, dt$$

The block - diagram of such unit is shown in Fig. 2b. It contains a counter (a logical element LE (a misalignment element) and a set point generator SG which sends a signal proportional to the desired value of the frequency

Thus the above analysis leads to these recommendations as to the choice of design technique for units that would ensure various corrective actions in systems with frequency as input signal.

I To obtain a signal of frequency deviation from its desired value we can use any of the above techniques.

1. The continuous technique whereby frequency multipliers and frequency demodulator have to be used yields a static accuracy for frequency obtained equal to 0.5 to 1 per cent at phase error of 10 to 12^0 in the frequency band $0 \le \frac{\Omega}{f_0} \le 0.1$ while the modulation capability is $\frac{f_\Omega}{f_0} \le 0.2$.

2. A high static accuracy ($\delta \le 0.001$) can be obtained both by the integrated technique of frequency $f(t)$ measurement and by measurement of period $\frac{1}{f(t)}$. The latter technique is faster. If it is changed successively each period can be used within the frequency band $0 \le \frac{\Omega}{f_0} \le 0.1$ at relatively large magnitude of modulation. However, to achieve a high static accuracy the appropriate linearization is required.

3. The integrated technique can be used when the frequencies $f(t)$ and $\Omega(t)$ are conmeasurable only with the provision that frequency multipliers with factor N and low modulation depth $(\frac{f_\Omega}{f_0} \le 0{,}1)$ are used. In practice the frequency band is $0 \le \frac{\Omega}{f_0} \le 0{,}1 N\delta$

II To obtain a signal that would be proportional to the first derivative with conmeasurable frequencies $f(t)$ and $\Omega(t)$

both continuous differentiators in combination with frequency multiplier and demodulator and fast digital differentiators. Both are approximately equal in terms of statical accuracy and dynamic error.

1. When analog differentiation and frequency multiplier with demodulator are employed the frequency band is bounded by the quantities $0 \le \Omega T_1 \le 0{,}2$ or $0 \le \frac{\Omega}{f_0} \le$ 0.2. The static accuracy of such differentiator is of the order 1 to 2 per cent. The dynamic error does not exceed 10°.

2. It is practical to design a fast digital differentiator when $0 \le \frac{\Omega}{f_0} \le 0{,}1$ as a unit that would calculate the difference between two periods of frequency $f(t)$ measured at time instances nT . The cycle time T has to be adjustable to the measurements range for $\Omega(t)$. A digital differentiator can operate within the range $0.02 \le \Omega T \le 0.2$, where the lower bound is made up by quantization in level, and the upper bound by the permissible phase error.

III. To obtain a signal proportional to the second derivative is difficult no matter whether digital or analog techniques are employed. The accuracy of such a signal is not high, phase errors considerable. Selection of a technique depends on the equipment employed to calculate the deviations and the first derivative.

No matter whether the frequencies $f(t)$ and $\Omega(t)$ are conmeasurable or not there is no point in designing digital corrective units that would calculate derivatives upwards of the second one.

IV. To obtain a signal proportional to an integral of the error it is advisable to use digital integrated frequency measurements techniques. This ensures a high accuracy of monitoring and control no matter whether the integrator employed is continnous or digital.

1. A digital integrator practically enables operation within the range $0 \le \frac{\Omega\theta}{2} \le 0.2$ at $\frac{f_B}{f_0} \le 0.5$.

2. A digital integrator enables operation within the range $0 \le \frac{\Omega\theta}{2} \le$ 0.2 at $\theta \le T$.

The integrated technique can also be recommended for design of corrective units discussed above for the case where the frequencies $f(t)$ and $\Omega(t)$ are not conmeasurable.

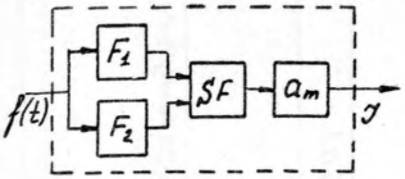
REFERENCES

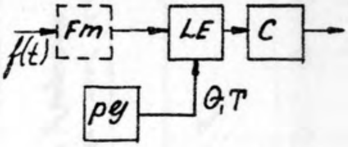
1. Агейкин Д.М. и др.Датчики контроля и регулирования. Машиностроение, I965.

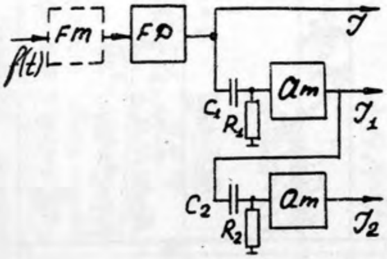
2. Fateeva Ye.A. Automation and Remote Control, No. 2, 1967.

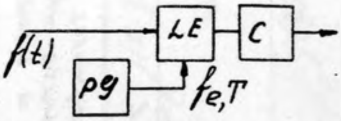
3. Krug Ye.K. Automation and Remote Control, No. 8,1967.

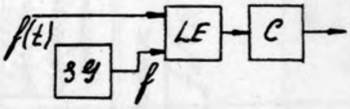
4. Круг Е.К., Александриди Т.М., Дилигенский С.Н. Цифровые регуляторы. "Энергия", I966.


f(t)
F1
F2
SF
am
J


f(t)
Fm
LE
C
pg
Θ,T


f(t)
Fm
FD
J
C1
R1
am
J1
C2
R2
am
J2


f(t)
LE
C
pg
fe,T


f(t)
LE
C
zg
f

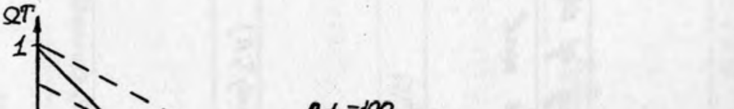

ΩT
1
0,4
0,2
0,1
0,04
0,02
0,01
Ad2=100
Ad2=50
Ad2=20
Ad1=100
Ad1=50
Ad1=20
Δ(ΩT)1
Δ(ΩT)2
Ap
100
200
400
1000
2000
4000
10000
20000

Table 1.

Idealized equations of the analog correcting units.	Continuous processing of the frequence signal.	
	Frequency responses of the real analog correcting units	Frequency responses of the real anolog correcting units when a frequence multiplier is used.
$y(t)=K_p\,x(t)$	$\frac{K}{\sqrt{1+(T_F\Omega)^2}}\,e^{-j\,\mathrm{arctg}\,T_F\Omega}$	$K e^{-j\left(\frac{\Omega}{f_0}\right)}$
$y(t)=T_{d_1}\frac{dx(t)}{dt}$	$\frac{K_1T_1\Omega}{\sqrt{(1+T_F^2\Omega^2)(1+T_1^2\Omega^2)}}\,e^{-j\left(-\frac{\pi}{2}+\mathrm{arctg}\,T_F\Omega+\mathrm{arctg}\,T_1\Omega\right)}$ $T_{d_1}=K_1T_1 \qquad T_1=R_1C_1$	$\frac{K_1T_1\Omega}{\sqrt{(1+T_1^2\Omega^2)}}\,e^{-j\left(-\frac{\pi}{2}+\frac{\Omega}{f_0}+\mathrm{arctg}\,T_1\Omega\right)}$ $T_{d_1}=K_1T_1 \qquad T_1=R_1C_1$
$y(t)=T_{d_2}^2\frac{d^2x(t)}{dt^2}$	$\frac{K_1K_2T_1T_2\Omega^2}{\sqrt{(1+T_F^2\Omega^2)(1+T_1^2\Omega^2)(1+T_2^2\Omega^2)}}\,e^{-j\left(-\pi+\mathrm{arctg}\,T_F\Omega+\mathrm{arctg}\,T_1\Omega+\mathrm{arctg}\,T_2\Omega\right)}$ $T_{d_2}^2=K_1K_2T_1T_2 \qquad T_2=R_2C_2$	$\frac{K_1K_2T_1T_2\Omega^2}{\sqrt{(1+T_1^2\Omega^2)(1+T_2^2\Omega^2)}}\,e^{-j\left(-\pi+\frac{\Omega}{f_0}+\mathrm{arctg}\,T_1\Omega+\mathrm{arctg}\,T_2\Omega\right)}$ $T_{d_2}^2=K_1K_2T_1T_2 \qquad T_2=R_2C_2$

Table 2

Idealized equations of the digital correcting units.	Integral processing of the frequence signal.	
	Frequency responses of the real digital correcting units.	Frequence responses of the real digital correcting units when a frequence multiplier is used.
$y(t)=K_I\sum_1^n x[nT]$ K_I $nT<t<(n+1)T$	$\frac{f_A\cdot\theta}{\Omega T}e^{j\left(-\frac{\pi}{2}-\frac{\theta\Omega}{2}\right)}$	$\frac{f_A\cdot N\theta^1}{T\Omega}e^{-j\frac{\pi}{2}}$ $\theta'=\frac{\theta}{N}$
$y(t)=K_p x[nT]$ $nT<t<(n+1)T$	$f_A\cdot\theta\, e^{j\left(-\frac{T\Omega}{2}-\frac{\theta\Omega}{2}\right)}$	$f_A\cdot N\theta' e^{-j\frac{T\Omega}{2}}$
$y(t)=K_1(x[nT]-x[(n-1)T]=$ $=K_1\Delta_1 x[nT]$ $nT<t<(n+1)T$	$f_A\theta\cdot T\Omega\, e^{j\left(\frac{\pi}{2}-\Omega T-\frac{\theta\Omega}{2}\right)}$	$f_A N\theta'\cdot\Omega T e^{j\left(\frac{\pi}{2}-\Omega T\right)}$
$y(t)=K_2(\Delta x[nT]-\Delta x[(n-1)T])=$ $=K_2\Delta_2 x[nT]$ $nT<t<(n+1)T$	$f_A\cdot\theta(T\Omega)^2 e^{j\left(\pi-\frac{3}{2}T\Omega-\frac{\theta\Omega}{2}\right)}$	$f_A N\theta'(\Omega T)^2 e^{j\left(\pi-\frac{3}{2}\Omega T\right)}$
$y(t)=K_3(\Delta_2 x[nT]-\Delta_2 x[(n-1)T]=$ $=K_3\Delta_3 x[nT]$ $nT<t<(n-1)T$	$f_A\theta(T\Omega)^3 e^{j\left(\frac{3}{2}\pi-2T\Omega-\frac{\theta\Omega}{2}\right)}$	$f_A N\theta'(\Omega T)^3 e^{j\left(\frac{3}{2}\pi-2\Omega T\right)}$

Table 3

The correcting signals.	Frequency singnal processing when its period is measured.	
	Frequency responses of the digital correcting units when each period is measured.	Frequency responses of the digital correcting units when period is measured at the nT.
Integral	$\frac{f_e f_A}{f_0^2\left(\frac{\Omega}{f_0}\right)} e^{j\left(-\frac{\pi}{2}-\frac{\Omega}{2f_0}\right)}$	$\frac{f_e f_A}{f_0^2 \Omega T} e^{-j\frac{\pi}{2}}$ $\frac{f_A}{f_0} < 0{,}1$
Deviation.	$\frac{f_e f_A}{f_0^2} e^{-j\frac{\Omega}{f_0}}$	$\frac{f_e f_A}{f_0^2} e^{-j\frac{T\Omega}{2}}$
The 1st derivative	$\frac{f_e f_A}{f_0^2} \frac{\Omega}{f_0} e^{j\left(\frac{\pi}{2}-\frac{3\Omega}{2f_0}\right)}$	$\frac{f_e f_A}{f_0^2} \Omega T e^{j\left(\frac{\pi}{2}-\Omega T\right)}$
The 2nd derivative.	$\frac{f_e f_A}{f_0^2}\left(\frac{\Omega}{f_0}\right)^2 e^{j\left(\pi-\frac{2\Omega}{f_0}\right)}$	$\frac{f_e f_A}{f_0^2}(\Omega T)^2 e^{j\left(\pi-\frac{3}{2}\Omega T\right)}$
The 3rd derivative	$\frac{f_e f_A}{f_0^2}\left(\frac{\Omega}{f_0}\right)^3 e^{j\left(\frac{3}{2}\pi-\frac{5\Omega}{2f_0}\right)}$	$\frac{f_e f_A}{f_0^2}(\Omega T)^3 e^{j\left(\frac{3}{2}\pi-2\Omega T\right)}$

67.1

A GENERALIZED SYNTHESIS LINEAR MULTIVARIABLE SYSTEMS

M.V. Meerov, R.T. Yanushevsky
Institute of Automation and Telemechanics

Moscow

USSR

Introduction

The paper will describe synthesis of multivariable control systems with minimal integrated squared performance functional. The great number of papers is this field can be divided into two mainstreams. First, a class in its own right is made up by problems where H.Wiener's mathematical tools are used.[1-3] (We do not distinguish deterministic and stochastic statements of optimal problems of this kind so close they are). Solutions to these problems have to yield the parameters of a transfer matrix function (they are the directly varied values of a control system that has not been excited initially and reacts in a certain way (by the optimality index to disturbances of a known kind). Second, the problem (the initial coordinates of the plant are assumed to be arbitrary) was stated in a more general way.[4-7] Their solutions is not related to a solution to the Wiener - Hopf equation which forms the basis of the papers of the first school; the dynamic equation are given in the phase (state) space; the directly varied values to solve the optimality problem are the plant coordinates and control signals.

A description of the plant dynamics in transfer functions defines just the controlled observed part of the plant (of course, if there are also uncontrolled and unobserved parts). The input and output coordinates of the plant are specific physical quantities. A description in terms of state space is more complete, however, the state vector coordinates (phase coordinates) proper are abstract quantities related to the output coordinates of the plant (controlled variables) through a certain transformation matrix. If it is remembered that for a large class of multivariable plants the quality of the process is defined by a generalized index which is a functional of output coordinates rather than of the plant state vector constituents, then[4-7] the optimality problem in the form of has no physical sense. (True, if the vector of output coordinates is expressed in terms of state vector, the synthesis problem can be solved as in[4-7] for the plant phase coordinates vector; however, a reverse transition to directly controlled quantities often non-elementary and depends on the state space basis chosen). For synthesis of multivariable systems is important that equations of plants that are sufficiently complex are found experimentally as weight (transfer) matrix function. Therefore when the approach of [4-7] is employed, one has also to find the dynamic system which has, for the given transfer matrix, the lowest order differential equations. This seems to suggest that practical use of results obtained in [4-7] for synthesis of complex multivariable systems is limited; in this respect the approach of[1-3] is preferable. The generalization of the approach in [8] for the case of non-zero initial conditions of the system needs a stricter argumentation.

The approach described in the paper is not a direct extension of the above papers and differs that the problem is stated in more general terms: assuming that the plant dynamic equations are given in integrated form for controlled variables, a multivariable system is synthetized under arbitrary initial conditions, the disturbances and desired signals to be reproduced. The solution suggested yields expressions for transfer matrix functions of the optimal system that are convenient for computers. We simplified the solution procedure for certain types of multivariable plants.

Also, the integrated square form is, as a rule, an indirect performance index for a control system. An engineer is interested more directly in accuracy of control, control time, overshooting, etc. The latter is not normally incorporated in the performance criterion but should be monitored if the requirements to performence expressed as the optimality index are not rigid; in other words the performence functional integrand factors have to be chosen so as to satisfy the above quality estimates.Therefore we will also study the effect of performence functional factors on the properties of an optimal system.

2. Statement of the problem and basic relations

Assume that we have a linear multivariable plant whose dynamic state is described by this set of equations

$$y(t)=Z(t)z(0)+\int_0^t W(t-\tau)u(\tau)d\tau+f(t) \qquad (2.1)$$

where x(t) is plant output coordinates vector; u(t) is the control actions vector; f(t) is the disturbances vector; W(t) is the plant weight matrix function; $Z(t), z(0)$ are the fundamental matrix and the initial conditions vector respectivly; they determine the free constituent of the solution to (2.1).

It is required to find a control low to minimize the functional:

$$I=\frac{1}{2}\int_0^\infty\{(A(p)[y(t)-y_z(t)])^*(A(p)[y(t)-y_z(t)])+[c(0)u(t)]^*[c(0)u(t)]\}dt \qquad (2.2)$$

where $y_z(t)$ is the desired vector; $A(p)$- the diagonal constant operating matrix;

$C(0)$- a constant factor;

$(^*)$- the transposition symbol.

The actions $y_z(t), f(t)$ are assumed vanishing[9] continuous time functions, the intergrand elements are taken non-negative. (The optimal problem is assumed to have a sense).

As a synthesis problem the problem just stated is to design a closed-loop control system, that would reproduce the desired signals $y_z(t)$ and counteracts the disturbances $f(t)$ with efficiency that depends on the factors of the matrix $A(p)$ and $C(0)$.

The increment of functional (2.2) ΔI (its first variation) due to $y(t)=y^o(t)+\Delta y(t)$, $u(t)=u^o(t)+\Delta u(t)$ (where $u^o(t)$, $y^o(t)$ are the optimal control actions and changes of output coordinates caused by them; $\Delta y(t) = \int_0^t W(t-\tau)\Delta u(\tau)\, d\tau$) equals

$$\Delta I = \int_0^\infty \{ (\int_0^t \Delta u^*(\tau) W^*(t-\tau)\, d\tau) A(-p) A(p)(y^\circ(t) - y_z(t)) + \Delta u^*(t) c^2 u^\circ(t) \} dt \qquad (2.3)$$

By changing the intergration order in the first item of integrand (2.3)

$$\Delta I = \int_0^\infty \Delta u^*(\tau) (\int_\tau^\infty W^*(t-\tau) A(-p) A(p)(y^\circ(t) - y_z(t))\, dt + c^2 u^\circ(\tau)) d\tau$$

we will obtain since the first variation of the functional is zero this expression for optimal control *)

$$u^\circ(t) = \frac{1}{c^2} \int_t^\infty W^*(\tau - t) A(-p) A(p) (y_z(\tau) - y^\circ(\tau))\, d\tau . \qquad (2.4)$$

Let us use further on Laplace transforms; the functions $y^\circ(t)$, $u^\circ(t)$ will be analyzed in the plane of complex variable s .Then the expression for optimal control (2.4) is

$$U^\circ(s) = \frac{1}{c^2} \{ W^*(-s) A(-s) A(s) (Y_z(s) - Y^\circ(s)) \}_+ \qquad (2.5)$$

where the symbol ($-s$) means a Laplace transform of the function which exists at $t < 0$; the fraces with the positive sign denote the positive time terms the corresponding expression.With this notation eq. (2.5) incorporatesrealizability of the control system.

Having obtained a Lạplace transform for eq. (2.1) and introduced this into (2.5) we shall have

$$\{ [E + \frac{1}{c^2} W^*(-s) A(-s) A(s) W(s)] U^\circ(s) \}_+ = \qquad (2.6)$$
$$= \frac{1}{c^2} \{ W^*(-s) A(-s) A(s) (Y_z(s) - F(s)) \}_+ - \frac{1}{c^2} \{ W^*(-s) A(-s) A(s) Z(s) \}_+ z(0) .$$

The matrix $E + \frac{1}{c^2} W^*(-s) A(-s) A(s) W(s)$ is a Hermit matrix and can be represented [10-12] as a product of matrices $H^*(-s) H(s)$.

*) The minimization of the functional $I = \frac{1}{2} \int_0^\infty \{ [A(p)(y(t) - y_z(t))]^* \cdot [A(p)(y(t) - y_z(t))] + [c(p) u(t)]^* \cdot [c(p) u(t)] \} dt$

where $c(p)$ is the operator poljnomial is easily reducible to our problem analyzed by introduction of a dummy control $v(t) = c(p) u(t)$ and its elimination from the result.

Therefore after replacing thus the left-hand side of (2.6) we will multiply both sides of the expression by $H^{*-1}(s)$ [11] and after elementary transformations we will finally obtain an expression for the optimal control law

$$U^{o}_{(s)} = \frac{1}{c^2} H^{-1}_{(s)} \left\{ H^{*-1}_{(-s)} W^{*}_{(-s)} \mathcal{A}_{(-s)} \mathcal{A}_{(s)} \left[Y_z(s) - F(s) - Z(s)\, z(0) \right] \right\}_{+} \qquad (2.7)$$

By substituting (2.7) into (2.1) we shall have

$$Y^{o}_{(s)} = \left(Z_{(s)} - W_{(s)} H^{-1}_{(s)} \left\{ \frac{1}{c^2} H^{*-1}_{(-s)} W^{*}_{(-s)} \mathcal{A}_{(-s)} \mathcal{A}_{(s)} Z_{(s)} \right\}_{+} \right) z(0) + \qquad (2.8)$$

$$+ F(s) + W(s) H^{-1}_{(s)} \left\{ \frac{1}{c^2} H^{*-1}_{(-s)} W^{*}_{(-s)} \mathcal{A}_{(-s)} \mathcal{A}_{(s)} \left[Y_z(s) - F(s) \right] \right\}_{+} .$$

It follows from (2.7) and (2.8) that the stability of the system depends on the stability of $H^{-1}(s)$ [1] which is feasible [10-12]. One can easily see that there is just one solution to (2.7), (2.8) although a Hermit matrix can be decomposed with the accuracy to a unitary matrix. This easily follows from replacement of $H(s)$ in (2,7) (2.8) by the matrix $H_1(s) = QH(s)$ where Q is a unitary matrix and from $Q^{*}Q = E$.

To solve the synthesis problem the vector $z(0)$ should be eliminated from expressions (2.7), (2.8). Elementary transformations will yield a solution to the synthesis problem as

$$U^{o}_{(s)} = -N_1(s)\left(Y^{o}_{(s)} - F(s)\right) + \left[E + N_1(s) W(s)\right] H^{-1}_{(s)} \left\{ M(-s) \left[Y_z(s) - F(s) \right] \right\}_{+} \qquad (2.9)$$

where

$$M(-s) = \frac{1}{c^2} H^{*-1}_{(-s)} W^{*}_{(-s)} \mathcal{A}_{(-s)} \mathcal{A}(s) , \qquad (2.10)$$

$$N_1(s) = H^{-1}_{(s)} \left\{ M(-s) Z(s) \right\}_{+} \cdot \left(Z(s) - W(s) H^{-1}_{(s)} \left\{ M(-s) Z(s) \right\}_{+} \right)^{-1} . \qquad (2.11)$$

If the number of control signals equals that of controlled variables, expression (2.10) simplifies to the form

$$M(-s) = H^{*-1}(-s) - H(s). \qquad (2.12)$$

The free terms in the plant equations (2.1) contains an initial conditions vector whose dimension in solutions to the synthesis problem is assumed the same as for the vector of controlled variables; [redacted] generally this needs an explanation. If $y(t)$ is the plant state vector, then $z(0) = y(0)$ and (2.9), (2.12) present a solution to the analytic design problem other than the one in [5,6] ($y_z(t)$ is assumed zero). In a general case the vector of controlled variables is related to the state vector through a certain matrix, while the components of the vector $z(0)$ equal to certain linear combinations of the state vector components (depending on the basis chosen) and with respect to the controlled variables vector are equal to linear combinations of their initial values $y(0)$ and their derivatives. In order not to deal with the generalized transfer matrix, when $N_1(S)$ is found, the initial values of just senior derivatives of controlled variables are assumed non-zero. The problem stated is not narrowed, since the esistence of a solution to the synthesis problem at any initial conditions follows from the above case where $z(0)=y(0)$ is the plant state vector. This should be borne in mind when $Z(s)\,z(0)$ are written in (2.11), though, as the discussion below will show, expression (2.11) may be not used in finding $N_1(S)$.

Expression (2.9) shows that the optimal controller should contain channels for disturbances and control signals. Let us note that when the control system is designed the disturbances cannot be always measured directly. Then to find the transfer matrix function of the optimal system that wolld maximally counteract the disturbances we should represent $F(s) = f(s)\,e$ where $f(S)$ is a diagonal matrix that consists of zeroes and poles of disturbances, e is the column vector of their amplitude; e should be eliminated from expressions (2.7), (2.8). Then with zero initial conditions we have

$$U^{o}(s) = -N_1^{(1)}(s)\,Y^{o}(s) + \left[E + N_1^{(1)}(s)W(s)\right]H^{-1}(s)\left\{M(-s)Y_{\tau}(s)\right\}_{+} \qquad (2.13)$$

where

$$N_1^{(1)}(s) = H^{-1}(s)\left\{M(-s)f(s)\right\}_{+}\cdot\left[f(s) - W(s)H^{-1}(s)\left\{M(-s)f(s)\right\}_{+}\right]^{-1}. \qquad (2.14)$$

Synthesis of a linear control system was studied for systems with one controlled variable. The system had to reproduce $Y_{\tau}(s)$ (the initial conditions of the system coordinates are assumed zero) with the performance functional of the form of eq. (2.2). A similar result can be easily obtained for multivariable plants assuming that in (2.7), (2.8) $Y_{\tau}(s) = y_{\tau}(s)\,e$ ($y_{\tau}(s)$ is a diagonal matrix that consists of zeroes and poles of adjustments) and eliminating the vector e from these expressions. Before that eq.(2.8) is written for the difference $Y_{\tau}(s) - Y^{o}(s)$.

$$U^{o}(s) = N_1^{(2)}(s)\left[Y_{\tau}(s) - Y^{o}(s)\right] + N_1^{(2)}(s)F(s) - \qquad (2.15)$$
$$- \left[E + N_1^{(2)}(s)W(s)\right]H^{-1}(s)\left\{M(-s)F(s)\right\}_{+}$$

where

$$N_1^{(2)}(s) = H^{-1}(s)\left\{M(-s)y_{\tau}(s)\right\}_{+}\cdot\left[y_{\tau}(s) - W(s)H^{-1}(s)\left\{M(-s)y_{\tau}(s)\right\}_{+}\right]^{-1}. \qquad (2.16)$$

These expressions simplify readily in case where the numbers of control signals and controlled variables are equal.

3. Calculation procedure

The above discussion has shown that the calculating scheme of our problem reduces to a number of algebraic operations on operator matrices in the plane of complex variable S. The parameters of the optimal system are calculated by formulas (2.9) - (2.16). The most difficult thing in calculation is to decompose the Hermit matrix $E + \frac{1}{C^2} W^*(-s) A(-s) A(s) W(s)$. The decomposition procedure has been described in [10-12]. We have to introduce still another way of finding the matrix $N_1(s)$ since, on the one hand, the synthesis at $Y_\tau(s) = F(s) = 0$ has a separate solution [5,6], and on the other hand, the solution is a starting point for finding the operators of relations for adjustments and disturbances. The approach suggested is based on the similarity between expression (2.7) for optimal control and the corresponding expression for the diagram of Fig. 1.

$$U^o_{(s)} = -\left[E + N_1(s) W(s)\right]^{-1} N_1(s) Z(s)\, z(0). \tag{3.1}$$

A comparison of (2.7) and (3.1) shows that there is a matrix

$$H(s) = E + N_1(s) W(s) \tag{3.2}$$

for which the following expression is valid

$$E + \frac{1}{c^2} W^*(-s) A(-s) A(s) W(s) = E + W^*(-s) N_1^{-*}(-s) + \\ + N_1(s) W(s) + W^*(-s) N_1^{-*}(-s) N_1(s) W(s). \tag{3.3}$$

Assuming a kind of matrix $N_1(S)$ we can make a set of nonlinear equations to find its indefinite factors. Similarly the calculation procedure can be designed for the transfer matrix function of the open-loop system $G(s) = N_1(s) W(s)$. The calculation procedure simplifies considerably when applied to separate classes of multivariable plants where the specific form of their transfer matrices is of importance. For illustration purposes we propose to discuss synthesis of multivariable plants with intra-group symmetry.

4. Solution for a class of multivariable plants

We will discuss only those plants where groups with identical parameters and relations and which are known as plants with intra-group symmentry [13]. Those groups represent in their turn a subclass and are called unitype bounded plants [14]. We shall start with these plants. That the solution to the optimality problem is simplified for the performance functional (2.2) (the dimensionality of vectors $y(t), y_r(t), f(t), u(t)$ is n, $A(p)$ is the diagonal matrix with similar elements $a(p)$) for these plants follows from the description of the two types of the system movement in [15].

a) overall natural described as

$$y_\Sigma(s)=\sum_{i=1}^{n} y_i(s), \; u_\Sigma(s)=\sum_{i=1}^{n} u_i(s), \; y_{\Sigma r}(s)=\sum_{i=1}^{n} y_{ri}(s), \; f_\Sigma(s)=\sum_{i=1}^{n} f_i(s). \quad (4.1)$$

b) overall relative described as

$$y_{i\Sigma}(s)=\sum_{j=1}^{n}(y_i(s)-y_j(s)), \quad u_{i\Sigma}(s)=\sum_{j=1}^{n}(u_i(s)-u_j(s)), \quad (4.2)$$

$$y_{ri\Sigma}(s)=\sum_{j=1}^{n}(y_{ri}(s)-y_{rj}(s)), \quad f_{i\Sigma}(s)=\sum_{j=1}^{n}(f_i(s)-f_j(s)).$$

It follows naturally that

$$y_i(s)=\frac{1}{n}(y_\Sigma(s)+y_{i\Sigma}(s)), \quad u_i(s)=\frac{1}{n}(u_\Sigma(s)+u_{i\Sigma}(s)), \quad (4.3)$$

$$y_{ri}(s)=\frac{1}{n}(y_{r\Sigma}(s)+y_{ri\Sigma}(s)), \quad f_i(s)=\frac{1}{n}(f_\Sigma(s)+f_{i\Sigma}(s)).$$

$$(i=1,\dots,n).$$

When the transform of eq. (4.1), (4.2) is used, the performance functional decomposes into a number of functionals

$$I=\frac{1}{n}I_\Sigma+\frac{1}{n^2}\sum_{i=1}^{n} I_{i\Sigma},$$

$$I_\Sigma=\frac{1}{2}\int_0^\infty\{[a(p)(y_\Sigma(t)-y_{r\Sigma}(t))]^2+[cu_\Sigma(t)]^2\}dt, \quad (4.4)$$

$$I_{i\Sigma}=\frac{1}{2}\int_0^\infty\{[a(p)(y_{i\Sigma}(t)-y_{ri\Sigma}(t))]^2+[cu_{i\Sigma}(t)]^2\}dt. \quad (4.5)$$

Instead of eqs. (2.1), (2.5) we have

$$y^{o}_{\Sigma}(s) = y_{cb\Sigma}(s) + W_{\Sigma}(s)\,u^{o}_{\Sigma}(s) + f_{\Sigma}(s) \tag{4.6}$$

$$y^{o}_{i\Sigma}(s) = y_{cbi\Sigma}(s) + W_{i\Sigma}(s)\,U^{o}_{i\Sigma}(s) + f_{i\Sigma}(s) \tag{4.7}$$

where

$$W_{\Sigma}(s) = w(s) + (n-1)\,\ell(s)\,,\quad W_{i\Sigma}(s) = w(s) - \ell(s)\,,$$

$w(s)$ is the transfer matrix of the plant diagonal elements; $\ell(s)$ - the transfer matrix of cross relations; $y_{cb\Sigma}(s)$, $y_{cbi\Sigma}(s)$ - are the free components of the transfer process for various types of motion. As has been shown in[15], a solution to the initial optimality problem is equivalent to minimization of functionals in eqs. (4.4) (4.5) by eq. (4.6)(4.7). The result thus obtained is extended to multivariable plants with intra-group symmetry[15].

Analysis of the two types of movement of separate groups within a multivariable system with intra-group symmetry and of the appropriate control signals makes it possible to use, instead of the initial set of equations and a performance criterion, the equivalent equations of motion of the generalized coordinates introduced and the quality functional appropriately transformed. To find in the transformed functional those constituents that are related to the above types of motion and to analyse the optimization problem for each type of motion means to solve the initial problem. The transformation suggested makes it possible to replace optimization of a multivariable system by equivalent optimizations of nonrelated lower-order systems. This simplifies the calculation procedure substantially and can be treated as decomposition for the given class of plants.

5. The structure and properties of the optimality problem

From the above expressions for solution of the synthesis problem follows that the optimal system can generally be a combined one with channels for disturbances and desired signals (Fig. 1). The various structural versions for the case where direct measurement of disturbances is impossible as well for extension of the approach in [1,3] (a narrower approach than in this paper) to multivariable plants are represented in Figs 2 and 3 respectively. ($N_2(s)$, $N_3(s)$ are the operators for disturbances and desired signals). Unlike the scheme of Fig. 1, the operators $N_1^{(1)}(s)$, $N_1^{(2)}(s)$ depend substantially on the form of $F(s)$, $Y_\tau(s)$. The stability of the optimal system as follows from the relations given in the paper is ensured by the stability of zeroes in $H(s)$. Therefore [10–12] a stable control system can be constructed for a broad class of linear multivariable plants with the exception of plants which contain eliminable unstable zeroes and poles. One can easily see that the presence of the latter in the transfer matrix function of the plant (as distinct from the eliminable stabel zeroes and poles which do not affect synthesis) lead to the existance of an unstable component in the solution to the optimality problem.

With the given $A(p)$, C the equation

$$\left| E + \frac{1}{C^2} W^*(-s) A(-s) A(s) W(s) \right| = 0 \qquad (5.1)$$

for the optimal system is similar to a characteristic equation [16,17]. Because the zeroes and poles (5.1) are symmetrical in the plane of a complex variable with respect to the origin of coordinates, assuming $-s^2 = z$ one can study it by techniques known in the control theory [16,17]. The analysis was to find the effect of factors $A(p)$, C on the position of zeroes and poles of equation (5.1) i.e., in the final analysis, on the properties of an optimal system (The condition of reality of optimal system roots based on [16], of oscillatory index based on [17] etc). This is the chief factor for selection of factors in the integrand of performance functional when specific problems are solved.

An important performance criterion is the accuracy of control related in the final analysis, to gains in separate channels of a multivariable system ($f(t)$, $y_2(t)$ are assumed to be disappearing time functions with the convergence condition of eq. (2.2); However, a replacement of a non-disappearing function by a disappearing one with infinite decreasing time is quite justifiable for practical purposes). The above procedure of optimal system parameters calculation makes it possible to estimate the gains of a control system transfer matrix function while no solution to the optimality problem is required. It is just necessary to assume $s=0$ in eq. (3.3). Then

$$\frac{1}{c^2} W^*(0) A^2(0) W(0) = W^*(0) N^*(0) + N(0) W(0) + W^*(0) N^*(0) N(0) W(0). \quad (5.2)$$

If the gains of diagonal elements $N(0)W(0)$ are large (this is a property of high performance systems), the magnitude of $N(0)W(0)$ by which it is easy to estimate the properties of the system in steady state, can be assumed close to $A(0)/c$. The gains of channels in a multivariable system are easily seen to increase infinitely with the increase of the matrix $N(0)W(0)$ elements magnitude. In this case we come to linear systems with large gain [16].

6. Conclusions

We have discussed synthesis of multivariable control systems with performance functional in integrated square form. It has been shown that the optimal system has to be a combined one with channels for disturbances. Expressions for systems parameters are presented. An optimal system synthesis procedure has been presented that would maximally counteract the disturbances if these cannot be measured. The specific features of solution for multivariable plants with intra-group symmetry have been found and the effect of the functional factor on the properties of the optimal system discussed.

REFERENCES

I. Wiener N.L., Extrapolation, Interpolation and Smoothing of Stationary Time Sεries, John Willey, N.Y., 1950.

2. Катковник В.Я., Полуэктов Р.А. Многомерные дискретные системы управления, Изд-во "Наука", I966г.

3. Чанг С.Л., Синтез оптимальных систем автоматического управления, Изд-во "Машиностроение", I964г.

4. kalman R.E., A New Approach to Linear Filtering and Prediction Problems, Trans. ASME, J.Basic Engineering, March 1960.

5. Зубов В.Н., К теории аналитического построения регуляторов, Автоматика и телемеханика, т.XXIУ, №8, I963г.

6. Летов А.М., Аналитическое конструирование регуляторов, I-У. Автоматика и телемеханика, т.XXI, №4,5,6, I960г. т.XXП, I96I; т.XXШ, I962.

7. Красовский А.А., Интегральные оценки моментов и синтез линейных систем, Автоматика и телемеханика, т.XXУШ, №I0, I967г.

8. Willis B.H., Brockett R.W., The Frequency Domain Solution of Regulator Problem, IEEE, Trans., AC-10, No.3, 1965.

9. Четаев Н.Г., Устойчивость движения, Гостехиздат, I955г.

I0. Brockett R.W., Mesarovic M.D., Synthesis of Linear Multivariable System, Applic. and Ind., No.62, 1962.

II. Davis M.C., Factoring the Spectral Matrix, IEEE Trans., AC-8, No. 4, 1963.

I2. Youda D.C., On the Factorization of Rational Matrix, IRE Trans. JT-7, No.3, 1961.

I3. Цукерник Л.В., Устойчивость связанной системы автоматического регулирования при внутригрупповой симметрии, Изд-во АН СССР, ОТН, Энергетика и автоматика, №4, I959.

I4. Морозовский В.Т., К теории однотипных связанных систем автоматического регулирования с симметричными связями, Автоматика и телемеханика, т.ХХП, №3, I96I.

I5. Янушевский Р.Т., Об аналитическом конструировании одного класса многосвязных систем, Автоматика и телемеханика, т.ХХУШ, №8, I967.

I6. Мееров М.В., Синтез структур систем автоматического регулирования высокой точности, Изд-во "Наука", I967.

I7. Фельдбаум А.А., Электрические системы автоматического регулирования, Оборонгиз, I957.

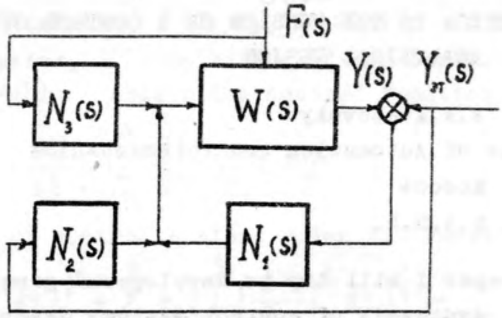


Fig. 1. Block-diagram of a combined optimal system.

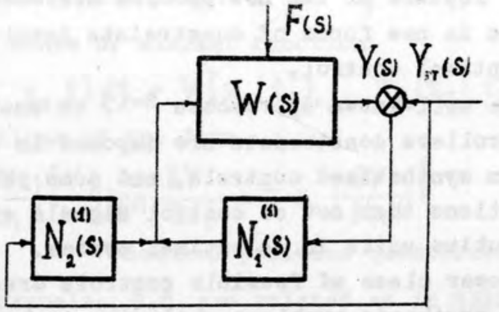


Fig. 2. Block-diagram of an optimal system for where disturbances are impossible to measure.

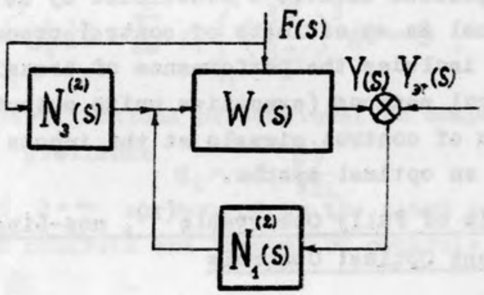


Fig. 3. Block-diagram of an optimal system without a channel for control signals.

68.3

A NEW SOLUTION TO THE PROBLEM OF A CONTROL SYSTEM ANALYTICAL DESIGN

A.A.Krasovsky

Institute of Automation and Telemechanics
Moscow
U.S.S.R.

In this paper I will try to develop and generalise the analytical synthesis of control systems described in articles[1-5] and books [6,7].

The basic feature of the new problem statement of those papers is in new forms of constraints involved in synthesis of optimal control.

Unlike the well-known approaches [8-13] to analytical design of controllers constraints are imposed in this discussion both on synthetized controls, and some phase coordinates functions that act as control signals at the inputs of executive units in an optimal system.

This narrower class of feasible controls drastically simplifies the synthesis problems while the technological sense and practical value of new techniques are preserved.

This paper will also show that the synthesis problem can also be simplified without a constraint by using a special functional as an estimate of control processes. This functional includes the performance of transitory processes and control actions (executive units output) as well as the operation of control signals at the inputs of executive units in an optimal system.

1. Synthesis of Fully Observable[x], non-Linear Plant Optimal Controls

If there is a plant

$$\dot{x}_i + F_i(x_1, x_2, \ldots, x_n, t) = u_i \qquad (1.1)$$
$$(i = 1, 2, \ldots, n)$$

[x] Where all phase coordinates can be measured and used.

and there is a function, $V(x_1, x_2, \dots, x_n, t)$, whose complete derivative, $\dot{V}$ by virtue of non-controlled plant equation ($u_i = 0$) equals the desired function $Q(x_1, x_2, \dots, x_n, t)$

$$\dot{V} = \frac{\partial V}{\partial t} - \sum_{i=1}^{n} \frac{\partial V}{\partial x_i} F_i = -Q \qquad (1.2)$$

then for a set of control actions under the constraints

$$\sum_{i=1}^{n} \frac{1}{K_i} \Big(\int_{t_1}^{t_2} |u_i|^p dt \Big)^{\frac{2}{p}} + \sum_{i=1}^{n} K_i \Big(\int_{t_1}^{t_2} \Big| \frac{\partial V}{\partial x_i} \Big|^q dt \Big)^{\frac{2}{q}} = \qquad (1.3)$$

$$= C\left[x_1(t_1), \dots, x_n(t_1), t_1\right], \qquad \frac{1}{p} + \frac{1}{q} = 1, \ p \geq 1$$

optimal in the sense of minimal function

$$I = \int_{t_1}^{t_2} Q(x_1, \dots, x_n, t)\, dt + V[x_1(t_2), \dots, x_n(t_2), t_2]$$

are control actions of the form

$$u_i = -K_i \Big| \frac{\partial V}{\partial x_i} \Big|^{q-1} \operatorname{Sign} \frac{\partial V}{\partial x_i} = \mp K_i \Big(\frac{\partial V}{\partial x_i} \Big)^{\frac{q}{p}} \qquad (1.4)$$

Here K_i, p, q desired positive quantities; K_i are gains of channels; p, q are related as in Hölder inequality. At $p = q = 2$ eq. (1.3) corresponds to weighed sums of actions at inputs and outputs of executive units for the period $t_2 - t_1$

$$\sum_{i=1}^{n} \frac{1}{K_i} \int_{t_1}^{t_2} u_i^2\, dt + \sum_{i=1}^{n} K_i \int_{t_1}^{t_2} \Big(\frac{\partial V}{\partial x_i} \Big)^2 dt = C \ ,$$

and optimal control actions proportional to components of function V gradients

$$u_i = -K_i \frac{\partial V}{\partial x_i}$$

The values $q = 1$, $p = \infty$ correspond to the given weighed sum of maximal squared controls and "flows" of controls (the quantities $\int_{t_1}^{t_2} \big| \frac{\partial V}{\partial x_i} \big| dt$).

Optimal controls in this case are of relay type.

$$u_i = -K_i \operatorname{Sign} \frac{\partial V}{\partial x_i}$$

At $p = 1$; $q = \infty$ the constraints is a desired weighed sum of squared controls "flows" $\int_{t_1}^{t_2} |u_i|\, dt$ and squared maximal values of control actions. Optimal control actions have a threshold level ($u_i = 0$ .at $\left|\frac{\partial V}{\partial x_i}\right| < 1$).

A sufficiently simple proof of the above theorem is this. A derivative function, V by plant equations $(u_i \neq 0)$ equals

$$\dot{V} = \frac{\partial V}{\partial t} - \sum_{i=1}^{n} \frac{\partial V}{\partial x_i} F_i + \sum_{i=1}^{n} \frac{\partial V}{\partial x_i} u_i = -Q + \sum_{i=1}^{n} \frac{\partial V}{\partial x_i} u_i .$$

Integrating over the interval $t_2 - t_1$ we will have

$$V[x_1(t_2), \ldots, x_n(t_2), t_2] - V[x_1(t_1), \ldots, x_n(t_1), t_1] = \\ = -\int_{t_1}^{t_2} Q\, dt - \sum_{i=1}^{n} \int_{t_1}^{t_2} \left(\frac{\partial V}{\partial x_i}\right)(-u_i)\, dt$$

and

$$I = V[x_1(t_1), \ldots, x_n(t_1), t_1] - \sum_{i=1}^{n} \int_{t_1}^{t_2} \left(\frac{\partial V}{\partial x_i}\right)(-u_i) dt. \qquad (1,5)$$

The minimal I corresponds to the maximum of

$$\sum_{i=1}^{n} \int_{t_1}^{t_2} \left(\frac{\partial V}{\partial x_i}\right)(-u_i)\, dt \leq \sum_{i=1}^{n} \int_{t_1}^{t_2} \left|\frac{\partial V}{\partial x_i}\right| |u_i| dt$$

By Hölder inequality

$$\int_{t_1}^{t_2} \left|\frac{\partial V}{\partial x_1}\right| |u_i| dt \leq \left(\int_{t_1}^{t_2} \left|\frac{\partial V}{\partial x_i}\right|^q dt\right)^{\frac{1}{q}} \cdot \left(\int_{t_1}^{t_2} |u_i|^p dt\right)^{\frac{1}{p}} \qquad (1.6)$$

Therefore $$\sum_{i=1}^{n} \int_{t_1}^{t_2} \left|\frac{\partial V}{\partial x_i}\right| |u_i| dt \leq \sum_{i=1}^{n} \eta_i \zeta_i ,$$

where $$\eta_i = \left(\int_{t_1}^{t_2} \left|\frac{\partial V}{\partial x_i}\right|^q dt\right)^{\frac{1}{q}} , \quad \zeta_i = \left(\int_{t_1}^{t_2} |u_i|^p dt\right)^{\frac{1}{p}} .$$

The equality sign is valid if and only if $|u_i|^p$ is proportional to $\left|\frac{\partial V}{\partial x_i}\right|^q$. Eq. (1.3) with these notations has the form

$$\sum_{i=1}^{n} \frac{\zeta_i^2}{K_i} + K_i \eta_i^2 = C \qquad (1.7)$$

The maximum of the quantity $\sum_{i=1}^{n} \eta_i \zeta_i$ with constraint of eq. (1.7) is reached when $\zeta_i = K_i \eta_i$. This is easily seen when solving an elementary problem for a function extremum.

Relations $|u_i|^p \equiv \left|\frac{\partial V}{\partial x_i}\right|^q$ and $\zeta_i = K_i \eta_i$ i.e.

$$\left(\int_{t_1}^{t_2} |u_i|^p dt\right)^{\frac{1}{p}} = K_i \left(\int_{t_1}^{t_2} \left|\frac{\partial V}{\partial x_i}\right|^q dt\right)^{\frac{1}{q}}$$

are simultaneously satisfied at

$$u_i = -\kappa_i \left|\frac{\partial V}{\partial x_i}\right|^{\frac{q}{p}} \operatorname{Sign}\frac{\partial V}{\partial x_i} = -\kappa_i \left|\frac{\partial V}{\partial x_i}\right|^{q-1} \operatorname{Sign}\frac{\partial V}{\partial x_i}$$

These controls make the right-hand side of relation (1.6) reach the maximal possible value and simultaneously equal to the left-hand side.

Thus controls (1.4) minimize the functional at constraints (1.3).

At p=q=2 the optimality of the above control action is easily proved by Bellman functional equation, i.e. by the normal procedure of optimal control synthesis.

The above control actions are also optimal for the case of no constraint and where a special functional,that describes the action of control signals in an optimal system, is used instead. We will show this by proving that control actions for plant (1.1),optimal in the sense of the minimal functional

$$I=\int_{t_1}^{t_2} Q\,dt + \frac{1}{4}\int_{t_1}^{t_2}\sum_{i=1}^{n}\kappa_i\left(\frac{\partial V}{\partial x_i}\right)^2 dt + \int_{t_1}^{t_2}\sum_{i=1}^{n}\frac{u_i^2}{\kappa_i}dt + V\left[x_1(t_2),..,x_n(t_2),t_2\right] \quad (1.8)$$

are of the form

$$u_i = -\frac{1}{2}\kappa_i\frac{\partial V}{\partial x_i}\ , \quad (1.9)$$

where V is still determined by linear partial derivative eq. (1.2)

Functional (1.8) is the estimate of the transitional processes performance (term $\int_{t_1}^{t_2} Q\,dt$), control actions (term $\sum_{i=1}^{n}\frac{1}{\kappa_i}\int_{t_1}^{t_2}u_i^2 dt$) as well as of the current values of deviations (term $\sum_{i=1}^{n}\kappa_i\int_{t_1}^{t_2}\left(\frac{\partial V}{\partial x_i}\right)^2 dt$).

Let us present U_i in the form

$$u_i = -\frac{1}{2}\kappa_i\frac{\partial V}{\partial x_i} + \delta u_i$$

where δu_i are arbitrary variations of control.

The complete derivative $\dot{V}$ by control plant equation will be

$$\dot{V} = \frac{\partial V}{\partial t} + \sum_{i=1}^{n}\frac{\partial V}{\partial x_i}\left(-F_i - \frac{1}{2}\kappa_i\frac{\partial V}{\partial x_i} + \delta u_i\right) = -Q - \frac{1}{4}\sum_{i=1}^{n}\kappa_i\left(\frac{\partial V}{\partial x_i}\right)^2 +$$

$$+\sum_{i=1}^{n}\frac{\partial V}{\partial x_i}\left(-\frac{1}{4}\kappa_i\frac{\partial V}{\partial x_i} + \delta u_i\right) = -Q - \frac{1}{4}\sum_{i=1}^{n}\kappa_i\left(\frac{\partial V}{\partial x_i}\right)^2 +$$

$$+\frac{1}{2}\sum_{i=1}^{n}\frac{\partial V}{\partial x_i}(u_i+\delta u_i)=-Q-\frac{1}{4}\sum_{i=1}^{n}k_i\left(\frac{\partial V}{\partial x_i}\right)^2-\sum_{i=1}^{n}\frac{u_i^2}{k_i}+\sum_{i=1}^{n}\frac{\delta u_i^2}{k_i}$$

By integrating over the time interval t_2-t_1 and using eq.(1.8) for the functional I we have

$$I=V\left[x_1(t_1),\dots,x_n(t_1),t_1\right]+\sum\frac{1}{k_i}\int_{t_1}^{t_2}\delta u_i^2\,dt$$

Hence follows directly that I is minimal at $\delta u_i=0$ and controls of eq.(1.9) are indeed optimal. There is no other solution.

If V and Q are positive definite functions, V-function is a Lyapunov's function for a non-controlled plant and optimal controls (1.4), (1.9) ensure stability at a non-disturbed system state $(x_1=\dots=x_n=0)$ no matter how high gains of channels k_i would be. This follows from the relation

$$\dot{V}=-Q+\sum_{i=1}^{n}\frac{\partial V}{\partial x_i}u_i=-Q-\sum_{i=1}^{n}k_i\left|\frac{\partial V}{\partial x_i}\right|^{q}.$$

To obtain explicit optimal controls, function V has to be found such that would satisfy eq. (1.2), All advantages over conventional approaches to analytical design follow from the fact that eq.(1.2) is linear and in the known solution a similar equation is non-linear (contains the term $\frac{1}{4}\sum_{i=1}^{n}k_i\left(\frac{\partial V}{\partial x_i}\right)^2$).

It is required to find a solution to eq.(1.2) independent of boundary conditions and termed "a forced solution". That this solution is recommended is due to the fact that, firstly, normally there are no factors to set the boundary conditions for V; secondly, normally transition processes have to be made optimal no matter when they are excited; in other words, the solution has to be independent of the initial time instants. Generally these requirements are met only by a "forced solution."

For a linear plant and a squared functional (Q is the desired quadratic form) the solution must be in the form of quadratic space coordinates.

For factors of this quadratic form there are $\frac{1}{2}n(n+1)$ linear differential equations.[4,7]

In case of a passive, linear or non-linear, plant, the role of the function V can be played by the plant overall power which by the law of energy preservation satisfies eq.(1.2) (Q in this case is a dissipative function).

With a view to finding optimal controls for a more general case, a non-linear plant of eq.(1.1) with analytical functions F_i, let us present these functions and desired function Q as power series

$$F_i = \sum_j a_{ij} x_j + \sum_{j,k} b_{ijk} x_j x_k + \sum_{j,k,l} c_{ijkl} x_j x_k x_l + \dots \qquad (1.10)$$

$$Q = \sum_{i,j} \beta_{ij} x_i x_j + \sum_{i,j,k} \gamma_{ijk} x_i x_j x_k + \sum_{i,j,k,l} \delta_{ijkl} x_i x_j x_k x_l + \dots$$

Here the plant coefficients, $a_{ij}, b_{ijk}, \dots$ and functional $\beta_{ij}, \gamma_{ijk}, \dots$ are generally time functions.

Summation with respect to all indices is made from 1 to n. The coefficie are independent of the order of indices. This is true for indices starting from the second one.(for $a_{ij}, b_{ijk}, \dots$).

Function V will be also determined in the form of a power series

$$V = \sum_{i,j} A_{ij} x_i x_j + \sum_{i,j,k} B_{ijk} x_i x_j x_k + \sum_{i,j,k,l} C_{ijkl} x_i x_j x_k x_l + \dots \qquad (1.11)$$

where the coefficients $A_{ij}, B_{ijk}, C_{ijkl}, \dots$ (of required optimal controls) are likewise generally time functions.

Substituting expressions for $\frac{\partial V}{\partial x_i}$, F_i, Q into eq.(1.2) and equalling the coefficients at similar products of coordinates, we find

$$\dot{A}_{ij} - \sum_{p=1}^{n} (a_{pi} A_{pj} + a_{pj} A_{pi}) = -\beta_{ij}$$
$$(i,j = 1,2,\dots,n)$$

$$\dot{B}_{ijk} - \sum_{p=1}^{n} (a_{pi} B_{pjk} + a_{pj} B_{pik} + a_{pk} B_{pij}) =$$
$$= -\gamma_{ijk} + \sum_{p=1}^{n} (b_{pjk} A_{pi} + b_{pik} A_{pj} + b_{pij} A_{pk}) \qquad (1.12)$$
$$(i,j,k = 1,2,\dots,n)$$

$$\begin{aligned}\dot{C}_{ijk\ell} - \sum_{p=1}^{n} (a_{pi} C_{pjk\ell} + a_{pj} C_{pik\ell} + a_{pk} C_{pij\ell} + a_{p\ell} C_{pijk}) = \\ = -\delta_{ijk\ell} + \sum_{p=1}^{n} (b_{pij} B_{pk\ell} + b_{pik} B_{pj\ell} + b_{pi\ell} B_{pjk} + \\ + b_{pjk} B_{pi\ell} + b_{pj\ell} B_{pik} + b_{pk\ell} B_{pij}) + \\ + \sum_{p=1}^{n} (c_{pjk\ell} A_{pi} + c_{pik\ell} A_{pj} + c_{pij\ell} A_{pk} + c_{pijk} A_{p\ell}) \\ (i, j, k, \ell = 1, 2, \dots, n)\end{aligned}$$

. . . .

For a stationary non-linear plant and a stationary functional (F_i, Q are not explicitly time-dependent), coefficients $A_{ij}, B_{ijk}, C_{ijk\ell}, \dots$ are constant, while linear differential equations (1.12) turn into the algebraic linear equations

$$\begin{gathered}\sum_{p=1}^{n} (a_{pi} A_{pj} + a_{pj} A_{pi}) = \beta_{ij} \\ \sum_{p=1}^{n} (a_{pi} B_{pjk} + a_{pj} B_{pik} + a_{pk} B_{pij}) = \gamma_{ijk} - \sum_{p=1}^{n} (b_{pjk} A_{pi} + b_{pik} A_{pj} + b_{pij} A_{pk})\end{gathered} \tag{1.13}$$

. . . .

Since coefficients $A_{ij}, B_{ijk}, C_{ijk\ell} \dots$ are independent of the order in which indices are written, first group of (1.12) eqs and (1.13), contains $\frac{n(n+1)}{2}$ unknown A_{ij} and equations, second group $\frac{n(n+1)(n+2)}{3!}$ unknown B_{ijk} and equations, etc.

The first group of eqs. (1.12) in a matrix form looks like

$$\dot{A} = Aa + a^* A - \beta \tag{1.14}$$

where $A = \|A_{ij}\|, a = \|a_{ij}\|, \beta = \|\beta_{ij}\|$ are square matrices, a * is a transpose matrix

Then, for a stationary case

$$Aa + a^* A = \beta \tag{1.15}$$

Eqs.(1.14), (1.15) coincide with the known[1-7] equations for linear plant optimal control coefficients.

Assuming the terms $\sum_{j=1}^{n} A_{ij} x_j$ linear in optimal control arguments

$$\frac{\partial V}{\partial x_i} = 2 \sum_{j=1}^{n} A_{ij} x_j + 3 \sum_{j,k=1}^{n} B_{ijk} x_j x_k + 4 \sum_{j,k,\ell=1}^{n} C_{ijk\ell} x_j x_k x_\ell + \dots \tag{1.16}$$

the following result is formulated.

Linear terms of non-linear plant optimal control variables are equal to linear plant optimal control arguments which coincide with the first approximation of the linear plant under consideration.

Thanks to their structure algebraic equations can be solved successively by groups. The first group is solved to yield factors A_{ij} of linear terms in optimal control arguments, then the second group is solved to yield coefficients B_{ijk} of second order terms then coefficients C_{ijkl} of third order terms are found, etc.

For a non-stationary plant or for time-varying factors of the given functional $\beta_{ij}, \gamma_{ijk},...$ a solution to linear differential equations (1.12) is sought. A specific (forced) solution to non-homogeneous linear equations has to be found. This also follows from the fact that when coefficients of the plant and the functional coefficients rate of change go to zero, the solution should tend to a solution to algebraic equations (1.13).

Solutions to differential equations (1.12) and algebraic equations (1.13) have to be found successively by groups. With factors of the plant and the functional changing slowly enough the iteration technique,whereby solutions to respective algebraic equations (1.13) are used as zero approximation, is convenient to find solutions to each group of equations. Though conceptually this procedure of optimal controls determination is clear,to find the factors practically may for a high order plant prove somewhat hard due to cumbersome control systems (1.12), (1.13). Therefore optimal control factors are better expressed as integrated estimates of first approximation equations weights. In Refs[1-7] expressions of the kind were obtained for stationary stable linear plant optimal control synthesis. The entire procedure of finding the optimal controls was thus simplified substantially.

2. Optimal Control Factors Expressed Through Integrated Estimates of the Plant Linear Model Weights.

Assume that the non disturbed state of a non-controlled plant ($u_1 = u_2 = \ldots = u_n = 0$) is asymptotically Lyapunov stable while the plant and the functional are stationary.

If the non-disturbed state is not stable, then a substitution of phase coordinates, e.g. $x_i^* = \exp(-\lambda t) x_i$ where λ is the plant characteristic number, can reduce the plant to a case of a Lyapunov-stable plant. The constraints are, of course, also to be transformed.

Weight functions which correspond to the first approximation equations, or weight functions of the plant linear model, are described by these equations.

$$\dot{W}_i^q + \sum_{p=1}^{n} a_{ip} W_p^q = 0 \quad , \quad W_i^q(0) = \varkappa_{iq} \, , \qquad (2.1)$$
$$(i, q = 1, 2, \ldots, n)$$

Heie W_i^q is the weight function which corresponds to the response of the i-th output of the linear model to δ , a pulse fed to the q-th input, (or to the initial unitary deviation in the q-th output); $\varkappa_{iq}$ is the Kronecker symbol:

$$\varkappa_{iq} = \begin{cases} \text{at } 1 \ i = q \\ \text{at } 0 \ i \neq q \end{cases}$$

From the identities

$$\frac{d}{dt}(W_i^q W_j^z) = \dot{W}_i^q W_j^z + W_i^q \dot{W}_j^z$$
$$\frac{d}{dt}(W_i^q W_j^z W_k^s) = \dot{W}_i^q W_j^z W_k^s + W_i^q \dot{W}_j^z W_k^s + W_i^q W_j^z \dot{W}_k^s$$
$$\frac{d}{dt}(W_i^q W_j^z W_k^s W_l^f) = \dot{W}_i^q W_j^z W_k^s W_l^f + W_i^q \dot{W}_j^z W_k^s W_l^f +$$
$$+ W_i^q W_j^z \dot{W}_k^s W_l^f + W_i^q W_j^z W_k^s \dot{W}_l^f .$$

. . . .

and eq. (2.1) follows that

$$\frac{d}{dt}(W_i^q W_j^z) + \sum_{p=1}^{n} (a_{ip} W_p^q W_j^z + a_{jp} W_i^q W_p^z) = 0$$
$$\frac{d}{dt}(W_i^q W_j^z W_k^s) + \sum_{p=1}^{n} (a_{ip} W_p^q W_j^z W_k^s + a_{jp} W_p^z W_i^q W_k^s + a_{kp} W_p^s W_i^q W_j^z) = 0 \qquad (2.2)$$
$$\frac{d}{dt}(W_i^q W_j^z W_k^s W_l^f) + \sum_{p=1}^{n} (a_{ip} W_p^q W_j^z W_k^s W_l^f + a_{jp} W_p^z W_i^q W_k^s W_l^f +$$

$$+ a_{kp} W_p^s W_i^q W_j^r W_l^f + a_{lp} W_p^f W_i^q W_j^r W_k^s) = 0$$

. . . .

Due to the condition of stability all weights vanish at $t=\infty$. By integrating eq. (2.2) within the range from 0 to ∞ and introducing this notation for integrated estimates of weights

$$J_{ij}^{qr} = \int_0^\infty W_i^q W_j^r dt \ ; \quad J_{ijk}^{qrs} = \int_0^\infty W_i^q W_j^r W_k^s dt \ ;$$
$$J_{ijkl}^{qrsf} = \int_0^\infty W_i^q W_j^r W_k^s W_l^f dt \ ; \ \ldots\ldots\ldots$$

we obtain

$$\sum_{p=1}^{n} (a_{ip} J_{pi}^{qr} + a_{ip} J_{ip}^{qr}) = \varkappa_{iq} \varkappa_{jr}$$
$$\sum_{p=1}^{n} (a_{ip} J_{pik}^{qrs} + a_{jp} J_{pik}^{rqs} + a_{kp} J_{pij}^{sqr}) = \varkappa_{iq} \varkappa_{jr} \varkappa_{ks} \qquad (2.3)$$
$$\sum_{p=1}^{n} (a_{ip} J_{pjkl}^{qrsf} + a_{jp} J_{pikl}^{rqsf} + a_{kp} J_{pijl}^{sprf} + a_{lp} J_{pijk}^{fqrs}) = \varkappa_{iq} \varkappa_{jr} \varkappa_{ks} \varkappa_{lf}$$

. . . .

Comparison of left-hand sides of eqs. (1.13), (2.3) reveals their identity but for the fact that eqs.(2.3) include factors of a plant transpose linear model.

The matrix of a transpose model weights is equal to the matrix of initial linear model weights. Therefore

$$\overset{*}{J}{}_{ij}^{qr} = J_{qr}^{ij} \ ; \quad \overset{*}{J}{}_{ijk}^{qrs} = J_{qrs}^{ijk} \ ; \quad \overset{*}{J}{}_{ijkl}^{qrsf} = J_{qrsf}^{ijkl} \ ; \ \ldots\ldots \qquad (2.4)$$

where asterisks denote integrated estimates of a transposed linear model weights.

By comparing eqs. (2.1) and (2.3) substituting eq.(2.4) into eq.(2.4) we obtain the following expressions for optimal control coefficients.

$$A_{ij} = \sum_{q,r} \beta_{qr} J_{qr}^{ij}$$
$$B_{ijk} = \sum_{q,r,s} \Big[\gamma_{qrs} - \sum_{p} (b_{prs} A_{pq} + b_{pqs} A_{pr} + b_{pqr} A_{ps}) \Big] J_{qrs}^{ijk} \qquad (2.5)$$
$$C_{ijkl} = \sum_{q,r,s,f} \Big[\delta_{qrsf} - \sum_{p} (b_{pqr} B_{psf} + b_{pqs} B_{prf} + b_{pqf} B_{prs} +$$
$$+ b_{prs} B_{pqf} + b_{prf} B_{pqs} + b_{psf} B_{pqr}) -$$

$$-\sum_{p}\left(c_{prs\ell}A_{pq}+c_{pqs\ell}A_{pr}+c_{pqr\ell}A_{ps}+c_{pqrs}A_{p\ell}\right)\Big]J_{qrs\ell}^{ijk\ell}\ .$$

. . . .

The first of these formulas was obtained in Ref.[1] Using these optimal controls coefficient very conveniently found successively on digital computers and analog computers with manual or digital computer operated calculations. To obtain linear terms coefficients of optimal control arguments it is sufficient to find integrated second-order estimates of weights, while to obtain coefficients of second order terms third order integrated estimates have to be found, etc.

Then by eq.(2.5) only multiplications and summations are needed.

In certain cases integrated estimates of weights can actually be found analytically (theoretically this is always possible) and optimal coefficients as a formula.

Integrated estimates of weights of a linear model with fixed coefficients can be used with advantage in finding optimal to controls for a non-stationary linear plant provided that appropriate eqs.(1.12) are solved by the iteration technique (the plant and functional coefficients are assumed to change slowly enough).

Indeed, zero approximations for the desired coefficient are found in the iteration technique by a set of linear controls (1.13) similar to the one used in the stationary case with the difference that coefficients $a_{ij}, b_{ijk}, \dots, \beta_{ij}, \gamma_{ijk}, \dots$ are slowly changing time functions. Thus zero approximations of optimal controls coefficients can at any given time be found by eq.(2.5) where all the plant and functional coefficients are taken for the given time while integrated estimates are taken for fixed values of a_{ij} which correspond to the given time.

Corrections of first approximations by eq.(1.9) will be found by equations similar to eq.(1.10) with zero approximation derivatives in the right-hand side; corrections of the second approximation are similarly expressed by derivatives of the first order approximation, etc. If an upper index is introduced for the ν-th approximation (ν), then by eq.(2.5) these approximations are expressed by the formulas

$$A_{ij}^{(\nu+1)}=\sum_{q,r} \dot{A}_{qr}^{(\nu)} J_{qr}^{ij} \tag{2.6}$$

$$B_{ijk}^{(\nu+1)}=\sum_{q,r,s}\left[\dot{B}_{qrs}^{(\nu)}-\sum_{p}\left(B_{prs}A_{pq}^{(\nu+1)}+B_{pqs}A_{pr}^{(\nu+1)}+B_{pqr}A_{ps}^{(\nu+1)}\right)\right]J_{qrs}^{ijk}$$

. . . .

Here as before the integrated estimates are computed for fixed ("frozen") values of the plant coefficients that correspond to the desired time instant. Time derivatives can be numerically found by values computed for sufficiently close time instants.

Thus, for a non-stationary non-linear plant, programs of optimal control coefficients variations in time can be found through integrated estimates of the first approximation linear model weights functions with "frozen" coefficients as well as by multiplications, summations and numerical differentiations. These operations can be performed on a digital computer, or an analog computer coupled to a digital computer or manual calculations.

References

1. Krasovsky, A.A. Automation and Remote Control No.10, 1967.
2. Krasovsky, A.A. Automation and Remote Control No.12, 1967.
3. Krasovsky A.A. Automation and Remote Control No.1, 1968.
4. Krasovsky A.A. Automation and Remote Control No.2, 1968.
5. Красовский А.А. Аналитическое конструирование систем управления нелинейными пассивными объектами, Известия АН СССР. Техническая кибернетика, I968г.

6. Красовский А.А. Статистическая теория переходных процессов в системах управления, Издательство "Наука", 1968 г.

7. Красовский А.А. Аналитическое конструирование систем управления полетом, Издательство "Машиностроение", 1968 г.

8. Letov A.M. Automation and Remote Control, Nos 4,5, 6, 1966.

9. Letov A.M. A Review Parer to the Second IFAC Congress

10. Зубов В.И. Теория оптимального управления, Издательство "Судостроение", 1966 г.

11. Krasovsky A.A., Letov A.M. Automation and Remote Control, No. 6, 1962.

12. Красовский А.А., Моисеев Н.Н. Теория оптимальных управляемых систем, Известия АН СССР, Техническая кибернетика, №5, 1967 г.

13. Roman Kulikowski. Procesy optymalne i adaptacyjne w ukladach regulacji automatycznej

14. Krasovsky A.A. The Problem of Analytical Design Controllers Extended to the Case of the Given Action of Controls and Control Signals(To be published in Automation and Remote Control in 1969)

68.4

LINEAR AND NONLINEAR SOLUTIONS FOR THE LETOV-KALMAN'S OPTIMUM SYNTHESIS PROBLEMS WITH APPLICATIONS TO LINEAR PLANTS

K. Bela

Institute of Energetics of the Academy of Sciences of the Socialist Republic of Rumania

Bucharest

University of Kraiova

Rumania

Introduction

The present paper considers some approaches to Letov problem of determining optimum feedback control laws for convex integral performance idexes with bounded time interval, including time-optimal problem. The solution is brought to establish the nonlinear invariant structures with nonlinear regulators detached from, as it is shown in Figure 1. It appears then that the inverse optimal control problem, as given by Kalman[2], can be generalized to nonlinear control systems. However, in this subject only introductury steps are mentioned.

First part of the work is devoted to time-optimal systems, for which, according to Boltianskii[3], an algorithm of optimal nonlinear control law synthesis using Letov scheme /Fig 1/ is given. In the second part, one describes a nonlinear control law synthesis procedure for optimal problem with convex integral performance index. With that, according to V.M. Popov[4], some relations between the Riccati equation and the adjoint equations for optimal system are generalized.

1. Synthesis algorithm for time-optimal problem.
   Let us condsider a problem of optimum transition of a linear process

$$\dot{x} = A(t)x + b(t)u \qquad /1.1/$$

from an initial state $x(t_0) = x_0$ to the terminal one $x(t_1) = x_1$

Here $A(t)$ is a real, quadratic $n \times n$ matrix, and $b(t)$ is a n vettor, both absolutely integrable in any finite time interval: $u(t)$ is a scalar, piece-wise continuous function taking its values in a closed control domain containing zero control. For simplicity, the control domain will be

$$|u(t)| \leq 1 \qquad /1.2/$$

It is known, that Pontryagin's maximum principle[5] supplies necessary and sufficient information to solve completely this problem. Indeed a final solution to this problem reduces to integrate a system:

$$\dot{x} = A(t)x + b(t)\,\mathrm{sign}\left[b(t)y(t)\right] \qquad /1.3/$$

$$\dot{y} = -A'(t)y \qquad /1.4/$$

with boundary conditions

$$x(t_0) = x_0 \;, \quad x(t_1) = x_1$$

Number of methods for solving such problem has been presented: these are mainly approximate methods that enable to determine an initial vector $y(t_0) = y_0$, guaranteeing that the optimal trajectory attains the point $x_1 = 0$

Here, a complete algorithum for determining an optimal control law $u = u(x)$ is proposed. It enables to obtain a full information on a topology of state space X, on number of switchings, and on a nonlinear regulator optimising the transient process.

The algorithm starts from the fact, that a controllability domain of the state space is decomposed, according to number of switchings, into manifolds that are necessary for optimal transition from any point x_0 of a given manifold to the terminal point $x_1 = 0$. For all manifolds, one gives equations of the optimal control and optimal trajectory, in initial as well as in transient regime.

First we denote the terminal point as a zero-dimensional manifold

$$v_0 = \{0\} \qquad /1.6/$$

which is attained by all optimal trajectories.

In this problem, the manifold /1.6/ can be attained by two ways only, described by equations

$$\dot{x} = A(t)\,x \pm b(t) \qquad /1.7./$$

with boundary conditions

$$x(\tau_1) = \xi_1 \quad , \quad x(t_1) = 0$$

Solving equations /1.7/ with /1.8/ yields:

$$\xi_1 = \mp \int_{\tau_1}^{t_1} \Phi^{-1}(\tau, \tau_1)\, b(\tau)\, d\tau \qquad /1.9/$$

Where $\Phi(\tau, \tau_1)$ - transition matrix of the homegenous equation, with initial condition $\Phi(\tau_1, \tau_1) = E$ - unit matrix.

On the other hand, solving equation /1.4/

$$y(t) = \Psi(t, \tau_1)\; y(\tau_1) \;\; ; \quad \Psi(\tau_1, \tau_1) = E \qquad /1.10/$$

one gets a linear form with respect to vector $y(\tau_1)$ components:

$$b'(t)\, y(t) = b'(t)\, \Psi(t, \tau_1)\, y(\tau_1) \qquad /1.11/$$

The points $\xi_1 = \xi_1(\tau_1)$, $\tau_1 \leq t_1$ from which it is possible to attain the manifold /1.6/ via the trajectory of equation /1.7/ without switching i. e. the points for which it is possible to find such initial vector $y(\tau_1)$, $-\infty < \tau_1 \leq t_1$, than in a whole interval $[\tau_1, t_1]$ the linear form /1.11/ not vanishes and has a sigh corresponding to that of equation /1.7/ according to equation /1.3./, generate a one - dimensional manifold $V_1 = V_1^{+} \cup V_1^{-}$:

$$V_1^{\pm} = \left\{ \xi_1 \middle| \xi_1(\tau_1) = \mp \int_{\tau_1}^{t_1} \Phi^{-1}(\tau, \tau_1)\, b(\tau)\, d\tau \, ; \;\; b'(t)\, y(t) \gtrless 0 \quad \forall t \in [\tau_1, t_1] \right\} \qquad /1.12/$$

In this definition, upper signs correspond to the branch V_1^+ , and lower - to V_1^-

Since the manifold V_1 depends upon one parameter τ_1 , it represents geometrically one-dimensional set, i.e. a curve in n-dimensional space. It can be immediately seen that V_0 belongs to V_1 as a point for $\tau_1 = t_1$.

Let us suppose that $x(t_0) = x_0 \in V_1$, i.e. that one can find such $\tau_1 = t_0$ that $\xi_1 = x_0$, upper or lower sign in formal definition /1.12/ being preserved. Then the optimal trajectory takes a form

$$x^*(t) = \Phi(t,t_0)\left[x_0 \pm \int_{t_0}^{t} \Phi^{-1}(\tau,t_0)\, b(\tau)\, d\tau\right] \qquad /1.13/$$

Let us also suppose to have found a vector $y(t_0)$ that satisfies appropiate conditions in /1.12/ and generates a solution to the adjoint system

$$y^*(t) = \Psi(t_1\ t_0)\, y^*(t_0) \qquad /1.14/$$

From equations /1.13/ and /1.14/, taking into account the known relation between the transition matrices of the adjoint systems

$$\Phi(t,t_0)\,\Psi'(t,t_0) \quad = E \qquad /1.15/$$

one can find a relation

$$y^* = y^*(x^*(t),t) \qquad /1.16/$$

which leads to ordinary relay-type systems with additional nonlinearities. As it will be seen later such relations, and corresponding optimal control structures, may be easily found for autonomous systems to be optimized.

To determine vector $y(t_0)$, we may assume that at $t_0 = \vartheta_1$ a switching occurs, leading the system's state x to the manifold V_1.

The manifold V_1 can be attained, with one switching, only from the manifold V_2. Denoting by V_2^+ and V_2^- two branches of this manifold, that are continuouly extended after switching to V_1^+ and V_1^- respectively, we may establish the following formal definition

$$V_2 = V_2^+ \cup V_2^- ; \quad V_2^\pm = \Big\{ \xi_2 \,\Big|\, \xi_2 = \pm \int_{\tau_2}^{\vartheta_1} \Phi^{-1}(\tau,\tau_2) b(\tau)\, d\tau + \qquad /1.17/$$
$$+ \Phi^{-1}(\vartheta_1,\tau_2)\xi_1^\pm ; \xi_1^\pm \in V_1^\pm ; b'(\vartheta_1)y(\vartheta_1) = 0 ; \tau_2 \leq \vartheta_1 \leq t_1 ; b'(t)y(t) \gtrless 0 \ \forall t \in [\tau_2, \vartheta_1) \Big\}$$

The definition shows that the manifold V_2 represents a surface in n-dimensional state space, which is generated by two parameters: τ_2, ϑ_1. Letting $\tau_2 = \vartheta_1$ one obtains a curve V_1 in the manifold V_2.

Repeating the same procedure, all manifolds up to V_{n1} are obtained:

$$V_{n1} = V_{n1}^+ \cup V_{n1}^- ; \ V_{n1}^\pm = \Big\{ \xi_n^\pm \,\Big|\, \xi_n^\pm = \pm(-1)^n \int_{\tau_n}^{\vartheta_{n-1}} \Phi^{-1}(\tau,\tau_n) b(\tau)\, d\tau + \Phi^{-1}(\vartheta_{n-1},\tau_n)\xi_{n-1}^\pm ,$$
$$\xi_{n-1}^\pm \in V_{n-1}^\pm ; b'(\vartheta_{n-1})y(\vartheta_{n-1}) = 0 , \tau_n \leq \vartheta_{n-1} \leq \vartheta_{n-2} ; (-1)^n b'(t)y(t) \gtrless 0 \ \forall t \in [\tau_n, \vartheta_{n-1}) \Big\} \qquad /1.18/$$

the latter representing a n-dimensional set in the state space, determined by the parameters $\vartheta_1, \vartheta_2, \ldots, \vartheta_{n-1}, \tau_n$.

It is clear, that for $\tau_n = \vartheta_{n-1}$ the manifold V_{n-1} is set off in the manifold V_{n1}, and for $\vartheta_{n-1} = \vartheta_{n-2}$ the manifold, V_{n-2} is set off in V_{n-2}, and so forth.

We'll show now how one can make use of the above manifolds. Let for instance $x_o \in V_{n1}$. Then $\tau_n = t_o$ and hence $\xi_n = x_o$. If $\vartheta_{n-1}, \vartheta_{n-2}, \ldots, \vartheta_1$ are known, then the points of the optimum trajectory transition from each manifold to the next one, are uniquely determined in order of decreasing dimensionality $\xi_{n-1}, \xi_{n-2}, \ldots \xi_1$. On the last segment of the trajectory, t_1 is determined from $x^*(t_1) = 0$.

In order to determine switching time instants $\vartheta_{n-1}, \vartheta_{n-2}, \ldots. \vartheta_1$, the following n equations are set together in a system :

$$\xi_n^\pm(\tau_n) = \pm(-1)^n \int_{t_o}^{\vartheta_{n-1}} \Phi^{-1}(\tau,t_o) b(\tau)\, d\tau \pm (-1)^{n-1} \Phi(\vartheta_{n-1},t_o) \int_{\vartheta_{n-1}}^{\vartheta_{n-2}} \Phi^{-1}(\tau,\vartheta_{n-1}) \cdot \qquad (1.19)$$
$$\cdot b(\tau)\, d\tau \pm \ldots \mp \Phi(\vartheta_{n-1},t_o)\Phi(\vartheta_{n-2},\vartheta_{n-1}) \ldots \Phi(\vartheta_1,\vartheta_2) \int_{\vartheta_1}^{t_1} \Phi^{-1}(\tau,\vartheta_1) b(\tau)\, d\tau = x_o$$

which is obtained directly from the first equations appearing in the formal definitions of V_{n1}, V_{n-1}, ... V_1, . after the vectors $\xi_{n-1} \ldots. \xi_1$, entering there linearly, have been eliminated from.

A solution to the equation's system /1.19/ has to satisfy the following conditions:

$$a/ \quad t_o \leqslant \vartheta_{n-1} \leqslant \vartheta_{n-2} \leqslant \dots\dots\dots \leqslant \vartheta_1 \leqslant t_1 \qquad /1.20/$$

that result from a correct order of transitions of the optimal trajectory $x^*(t)$, $t \in [t_o, t_1]$, from $x_o \in V_{n1}$ through V_{n1} , V_{n2} ,........ V_1 to $x_1 = 0$.

Now a vector $y^*(t_o)$ orientation can be determined from switching conditions:

$$b'(\theta_i)\, \Psi(\vartheta_i, t_o)\, y^*(t_o) = 0 \; ; \quad i = 1, 2, \dots, n-1 \qquad /1.21/$$

which represent a system of n - 1 homogeneous equations with n unknown components of the vector $y^*(t_o)$. Next, we calculate:

$$y^*(t) = \bar{\Psi}(t, t_o)\, y^*(t_o) \qquad /1.22/$$

and verify if:

b/ there is no other solution to $b'(t)\, y^*(t) = 0$ except the one of equation /1.21/.

c/ a sign of the linear form $b'(t)\, y^*(t)$ changes in a correct manner at the switching instants.

If all above conditions are satisfied with a strict suitableness of the manifold branches sequence, then the optimum control is found, and the synthesis of the optimal feedback control law can be at once accomplished. Moreover, relations of type of /1.16/ can be established along with a nonlinear structure of the optimal controller.

Equations /1.21/ remain valid also in the case when $x_o \in V_j$, $j \leqslant n-1$: then, $t_o = \vartheta_{n-1} = \vartheta_{n-2} = \dots = \vartheta_j$, $1 \leqslant j \leqslant n-1$ is a multiple root of the equation /1.19/, and the left-hand sides of the first n - j equations in /1.21/ reduce to the time-derivatives at $t = t_o$ of a corresponding linear form.

If conditions a) b) c) are satisfied in the whole controllability domain $\mathbf{X}_u = \{ x_o \mid \forall x_o \exists u \in [-1,1] \;\; x_o \to x_1$ in finite time $\}$ then the manifold is extended to the whole controllability domain: $V_{n1} = X_u$. In particular, for autonomous system (A and b - constant), such an assertion.

reduces to the theorem of n intervals[8]: then, if all eigenvalues of matrix A possess negative real parts, the controllability domain is identical with the whole space X.

Nevertheless, in number of cases, the domain V_{n1} occupies some part only of the controllability domain X_u.
In remaining part $X_u \setminus V_{n1}$, the above conditions are not satisfied. Indeed, the equation:

$$b'(t)y^*(t) = 0 \qquad /1.23/$$

in the condition b/ possesses also the roots different from those obtained via equation /1.19/. In this case, one continues to decompose the domain X_u into manifolds V_{n1}, V_{n2}, V_{n3}, V_{n4}, ... representing n dimensional sets of initial states $x_0 \in X_u$ for which the optimum control contains n - 1, n, n + 1, n + 2, ... switchings. Such manifolds are determined one after another by succesive integrations of equations /1.3/ and /1.4/, taking into account the condition /1.5/ and the number of switchings.

All manifolds mentionned above are then utilized.

For V_{n2}, as an example, we have:

$$V_{n2} = V_{n2}^{+} \cup V_{n2}^{-} \ ; \ V_{n2}^{+} = \Big\{ \xi_{n2} \,\Big|\, \xi_{n2}(\tau_{n2}) = \pm(-1)^{n+1} \int_{\tau_{n2}}^{\vartheta_{n1}} \Phi^{-1}(\tau, \tau_{n2})\, b(\tau)\, d\tau + \Phi^{-1}(\vartheta_{n1}, \tau_{n2})\, \xi_{n1}^{\pm},$$

$$\xi_{n1}^{\pm} \in V_{n1}^{\pm} ;\ b'(\vartheta_{n1})\, y(\vartheta_{n1}) = 0 ,\ \tau_{n2} \le \vartheta_{n1} \le \vartheta_{n-1} ;\ (-1)^{n+1} b'(t)\, y(t) \lessgtr 0,\ \forall t \in [\tau_{n2}, \vartheta_{n1}) \Big\} \qquad /1.24/$$

From the relation = x_0, a system of n equations is obtained, white from equation /1.21/ along with an equation for $_{n1}$ resulting from /1.24/, after the vector y t_0 has been eliminated from, we get:

$$\det\left[\Psi'(\vartheta_1, t_0)\, b(\vartheta_1) \ \ \Psi'(\vartheta_2, t_0)\, b(\vartheta_2) \ldots \Psi'(\vartheta_{n1}, t_0)\, b(\vartheta_{n1}) \right] = 0 \qquad /1.25/$$

So, a n + 1 - sth equation for determining of switching instants t_1, ϑ_1, ϑ_2, ..., ϑ_{n-1}, ϑ_{n1}, is obtained.

If a solution to this system satisfies conditions a) b) c) extended to V_{n2}, then indeed $x_0 \in V_{n2}$, and the optimum synthesis problem is completely solved. Else, one switching more is introduced and the computations are repeated assuming that $x_0 \in V_{n3}$

This procedure continues until conditions a) b) c) are completely satisfied.

Successive manifolds V_{n1}, V_{n2}, V_{n3}, ... are separated by hypersurfaces of n - 1 - sth dimension, that are obtained via a point-transformation of manifold V_{n-1} during system equations integration. Such hypersurfaces are, except V_{n-1}, the cages of the second kind [3].

Between the positive and the negative subsets of the manifolds V_{n1}, V_{n2}, V_{n3}, ..., the hypersurfaces of the singular trajectories are placed, that form with the switching surfaces the angles equal to zero.

We'll establish a logic scheme of the described algorithm, assuming the matrices A and b are constant.
It is known, that for every real constant quadratic matrix A there exists a quadratic nonsingular matrix T such, that [9]

$$A = T^{-1} KT, \quad K = \begin{bmatrix} \ddots & & & \\ & \delta_i & \omega_i & \\ & -\omega_i & \delta_i & \\ & & & \ddots \delta_j \ddots \end{bmatrix}, \quad \begin{matrix} i = 1,2,\ldots,m \\ j = 2m+1,\ldots,n \end{matrix}$$

where δ_i, ω_i - real and imaginary part of the complex conjugate eigenvalues of the matrix A, δ_j - real eigenvalues of A. Since the matrix T represents a nonsingular affine transformation of the state space $X = \{x\}$ into some space $Z = \{T\,x\}$ having the same qualitative structure as X, but in altered form , we'll consider that such a transformation has been performed, and that A is in canonical form A = K.

In order to simplify the algorithm, we'll distinguish two parts: 1) $x_0 \in V_{n1}$, considering that V_{n1} contains all manifolds V_k, k = 0 , 1,... n - 1. 2) $x_0 \notin V_{n1}$.
If all eigenvalues of the matrix A are real, then, according to the theorem of n-intervals, the part 1 exhausts all possibilities, since $V_{n1} = X_u$

Computational procedure for the part 1 consists of equations:

$$\pm\left[e^{-Kt_1} - 2e^{-K\theta_1} + 2e^{-K\theta_2} - \ldots + (-1)^{n-1} 2e^{-K\theta_{n-1}} + (-1)^n E\right]K^{-1}b = x_0 \qquad /1.27/$$

which result from equations /1.19/ for an autonomous system and describe completely the switching instants ϑ_{n-1}, $\vartheta_{n-2}, \dots, \vartheta_1, t_1, (t_0 = 0)$. The procedure consists also of a formula giving at once the vector y(o) for $y_1(0) = 1$:

$$y(0) = \left[\left(e^{-K\vartheta_{n-1}}b \quad e^{-K\vartheta_{n-2}}b \dots e^{-K\vartheta_1}b \quad I_l\right)'\right]^{-1} I_l \qquad /1.28/$$

where I_1 - vector of zero elements except 1-th element equal to 1. If two or more numbers $\vartheta_{n-1}, \vartheta_{n-2}, \vartheta_{n-3}, \dots$ are equal one to another, then the corresponding columns of the transposed matrix in formula /1.28/ are multiplied by $K^0 = E$, K^1, K^2 etc. Taking into account /1.28/, one can immediately determine a vector

$$y^*(t) = \left\{\left[e^{K(t-\vartheta_{n-1})}b \quad e^{K(t-\vartheta_{n-2})}b \dots e^{K(t-\vartheta_1)}b \quad e^{Kt} I_l\right]'\right\}^{-1} I_l \qquad /1.29/$$

To verify the results, one examines if the conditions:

a) $0 \leq \vartheta_{n-1} \leq \vartheta_{n-2} \leq \dots \leq \vartheta_1 \leq t_1$ all real

b) $b'y^*(t) \neq 0, \ \forall t \neq \vartheta_k, \ t \in [0, t_1], \ k = 1, 2, \dots n$

c) $$\left.\begin{cases} (-1)^{k-1} b'y^*(t) > 0, & \forall x_0 \in V_k^+, \ t \in (\vartheta_k, \vartheta_{k-1}) \\ (-1)^{k-1} b'y^*(t) < 0 & \forall x_0 \in V_k^-, \ t \in (\vartheta_k, \vartheta_{k-1}) \end{cases}\right\} \begin{matrix} \vartheta_0 = t_1, \\ \vartheta_n = t_0 = 0 \end{matrix} \qquad /1.30/$$

are satisfied. If it is the case, then the switching points ξ_k of the optimal trajectory $x^*(t)$ are determined by the formula:

$$\xi_k^{\pm} = \pm(-1)^k \left[E - 2e^{-K\vartheta_{k-1}} + 2e^{-K\vartheta_{k-2}} - \dots + (-1)^{k-1} 2e^{-K\vartheta_1} + (-1)^k 2e^{-Kt_1}\right] K^{-1} b \qquad /1.31/$$

$$k = 1, 2, 3, \dots, n$$

The optimal trajectory in each interval $[\vartheta_k, \vartheta_{k-1}]$ takes a form:

$$x^*(t) = e^{K(t-\vartheta_k)}\left[\xi_k \pm (-1)^{k-1} K^{-1} b\right] \mp (-1)^{k-1} K^{-1} b \qquad /1.32/$$

and the adjoint vector is given by

$$y^*(t) = e^{-K'(t-\vartheta_k)} \eta_k \qquad /1.33/$$

when it is denoted $\eta_k = y^*(\vartheta_k)$. The equations /1.32/ can be solved with respect to n unknown elements of the matrix $e^{K(t-\vartheta_k)}$ as functions of $x^*(t)$: the solution is unique due to the linearity of equations /1.32/. Next, from matrix $e^{K(t-\vartheta_k)}$, a matrix $e^{-K'(t-\vartheta_k)}$ can be obtained. It follows that the equation /1.33/ takes a form:

$$y^* = y^*(x) \qquad /1.34/$$

which is unique. For instance, in the case of real eigenvalues of the matrix A, i.e. for $m = 0$ in the canonical matrix /1.26/

$$y_j^* = \frac{\xi_k^j \pm (-1)^{k-1} \frac{b^j}{\delta_j}}{x^{j*} \pm (-1)^{k-1} \frac{b^j}{\delta_j}} \eta_{kj}, \quad x^* \in V_k \qquad \begin{matrix} j = 1,2,\dots,n \\ k = 1,2,\dots,n \end{matrix} \qquad /1.35/$$

where it is denoted $x = [x^j]$, $y = [y_j]$, $b = [b^j]$, $j = 1,2,\dots n$. Now, we can construct a linear form:

$$b'y^* = \sum_{j=1}^{n} b^j y_j^*(x^{j*}) \qquad /1.36/$$

which closes the loop of the time-optimal control giving a system with two nonlinearities /fig. 2/. Such a structure is distinct from other known schemas[10] by the fact, that the digital device doesn't appear in the control loop.

This is the end of the first part of algorithm. The second one is necessary only in a case when the matrix A possesses complex eigenvalues and the conditions /1.30/ in the first part of algorithm are not satisfied.

For the second part, the equations /1.27/ are replaced by the equations:

$$\pm\left[e^{-Kt_1} - 2e^{-K\vartheta_1} + \dots + (-1)^{n+N-1} 2e^{-K\vartheta_{n+N-1}} + (-1)^{n+N} E\right] K^{-1} b = x_0 \qquad /1.37/$$

involving n + N unknowns $\vartheta_{n+N-1}, \vartheta_{n+N-2}, \dots, \vartheta_1$, $\vartheta_0 = t_1$. To determine completely these unknowns form n equations /1.37/, N equations analogous to /1.25/ are taken:

$$\det \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & e^{-\delta_i\vartheta_k}\cos\omega_i\vartheta_k & e^{-\delta_i\vartheta_k}\sin\omega_i\vartheta_k & \dots & e^{-\delta_j\vartheta_k} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & e^{-\delta_i\vartheta_{n-1+q}}\cos\omega_i\vartheta_{n+q} & e^{-\delta_i\vartheta_{n-1+q}}\sin\omega_i\vartheta_{n+1+q} & & e^{-\delta_j\vartheta_{n-1+q}} & \end{bmatrix} = 0 \qquad (1.38)$$

Formulas /1.28/ and /1.29/ remain valid for the second part. The verification of results is accomplished with the aid of conditions similar to /1.30/, extended to all n + N roots of equations /1.37/ and /1.38/.

By repeated use of the algorithm for N = 1,2, ... , we can find such a number N, for which all above conditons are satisfied.

In this way, a number of switchings is found, and the procedure is terminated by the formulas /1.31/ - /1.34/.

The described algorithm is represented by a flow-diagram in Figure 3.

All above reasoning for the scalar control can be easily extended to the case of vector-valued control for linear plants, and to several classes of nonlinear plants.

2. Nonlinear synthesis of the optimal system for integral convex performance indexes.

We'll consider a problem of control u (x) optimization for a linear plant

$$\dot{x} = A(t)x + B(t)u \qquad /2,1/$$

where A(t) and B(t) - are real, continuous n x n and n x m matrices respectively, defined in a known time interval $[t_0 , t_1]$, An initial state is given:

$$x(t_0) = x_0 \qquad /2,2/$$

A m-dimensional control u steers the plant from x_0 to a terminal state

$$x(t_1) = x_1 \qquad /2,3/$$

terminal time t_1 being considered fixed.

The performance index is

$$I(u) = \frac{1}{2}\left[f(x(t_1)) + \int_{t_0}^{t_1}(x'G(t)x + u'H(t)u)\,dt\right] \qquad /2,4/$$

where $f(x_1(t_1))$ - a given continuous function, G (t) and H (t) continuous real quadratic symmetric matrices n x n and m x m respectively, in the interval $[t_0 , t_1]$.
Moreover it is assumed that G (t) - positive - semidefinite, H (t) - positive--definite matrix,

We'll say the control u (t) is admissible, if it guarantees the plant's transition from the state x_0 to the state x_1 with bounded value of I (u): we'll say the control is optimal, when the functional I (u) takes its minimal /or maximal/ value.

In the work[11] some fundamental theorems on the existence and uniqueness of the optimal control are given. In particular, it is shown that the unique solution to the system of equations:

$$\dot{x} = A(t)\,x - B(t)\,H^{-1}(t)B(t)\,y$$
$$\dot{y} = -G(t)x - A(t)\,y \qquad /2,5/$$

with boundary conditions /2.2/ and

$$y(t_1) = \mathrm{grad}\ f(x(t_1)) \qquad /2,6/$$

determines the optimal trajectory x (t) and corresponding optimal control

$$u^*(t) = -H^{-1}(t)\,B'(t)\,y^*(t) \qquad /2,7/$$

For several particular forms of f(x) the optimal control is given by

$$u^* = -H^{-1}(t)\,B'(t)\,R(t)\,x^* \qquad /2.8/$$

where R (t) – is a solution to the Riccati matrix equation:

$$\dot{R}(t) + R(t)B(t)H^{-1}(t)B'(t)R(t) + R(t)A'(t) + A(t)R(t) + G(t) = 0 \qquad /2.9/$$

with a boundary condition depending upon the function f (x) For example, with $f(x) = x'Fx$, where F – quadratic positive – semidefinite matrix, the boundary condition takes a form [11,12]

$$R(t_1) = F \qquad /2,10/$$

If the terminal state $x(t_1)$ has to be placed in a given smooth convex set

$$g_s(x_1) = 0, \quad s = 1,2,\dots p: \ p \leqslant n \qquad /2.11/$$

then the boundary conditions are determined by the formula[13,14]

$$y(t_1) = -\lambda\,\mathrm{grad}\,f(x_1) - \sum_{s=1}^{p} \mu_s\,\mathrm{grad}\,g_s(x_1) \qquad /2.12/$$

where λ and μ_s – nonnegative numbers with one of them equal to 1. E.g. for $\lambda = 1$, equations /2.11/ and /2.12/

completely determine the vector $y(t_1)$ and the multipliers s.

In particular, if we set $f = 0$, and if the equations /2.11/ are linear

$$x^s(t_1) = 0: \quad S = 1,2,\dots,p: \quad p \leq n$$

then we obtain

$$E_1 x_1 = 0$$
$$E_2 y_1 = 0 \qquad /2.13/$$

where E_1 - a diagonal matrix with p first elements equal to 1, and the n-p remaining - to zero, $E_2 = E - E_1$.

For $E_1 = 0$ the terminal state /2,3/ is free and the optimal control is obtained in a form of /2.8/, the Riccati equation /2.9/ being integrated with the condition /2.10/, where $F = 0$. For Riccati equation integrating, a method of successive approximation [15] can be applied. We'll show that the matrix R (t) can be obtained by integrating the adjoint linear equations system.

Let $\Phi(t)$ and Ψ (t) denote the transition matrices for equation /2.5/, i.e.

$$\dot{\Phi} = A(t)\Phi - B(t)H^{-1}(t)B'(t)\Psi$$
$$\dot{\Psi} = -G(t)\Phi - A'(t)\Psi \qquad /2.14/$$

A general solution to such a system takes a form

$$\Phi(t) = \Phi_1(t)C_1 + \Phi_2(t)C_2$$
$$\Psi(t) = \Psi_1(t)C_1 + \Psi_2(t)C_2 \qquad /2.15/$$

where $\Phi_i(t)$, $\Psi_i(t)$, $i = 1,2$ - are linearly independent particular solutions to the equations /2.14/, and C_1, C_2 are real quadratic matrices of integration constans.

A solution to system /2,5/ can be represented now as:

$$x = \Phi_1(t)D_1 + \Phi_2(t)D_2 \qquad /2.16/$$
$$y = \Psi_1(t)D_1 + \Psi_2(t)D_2$$

where D_1, D_2 - vectors which can be uniquely determined from given boundary conditions, like e.g. /2.2/ and /2.13/

Let us consider a matrix

$$R(t) = \Psi(t)\Phi^{-1}(t) \qquad /2.17/$$

It can be easily verified, that if $\Phi(t)$ and $\Psi(t)$ represent a solution to system /2.14/, then R(t) is determined by the riccati equation /2.9/, the boundary conditions for R (t) being completely specified by those for $\Phi(t)$ and $\Psi(t)$. An inverse transformation leads to the equation:

$$\Phi^{-1}[\dot{\Phi} - A\Phi + BH^{-1}B'\Psi] + \Psi^{-1}[\dot{\Psi} + G\Phi + A'\Psi] = 0 \qquad /2.18/$$

which is equivalent to system /2.14/ if the matrices $\Phi(t)$ and $\Psi(t)$ are nonsingular $\forall\, t \in [t_0, t_1]$. Indeed, the matrix R (t) remains unchanged if the matrices $\Phi(t)$ and $\Psi(t)$ are multiplied by an arbitrary, nonsingular constant matrix C. Equation /2.18/ remains valid in this case. It follows that the matrices $\Phi(t)$ and $\Psi(t)$ represent a solution to the system of linear equations /2.14/, both being determined precisely up to an arbitrary nonsingular, constant matrix C. In this way, we can replace the matrix C_1 in formula /2.15/ by, for example, the unit matrix E. Then the matrix C_2 is completely determined by the boundary condition for the matrix R (t) , in particular by /2.10/.

Description of the matrix R (t) with the aid of the adjoint linear equations system /2.14/ enables to determine R (t) as a function of x^*, y^* and time t in non-autonomous case . We present here a method for establishing such a function in autonomous case [17,18].

Let Q be a quadratic nonsingular matrix, which relies the homogenous system's matrix /2.14/ to a quasi-diagonal matrix

$$\begin{bmatrix} K & 0 \\ 0 & K' \end{bmatrix} = Q \begin{bmatrix} A & -BH^{-1}B' \\ -G & -A' \end{bmatrix} Q^{-1}$$

where K is in the form of /1.26/. It is known[19], that such a matrix is composed of eigenvectors of the system /2.14/ matrix. Now, a particular solutions $\Phi_i(t)$, $\Psi_i(t)$ to the system have a form:

$$\begin{aligned} \Phi_1(t) &= Q_{11}\, e^{K(t-t_0)} & \qquad \Phi_2(t) &= Q_{12}\, e^{-K'(t-t_0)} \\ \Psi_1(t) &= Q_{21}\, e^{K(t-t_0)} & \qquad \Psi_2(t) &= Q_{22}\, e^{-K'(t-t_0)} \end{aligned} \qquad /2.19/$$

where Q - quadratic submatrices of the matrix Q.

$$Q = \left[\begin{matrix} \overbrace{Q_{11}}^{n} & \overbrace{Q_{12}}^{n} \\ Q_{21} & Q_{22} \end{matrix}\right] \begin{matrix} \}n \\ \}n \end{matrix} \qquad /2.20/$$

Then the matrices $e^{K(t-t_0)}$ and $e^{-K'(t-t_0)}$ are uniquely determined from /2.16/, as the functions of x , y and initial conditions x_0:

$$e^{K(t-t_0)} = L(x^*, y^*, x_0), \quad e^{-K'(t-t_0)} = L'^{-1}(x^*, y^*, x_0) \qquad /2.21/$$

Next, a relation

$$LL^{-1} = E$$

enables to determine the vector y^* as a function of x^* , x_0 and t_1:

$$y = y^*(x^*_; x_0, x_1) \qquad /2.22/$$

The above representation is unique if one requires that the terminal conditions are satisfied. The relation /2.22/ along with the formula /2.7/ define a nonlinear invariant control law, for the problem considered, represented by a bloc diagram shown in Figure 4.

It is clear that if the optimal synthesis result can be reduced to the linear control law /2,8/ then the matrix R(t) is represented as a function of x and x_0.

The above results can be easily extended to a more general case of the performance index, when a time derivative of the control appears under integration operator [18]. Then, a nonlinear controller with an integrating device is obtained, as shown in Fig. 5.

In particular, for $t_1 = \infty$, the nonlinear controllers become linear and are identical to those obtained by Letov and Kalman method.

References

1. A.M. Letov Analytic contructing of regulators, pt. II, III, Artomatika i Telemekhanika, 4, 5, 6, 1960, 4, 1961, 11, 1962.
2. R.E. Kalman, When is a Linear Control System Optimal?, J. of Basic Engineering, Transanctions of the ASME, Seriss D, March, 1964.
3. V.G. Boltianskii Mathematical Methods of Optimal Controll /in Russian, book/, Fizmatgiz, Moscow, 1966.
4. V.M. Popov, Stabilitates sistemelor automate cu parametri variabili in timp, Studii si cercetări de energetică si electrotehnică, 2,1965
5. L.S. Pontryagin et al "Mathematical Theory of Optimal Processes" J. Wiley 1962
6. L.W. Neustadt, Synthesing Tine Optimal Control Systems, Journal of Math. Analysis and Applications, 1, 1960.
7. J.H. Eaton, An Iterative Solution to Time Optimal Gontrol, Journal of Math. Analysis and Applications, 5, 1962.
8. A.A. Feldbaum "Optimal Control Systems" Academic Press 1965
9. F.R. Gantmacher "The Theory of Matrices" vols I, II, Chelsea, New York 1959.
10. T.A. Hawkes, Systemes de commande en temps minimal, Sutomatisme, 4, 1966.
11. E.B. Lee, L. Markus, Foundations of Optimal Control Theory, John Wiley, and Sons, Inc., New-York, London, Sydney, 1967.
12. R.E. Kalman, Contributions to the Theory of Optimal Control. Bol. Soc. Mat. Mexicana, 5, 102-119, 1960.
13. L.I. Rozonoer " Pontryagin's Maximum Principle in the Optimum Systems Theory", Avtomatika i Telemekhanika 10, 11, 12, 1959
14. J.T. Tou, Modern Control Theory, McGraw-Bill Book Company, New-York, 1964.

15. V.I. Zubor "Concerning the theory of the analytic constructing of regulators" Avtomatika i Telemekhanika, 8, 1963

16. V.F. Krotor "Approximate synthesis of the optimum control. Avtomatika i Telemekhanika, 11, 1964.

17. V. Răsvan, O sinteză optimală pe baza unui criteriu integral patratic, Studii si cercetări de energetică si electrotehnică, 2, 1969. In Rumanian

18. C. Belea, Asupra unor proprietăti ale sistemelor adjuncte si utilizării lor in sinteza optimală, Lucrările Institutului de energetică al Academiei R.S.R,- Bucuresti 1968. In Rumanian

19. L.S. Pontryagin "Ordinary D.Herential Equations" /in Russian, book/. Izd. Nauka, Moscow, 1965.

20. R.S. Bucy, Global Theory of the Riccati Equation, Journal of Computer and System Sciences, 4, 1967.

69.7

SELF-ORGANIZATION OF AN EXTREMAL CONTROL SYSTEM

Ivakhnenko A.G., Khrushcheva N.V., Neskhodovsky V.I.

Institute of Cybernetics

Kiev, USSR

> "By the difficulties of solution and implications for science and practice the attack on the problem of self-organization is comparable to that on the mistery of the atom nucleus. In the same way as the first half of the twentieth century will go down into history as the era of fundamental discoveries in nuclear physics, the second half will hopefully see a solution to self organization, the central problem of cybernetics.
> (A.Ya.Lerner. Foreword to "Principles of self-organization", Moscow, 1966, in Russian)

Definitions of "self-learning" and "self-organization"

The term "self-organization" implies the process of spontaneous improvement of organization, i.e. decrease in the entropy of a system made by a number of interrelated elements under the action of the environment or its own positive feedbacks. A narrower concept is "self-learning" or "adaptation" which usually implies just gradual change in adjustible parameters of the system (factors of equations). Self-organization differs in the kind of actions and in that the structure can be changed.

An example of a self-learning algorithm of a pattern-recognition system can be self-division of the set of input images into compact groups[1,2]. A human teacher has only to inform the system on the identification of the pattern of each group accepted in the human society, otherwise, though it will recognize images, the system will have to identify the patterns by itself, and these identifications can coincide with those accepted by mere conjecture.

The first experiments on self-learning of recognition systems involved the action of positive feedbacks (the system inputs rather than outputs). The process of prototypes learning is then akin to bumping an unstable body. Without human interference the system selects the possible classifications of images into patterns with almost equal probability[3]. Also, inputs and outputs of a system can be used in combination for self-learning[2].

Self-organization as well as self-learning are both affected by external actions or feedbacks. However, whereas in self-learning one can always trace the lines of actions transmission (inputs and outputs) this is nearly imposeible in self-organization since the actions affect a set of uniform, in a certain sense, elements of the system. A human designer is to select non-linear characteristics of these integral actions (e.q. non-linearity of income tax) and transmit certain "elementary algorithms" of the action to particles of the system (e.g. companies).

The well-known difficulties inherent in complex systems are not encountered in self-organizing systems with their "integral actions" and "elementary algorithms". As example we can mention padking by vibration when all parts are acted upon simultaneously.

Below we will describe self-organization of an extremal control system, an engineering problem. The process of self-organization leads to an ordered positioning of a recognizing system prototype set ("pole gas") each of which has its own "elementary algorithm" and is controlled by "integral actions".

A combined System of Extremal Control with the Recognising System As Corrector. Constraints and Field of Application

We have spoken above on self-learning of recognising systems be cause the subject of this paper is self-organization of a combined extremal control system which consists of an open-loop (OL) and its corrector (C); we propose to use the recognising system for the latter purpose (Fig. 1).

Let us find the constraints on the problem. We deal with both unimodal and multimodal extremal characteristics (hills) which satisfy the following requirement: the optimal characteristic of the control plant ("a set of desired states") in the region of operating conditions is a sufficiently smooth line 0_2^1 0_2^{II} which can be approximated by a linear-piecewise function (Fig.2). It is assumed that multidimensional problems are reducible, by a technique termed divergence, to one-dimensional problems solved in the space of three variables, and where, is a generalized disturbing action, is a generalized control action, is a generalized quality factor (that would incorporate both the extremum factor and the values of disturbances, i.e., for instance, in a simplest case we will have for a linear extremal characteristic

$$\varphi = \tau_0 + \tau_1 \mu + \tau_2 \lambda + \tau_3 \mu^2 + \tau_4 \lambda^2 + \tau_5 \mu \cdot \lambda$$

and the optimal characteristic 0_2^I 0_2^{II} is determined by the expression $\frac{\partial \varphi}{\partial \mu} = 0$ or $\mu = K_0 + K_1 \lambda$, where $K_0 = -\frac{\tau_1}{2\tau_3}$ $K_1 = -\frac{\tau_5}{2\tau_3}$

Let us require that in all points of that characteristic $\Psi=1$. Excluding μ we will have (along the line 0_2^I e_2^{II})

$$\varphi = \beta_0 + \beta_1 \lambda + \beta_2 \lambda^2,$$

where

$$\beta_0 = \tau_0 - \frac{\tau_1^2}{4\tau_3}, \quad \beta_1 = \tau_2 - \frac{\tau_1 \cdot \tau_5}{2\tau_3}, \quad \beta_2 = \tau_4 - \frac{\tau_5^2}{4\tau_3}.$$

Evidently if this requirement is to be valid we have to use the converter

$$\Psi = \frac{\varphi}{\beta_0 + \beta_1 \lambda + \beta_2 \lambda^2} = f(\varphi, \lambda)$$

With such conversion we will have in each point of the hill and at its summit $\Psi = 1$. To be more exact

$$\Psi = \frac{\tau_0 + \tau_1 \mu + \tau_2 \lambda + \tau_3 \mu^2 + \tau_4 \lambda^2 + \tau_5 \mu \lambda}{\beta_0 + \beta_1 \lambda + \beta_2 \lambda^2} = 1 - (-K_0 - K_1 \lambda + \mu)^2.$$

Practically such a converter contains tables (charts) which denote changes in requirements to the magnitude of the extremum index as a function of the range of the quantity. E.G. at one kind of one we can way "good enough", i.e. $\Psi = 1$ if iron content in waste will be 10%. For another kind, , the same evaluation will be given if $\varphi = 8\%$ etc.

Adaptation processes are caused by immeasurable additive disturbances which lead to drift and turn of the extremum hill. In our example

$$\psi = 1-\{-[K_0+N_0(t)]-[K_1+N_1(t)]\lambda+\mu\}^2,$$

where $N_0(t)$, $N_1(t)$ - are slowly changing disturbances (shift and turn).

To eliminate the effect of transient processes the quality factor sensor should have a certain averaging or the inertia of the plant should be compensated by precompensators[5] which does not lead to much trouble in measuring circuits.

Consequently we believe that the quantities, ψ, μ, λ are measurable and it is sufficient to evaluate in a two-point system, e.g.

$1 \geq \psi \geq 0,8$ - "good enough"

$\psi < 0,8$ - "needs regulation"

The experience, however limited, accumulated in comparison of various adaptive extremal control systems with the system suggested here shows that it is in this very frequent case that this system is competitive. If the extremum factor can be measured precisely enough, then the advizability of self-organization and recognition systems is doubtful. In many complex situations we can only say that "everything is alright" or "there's something wrong". In other words we differ two quality levels. The system described is recommended in such cases.

The Open-Loop Part of the System and "Tethers"

The open-loop part of the system is a functional converter with a characteristic which is easily shifted, turn and even change. At optimal adjustment the characteristic of the open-loop part should correspond to optimal adjustment of the plant. In the above example we require a linear relation

$$\mu = d_0 + d_1\lambda,$$

where $d_0 = K_0 + N_0(t)$, $d_1 = K_1 + N_1(t)$.

In more complicated cases the characteristic of the plant is expressed by the polinomial

$$\mu = [K_0+N_0(t)]+[K_1+N_1(t)]\cdot\lambda+[K_2+N_2(t)]\cdot\lambda^2+\dots+[K_n+N_n(t)]\cdot\lambda^n,$$

Then the characteristic of the open loop part will be

$$\mu = d_0 + d_1\lambda + d_2\lambda^2 + \dots + d_n\lambda^n.$$

Self-learning (adaptation) consists in making the OL part characteristic coefficients "follow" changes in coefficients of the optimal response

i.e. so that

Such an adaptation is possible in the principle only when at least two points on the surface of the extremum hill which is an even function can be measured. Ref. [6] has shown that at even characteristic search is essential at the plant. The system described here is nevertheless termed searchless because instead of periodic changes in control actions as a function of time small "peaks" employed by way of search steps are super imposed on the OL response[6]. The peaks amplitude and distribution is selected according to the shape of the hill and distribution of tethers so as to ensure maximal speed of the system. Peaks are unnecessary if there is no corrector.

In practice the open-loop part is made by a key matrix where keys open as a function of the position of the "meaningful bit" in the unitary code of the generalized disturbance as well as on the number of the flip-flop which responds in ring pulse counters[7]. Another structure of a controlled functional converter is a logic circuit with threshold elements [8].

A Recognising Corrector (First Version)

A recognising system is known to be a logic device which in order to classify input signals (generally called "images") into classes (or patterns) compares certain measures of these signals proximity to prototype (or reference) signals which are formed in the system during learning.

We will describe below the actions of two versions of recognising systems which are used to correct the OL characteristic special attention is directed to the prototype set self-organization processes. In the first version of the corrector input signals (features) are made by the coordinates of the representing point which corresponds to the "state" of the plant at a given instant of time. Fig.3. shows that the plane can be divided into three areas or "situations":[x)]

I. Regulate, decrease μ.

[x)] The terms "state" and "situation" are analogous to "image" and "pattern" used in recognition of graphical images.

II. Good enough, steady.
III. Regulate, increase μ.

A recognising system is a very flexible and easy to adjust model of the plant. In case the optimal response is a direct line a recognising system with just three point prototypes is sufficient

$$\alpha_1(\mu_1, \lambda_1),\ \alpha_2(\mu_2, \lambda_2),\ \alpha_3(\mu_3, \lambda_3)$$

The system computers the measure of proximity between the input signal and the prototypes, e.g. in this way

$$\Sigma_1 = |\mu-\mu_1| + |\lambda-\lambda_1|,\ \Sigma_2 = |\mu-\mu_2| + |\lambda-\lambda_2|,\ \Sigma_3 = |\mu-\mu_3| + |\lambda-\lambda_3|.$$

Then the comparator selects the minimal of the three situations and thus indicates the situation. For instance the representing point of Fig.3 will be associated with situation 1. With the situation known it is easy to indicate in what direction should be changed to reach situation II which is the control objective.

A recognising system indicates a situation correctly only if the boundaries of the "attraction area" of the poles $\Sigma_1 = \Sigma_2$ and $\Sigma_2 = \Sigma_3$ coincide with the boundaries of the situation II (shown in Fig.3 as dotted lines). When the boundaries are direct lines it is sufficient to make the poles "normal" as shown in Fig.3b. The process of pole poritioning and maintaining the position when the "extremum hill" changes and shifts is precisely the process of the recognising system self-organization.

Weighted shift Three-Pole Self-Organization

A possible three pole self-organization is shown in Fig.3. If the process is successful the pole must be in the so-called "normal position of Fig.3b.

First two extreme poles are positioned in the left-hand upper and right-hand lower angles of the plane and the system is left to itself and extermal disturbances. A spontaneous random process of the poles self-positioning sets on. The feedback sensor gives only one of two commands "regulate" or "good enough" steady". The recognising system has three prototypes and therefore three outputs: "regulate, decrease μ "; "Good enough, steady" and "regulate, increase μ ". There must be either an in consistency or a consistency between the commands of the sensor and the recognising system.

In the first case, one of the extreme poles which is the closest to the representigg point (and this is the effect of the open-loop part characteristic peak on selection of pole turn) makes a step in its direction (Fig.3a and 3c), while the second extreme pole moves in parallel in the reverse direction. The step decreases with the distance by the exponential smoothing law and recommendations of stochastic approximation. The middle pole is maintained by exponential smoothing in the "gravity center" (or the middle point) of the good operation state set.

In the second case the extreme poles do not move. The system is allowed to control (i.e. to correct the position or open-loop characteristic shape) only after a long enough consistency between the outputs of the sensor and the recognising system (Fig. 3b).

The prototype learning method suggested by us was later termed "weighted shift technique"[9]. A salient feature of this technique is that two prototypes (poles) respond to the representing point; the "correct" (as indicated by the teacher or outputs of the system itself) moves toward the representing point and the "incorrect one moves away from it (Fig. 4).

In such a technique a multidimensional stochastic approximation of all coordinates of the poles.

$$K_N = K_{N-1} + \gamma_N (y - K^T_{N-1}) X,$$

where

$$y = h^T X\ , \quad K = \begin{bmatrix} K_0 \\ K_1 \\ \vdots \\ K_n \end{bmatrix} \quad , \quad X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \ , \quad \gamma_N = const$$

is replaced by one-dimensional approximation

$$\tau_{i(K+1)} = \tau_{iK} + \gamma_K (x_{iK} - \tau_{iK}),$$

because the direction of motion has been found unambiguously.

Theorem on "Pole Gas Self-Organization" Processes Stability

A set of prototypes can be conveniently regarded as a certain "pole gas" whose particles interact between themselves and with the environment. Elementary interaction algorithms are selected by a human operator. They can be chosen so as to make the process converge. Below we will formulate a theorem for pole gas self-organization process stability.

For self-organization of the pole gas a number of algorithms can be suggested; the theorem formulation will change accordingly. As example we will recall one of the known formulation. This is valid for a prototype (pole gas) set from a system whose action is explained in Fig. 5.

The poles of the recognising system are either rigidly fixed (those that the region of possible operating conditions) or are rigidly connected to vertices of tethers or are free. Both in the "upper" and "lower" sets of poles separate poles interact with each other, with the rigidly fixed poles and with poles whose coordinates coincide with those of opper (for the "lower set of lower) tethers of the open-loop part characteristic.

The interaction law for two adjacent particles of the gas is their "elementary algorithm". The sign of interaction changes for the opposite with respect to the pole, whose coordinates coincide with the representing point when the two-point evaluation sensor indicates that regulation is required. In case where the representing point lies in the region of the plant good operation the interaction force has the same sign as most of the particles.

The position of each particle depends on a sum of effects of elementary algorithms for all its adjacent particles from the same set. If there is a change in just the sign of just one elementary algorithm causes a shift of the free particle and consequently of the open-loop part response middle line. The poles shift in careful steps and the characteristic at any speed.

Theorem. In order that under the action of the pole gas particles motion the open-loop response be maintained within the system good operation zone it is sufficient that:

1) elementary algorithms correspond to mutual repulsion of particles of the pole gas and attraction with respect to the pole in the representing point which has been outside the plant good operation zone;

2) the distance between the representing point and the boundaries of that zone should not exceed the height of peaks;

3) the probability of each disturbance should be non-zero.[x)]

[x)] The third condition is met practically always because the teeth are adjusted in the operating range of the characteristic, or in other words, correspond to the values of actual disturbances. To facilitate the self-organization process the peaks should correspond to most frequent values of disturbances.

When the pole gas is programmed on computers the mutual attraction of particles is conveniently replaced by the requirement that equal distances should be maintained between them and thus the position of each pole can be computed, The process of each particle positioning is very much akin to that process of successive stochastic approximation which takes place in analog-to-digital converters, in self-learning sensors[4] and other devices which use successive information accumulation in time.

Thus in this case we have realization of stochastic approximation algorithms[10]. The proof of the theorem does not differ from that of stochastic approximation process convergence[11].

Recognising Corrector (Second Version)

In the second version of the corrector the features are made by several recent values of the qualiry criterion . Measured in different vertices of the open-loop response tethers. The minimal number of vertices to be considered is three; however when there is a dispersion in the sensor readings that number should be increased to ten to fifteen.

Then as distinct from the preceding discussion the "state" is the mutual position of the plant optimal response and the OL part response. Each "state" of the system has a corresponding code of the quality index values; some examples are shown in Table 1. A set of feasible states can be divided into, e.g. six basic "situations" shown in the same table. In the recognising corrector six codes of prototypes (references) have been introduced by means of learning(shows). As a measure of proximity the system computers scalar products of input signals and codes of prototypes; the result is easily seen to be independent of what vertices were hit by the representing point recently.

Table 1 shows examples of one-to-one correspondence between input states and the prototypes recorded in the recognising system. All other states of the plant will be associated with one or several of these prototypes by the system. The operational algorithm will be explained by the following example.

Let, e.g. the state of the system be described by the following code (three "current" vertices are involved):

We find five scalar products of the input signal by five first prototypes

$$\sum_1 = 0 -1\ 00 +1 +1\ 000000 = +1$$

$$\sum_2 = 0 +1\ 00 -1 -1\ 000000 = -1$$

$$\sum_3 = 0 +1\ 00 +1 -1\ 000000 = +1$$

$$\sum_4 = 0 -1\ 00 +1 -1\ 000000 = -1$$

$$\sum_5 = 0 +1\ 00 +1 -1\ 000000 = +1$$

The system makes a decision according to the following rule: if this situation occurs completely or with other situations if there is no constituent which corresponds to the given situation.

In this example the following decision will be made: the OL response is shifted upwards ($\sum_1 = +1$) and turned counter clockwise ($\sum_3 = +1$) but in the operating condition region there are no large deviations ($\sum_5 = +1$). The appropriate command is sent to the unit which corrects the position of the OL response. The correction process is accomplished slowly until there is a change in input signals of the recognising system. A code made exclusively of signals "-1" means that the OL response has gone far from the region of good operation. The amplitude of all peaks will increase until the position of the optimal response has been found. In case of contradictory recommendations (e.g., turn the response clockwise and counter clockwise) the system awaits additional data until the conflict settles.

Self-Organization Processes Simulation

So far simulation of self-organization processes such as self-settling of pole gas particles in normal operation have been completed for simple cases. Self-settlement of three poles has also been simulated[4, 11,12]. By simple graphical construction the position of poles was found after each "operating cycle". Disturbances were imitated by random numbers tables or by random numbers generator with a specified probability distribution (usually even). The plant characteristic was approximated by a second or third order polinomial. The quality index was quantized for two levels. At the start of the process the extreme poles are made as far apart from the open-loop part characteristic as possible.

During the "operation" after each step of poles the middle line of the open-loop part was determined as the locus of points equidistant from two extreme poles. Simulation showed that the process converges: after a certain number of steps the open-loop part characteristic was found inside the good operation zone.

A similar simulation for nine poles (which made it possible to approximate non-linear boundaries of good operation by three stretches of a direct line) was accomplished by T.Gergei[12]. Computations were performed at various initial positions of poles.

Self-organization of pole gas was simulated physically. Floats with magnetized cores inside floated freely on the surface of water. As the "operation" proceeded some of them (which at this cycle found themselves closer to the representing point than the others) were fixed rigidly. The line of fixed floats described the open-loop characteristic with increasing precision as time passed. Due to the mutual repulsion the relative position of other "poles" changed at each cycle.

Despite the fact that (as follows from the above discussion) simulation of self-organising processes was not made in sufficient detail, the investigations completed is conclusive as far as applicability of the pole interaction algorithms for a division of the state space into situations that would make possible control over the position of OL response middle line.

Application

A combined system of extremum control intended for control over iron or dressing has been made in the Institute of Cybernetics Academy of Sciences of the Ukrainian Soviet Socialist Republic. The wet magnetic dressing unit is a multidimensional (three disturbances and three controls, one quality index) plant with an extremal response. Control is made by the minimum quality index. The first stage will be operation in the "advisory capacity"[14].

The system is digital, actions are quantized for three four levels. The quality index is quantized for two levels. In the open-loop part of the system convergence and divergence matrices are employed.

The corrector consists of a feature selector and a recognising system ("Alpha" type). The latter system classifies all possible plant states into five situations. A separate characteristic $\mu = f(\lambda)$ in the open-loop part corresponds each situation,

The state is characterized by a set of values of in three tethers of the open-loop part characteristic. Information on the current coordinates of the plant (disturbances and the quality index) is coded by the feature selector.

The system is made of semi-conducting modules of a "MIR" computer. Tests have shown that the system is operable. We hope to make the system self-learning (at present the recognising system is trained in prototypes of situations before the actual operation) and make the structure of the open-loop part more flexible.

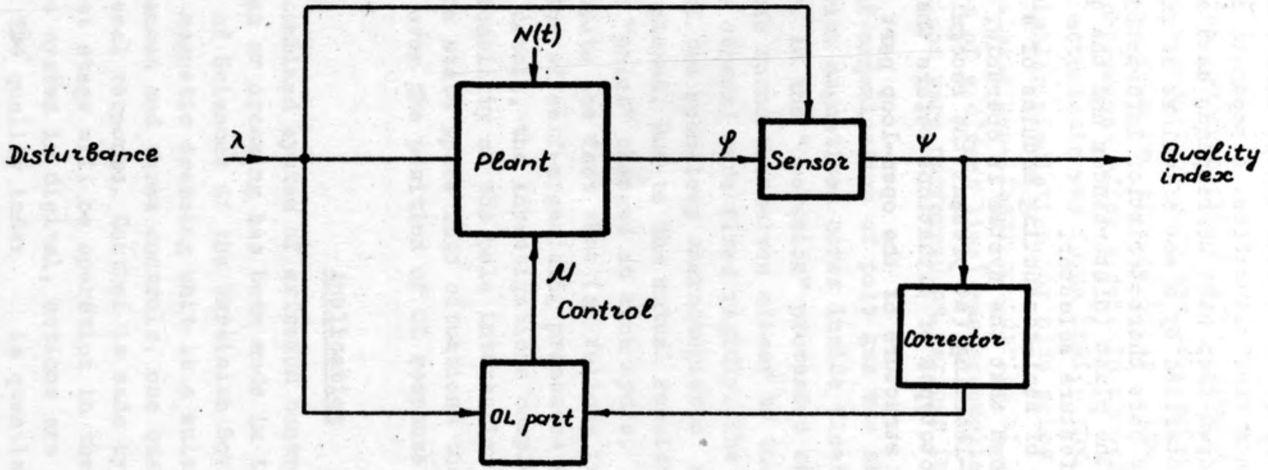


Fig. 1. Schematics of a combined extremal control system with two possible values of the quality factor.

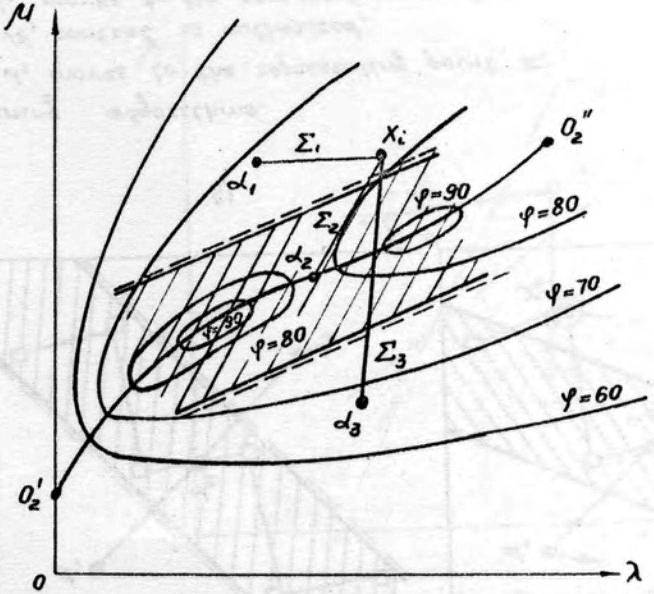


Fig. 2. An example of a bimodal problem with smooth optimal response $O_2' O_2''$ of an extremal control plant, x_i is the representing point, $\alpha_1, \alpha_2, \alpha_3$ are the prototypes (of a pole).

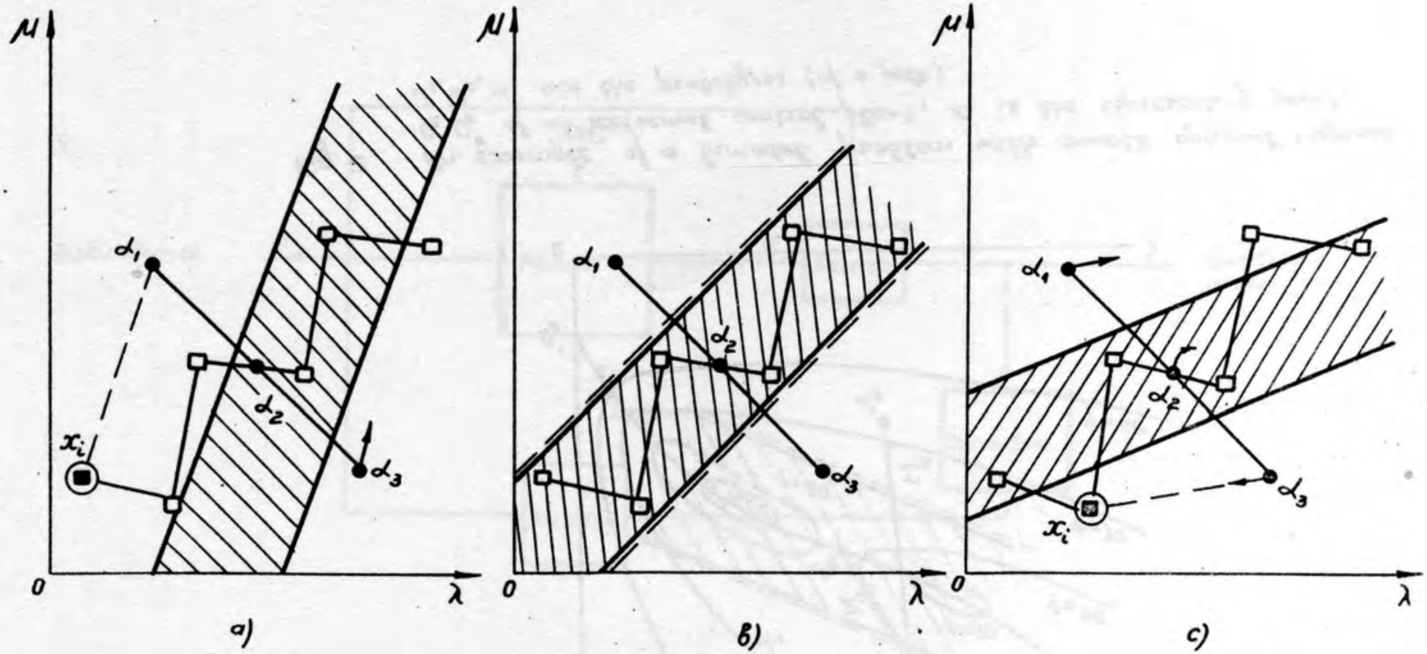


Fig. 3. Pole self-learning algorithms.

a) The first pole α_1 moves to the representing point x_i;

б) Poles do not move, control is authorized;

c) The third pole α_3 moves to the representing point x_i.

The situation of a „good enough" operation is shaded.

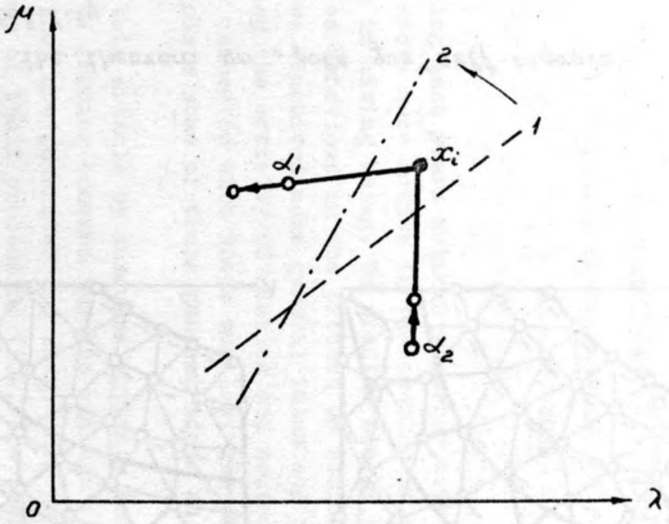


Fig. 4. Weighted shift method (a multi-dimensional approximation is reduced to a one-dimensional approximation)

1. Initial position of a boundary between situations.
2. Position of the boundary after the poles hawe moved to positions shown as dotted line.

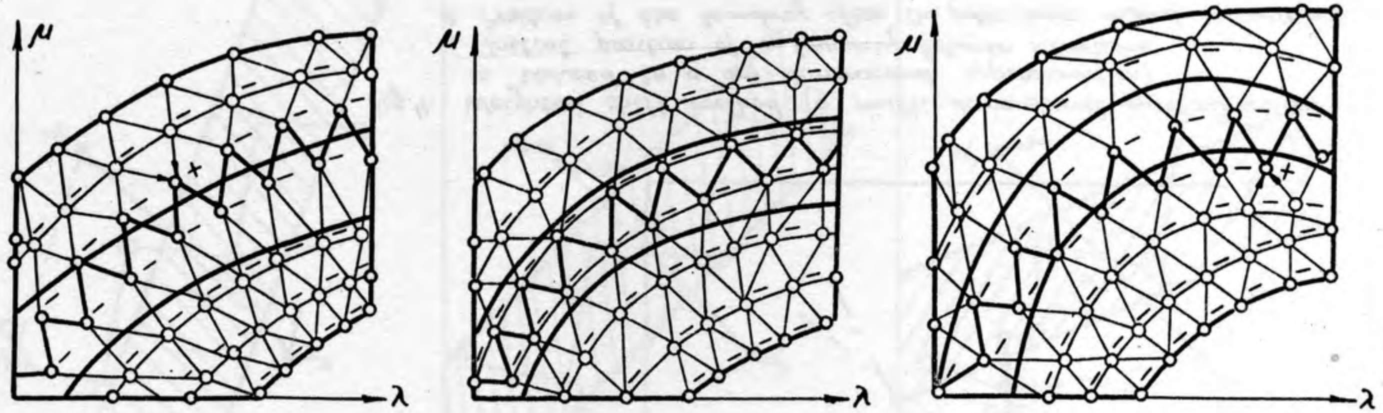


Fig. 5. Formulation of the theorem on „pole gas" self-organization process stability.

70.2

SIMPLEST SEARCH MECHANISM FOR MUSCLE ACTIVITY CONTROL

M.A. AISERMAN, E.A. ANDREEVA

Institute of Automatics and Telemechanics

Moscow, USSR

During last years authors together with a small group of collaborators were active in the problem of control of muscle activity in living being. We studied single muscle's and group's of muscles behaviour in the process of search activity and step by step we conceived some general ideas and model representations describing the activity of muscles and neuronal organizations directly connected with these muscles. The aim of this report is to explain some of these general ideas and model representations.

In this article by muscular activity control we understand the law of forming command impulse volleys (i.e. under what conditions they are generated and where they are directed) and the techniques of their processing (i.e. the nature and arrangement of muscular response to those pulses) with a view to achieving a certain objective necessary for the organism.

Even the limited research conducted by the group are sufficient to see that the muscles that solve the search and movement tasks are controlled by different mechanisms under different conditions. In our work we made an attempt to isolate one of these mechanisms and to investigate it under condition when it works separately from other mechanisms. The mechanism under study was named simplest search mechanism (SSM for short). This mechanism, is used by the organism in cases when certain tensions or joint angles are to be strictly maintained.

§1. A review of experimental results upon which the concepts of the simplest search mechanism (SSM) are based.

Facts on which our ideas about SSM are based came from our own experiments which yielded the possibility to observe muscle activity both in artificial conditions organized by means of external feedback loop which helped us to observe in what way the brain solves the search problem and in natural conditions. In all experiments we recorded both electromyogram (EMG) signals coming from surface electrodes and the so-called enveloping EMG which came as a result of EMG filtering in a detector, a lag with transfer function $\frac{K}{T_p+1}$ and a catode repeater.

1°. Experiments on maintaining minimum pain stimulus depending on muscle tension

In these experiments [1] the electric pain stimulus was artificially organized in the external feedback circuit in such a way that it depended only on the enveloping EMG of one or two muscles.

There was only one value of the enveloping EMG when the pain stimulus was minimal (in some cases this value was equal to zero). Accordingly in the cases of one muscle there exists on the plane "value of enveloping-time" (fig. 1) a horizontal line corresponding to the minimal pain stimulus.

Fig. 1 shows the experimental enveloping curve progress. It decreases till it reaches some horizontal line under the line corresponding to the minimum of the pain stimulus. After reaching this horizontal line it jumps up only to begin to fall down again. These jumps of enveloping EMG are called further splashes. The experiment shows that the amplitudes of splashes are random but there exists two typical amplitudes of splashes - a great one and a small one. If we delete the righthand side of a parabolic pain stimulus characteristic curve the process will be the same but in this case only small splashes remain. If the position of the line of minimal pain stimulus were suddenly changed (points a and b of fig. 2) the process would be the same but the position of the line of the splashes starts' varies remaining always under the line of the pain stimulus minimum and parallel to it. As a result the general electric activity of the muscles (and the muscle tension aswell) always oscillates near the value corresponding to the minimum of pain stimulus searching and following this value.

This search process remains the same both in case when the minimal pain stimulus line is displaced continuously and when a change of some other factors during the process takes place. In case when the pain depends on two or more muscles the process is the same and the splashes of these muscles begin simultaneously.

In all cases it was observed that splashes start each time when the rising pain stimulus reaches some level and they don't start if this level were reached during pain stimulus decrease.

It was shown that simultaneously with the splashes of the muscles under consideration from which the pain stimulus depends there appeared splashes of other "outside" muscles but the moments of those splashes starts' are not regular. That means that they do not follow the lines of corresponding pain-stimulus values.

2^o. Experiments on joint angle maintaining and on oscillating it with highest possible frequency

In this experiments [3] we recorded EMGs and their envelopes for a pair of antagonistic muscles of wrist or elbow joint and values of joint angle. The subject we experimented with saw the light spot of the oscilloscope screen which showed him his joint angle value and the gain was chosen so large that a change of joint angle of about 2 - 3 angle minutes threw the light spot off the screen. The aim of the subject was to keep the point on the screen. In cases when he could achieve this goal experiments showed (fig. 3) that joint angle oscillates at a frequency about 7 - 10 cycles/sec. On the enveloping EMGs of both antagonistic muscles typical splashes can be clearly distinguished and they are of the same kind as those observed in experiments with the pain stimulus (see above).

In case of joint angle experiments these splashes start each time when absolute value of angular velocity of the increasing joint angle reaches some level. Every time splashes of both antagonistic muscles start simultaneously but splashes of one of the muscles are always large and those of the second one are small. Large splashes occur by turns that is if the splash of one muscle at some moment is large that means that the next splash of this muscle would be small and the large one would be at its antagonist. Each time that muscle of the pair would have a large splash which stretches at that moment. As the frequency of the joint angle is 7-9 cycles/ sec and only one large splash of the muscle occurs in a period it is clear that the typical time interval between two large splashes is about 0,1 second.

If now we would ask the subject not to maintain the joint angle during this experiment but to oscillate the angle with highest possible frequency and with any convenient for him but visible amplitude the character of curves recorded would be practically the same as on fig. 3. But in this case the amplitude of the splashes would considerable increase with the result that the frequency will decrease a little and a change will be observed of the level of angular velocity absolute value at which the splashes occur. The picture repeates itself in experiment with different joint angle values and with some other changes in experiments' conditions.

Comparison of fig. 2 and 3 gives many reason to believe that muscles' activity control in these two different cases is organized by means of the same mechanism, that the splashes of the muscles are the main thing in this mechanism and that the mechanism acts organizing and changing the moments of the splashes' starts. These moments occur in case some value which is in some way a main value for experiments' conditions when increasing comes to some level. In the experiment with external loop this was pain stimulus,in the join experiment it was absolute value of angular velocity. Further we'll describe the model representation which shows in what way this mechanism can be realized. We are assured that the above described experiments represent examples of the work of such a mechanism.

§2. Simplest search mechanism (SSM)

The central parts of the SSM are a random interneuronal pool(RIP) and neuronal organization which we call function discomfort (FD for short). The aim of FD is to find necessary moments when volleys of impulses must be sent to RIP. The aim of RIP is to realize a splash of muscle activity when the volley of impulses comes to the muscle.

1°. Random interneuronal pool (RIP)

From our point of view the main parameter which may characterize the state of a muscle in every moment is the instant value N_{α} of the number of fired α-motoneurons. In spite of that basic role will be ascribed farther to random interneuronal network whereas α-motoneurons will be understood only as "output relays" of that interneuronal network (of RIP). We suppose that RIP has two inputs: the main input and the background input. The theory of RIP was developed by L.I. Rosonoer [4] who theoretically studied in what way number N_{u} of excited interneurons varied in time. Since the number of α-motoneurons is sufficiently high and the leads to them from interneurons are random, the number N_{α} is approximately proportional to N_{u}.

The main question is in what way RIP responds to the volley of impulses when it comes to one of the inputs. L.I. Rosonoer has shown in his works that during period when the volley of impulses excites the input the number N_{u} rises. After that N_{u} decreases not immediately but step by step (not necessarily monotonously) till it reaches the average value which corresponds to the autonomous behaviour of RIP and then it oscillates near that value (fig. 4). This reaction of RIP to a short volley of impulses is called farther splash of RIP. It is easy to understand that amplitude of splash and its duration are random. The splash of RIP causes a splash of muscle activity which we can see on enveloping EMG.

For SSM the splash is the simplest form of muscle activity, an "atom" of it. The work of the SSM depends on the moment, when the impulses volleys arouse the splashes and it is practically independent on the number of impulses in a volley and the parameters of impulses.

If we discuss the input impulses which are applied to the main input of RIP then the additional impulses coming to the background input are equal in their action to decreasing of the average level of the RIP thresholds' interneurons. As a result the number N_{u} increases and that means amplitude of splash increasing. In other words impulses which come to the main input cause the splash and impulses which come to the background input control the amplitude of the splash.

RIP can only arouse splashes when impulses come to main input. Moments when the splashes start depend fully on the moments when impulses are recieved while these moments depend on special neuronal organization which we call function-discomfort.

2°. Function-discomfort and the splash law

We were interested here only in the muscle activity. This article deals only with such muscular activity during which the brain coordinates and changes muscle tension so as to achieve a certain objective specified in advance. If the deviation from the objective reaches a certain threshold we learn about it. It is natural to suggest that in cases of the kind there is "something", which depends on receptor signals and when this "something" achieves a certain threshold that sensation of deviating from the objective arises. That something we denote as "function-discomfort", but we assume that along with the threshold just mentioned there is another, much lower threshold, the achievement of which does not yet make the sensation of deviation from the objective, but makes it imperative to use SSM to determine the times when command volleys are generated (see below).

The function-discomfort determines the moments when impulses are sent to the main input of some muscle's RIP. The general law by which these moments are determined and changed will be termed the law of command volley generation or because it causes a splash of N_{ω} the splash law: a volley of impulses arises every time when FD increasing comes to some level and it doesn't arise when the same level is reached when FD decreases.

A question naturally arises: where, i.e. to the RIP of what muscles the pulse volleys thus generated, are directed.

In a general case when there is no, generally speaking, information on what specific muscles cause the discomfort we assume - and the experiments confirm - that the volleys are directed to all skeletal muscles simultaneously and cause simultaneous, but random and different in size, splashes.

In those special cases when there is a definite information which muscles determine FD the command on a splash reception can only be directed to RIPs of those muscles.

3°. Simplest scheme of SSM

Fig. 5 shows simplest scheme of SSM. The volley of impulses coming to SSM causes a splash of SSM that is a splash of number N_α of fired α -motoneurons and as a result an increasing and subsequent fall of the muscles' tension till a new volley of impulses will reach SSM from FD and cause a new splash of N_α .

Because of inertia of muscle a splash of its tension will be more smooth and prolonged as compared to a splash of SSM (fig. 6). That means that the muscle in respect to a splash of SSM acts as an inertial smoothing filter. This simple scheme (fig. 5) is sufficient enough to give full explanation of all results which were obtained in experiments on maintaining minimum pain stimulus in case the stimulus depends on muscle tension (see §2 , 1°). In these experiments the pain stimulus plays the role of FD. Let's assume that in some moment $t = t_0$. the enveloping EMG determines a point a (fig. 7). In this moment conditions for splash don't exist so N_α (and the enveloping EMG aswell)falls down till it reaches value at point б where conditions of splash arise.

At this moment the splash goes from FD to SSM and a splash of SSM starts. During this splash N_α goes up and subsequently down as it was described before till it reaches again the y_1 value when a new volley will arise.

These conceptions fully explain the existence of small and large splashes, the transiton to the new level during one splash, the fact that new splashes always start at the line which always lies strictly parallel to the line $y = y_{min}$ and under it, the disappearance of large splashes when there is no right-hand side of FD characteristic and all other facts which were observed during experiments and described above.

4°. The role of spindles and some detalization of the SSM scheme. The work of the scheme when it controls pair of antagonistic muscles

The possibilities of SSM get wider when one takes in account the spindle reception of the muscles which was till now ignored in our previous discussion of SSM.

Let's remind the simplest things about spindle reception which will be used further. The muscle spindles are receptors which react on: 1) change of the length of a muscle ℓ and 2) simultaneously on changing of both ℓ and $\frac{d\ell}{dt}$. In this last case one can approximately assume that a spindle reacts on the linear combination $h = a\ell + b\frac{d\ell}{dt}$. The factor b is equal to zero when $\frac{d\ell}{dt} \leq 0$, i.e. the spindle reacts on the velocity of ℓ only when muscle is stretched. Further attention will be concentrated on the h signal only.

External commands from the brain on the spindle receptor are changing the factors a and b by means of innervation of special interfusal muscle fibres which are situated inside the spindles and in general are similar to usual extrafusal muscle spindles. The tension depends mainly on the number of fired special neurons which are called γ-motoneurons. The signal from the spindles which is proportional to the value of h acts on α-motoneurons and on the above-mentioned background input of the RIP. Signals which control the number of fired γ-motoneurons come from "above" and we assume that they come also from neuronal organization which we called FD. Fig. 8 shows the scheme of SSM with peculiarities brought in by spindle reception.

Let's assume now that the joint angle changes as a result of tension of a pair of muscles-antagonists (fig. 9). Suppose that each of these muscles is governed by an SSM mechanism of the same kind. There exists only one FD for both muscles which depends on value and velocity of the joint angle and the volleys which this FD organization are sending in the moment of splash occurance simultaneously to both muscles' RIP.

Let's assume now that at some moment the FD organization aroused these volleys. The result will be that the splashes of the SSM will not be equal because the splash depends not only from command volley but from the background signal. This background signal is depending on the value of h for each muscle of the pair of antagonists separately and that is why this signal is not equal for both muscles. The value of h depends on the position and velocity of the joint angle, i.e. on ℓ and $\frac{d\ell}{dt}$ for each muscle. Let's assume, for example, that at the moment of the volley occurance the factor h for the left muscle on the fig. 9 is much greater as compared to that of the right muscle. The result will be that splashes of both muscles will begin simultaneously but the splash of left muscle will be much higher that that of the right one. As a result the summary effort with which both muscles act on the bone will be directed to the left muscle. This inequality of splashes is even greater because of the above-mentioned reciprocal innervation which wasn't discussed (fig. 10). This full scheme of SSM for a pair of muscles antagonists fully explains all results of mentioned experiments (see §2, 2°) on accurate maintaining and quick oscillation of joint angle.

The FD in this case was the absolute value of velocity of joint angle, with result that command volleys appeared every time when absolute value of joint angle velocity increasing reaches a certain level,and the signals from muscles are lead to the background input of RIP of that muscle which in this moment stretches. The mechanism fully explains the simultaneity of the splashes of antagonistic muscles and the reason why great and small splashes come in turn. Tremor of the joint angle is a result of these splashes. Its main high frequency (about 10 cycles/sec) is determined by these splashes'frequency while the tremor's slow irregular component is due to the fact that angle control depends on the angle velocity and that value of splashes is random and irregular.

The mechanism described accounts for many other details which were observed during these experiments (see[3]). In particular it explains the coincidence of observed facts in the two cases when subject carries out two directly ppposite tasks: he has to maintain strictly quiet the position of the joint and contrarily he has to oscillate with the joint as quick as possible. In both cases the joint is controlled by the same mechanism which works strictly in the same way[+)] with the same FD - in both cases its role plays absolute value of joint angle velocity. From the point of view of this mechanism to keep quiet is to oscillate with the greatest possible frequency but with small amplitude. The amplitude depends only on the threshold Δ of the FD which must be reached for arising of the command volley of impulses. When the subject goes from maintaining quiet joint angle position to moving it with high frequency only a change of this threshold Δ of FD takes place.

The possibility to explain in detail by means of one scheme three different situations such as sustaining of the minimal pain stimulus, accurate maintenance of joint angle and maintenance of quick oscillations of joint angle gives us many reason to believe that the same mechanism can be of use in many other cases when the aim is to maintain some posture which corresponds to a minimum of discomfort of some kind.

In spite of the fact that it wasn't proved experimentally till now authors have reason to believe that the same SSM mechanism could in main features explain the simplest processes going on when we keep the vertical posture of our body.

[+)] In literature the mechanism of maintaining joint angle often is described as a usual control loop in which the signal from spindles is used as a signal of deviation of angle real value from the value of angle which must be maintained. Our discussed above experiments fully contradict this assumption because it can't explain splashes of the muscle which is stretched, existance of simultaneous splashes of muscles antagonists and the fact that this effect is enhanced[3] during the deafferentation of the muscle antagonists (see[3]).

REFERENCE

1. Zacharova L.M., Litvintzev A.I. Search activity of muscle under the condition of closing it with artificial feedback. "Automatics and Remote Control", No. 11, 1966.
2. Litvintzev A.I. Muscle searching activity under condition of shutting artificial feedback simultaneously along several muscles. "Automatics and Remote Control", No. 3, 1968.
3. Chernov V.I. Control of one muscle and a pair of muscle-antagonists under accurate search conditions. "Automatics and Remote Control", No. 7, 1968.
4. Rozonoer L.I. Random interneuronal nets. "Automatics and Remote Control", No. 4-6, 1969.

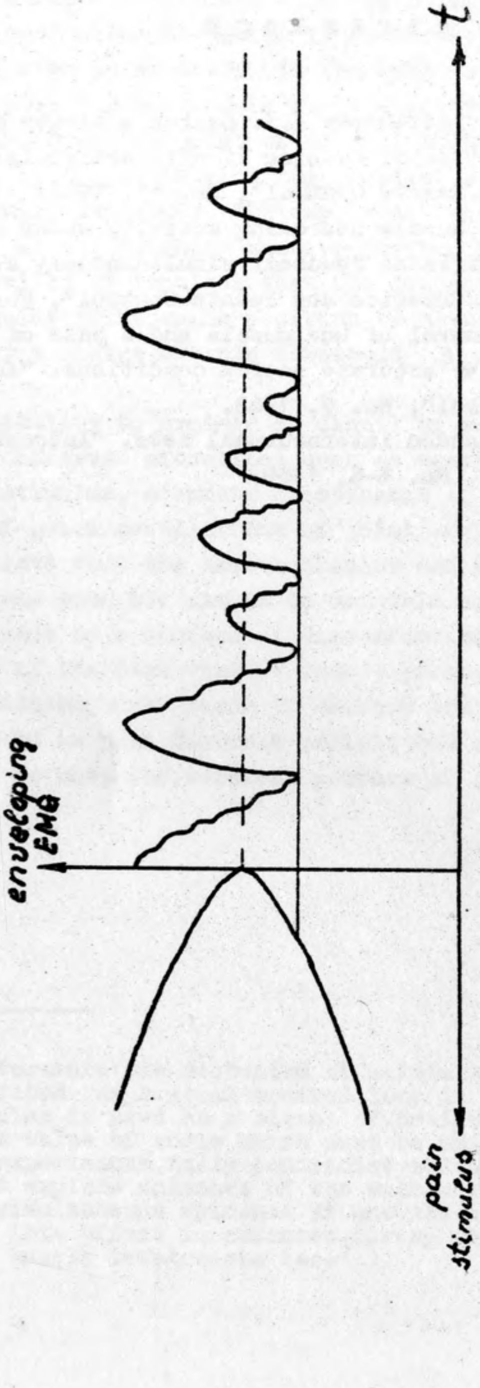


fig. 1

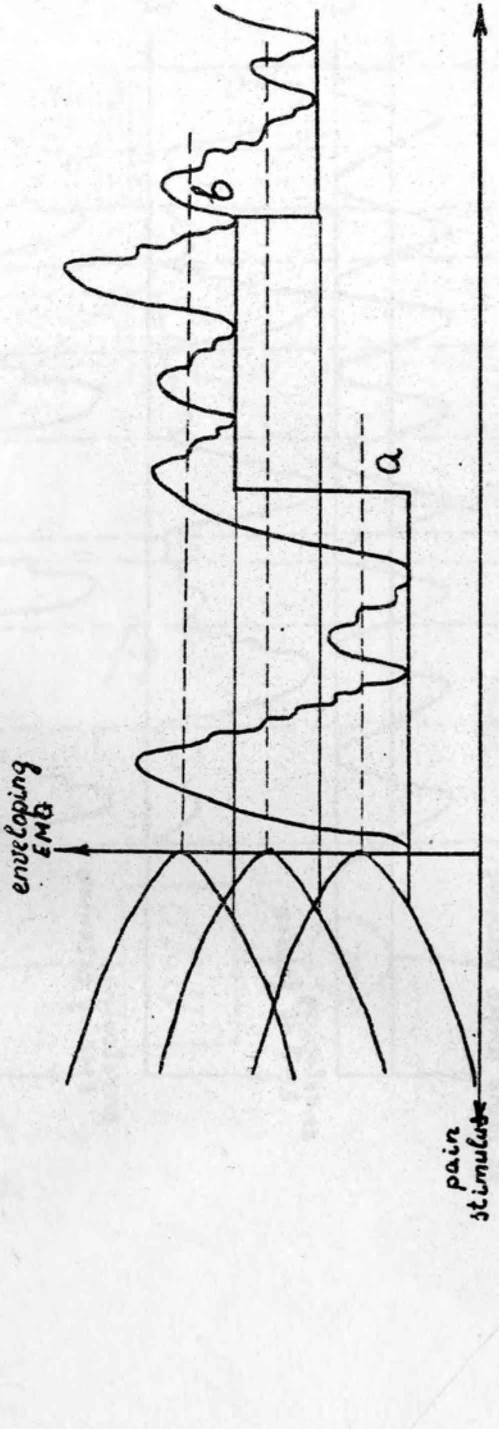


fig. 2

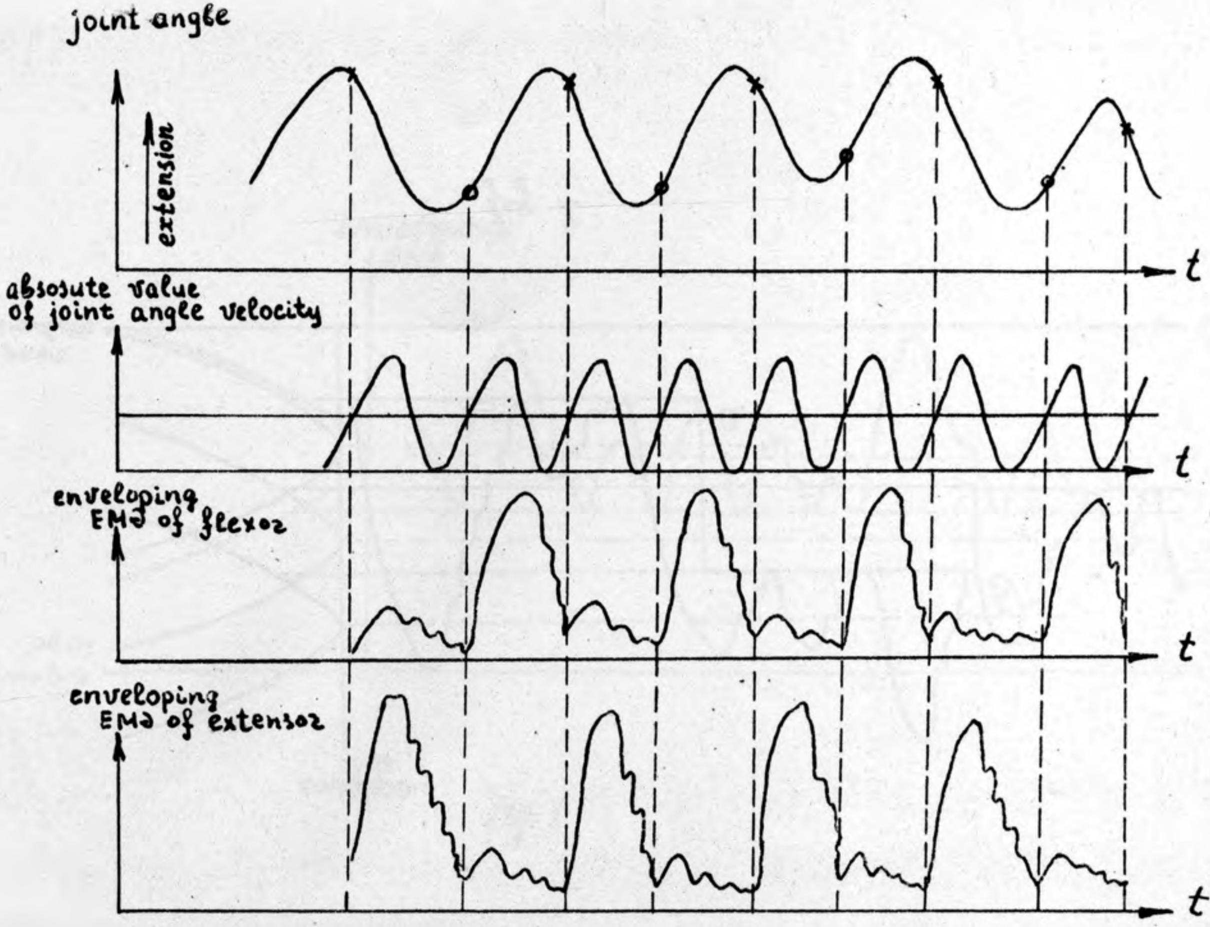

joint angle
extension
t
absosute value
of joint angle velocity
t
enveloping
EMG of flexor
t
enveloping
EMG of extensor
t

fig. 3

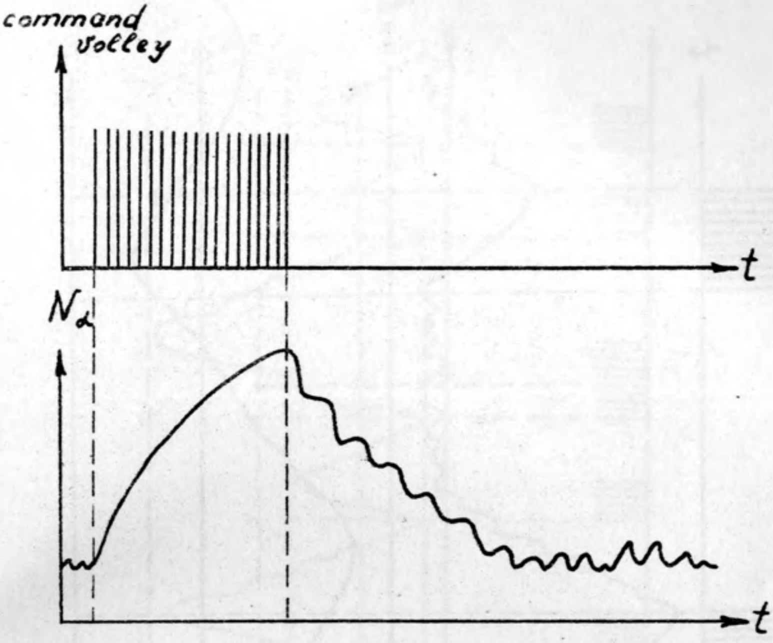


fig. 4

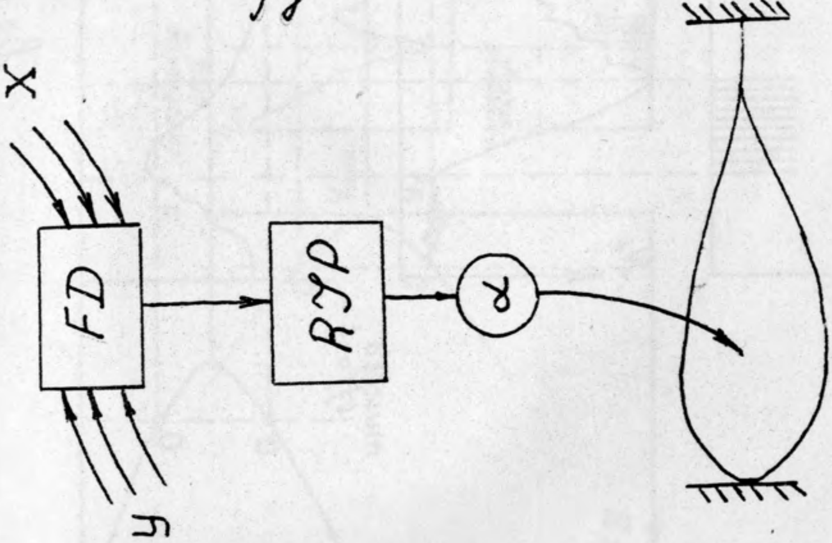


fig. 5

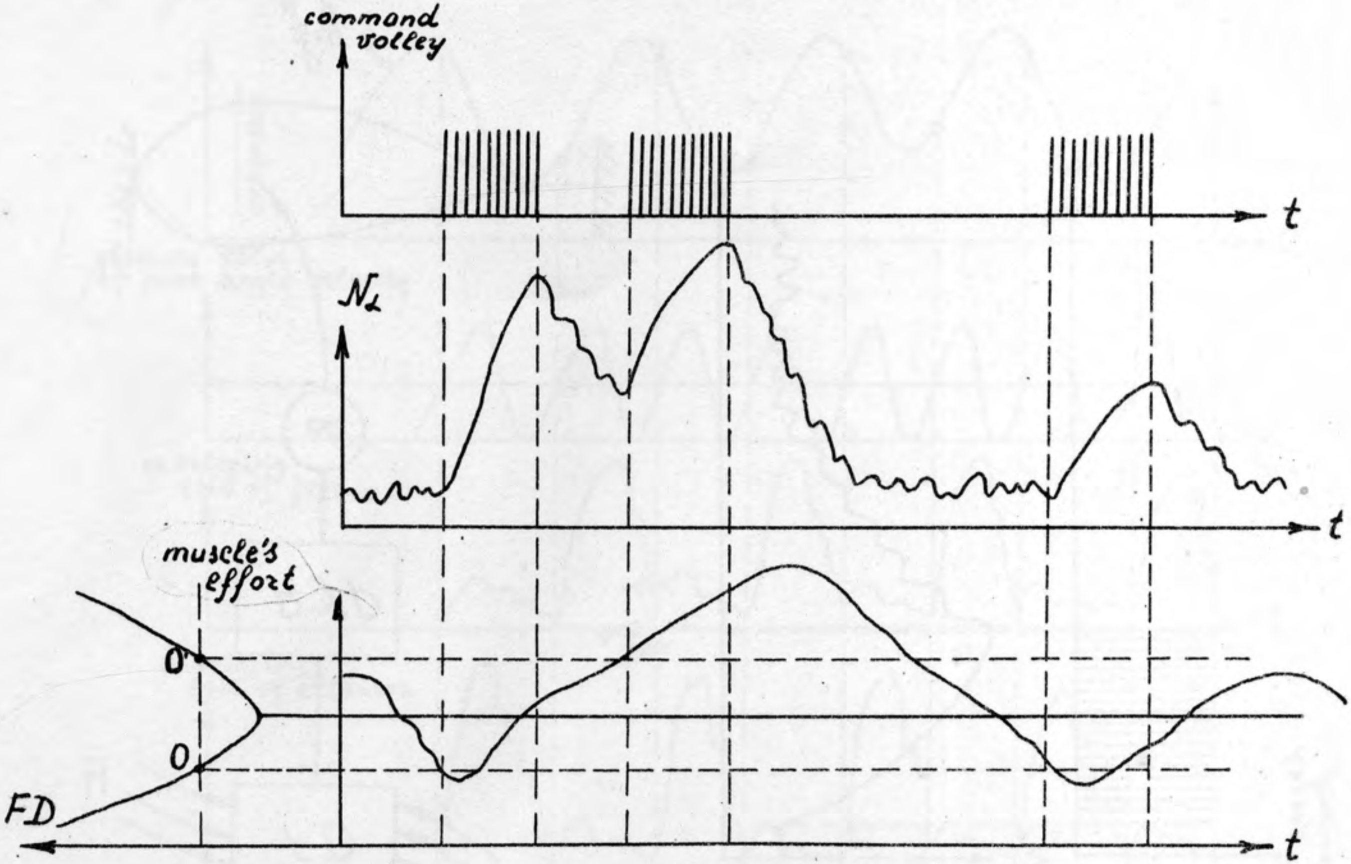


fig.6

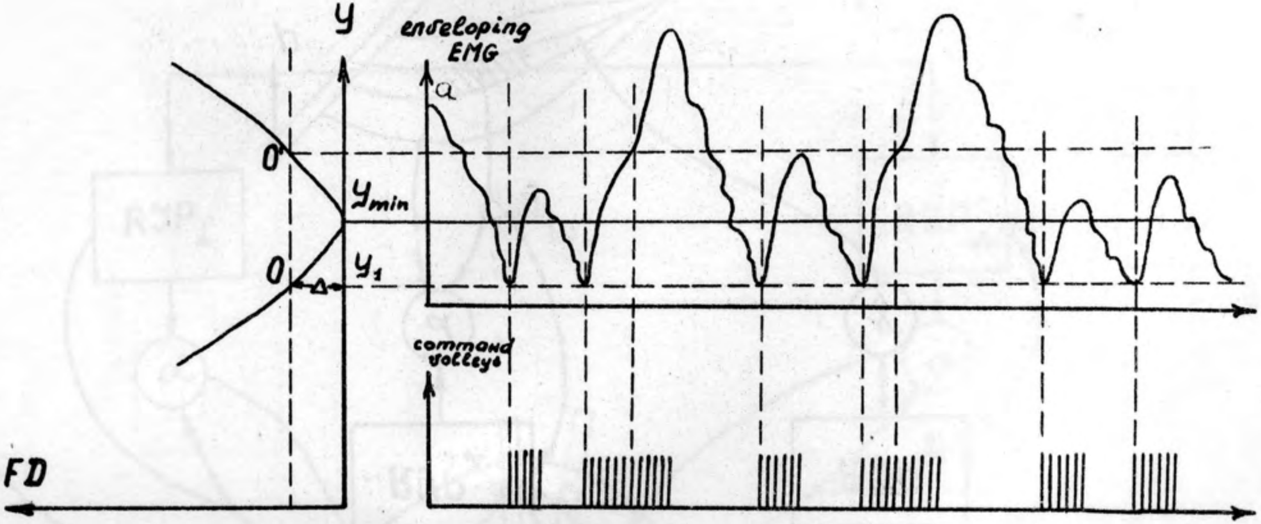


fig. 7

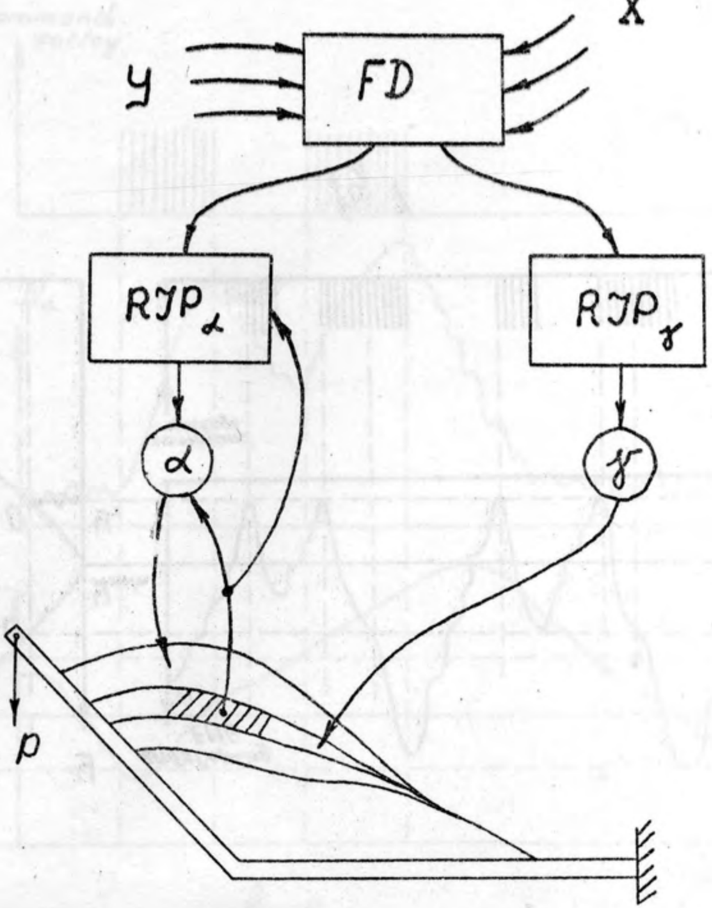


fig. 8

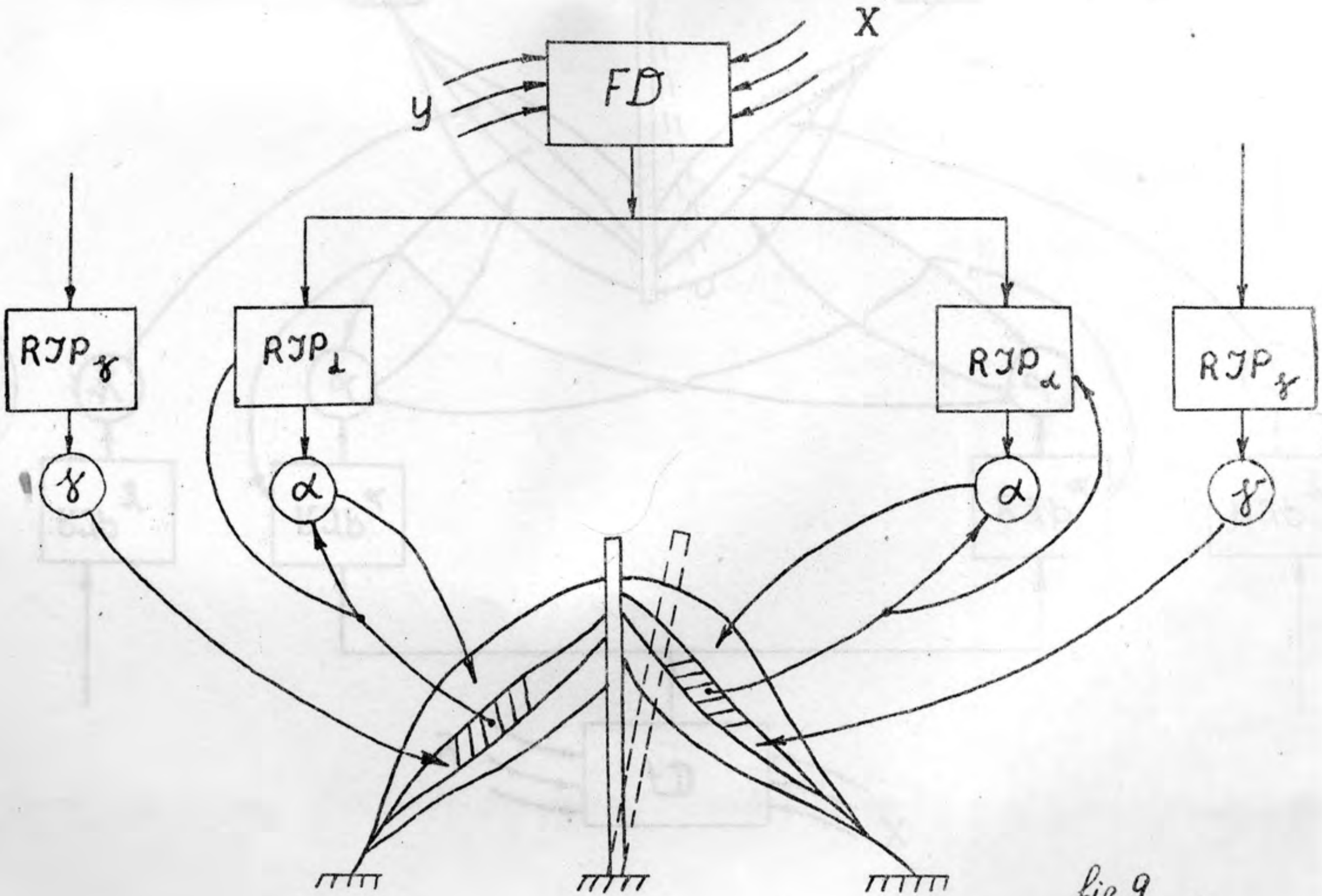

X
y
FD
RJP$_\gamma$
RJP$_\alpha$
RJP$_\alpha$
RJP$_\gamma$
γ
α
α
γ

fig. 9

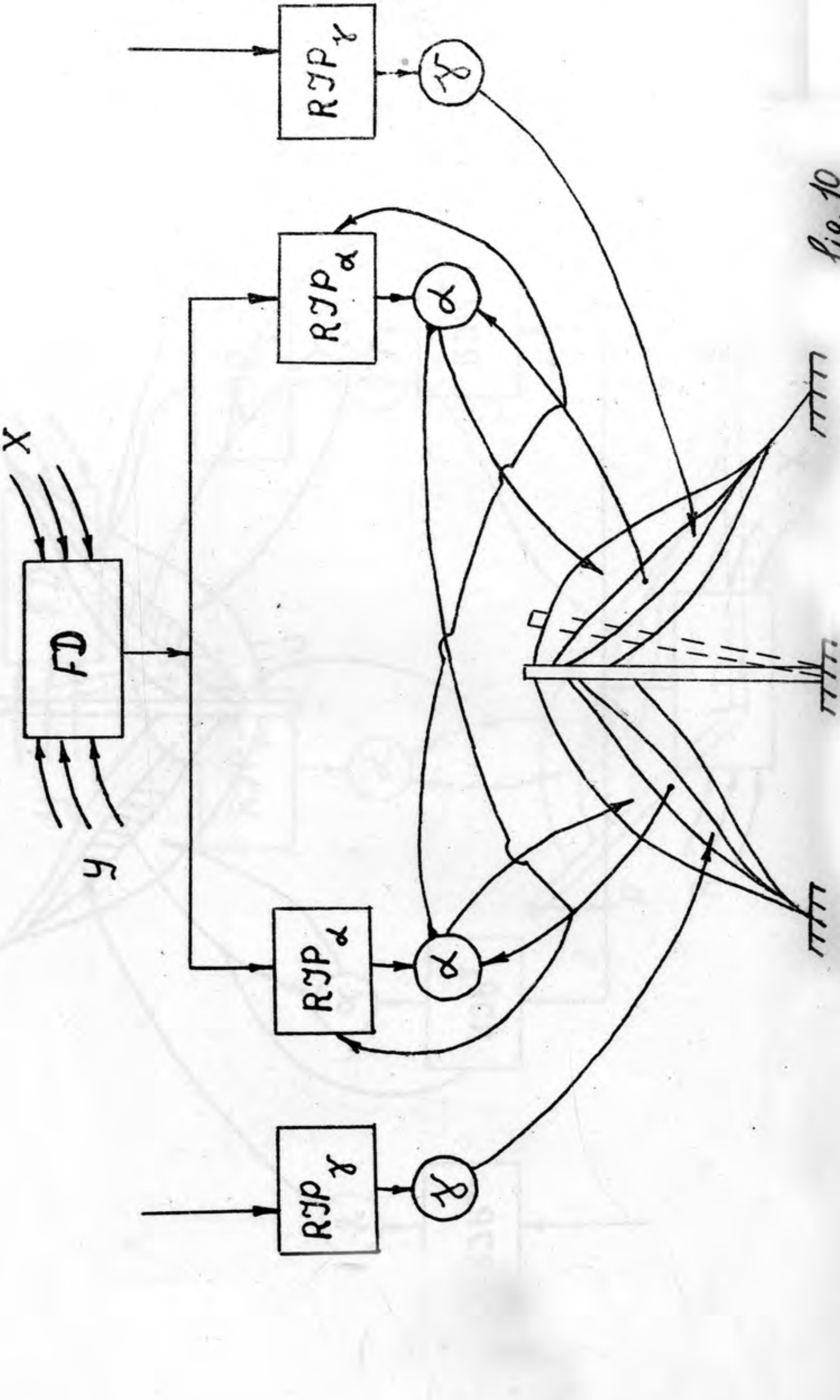

X
y
CD
RJP$_\gamma$
γ
RJP$_\alpha$
α
RJP$_\alpha$
α
RJP$_\gamma$
γ

fig. 10